

Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez

José Antonio Pérez-Gil, Salvador Chacón Moscoso y Rafael Moreno Rodríguez
Universidad de Sevilla

Una práctica habitual para obtener evidencia de validez de modelos teóricos basados en los resultados del Análisis Factorial Exploratorio -EFA-, (Arruda et. al., 1996), suele ser utilizar el Análisis Factorial Confirmatorio (CFA) para demostrar la validez de la estructura factorial previamente obtenida con EFA y por consiguiente la validez de las deducciones teóricas inferidas del mismo, es decir, los modelos obtenidos con EFA suelen ser validados con CFA. El objetivo de este artículo es incidir en la posible redundancia de este procedimiento usado en validez de constructo. Se intenta mostrar que generalmente CFA siempre validará las soluciones obtenidas previamente mediante EFA, dado que la estructura factorial resultante de EFA es una de las posibles soluciones que se derivan de la matriz de datos y, por consiguiente, es previsible que en otra muestra extraída de la misma población de referencia, se confirme la estructura obtenida en la anterior. En este sentido, el binomio EFA-CFA, puede ser considerado como un pseudoprocedimiento para la validación de constructo.

Construct validity: The use of factor analysis. The results of exploratory factor analysis (EFA) are usually used to deduce theoretical models (Arruda et. al., 1996) and then confirmatory factor analysis (CFA) is used to demonstrate the construct validity of those deductions. Therefore, models obtained from the former are usually validated using latter. The object of this study is to demonstrate the redundancy of this procedure in construct validation. It will be shown that, usually, CFA will validate solutions obtained using EFA. This article emphasizes that this type of procedure does not offer definitive information about the construct validity of the solutions obtained using exploratory factor analysis. In this sense, the EFA-CFA binomial could be considered as a pseudo construct validation.

En general, cuando tratamos con características físicas simples o de bajo nivel inferencial, como la longitud y el peso, éstas gozan de la ventaja de ser conceptos muy arraigados en la historia de la humanidad y ello permite establecer casi de modo automático o axiomático su conceptualización. Sin embargo, como señalan Thorndike y Hagen (1989), no ocurre lo mismo con características o conceptos menos ostensibles como pueden ser el de durabilidad de un cuerpo o aún más conceptos muy elaborados como la mayoría de los utilizados en psicología y otras disciplinas afines, que no son directamente observables y necesitan del apoyo de medidas indirectas o indicadores de los mismos para su medición. Casi todos tenemos una idea general acerca de lo que queremos decir cuando caracterizamos un concepto psicológico no observable o constructo hipotético, pero existen muchos aspectos específicos en los que podríamos estar en desacuerdo cuando tratamos de elaborar una definición precisa del mismo. Lo anterior pone de manifiesto que el problema de establecer acuerdos en lo que respecta a la validez de un concepto dado, es aún más complejo cuando consideramos atributos como los que trata la psicología (constructos o

conceptos no observables directamente). Dichos conceptos necesitan del acuerdo o consenso sobre sus significados para otorgarles el carácter de útiles y válidos desde un punto de vista científico, es decir, necesitan la implementación de procedimientos de validación. Es, en este marco de referencia, donde cobra su sentido más amplio el concepto de *validez de constructo*.

La validez de constructo es la principal de los tipos de validez, en tanto que «la validez de constructo es el concepto unificador que integra las consideraciones de validez de contenido y de criterio en un marco común para probar hipótesis acerca de relaciones teóricamente relevantes» (Messick, 1980; p.1015), en este mismo sentido (Cronbach, 1984; p.126) señala que «la meta final de la validación es la explicación y comprensión y, por tanto, esto nos lleva a considerar que toda validación es validación de constructo».

La génesis de la validez de constructo como un concepto integrador de validez hay que situarla en la primera versión de los *Standards for Educational and Psychological Testing* (APA, 1954) y en la publicación del influyente trabajo de Cronbach y Meehl (1955). Según estos autores, esta validez consiste en un análisis de la significación de las puntuaciones de los instrumentos de medida expresada en términos de los conceptos psicológicos asumidos en su medición. Como señala Martínez Arias (1995), este nuevo concepto de validez se empieza a percibir fundamental y básico y, aún más importante, como integrando a los anteriores enfoques de validez —heredados de las tradiciones empirista (validez criterial) y racionalista (validez de contenido)—, poniéndose así los ci-

Correspondencia: José Antonio Pérez-Gil
Facultad de Psicología
Universidad de Sevilla
41005 Sevilla (Spain)
E-mail: perezgil@psicoexp.us.es

mientos para este enfoque globalizador que va a echar por tierra la concepción tripartita de la validez y va a defender una concepción unificada de la misma, en la cual parece haber un consenso emergente acerca del papel central desempeñado por la validez de constructo (Moss, 1992).

Los trabajos de Cronbach (1980, 1982, 1988), Guion (1977, 1980), Loevinger (1957) y Tenopyr (1977) destacan por su apoyo a esta perspectiva integradora, mas la figura clave es la de Samuel Messick (1975, 1980, 1981, 1988, 1989, 1994, 1995). Messick (1995) afirma que *'la validez unificada integra consideraciones de contenido, criterio y consecuencias en un marco de referencia de constructo para la evaluación empírica de hipótesis racionales acerca del significado de las puntuaciones y de relaciones relevantes desde el punto de vista teórico, incluyendo las de naturaleza científica y aplicada'* (p.741).

De lo anterior, se puede establecer que es justamente el significado de las puntuaciones el que proporciona una base racional, permitiendo (1) juzgar la relevancia y representatividad del contenido del test (validez de contenido) y (2) establecer hipótesis predictivas (validez criterial) que, a su vez, pueden contribuir a arrojar luz acerca de la naturaleza del constructo. Dicho de otro modo, *'la validez de constructo subsume la relevancia y representatividad de los contenidos, así como las relaciones con los criterios, ya que ambas dan significado a las puntuaciones de los tests'* (Martínez Arias, 1995, p.335).

Por consiguiente, desde este planteamiento la validez de constructo puede ser conceptualizada en términos de un proceso científico de contrastación de hipótesis, donde entraría tanto lo empírico como los juicios racionales: las hipótesis serían las inferencias realizadas a partir de las puntuaciones de los tests y la validación el proceso de acumulación de evidencias que apoyen dichas inferencias, usos o interpretaciones de las puntuaciones del test. *'La validación de un test abarca todas las cuestiones experimentales, estadísticas y filosóficas por medio de las cuales se evalúan las hipótesis y teorías científicas'* (Messick, 1989, p. 14).

Entre los procedimientos o técnicas estadísticas utilizados para la contrastación de la validez de constructo destaca en mayor medida el Análisis Factorial (en adelante AF). En general, podemos decir que ésta es la técnica por excelencia utilizada para la validación de constructo.

Conceptualmente, AF presenta dos tipos o modalidades diferentes: Análisis Factorial Exploratorio (en adelante EFA) y Análisis Factorial Confirmatorio (en adelante CFA). Las diferencias entre ambas son numerosas, tanto desde una perspectiva teórica como matemática. La diferencia más importante está referida a que CFA se conduce principalmente por teorías sustantivas y por expectativas, mientras que EFA, como su nombre implica, principalmente es una técnica que, basada en los datos, intenta descubrir la estructura subyacente que éstos poseen (Bollen, 1989).

En general, podemos establecer que ambos tipos de procedimiento se corresponden con las dos grandes aproximaciones que solemos llevar a cabo para la definición de los constructos: la aproximación inductiva o exploratoria y la deductiva o confirmatoria.

La *aproximación inductiva o exploratoria* supone delimitar un número amplio de indicadores que supuestamente miden el constructo. Seguidamente estos indicadores son analizados mediante AFE para buscar patrones de relación entre los indicadores y, finalmente, *a posteriori*, «se pone nombre...» a esos patrones de relación, «definiendo» de esta forma el constructo. Esta aproxima-

ción trabaja desde las mediciones empíricas de los indicadores hacia la definición del constructo.

Esta manera de proceder en la perspectiva exploratoria puede tener implicaciones importantes sobre las inferencias de los modelos teóricos derivados de EFA porque, aunque éste se usa con el propósito de generar estructuras de modelos teóricos e hipótesis que se puedan contrastar empíricamente (Gorsuch, 1983), es realizado sin ninguna especificación previa sobre el modelo y, por lo general, no más que el número de factores y/o la consideración de si éstos están o no relacionados entre ellos (Gerbin & Hamilton, 1996). El procedimiento que sigue EFA supone extraer automáticamente, con criterios estadísticos, los factores, y, entonces, rota la solución inicial para obtener la estructura factorial más simple desde el punto de vista de su interpretación más significativa, siguiendo los criterios de parsimonia establecidos por Thurstone (1931,1947). En un sentido estricto, los criterios utilizados en este procedimiento no se ajustan adecuadamente a los contemplados por Messick (1995) y, por consiguiente, supone una aproximación débil a la definición y/o validación de un constructo.

En la *aproximación deductiva o confirmatoria* el constructo no se mide en el vacío. Está insertado en una teoría que dirige la propia definición del constructo indicando los comportamientos que pueden considerarse indicadores del constructo. La contrastación de las hipótesis estructurales que pone de manifiesto la teoría determinará la validez del mismo. Esta aproximación va desde la teoría hacia los hechos. En los procedimientos de CFA, por otra parte, se comienza con un modelo teóricamente plausible asumido para describir y/o explicar los datos empíricos. La construcción del modelo está basada o una información a priori sobre la naturaleza de la estructura de los datos, o bien en una teoría sustantiva en el campo de trabajo del que se trate. El procedimiento permite una valoración de la correspondencia de semejanza entre las características del concepto planteado y los datos obtenidos sobre dicho concepto a través de sus indicadores, es decir, evaluar la correspondencia teórica entre el constructo y el dato (Schmitt, 1995) y hacer inferencias sobre los constructos delimitados a partir de las muestras particulares de indicadores. En este sentido CFA se ajusta a los criterios señalados por Messick (1995) respecto a la validez de constructo y, por tanto, consideramos que CFA permite una aproximación fuerte a la definición y/o validación de un constructo.

Antes de que se desarrollasen las técnicas confirmatorias del AF, se utilizaba EFA tanto para fines exploratorios como confirmatorios. En la actualidad, aun cuando los procedimientos con CFA están muy desarrollados (existen en el mercado software potente para llevar a cabo CFA), todavía se sigue usando EFA con fines confirmatorios.

En los párrafos anteriores se ha descrito ya el uso del EFA con fines exploratorios. La utilización del EFA con fines confirmatorios implicaba tener en cuenta una estructura factorial basada en una teoría sustantiva, y aplicar AF en una muestra de datos para comprobar si la estructura resultante era coincidente o no con la estructura teórica.

Entendemos que, esta manera de proceder, es correcta en sí misma y se ajusta a los criterios de validez de constructo ya señalados, es decir, asumida a priori una estructura teórica, obtenemos una muestra representativa de la población y aplicamos el EFA. Si la estructura obtenida es coincidente con la estructura teórica hemos confirmado nuestro modelo teórico.

No obstante, la cuestión clave estaría en el caso contrario, es decir, qué ocurre si la estructura factorial obtenida con EFA no coincide con la estructura teórica de nuestro modelo. Generalmente solemos modificar nuestro modelo teórico y supeditarlo a la estructura obtenida (mediante la técnica de AF) con nuestros datos. Entendemos que en este caso, podemos estar incurriendo en un error fundamental. Dada las características propias de esta técnica (indeterminación factorial), la estructura factorial aportada es una solución plausible dentro del conjunto infinito de soluciones posibles que se pueden obtener con la misma. En este sentido, como mostraremos más adelante, no debemos dejar exclusivamente a la técnica la responsabilidad de nuestras posiciones teóricas.

Con frecuencia, los resultados del EFA se usan para definir constructos y por consiguiente para deducir modelos teóricos (Arruda et. al., 1996) y entonces se usa el CFA para demostrar la validez de constructo de aquellas deducciones. Considerando este argumento, los modelos obtenidos a partir del primero son validados usando el segundo. En general, las soluciones obtenidas usando EFA siempre serán validadas usando CFA. Por tanto podría producirse una tautología en esta manera de proceder.

El principal objetivo de este artículo es describir las implicaciones metodológicas de esta concepción y tratar de mostrar que hay una redundancia del procedimiento en la validación de constructo. Se ilustrará que, usualmente, las soluciones obtenidas usando EFA siempre serán validadas usando CFA.

Método

Modelo

Se llevó a cabo un estudio con técnicas de Monte Carlo para obtener muestras representativas a partir de cuatro supuestos modelos teóricos poblacionales con múltiples indicadores. Los modelos tenían seis o nueve indicadores en dos o tres factores respectivamente con tres indicadores por factor. Los pesos de los respectivos indicadores fueron de .7, .8 y .9, con las varianzas de los factores puestas a 1. En dos de ellos los factores estaban relacionados y en los otros dos no relacionados.

Muestras

Partiendo de las cuatro matrices poblacionales de varianzas-covarianzas (obtenidas de la combinación del número de factores (dos o tres) y de la correlación entre ellos (relacionados/no relacionados)) se extrajeron un total de 2400 muestras simuladas, de manera que hubo seiscientas muestras emparejadas para cada modelo. Es decir, se obtuvieron trescientas muestras emparejadas de cada modelo, 150 pares de muestras para aplicar EFA a una de las muestras de cada pareja, y, entonces, en cada una de las otras, se validó, utilizando CFA, la estructura factorial obtenida con EFA.

Instrumentos

Se utilizó el programa GENRAW (Jöreskog y Sörbom 1993) para generar las 2400 muestras, considerándose en cada caso las correspondientes matrices de varianzas-covarianzas poblacionales de cada uno de los modelos propuestos y especificándose que la distribución de las variables fuese normal.

Procedimiento

El procedimiento seguido supuso analizar cada conjunto de muestras de acuerdo a lo siguiente: a) en primer lugar, en una de las muestras de cada pareja se aplicó EFA utilizando Máxima Verosimilitud (ML) como procedimiento de extracción de los factores y especificando dos o tres factores y eligiendo VARIMAX u OBLIMIN como procedimientos de rotación según el correspondiente modelo analizado. Para ello se utilizó el módulo PROC FACTOR del programa SPSS (Spss INC, 1994); b) en segundo lugar, para validar la estructura factorial de pesos obtenidos como solución plausible con EFA/ML, en la otra muestra de cada pareja se llevó a cabo CFA/ML, mediante el programa LISREL8 (Jöreskog y Sörbom 1993), utilizando la matriz factorial de pesos obtenidas previamente en la primera muestra, es decir, se aplicó EFA a una de las muestras de cada pareja, y, entonces, en cada una de las otras se validó la estructura factorial obtenida previamente utilizando CFA.

Resultados

En primer lugar, la tabla 1 presenta el número y porcentaje de veces que la hipótesis nula fue aceptada utilizando el estadístico χ^2 como índice de bondad de ajuste.

En todos los modelos el porcentaje de muestras en las que se aceptó la hipótesis nula fue del 70%, excepto en el modelo 2 que fue superior (80%), quedando sin confirmar un porcentaje de muestras del 30%.

Es un hecho conocido que el estadístico χ^2 es sensible al tamaño de las muestras y a la violación del supuesto de normalidad multivariante de las variables observadas (Jöreskog y Sörbom 1989). Es por ello que, además de utilizar este índice, en la literatura sobre modelos de ecuaciones estructurales (SEM) se propongan diferentes índices de bondad de ajuste que corrigen este inconveniente del estadístico χ^2 (GFI, AGFI, RMR, RMSA). Entre los más conocidos y usuales están el índice de bondad de ajuste denominado GFI y una variante ajustada del mismo (AGFI) propuestos por Jöreskog y Sörbom (op. cit.).

<i>Tabla 1</i>			
Bondad de ajuste de la estructura factorial usando el índice χ^2			
Estadístico χ^2 (Índice de bondad de ajuste)			
Modelo 1: Dos factores incorrelacionados (ML/Varimax)			
H ₀ Aceptada		H ₀ No aceptada	
N° de muestras	Porcentaje	n° de muestras	Porcentaje
210	70%	90	30%
Modelo 2: Dos factores correlacionados (ML/Oblimin)			
H ₀ Aceptada		H ₀ No aceptada	
N° de muestras	Porcentaje	n° de muestras	Porcentaje
240	80%	60	20%
Modelo 3: Tres factores incorrelacionados (ML/Varimax)			
H ₀ Aceptada		H ₀ No aceptada	
N° de muestras	Porcentaje	n° de muestras	Porcentaje
210	70%	90	30%
Modelo 4: Tres factores correlacionados (ML/Oblimin)			
H ₀ Aceptada		H ₀ No aceptada	
N° de muestras	Porcentaje	n° de muestras	Porcentaje
210	70%	90	30%

El índice GFI tiene la ventaja de ser insensible al tamaño muestral; sus valores oscilan entre 0 y 1, aunque excepcionalmente puede tener valores superiores a 1. Los ajustes de modelos con valores superiores a .90 en este índice pueden considerarse aceptables. En la tabla 2 se presentan los resultados utilizando el índice GFI.

Puede observarse que en todos los contrastes se obtuvo un valor del índice GFI superior o igual a .96. En el 5% de las muestras de obtuvo un valor de .96, el 35% de las mismas aportó un valor de .97; de igual modo, otro 35% obtuvo valores de .98 y en un 25% de las muestras el valor de GFI fue de .99. En general, puede observarse una leve disminución de los valores de ajuste en función de la complejidad de modelo que se pretende ajustar. Así, podemos observar que, en términos generales, los modelos con dos factores aportan mejores valores del índice que en los modelos de tres factores.

El índice AGFI puede considerarse una variante del anterior en el sentido de que es el índice anterior ajustado en base a los grados de libertad del modelo que se contraste. Al igual que el índice anterior sus valores oscilan entre 0 y 1 y valores de ajuste superiores a .80 de este índice pueden considerarse buenos ajustes del modelo a los datos.

En la tabla 3 se presentan los resultados utilizando el índice AGFI. Puede observarse que, en todos los contrastes, se obtuvo un valor del índice AGFI superior o igual a .94. En el 2.5% de las muestras de obtuvo un valor de .94, siendo este valor el menos frecuente; el 22.5% de las mismas aportó un valor de .95; un 37.5% obtuvo valores de .96, siendo éste el valor más frecuente; en el 12.5% de las muestras de obtuvo un valor de .97, el 29% de las mismas aportó un valor de .98; y sólo un 5% obtuvo valores de .99. En general, puede observarse con mayor claridad que en el índice GFI se produce una disminución de los valores de ajuste en función de la complejidad de modelo que se pretende ajustar (modelos con mayor número de grados de libertad). Así, podemos observar que, en términos generales, los modelos con dos factores aportan mejores valores del índice que los modelos de tres factores (se puede observar como en los modelos de tres factores no se

alcanzan valores por encima de .97 y, en el modelo de tres factores correlacionados aparece el valor de .94 para el índice AGFI.

Discusión

En general, en todos los contrastes realizados se obtuvieron valores óptimos en los índices de ajustes usados para validar, con CFA, la solución plausible obtenida con EFA, y ello fue así independientemente de la complejidad de las estructuras característica de las poblaciones simuladas, esto es, dos o tres factores, relacionados/no relacionados.

Los resultados demuestran que, en general, las soluciones factoriales obtenidas usando EFA siempre son validadas usando CFA. Esto implica que la distribución muestral de las matrices de varianzas-covarianzas permite validar los resultados exploratorios a priori, es decir, es difícil encontrar diferencias analíticas significativas entre las matrices utilizadas para ambos procedimientos. En este sentido, si en una muestra particular la estructura factorial resultante, usando EFA, no es validada usando CFA con otra muestra obtenida de la misma población de referencia, implicaría problemas de sesgo que entendemos pueden deberse exclusivamente al error de tipo I inherente al procedimiento analítico en sí mismo y, en todo caso, en situaciones aplicadas podrían atribuirse además a problemas de representatividad de las muestras analizadas (comparaciones EFA-CFA en muestras no equivalentes).

En resumen, el análisis comparativo de los valores de ajustes en los distintos índices permite considerar como resultado de este trabajo que, en general, dos muestras obtenidas desde una misma población, si el procedimiento de extracción de las mismas es el adecuado, coincidirán en sus valores de ajuste, es decir, razonablemente la mayoría de las veces la decisión a tomar irá en el sentido de aceptar la hipótesis nula.

Estadístico GFI (Índice de bondad de ajuste)		
Modelo 1: Dos factores incorrelacionados (ML/Varimax)		
Valor	n° de muestras	Porcentaje
.99	150	50%
.98	90	30%
.97	60	20%
Modelo 2: Dos factores correlacionados (ML/Oblimin)		
Valor	n° de muestras	Porcentaje
.99	150	50%
.98	90	30%
.97	60	20%
Modelo 3: Tres factores incorrelacionados (ML/Varimax)		
Valor	n° de muestras	Porcentaje
.98	150	50%
.97	120	40%
.96	30	10%
Modelo 4: Tres factores correlacionados (ML/Oblimin)		
Valor	n° de muestras	Porcentaje
.98	90	30%
.97	180	60%
.96	30	10%

Estadístico AGFI (Índice de bondad de ajuste)		
Modelo 1: Dos factores incorrelacionados (ML/Varimax)		
Valor	n° de muestras	Porcentaje
.99	30	10%
.98	120	40%
.97	30	10%
.96	60	20%
.95	60	20%
Modelo 2: Dos factores correlacionados (ML/Oblimin)		
Valor	n° de muestras	Porcentaje
.99	30	10%
.98	120	40%
.97	30	20%
.96	90	30%
.95	30	10%
Modelo 3: Tres factores incorrelacionados (ML/Varimax)		
Valor	n° de muestras	Porcentaje
.97	60	20%
.96	120	40%
.95	120	40%
Modelo 4: Tres factores correlacionados (ML/Oblimin)		
Valor	n° de muestras	Porcentaje
.97	30	10%
.96	180	60%
.95	60	20%
.94	30	10%

Por último, entendemos que el trabajo que presentamos permite poner de manifiesto que este tipo de procedimiento (EFA-CFA) no añade una mejor información ni más definitiva sobre la validez de constructo de las soluciones (estructura factorial) obtenidas con EFA.

EFA asume implícitamente que el procedimiento puede o tiene capacidad para descubrir la estructura relacional que subyace a los indicadores y que la estructura que resulta de la aplicación de EFA es una evidencia de validez porque se asume que esta estructura resultante es coincidente con una estructura sustantiva soportada desde un marco teórico. Entendemos que este razonamiento es falaz porque se obvia la consideración de que toda justificación sustantiva sobre un marco teórico de referencia debe establecerse *a priori* y, por tanto, el hecho de «poner nombre...» *a posteriori* a los patrones de relación entre indicadores entra en conflicto con

los criterios de garantía de validez que no pueden asumirse a no ser que se plantee con nitidez el referente conceptual, y, a su vez, la validez de dicho referente vendrá dado por su correspondencia con las evidencias obtenidas, es decir, teorías previas y utilidad que presente dicho referente (Schmitt, 1995). La validez, por tanto, no debe venir dada por el procedimiento exclusivamente y, en este sentido, el binomio EFA-CFA para obtener validez de constructo podría considerarse como un pseudo-procedimiento para obtener dicha validez.

Entendemos que una aproximación exploratoria o proceso estrictamente inductivo en el que se parta de un total desconocimiento de lo que se quiere definir o validar, dará lugar a resultados azarosos y posiblemente inesperados, pues dependería completamente de circunstancias o preferencias arbitrarias y descontroladas.

Referencias

- American Psychological Association, American Educational Research Association and National Council on Measurement in Education (1954,1985). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Arruda, J.E.; Weiler, M.D.; Valentino, D.; Willis, W.G.; Rossi, JS; Stern, R.A.; Gold, S.M. & Costa, L. (1996). A Guide for Applying Principal Components Analysis and Confirmatory Factor Analysis to Quantitative Electroencephalogram Data. *International Journal of Psychophysiology*, 23 (1-2). 63-81
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Cronbach, L.J. (1949, 1960, 1970, 1984, 1990). *Essentials of psychological testing (1ª-5ª edición)*. New York: Harper.
- Cronbach, L.J. (1980). Validity on parole: How can we go straight? En Schrader, W.B. (Ed.) *New directions for testing and measurement, planning achievement: Progress over a decade*. Jossey Bass. San Francisco.
- Cronbach, L.J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey Bass.
- Cronbach, L.J. (1988). Internal consistency of test: Analyses old and new. *Psychometrika*, 53, 1, 63-70.
- Cronbach, L.J. y Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281 -302.
- Gerbin, D.W. y Hamilton, J.G. (1996). Viability of Exploratory Factor Analysis as a Precursor to Confirmatory Factor Analysis. *Structural Equation Modeling*, 3,2-72.
- Gorsuch, R.L. (1983). *Factor Analysis (2ª Ed.)*. Hillsdale, NJ: Erlbaum.
- Guion, R.M. (1977). Content validity: Three years of talk-what's the action?. *Public Personnel Management*, 6, 407-414.
- Guion, R.M. (1980). On trinitarian doctrine as of validity. *Professional Psychology*, 11, 385-398.
- Jöreskog, K.G & Sörbom, D. (1989). *LISREL7: User's reference guide*. Mooresville, IN: Scientific Software.
- Jöreskog, K.G & Sörbom, D. (1993). *LISREL8: Structural Equation Modeling with the SIMPLIS. Command language*. Chicago, IL: Scientific Software.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.
- Martínez-Arias, M.R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30,955 -966.
- Messick, S. (1980). Test validity and ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, 89, 575-588.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. En H. Wainer y H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. The specification and development of tests of achievement and ability. En R. L. Lino (Ed.), *Educational Measurement (3th edition)*. Washington, DC: American Council on Education.
- Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment*, 10, 1-9.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 15, 5-12.
- Moos, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229-258.
- Schmitt, F.F. (1995). *Truth: A primer*. Boulder, Colorado: Westview Press.
- SPSS Inc., M. J. Norusis (1994). *SPSS/PC+ for IBM=PC/XT/AT*. Chicago, IL: SPSS Inc.
- Tenopir, M.L. (1977). Content-construct confusion. *Personnel Psychology*, 30, 47-54.
- Thorndike, R. L. y Hagen, E.P. (1989). *Measurement and evaluation in Psychology and education (2ª edición)*. New York: Wiley.
- Thurstone, L.L., (1931). Multiple Factor Analysis. *Psychological Review*, 38, 406-427.
- Thurstone, L.L., (1947). *Multiple Factor Analysis*. Chicago, IL: University of Chicago Press.