

Interdisciplinaria

Centro Interamericano de Investigaciones Psicológicas y Ciencias Afines

lcalvo@conicet.gov.ar

ISSN (Versión impresa): 0325-8203

ARGENTINA

2005

María Cristina Richaud

DESARROLLOS DEL ANÁLISIS FACTORIAL PARA EL ESTUDIO DE ÍTEM
DICOTOMICOS Y ORDINALES

Interdisciplinaria, año/vol. 22, número 002

Centro Interamericano de Investigaciones Psicológicas y Ciencias Afines

Buenos Aires, Argentina

pp. 237-251

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Universidad Autónoma del Estado de México

DESARROLLOS DEL ANALISIS FACTORIAL PARA EL ESTUDIO
DE ITEM DICOTOMICOS Y ORDINALES
María Cristina **Richaud** *

Resumen

Los ítem de evaluación cualitativa (*dicotómicos u ordinales*) presentan problemas específicos para la utilización del *análisis factorial* en las pruebas de personalidad.

En el presente trabajo se revisan algunos desarrollos recientes en análisis factorial que resultan apropiados para este tipo de ítem. Estos desarrollos se han realizado en el contexto de dos modelos estadísticos: la *Teoría de la Respuesta al Ítem* y el modelo de las ecuaciones estructurales. También se verá la relevancia del escalamiento de ítem en el contexto de estos dos modelos. Se presentan algunos ejemplos de la utilidad de estos modelos para resolver cuestiones básicas, tales como la dimensionalidad de la escala y las propiedades generales de los ítem, la adecuación de las respuestas observadas y el funcionamiento diferencial del ítem a través de diferentes submuestras.

Palabras clave: Ítem dicotómicos u ordinales - análisis factorial - Teoría de la Respuesta al Ítem.

Abstract

Factor analysis has been used in formulating conceptual models in personality and personality assessment, as well as in

* Doctora en Psicología. Directora del Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental (CIIPME) y Miembro de la Carrera del Investigador Científico del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). E-Mail: mrichaud@conicet.gov.ar

Richaud

the process of construction of personality scales. *Factor analysis* assumes continuously measured interval level data. However, applications of the factor analysis model in the personality literature frequently are conducted using dichotomous or ordinal data obtained at the item level.

It has been proposed several solutions for studying dichotomous or ordinal data. Christoffersson (1978) introduced a method for factor analyzing dichotomous data using tetrachoric correlations. Muthén (1984) extended this method to provide a less computationally heavy approach.

Standard factor analysis implies two different levels of variables: unobserved factors, and observed indicators for those factors (items). The generalized least squares method to the factor analysis of dichotomous data requires one additional intermediate level between the observed data and the latent variable. Thus two levels of abstraction are involved in the analysis: observed dichotomous or ordered categorical items are linked to unobserved latent response variables via tetrachoric or polychoric correlations. These unobserved latent response variables then serve as the indicators for the factors. In this model the factors summarize the relations among latent variables rather than directly among observed variables.

Another method for the factor analysis of dichotomous or ordered categorical items is that of maximum likelihood. As in the case of the generalized least squares method, the maximum likelihood approach use tetrachoric correlations among items, but approximates a numerical integration of a distribution of observations, assumed to be normal, using weighted sums.

There exist also parallel analysis programs (Buja, & Eyuboglu, 1992; Horn, 1965) that produce data sets based in aleatory numbers normally distributed, generated by the computer (O'Connor, 2000).

Another manner of analysis of relationships between unobserved factors and observed dichotomous or ordinal data is that of applying *Item Response Theory* (IRT).

In conjunction with exploratory item-level factor analyses that address the underlying dimensionality of the item set, IRT and confirmatory item-level factor analyses are useful for the construction and validation of personality inventories.

Estudio de ítem dicotómicos y ordinales

Another important function of IRT is in the design of *appropriateness* indices that serve in evaluating validity scales, identifying those protocols that may be characterized by aberrant responding for a set of items.

IRT has also been used to develop the full-information item-level factor analysis (Bock, & Schilling, 1997) that directly works on response patterns and avoid the artifacts associated to *phi* and tetrachoric coefficients (McLeod, Swygert, & Thissen, 2001; Swygert, McLeod, & Thissen, 2001).

Summing up, it is necessary to elucidate implications of conceptual models of intelligence and personality assessment and their impact on how one approaches these data statistically (Panter, Swygert, Dahlstrom, & Tanaka, 1997). There are not standar methods nor models that one applies always and in every cases if one really want to obtain valid measures and assessment (Richaud de Minzi, 2005).

Key words: Dichotomous and ordinal items - factor analysis - Item Response Theory.

El análisis factorial ha sido y es muy utilizado en los estudios de evaluación de la personalidad, tanto en el proceso de construcción de escalas como en la puesta a prueba de las teorías. Sin embargo, en la mayoría de los casos no fueron tenidos en cuenta los supuestos necesarios para extraer inferencias apropiadas según el nivel de medición de los ítem estudiados. Si bien el análisis factorial supone la utilización de ítem a nivel intervalar, es muy frecuente encontrar análisis factoriales realizados con ítem dicotómicos u ordinales de tipo Lickert. Hay que tener presente que los análisis factoriales producen resultados significativos solamente si los datos utilizados son verdaderamente continuos y multivariadamente normales.

La aplicación empírica del análisis factorial exploratorio trata de identificar constructos o dimensiones subyacentes significativas, a través del estudio de la covariancia entre rasgos observados, respuestas, signos y síntomas. En el modelo del análisis factorial confirmatorio, las interrelaciones entre los indicadores y el constructo se especifican antes de poner a prueba

Richaud

el modelo y luego se evalúa si la covariancia observada se ajusta al modelo teórico preestablecido.

Si se revisan los estudios de personalidad donde se utiliza el análisis factorial estándar, se observa generalmente que se supone que los datos están en un nivel intervalar, de acuerdo con los supuestos estadísticos básicos del modelo del análisis factorial. Sin embargo, en realidad casi ninguno de los datos obtenidos cumple estos supuestos. A lo sumo, se aproximan a datos continuos sumando información de unidades, como respuestas a ítem en *clusters* o valores de escalas.

Se han propuesto varias soluciones para el estudio de ítem dicotómicos u ordinales.

Christoffersson (1978) ideó un método para analizar factorialmente los ítem dicotómicos usando correlaciones tetracóricas. Muthén (1984) extendió este método para proporcionar una forma computacional menos compleja.

El modelo del análisis factorial estándar implica dos niveles diferentes de variables: *factores inobservables* e *indicadores observables de estos factores* (ítem). *El método generalizado de los mínimos cuadrados* para el análisis factorial de datos dicotómicos requiere un nivel intermedio entre los datos observados y la variable latente. Este nivel corresponde a las variables latentes de respuesta, una para cada dicotomía observada, que son cortadas en puntos de umbral correspondientes a juicios observados como *verdadero-falso*. Recuérdese que la correlación tetracórica supone la relación entre dos variables latentes continuas que han sido dicotomizadas en determinados puntos umbral. El análisis factorial puede entonces aplicarse directamente porque se está trabajando con indicadores continuos aunque inobservables.

En este tipo de análisis hay dos niveles de abstracción: ítem observados dicotómicos u ordinales, que se relacionan con variables de respuesta no observables a través de las correlaciones tetracóricas o policóricas. Luego, estas variables de respuesta latentes sirven de indicadores de los factores. Por lo tanto, en este modelo los factores resumen las relaciones entre variables de respuesta latentes más que directamente entre las variables observadas.

Por otra parte, la factibilidad del método generalizado de los cuadrados mínimos está relacionada inversamente con el número de ítem observados: a medida que los ítem observados aumentan, disminuye la factibilidad debido al aumento del tamaño de la matriz de pesajes utilizada en el cálculo.

Otro método propuesto para el caso de los ítem dicotómicos u ordinales es el *de máxima verosimilitud* que opera bajo un principio de *datos per-*

Estudio de ítem dicotómicos y ordinales

didos que en el contexto del análisis factorial, puede verse como un intento de estimar *factor scores* contra los que se puede hacer una regresión de las variables observadas (Bock & Aitken, 1981; Bock & Lieberman, 1970).

Este método parte también de las correlaciones tetracóricas entre los ítem pero, a diferencia del método de los cuadrados mínimos, el método de máxima verosimilitud se aproxima a una función integral de una distribución de observaciones supuestamente normal, usando sumas ponderadas. En este método, a diferencia del primero en el que la complejidad aumenta a medida que aumenta el número de ítem, la complejidad numérica y computacional del análisis factorial aumenta, a medida que se incrementa el número de factores hipotéticos que subyacen a un determinado conjunto de ítem (Mislevy, 1986).

Existen también los llamados *programas de análisis paralelo* (Buja & Eyuboglu, 1992; Horn, 1965) que producen conjuntos de datos basados en números aleatorios distribuidos normalmente, generados por la computadora (O'Connor, 2000). Los autovalores de los datos brutos a nivel del ítem basados en información dicotómica o de tipo Likert, no se pueden comparar apropiadamente con los autovalores generados en el análisis paralelo, que se sustentan en números aleatorios de distribución normal. En lugar de ello, primeramente se debería determinar el número de factores o componentes, hallando los autovalores para la matriz de correlaciones tetracóricas o policóricas, y luego comparar estos autovalores contra los producidos aleatoriamente por la computadora. Y, de hecho, las subsecuentes factorizaciones se deberían hacer sobre la matriz de correlaciones tetracóricas o policóricas, y no sobre las correlaciones de Pearson.

El análisis paralelo también se puede hacer sobre permutaciones aleatorias de los datos directos o brutos, en que las distribuciones del *ítem/variables* son idénticas a aquellas provenientes de los datos reales. No se debería usar este método de análisis paralelo en datos a nivel de ítem, pues pueden surgir factores por la semejanza de la distribución de los ítem bajo análisis.

Teoría de la Respuesta al Ítem (TRI)

Otra forma de analizar las relaciones entre los indicadores y un factor subyacente, cuando los datos son dicotómicos u ordinales, es aplicando la Teoría de la Respuesta al Ítem (TRI).

La TRI se refiere al caso en que se supone que una dimensión subyacente simple genera un conjunto de respuestas observadas a un conjunto de

ítem dicotómicos. La probabilidad de responder a un determinado ítem de un conjunto es una función de dos componentes: los *parámetros del ítem* y un *parámetro para la persona*, que refleja dónde está localizada la misma en el atributo subyacente. Cada ítem de la escala puede representarse gráficamente con una curva o función no lineal (en general, una función logística) que caracteriza las propiedades específicas del ítem. En la evaluación educacional, la dimensión latente subyacente (θ) a un conjunto de ítem se llama *habilidad*, mientras que en la evaluación de la personalidad, describe la forma con que las personas responden a los ítem de un inventario, las características de estos ítem y el lugar que cada individuo tiene en la dimensión subyacente de personalidad.

Un supuesto común en los modelos TRI más usados es que los ítem que conforman el test miden solamente una habilidad. A esto se lo llama *supuesto de unidimensionalidad*. Un concepto relacionado con la unidimensionalidad es el de *independencia local* que significa que las habilidades o rasgos especificados en el modelo son los únicos factores que influyen en las respuestas de los examinados a los ítem del test. Este conjunto de habilidades representa el *espacio latente completo*.

Existen modelos de uno, dos o tres parámetros que se utilizan según las características del diseño. El modelo de un parámetro *b* llamado *dificultad o umbral del ítem* (Rash, 1960), describe la probabilidad de adherir a un determinado ítem y refleja el punto de inflexión de la curva en el cual la probabilidad de adherir al ítem o de rechazarlo, es .5. El parámetro indica el punto de transición entre la probabilidad de rechazar el ítem como *no autodescriptivo* a la de adherir al ítem como *autodescriptivo*, con valores bajos que indican alta adhesión o ítem no dificultosos.

El modelo de dos parámetros muestra que la probabilidad de que un respondiente seleccionado al azar adhiera al ítem, es una función del *parámetro umbral (b)* y del *parámetro a* que refleja el poder discriminativo del ítem. Los índices de discriminación del ítem muestran cuán fuertemente se relaciona un ítem particular al atributo subyacente, correspondiendo las inclinaciones más empinadas (las más grandes y positivas) a mayor diferenciación en determinados punto o puntos a lo largo de la dimensión del atributo que las de inclinación menos empinada. Los dos parámetros se relacionan con el concepto psicométrico estándar de dificultad (Panter, Swygert, Dahlstrom & Tanaka, 1997).

El modelo de tres parámetros agrega a los otros dos, un tercer *parámetro c*, llamado *de la asíntota más baja* y describe la probabilidad distinta de 0 de que respondientes con bajos niveles en el constructo subyacente puedan adherir a un ítem como verdadero o autodescriptivo.

Estudio de ítem dicotómicos y ordinales

Los parámetros *discriminación* y *dificultad del ítem* determinan conjuntamente cuán bien funciona el ítem (Reise & Henson, 2003).

Steinberg y Thissen (1995) realizaron dos análisis separados a través del modelo TRI de dos parámetros, de las dimensiones *Acción* y *Pensamiento* de orientación estado acción, usando datos de la Escala de Acción de Kuhl (1985). Los *parámetros de discriminatividad (a)* mostraron que algunos ítem estaban más relacionados con el constructo subyacente que otros y que los *parámetros umbral (b)* en la dimensión *Acción* también diferían a través de los ítem. El examen de las curvas correspondientes a cada ítem permitió a los investigadores determinar cuáles eran los ítem más útiles para la medición de cada dimensión y cuáles podían ser eliminados sin disminuir la precisión de la medición.

Conjuntamente con el análisis factorial exploratorio de ítem dicotómicos u ordinales, que determina la dimensionalidad subyacente del conjunto de ítem, la TRI y el análisis factorial confirmatorio son útiles en la construcción y validación de inventarios de personalidad. Ambos modelos proveen información sobre los parámetros de los ítem y sobre la estimación de los rasgos para cada respondiente. El supuesto de unidimensionalidad es restrictivo y puede producir problemas, pero las escalas pueden descomponerse en escalas unidimensionales más pequeñas antes del análisis o los ítem pueden ser reagrupados en *testlets* (Wainer & Kiely, 1987) que son grupos de ítem que se suponen unidimensionales (Panter et al., 1997).

Otra importante función de la TRI es el diseño de índices de lo que se considera una respuesta apropiada. En la evaluación de la validez de escalas de personalidad, se han incorporado estos índices para identificar protocolos que pueden ser caracterizados como aberrantes, cuando se responde a un conjunto de ítem. Las respuestas se consideran aberrantes o inapropiadas si sujetos con bajos niveles en el atributo responden a ítem que indican altos niveles en el mismo, o inversamente, si sujetos con altos niveles en el atributo no pueden responder a ítem bajos en el mismo (Reise & Waller, 1993; Swygert, Panter, Dahlstrom & Reise, 1996). Los *índices Iz* (Dragow, Levine & Williams, 1985) pueden emplearse para identificar valores inapropiados en un inventario de personalidad con ítem dicotómicos, en la medida que se satisfagan los supuestos de la TRI (Birenbaum, 1985).

En un estudio con el MMPI-2 (Swygert et al., 1996) fue difícil diferenciar con las escalas de validez, entre respondientes con altos niveles de psicopatología y los que tradicionalmente se describen como *inconsistentes*. La inconsistencia en una escala predijo la respuesta inconsistente en otra, resultando que el valor obtenido en la escala *Esquizofrenia (Sc)* era el mejor

Richaud

predictor del *valor Iz* del respondente. En este caso, los índices de lo que se considera apropiado fueron útiles en el estudio de la validez de las respuestas al ítem, al mismo tiempo que subrayaron que las respuestas no válidas en una escala de un inventario de personalidad tenían probabilidad de acompañarse de respuestas no válidas en otras escalas.

Los parámetros de los ítem de la TRI se basan en una *propiedad de invariancia* que permite aplicaciones importantes como la *función diferencial del ítem (Differential Item Function)*.

El modelo TRI supone una estructura subyacente unidimensional cuyos indicadores son los ítem. Este modelo permite poner a prueba que aspectos del modelo subyacente son invariantes a través de distintas poblaciones tales como género, edad o estatus clínico.

En el caso del funcionamiento diferencial del ítem, los respondentes provenientes de dos grupos tienen diferentes probabilidades de contestar al ítem, aun cuando los dos grupos estén en el mismo punto del atributo subyacente. Por ejemplo, mujeres y hombres con el mismo nivel de *Depresión estimada* deberían tener la misma probabilidad de responder de la misma manera a un ítem en particular. Si las probabilidades difieren por género, entonces el ítem muestra sesgo.

Thissen, Steinberg y Gerrard (1986) llevaron a cabo un estudio con el *Forced Choice Sex Guilt Inventory*. Los resultados indicaron que las mujeres y los hombres no solo obtenían medias diferentes en el inventario, sino que además algunos ítem tenían diferentes umbrales según sexo. Steinberg (1994) examinó la presencia de la función diferencial del ítem (DIF) debido a los efectos contextuales en la *Trait Anxiety Scale of the State-Trait Anxiety Inventory* (Spielberger, Gorsuch & Lushene, 1979) manipulando experimentalmente el orden del ítem y evaluando si las pendientes de los ítem (poder discriminativo) permanecían constantes a través de diferentes condiciones (Panter et al., 1997).

La TRI también fue aplicada para desarrollar el análisis factorial de información completa. Este método de análisis factorial desarrollado por Schilling y Bock (Bock & Schilling, 1997) trabaja directamente sobre los patrones de respuesta y evita los artefactos asociados a los coeficientes *phi* y tetracórico (McLeod, Swygert & Thissen, 2001; Swygert, McLeod & Thissen, 2001).

Leonelli, Chang, Bock y Schilling (2000) realizaron un análisis factorial exploratorio de información completa de ítem dicotómicos sobre la muestra normativa del MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989). Compararon los resultados del mismo con los de Johnson, Butcher, Null y Johnson (1984), quienes habían realizado un análisis de

Estudio de ítem dicotómicos y ordinales

componentes principales de un conjunto de 550 ítem de la versión previa del MMPI (Hathaway & McKinley, 1943). Aparte de los cambios de versión y las diferencias de muestra las diferencias esenciales entre los dos estudios se atribuyeron a la utilización del método de Schilling y Bock, ya que este último utiliza toda la información de las respuestas a los ítem, mientras que el análisis de componentes principales utiliza información parcial contenida en los coeficientes de correlación por pasos. El análisis de información completa retuvo todos los ítem incluidos inicialmente que se distribuyeron en 10 factores definidos con precisión y fáciles de interpretar.

En los últimos años se desarrollaron nuevos índices de bondad de ajuste para modelos unidimensionales de TRI con ítem dicotómicos. Orlando y Thissen (2000, 2003) propusieron agrupar a los examinados de acuerdo al número de puntajes correctos y luego desarrollaron una variación con frecuencias observadas basadas en el uso de expectativas posteriores (Stone, 2000a; Stone, 2000b; Stone, Mislevy & Mazzeo, 1994; Stone & Zhang, 2003). Además de tests de *ji* cuadrado, se propusieron otros métodos para evaluar el ajuste, como el basado en el Multiplicador de Lagrange (Glas, 1998, 1999, 2001; Glas & Falcon, 2003). También parece una técnica efectiva comparar datos observados con datos replicados basados en la estimación de parámetros de TRI (Hambleton & Han, 2004; Sinharay, 2004). Recientemente se desarrollaron procedimientos no paramétricos para evaluar la bondad de ajuste, la mayoría de los cuales comparan la Curva Característica del Ítem (ICC) estimada en forma no paramétrica con ICC basadas en el modelo paramétrico de interés (Douglas & Cohen, 2001; Ponocny, 2001, Wells & Bolt, 2004).

Los últimos desarrollos son los realizados por Lu y Lin (2005) que examinan la plausibilidad de la distribución de frecuencias de la habilidad de grupos de examinados, obteniendo cada valor posible del ítem y también los estudiados por Sinharay (2005) que emplea el método de control del modelo predictivo posterior (modelo bayesiano) (Rubin, 1984) realizando un gráfico del ítem en el que compara las proporciones de puntajes correctos observados y predichos de examinados con puntajes brutos diferentes.

Chernyshenko, Stark, Chan, Drasgow y Williams (2001) hicieron un estudio comparando el ajuste de varios modelos de TRI a dos instrumentos de evaluación de la personalidad. Se analizaron los datos de 13.059 individuos que respondieron a la versión de inglés estadounidense de la quinta edición del Cuestionario 16 PF de Personalidad (Cattell, 1995) y de 1.770 individuos que respondieron a los 50 ítem del *Big Five Personality* de Goldberg (1992). Se consideraron varios aspectos relacionados con el ajuste a los modelos TRI. Se examinaron dos de los modelos paramétricos más

Richaud

populares diseñados para ítem dicotómicos (los modelos logísticos de dos y tres parámetros) y un modelo paramétrico para ítem politómicos (el modelo de respuesta graduada de Samejima). También se examinaron las fórmulas no paramétricas de máxima verosimilitud para datos dicotómicos y politómicos que habían probado proporcionar buenos ajustes a diversos tests de habilidades cognitivas (Drasgow, Levine, Tsien, Williams & Mead, 1995). Los modelos logísticos de dos y tres parámetros ajustaron razonablemente bien a algunas escalas pero no a otras y el modelo de respuesta graduada no ajustó bien. Las fórmulas no paramétricas de modelos de puntaje alcanzaron el mayor ajuste a los modelos considerados.

Para concluir se presentó una revisión de los métodos existentes para resolver el problema del estudio de los ítem de las pruebas, especialmente en las de personalidad, que plantean cuestiones especiales en comparación con las de inteligencia y educación y se revisó con especial énfasis en el caso del empleo del análisis factorial con variables no intervalares, que es común en el trabajo de construcción de tests.

Los modelos de análisis factorial de ítem dicotómicos u ordinales fueron comparados con otros modelos psicométricos para datos dicotómicos.

La interpretación de los dos parámetros de la TRI: dificultad o umbral del ítem y poder discriminativo, en el marco del análisis factorial general indica que si se hipotetiza más de un atributo subyacente, los pesajes factoriales obtenidos por cada ítem reflejan el grado en que cada ítem mide cada uno de los factores subyacentes. Si se aplica el criterio de la estructura simple de Thurstone, se supone que aunque se obtenga una solución multifactorial, los ítem tenderán a pesarse en un solo factor. Los pesajes de los ítem en el factor pueden considerarse semejantes a los poderes discriminativos de los ítem para cada uno de los atributos subyacentes.

Al mismo tiempo, el parámetro dificultad del ítem de la TRI puede verse como un mapeo del atributo subyacente a los datos observados. Es decir que los modelos de dos parámetros de la TRI pueden considerarse como casos especiales unidimensionales de los modelos de análisis factorial. Esta forma de pensar los modelos de la TRI como una aplicación específica de la lógica del análisis factorial de ítem dicotómicos u ordinales puede ayudar a ver que no hay una diferencia sustancial entre los investigadores que trabajan con análisis factorial y aquellos que lo hacen con la TRI (Panter et al., 1997).

Finalmente, es necesario tener en cuenta las implicaciones de los modelos conceptuales acerca de la evaluación de la inteligencia o la personalidad y su impacto en la forma en que se tratan estadísticamente los datos obtenidos, respetando los supuestos que subyacen a las estadísticas utiliza-

Estudio de ítem dicotómicos y ordinales

das. Por ejemplo, si un investigador cree que un atributo particular está distribuido en forma continua, las opciones de respuesta que se le presentan al sujeto deben tratar de representar el escalamiento ordinal de un determinado ítem en un continuo subyacente. Sin embargo, siguiendo otro modelo conceptual del mismo fenómeno, el investigador puede estar interesado solamente en identificar la ausencia de una determinada observación en un estudio de caso control apareado. Mientras el modelo de evaluación se ha construido en términos de datos ordinales, el nuevo modelo conceptual está basado en una dicotomía. En este caso, el modelo de medición deberá ser consistente con el modelo conceptual y establecer un punto de corte basado en criterios externos al de los datos categóricos ordenados (Panter et al., 1997).

Es decir, no hay métodos ni modelos estándar que se apliquen siempre y en todos los casos si realmente se quieren obtener evaluaciones y mediciones válidas (Richaud de Minzi, 2005).

Referencias bibliográficas

- Birenbaum, M. (1985). Comparing the effectiveness of several IRT based appropriateness measures in detecting unusual response patterns. *Educational and Psychological Measurement, 45*, 523-533.
- Buja, A. & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509-540.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2*. Minneapolis: University of Minnesota Press.
- Bock, R.D. & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters. Application of an EM algorithm. *Psychometrika, 46*, 443-459.
- Bock, R.D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179-197.
- Bock, R.D. & Schilling, S. (1997). High-dimensional full-information item factor analysis. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (pp. 163-176). New York: Springer-Verlag.

Richaud

- Cattell, R.B. (1995). *Sixteen Personality Assessment*. Institute for Personality and Ability Testing, Inc.
- Chernyshenko, O.S., Stark, S., Chan, K.Y., Drasgow, F. & Williams, B. (2001). Fitting Item Response Theory Models to Two Personality Inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562.
- Christoffersson, A. (1978). Two-step weighed least squares factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- Drasgow, F., Levine, M.V., Tsien, S., Williams, B. & Mead, A. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19, 143-165.
- Drasgow, F., Levine, M. & Williams, E.A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67-86.
- Douglas, J. & Cohen, A.S. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234-243.
- Glas, C.A.W. (1998). Detection of differential item functioning using lagrange multiplier tests. *Statistica Sinica*, 8(1), 647-667.
- Glas, C.A.W. (1999). Modification indices for the 2-pl and the nominal response model. *Psychometrika*, 64, 273-294.
- Glas, C.A.W. (2001). Differential item functioning depending on general covariates. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 131-148). NY: Springer.
- Glas, C.A.W. & Falcon, J.C.S. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement*, 27(2), 87-106.
- Goldberg, L.R. (1992). The development of markers for the Big-Five Factor structure. *Psychological Assessment*, 4, 26-42.
- Hambleton, R.K. & Han, N. (2004, April). *Assessing the fit of IRT models: Some approaches and graphical displays*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.
- Hathaway, S.R. & McKinley, J.C. (1943). *The Minnesota Multiphasic Personality Schedule (revised)*. Minneapolis: University of Minnesota Press.

Estudio de ítem dicotómicos y ordinales

- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Johnson, J.H., Butcher, J.N., Null, C.H. & Johnson, K.N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology*, 47, 105-114.
- Kuhl, J. (1985). Volitional mediators of cognition-behavioral consistency: Self-regulatory processes and action versus state orientation. In J. Kuhl & J. Beckman (Eds.), *Action control: From cognition to behavior* (pp.101-128). Berlin: Springer-Verlag.
- Leonelli, B.T., Chang, C.H., Bock, R.D. & Schilling, S.G. (2000). Interpretation of a full-information item-level factor analysis of the MMPI-2: Normative sampling and nonpathognomonic descriptors. *Journal of Personality Assessment*, 74(3), 400-422.
- Lu, Y. & Lin, S. (2005). *Assessing fit of Item Response Theory Models*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal.
- McLeod, L.D., Swygert, K.A. & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 189-216). Mahwah, NJ: Lawrence Erlbaum.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.
- Orlando, M. & Thissen, D. (2000). Likelihood-based item-fit indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 24(1), 50-64.
- Orlando, M. & Thissen, D. (2003). Further investigation of the performance of 2 S X -: An item fit index for use with dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 27(4), 289-298.
- Panter, A.T., Swygert, K.A., Dahlstrom, W.G. & Tanaka, J.S. (1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment*, 68, 561-589.
- Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rash model. *Psychometrika*, 66, 437-460.
- Rash, G. (1960). *Probabilistic models for intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Richaud

- Reise, S.P. & Henson, J.M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81(2), 93-103.
- Reise, S.P. & Waller, N.G. (1993). Traitendness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology*, 65, 143-151.
- Richaud de Minzi, M.C. (2005). Articulación de la teoría psicológica y la teoría psicométrica [Articulation between psychologic theory and psychometric theory]. *Suma Psicológica*, 12(1), 7-21.
- Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, 12, 1151-1172.
- Sinharay, S. (2004). *Practical applications of posterior predictive model checking for assessing fit of unidimensional item response theory models*. ETS Paper presented at American Educational Research Association and the National Council on Measurement in Education. Proprietary work of ETS. Extraído de la World Wide Web el 6 de Junio, 2005: <http://www.ets.org/research/dload/NCME>
- Sinharay, S. (2005). *Bayesian Item Fit Analysis for Unidimensional Item Response Theory Models*. Unpublished manuscript. Proprietary work of ETS. Extraído de la World Wide Web el 6 de Junio, 2005: <http://www.ets.org/research/dload/NCME>
- Spielberger, C.D., Gorsuch, R.L. & Lushene, R.E. (1979). *Manual for the STAI*. Palo Alto, CA: Consulting Psychologists Press.
- Steinberg, L. (1994). Context and serial-order effects in personality measurement: Limits on the generality of measuring changes the measure. *Journal of Personality and Social Psychology*, 66, 341-349.
- Steinberg, L. & Thissen, D. (1995). Item response theory in personality research. In P.E. Shrout & S. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske*, (pp. 161-181). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stone, C.A. (2000a). Monte Carlo based null distribution for an alternative goodness-of fit test statistic in IRT models. *Journal of Educational Measurement*, 37(1), 58-75.
- Stone, C.A. (2000b, April). *Empirical power and Type I error rates for goodness-of-fit statistic based on posterior expectations and resampling-based inference*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Estudio de ítem dicotómicos y ordinales

- Stone, C.A. (2004). IRTFIT-RESAMPLE: A computer program for assessing goodness of fit of Item Response Theory models based on posterior expectations. *Applied Psychological Measurement*, 28(2), 143-144.
- Stone, C.A., Mislevy, R.J., & Mazzeo, J. (1994, April). *Classification error and goodness-of-fit in IRT models*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Stone, C.A. & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, 40(4), 331-352.
- Swygert, K.A., McLeod, L.D. & Thissen, D. (2001). Factor analysis for items or testlets scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 217-249). Mahwah, NJ: Lawrence Erlbaum.
- Swygert, K.S., Panter, A.T., Dahlstrom, W.G. & Reise, S. (1996). *The use of appropriateness indices in the MMPI-2*. Chapel Hill: University of North Carolina, L.L. Thurstone Psychometric Laboratory.
- Thissen, D., Steinberg, L. & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99, 118-128.
- Wainer, H. & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185-201.

*Centro Interdisciplinario de Investigaciones en
Psicología Matemática y Experimental (CIIPME)
Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET)
Tte. Gral. Perón 2158
(C1040AAH) Buenos Aires
República Argentina*

Fecha de recepción: 27 de abril de 2005
Fecha de aceptación: 24 de mayo de 2005