Anuario de Psicología
1994, nº *62*, 55-65
© 1994, Facultat de Psicologia
Universitat de Barcelona

Saturaciones factoriales e índices de discriminación en la Teoría clásica del test y en la Teoría de respuesta a los ítems

Pere Joan Ferrando
Universidad Rovira i Virgili

Se presenta un estudio relacional que trata de mostrar en forma integrada las equivalencias existentes entre las saturaciones en el modelo de un factor común, los índices de discriminación en el modelo de dos parámetros en la Teoría de Respuesta a los Ítems y las correlaciones ítem-total en el análisis de reactivos basado en el modelo clásico. Las relaciones obtenidas se ilustran con un ejemplo numérico.
Palabras clave: *Análisis factorial, teoría de respuesta a los ítems, análisis de reactivos, unidimensionalidad, fiabilidad.*

The aim of this paper is to show the relationships existent among the loadings in the one common factor model, the discrimination indexes in the two parameter Item Response Model and the item-total (biserial and point biserial) correlations in the classical item analysis. A numerical study which illustrate the equivalences is also presented.
Key words: *Factor analysis, Item Response Theory, Item Analysis, Unidimensionality, Reliability.*

Una de las etapas fundamentales en la elaboración de una escala psicométrica es la que habitualmente se denomina «análisis de ítems». En dicha etapa, utilizando una serie de indicadores, se evalúa el comportamiento de cada uno de los ítems o reactivos elaborados en etapas anteriores. Los resultados de dicha evaluación se utilizan para seleccionar aquel conjunto de ítems que formará la escala definitiva.

Si un psicólogo pretende actualmente construir un instrumento psicométrico, se encontrará con que existen diversos modelos teóricos para llevar a cabo el análisis de ítems. En términos muy generales, los principales índices usados en la evaluación son los mismos: discriminación y dificultad. Sin embargo, los conceptos y definiciones de tales índices pueden variar bastante de un modelo

Dirección del autor: Pere Joan Ferrando. Departamento de Psicología, Facultad de Educación y Psicología. Universidad Rovira i Virgili. Pza. Imperial Tárraco 1. 43005 Tarragona.

teórico a otro. Desde un punto de vista aplicado, el conocimiento de las relaciones entre los distintos enfoques dados a estos índices es importante por dos razones: 1) para elegir el modelo más apropiado en cada caso y 2) para evitar la utilización de información redundante.

Las relaciones entre los parámetros utilizados en los modelos de la teoría de respuesta a los ítems (TRI), el modelo de análisis factorial (AF) y los indicadores habituales en el análisis de reactivos basado en la teoría clásica del test (TCT), han sido tratadas repetidamente a lo largo del desarrollo de los primeros. Actualmente, el tratamiento más exhaustivo del tema corresponde a modelos basados en ítems binarios y un único rasgo latente y puede hallarse expuesto formalmente en Lord y Novick (1968) o en Takane y de Leeuw (1988). Exposiciones más conceptuales pueden encontrarse en Lord (1980) o en McDonald (1985) entre otros.

A pesar de que las relaciones entre los tres modelos se encuentran firmemente establecidas, lo cierto es que, desde la perspectiva aplicada, los resultados proporcionados por algunos procedimientos de análisis basados en ellos tienden a tratarse como indicadores totalmente distintos, lo que, en muchos casos, se traduce en exposiciones redundantes. Esta situación puede deberse a distintas causas, siendo quizá la más inmediata la distinta terminología utilizada, pero puede pensarse también en otras razones. En primer lugar, los estudios relacionales tienden al análisis de los modelos estructurales, lo que no se traduce directamente en una equivalencia a nivel de resultados. En segundo lugar no abundan los tratamientos conjuntos de los tres modelos sino que las comparaciones tienden a hacerse dos a dos.

En el presente trabajo se pretende exponer de una forma integrada las relaciones entre los indicadores de discriminación utilizados en los modelos de TRI, AF y TCT, en el caso particular de reactivos dicotómicos y un solo rasgo latente que se supone normalmente distribuido. A pesar de las restricciones, probablemente este caso sea aún el más usual en investigación psicométrica aplicada. Más que en un tratamiento formal de los modelos, se pretende incidir en un estudio de las equivalencias a nivel de indicadores que pueda resultar de utilidad en análisis de reactivos o construcción de escalas.

Modelo de dos parámetros y modelo de factor común

Consideraremos, en primer lugar, al modelo de ojiva normal bajo ciertas condiciones restrictivas. Sea «zy» una variable de respuesta en un ítem determinado «j» con distribución N (0,1). Dicha variable se halla influida linealmente por una variable latente (factor) «θ», también con distribución N (0,1). Para un sujeto «i» la expresión del modelo es:

$$(1) \qquad zy_{ij} = \beta_j \, \theta_i + e_{ij}; \; zy'_{ij} = \beta_j \, \theta_i$$

Donde «β» es el coeficiente de regresión estandarizado, que, en el caso de un sólo regresor, equivale a la correlación producto-momento entre «zy» y «θ».

Asumiendo los supuestos de independencia de los errores y homocedasticidad, la desviación típica de los errores en torno a la predicción, «s_e», se obtendrá mediante:

$$s_e = \sqrt{1 - \beta^2}$$

Supondremos ahora que «zy» es artificialmente dicotomizada en un nivel crítico «z_c», de forma que, si $z_y < z_c$, se asigna al ítem una puntuación de 1 y, en otro caso, de 0. Bajo los supuestos habituales en el modelo de regresión lineal, la probabilidad de obtener una puntuación de 1 en el ítem, condicionada a un determinado nivel en la variable latente, viene dada, después de hacer el cambio de variable: $z_e = (zy - zy')/s_e$, por la expresión siguiente.

$$(2) \qquad P/\theta = \int_{zp}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-1/2\, z_e^2}\, d\, z_e = \int_{zp}^{\infty} \varphi\,(z_e)\, dz_e$$

Donde

$$zp = \frac{z_c - zy'}{s_e} = \frac{z_c - \beta\,\theta}{\sqrt{1 - \beta^2}}$$

Si definimos ahora el índice de dificultad «b» como el nivel en la variable latente para el cual la probabilidad de acierto en el ítem es de 0.5, de (1) y del supuesto de distribución normal de los errores se deduce:

$$(3) \qquad z_c = \beta\, b; \qquad b = \frac{z_c}{\beta}$$

Definiendo ahora la variable «a» como:

$$(4) \qquad a = \frac{\beta}{\sqrt{1 - \beta^2}}$$

Es demostrable que: $zp = a\,(b - \theta)$ y, por tanto:

$$(5) \qquad \int_{zp}^{\infty} \varphi\,(z_e)\, dz_e = \int_{a(b-\theta)}^{\infty} \varphi\,(z_e)\, dz_e = \int_{-\infty}^{a(\theta-b)} \varphi\,(z_e)\, dz_e$$

La última expresión es la más conocida para el denominado modelo de ojiva normal. Representando en un gráfico la proporción de aciertos (probabilidad) correspondiente a cada nivel de «θ», se obtendría la curva característica del ítem (CCI) en forma de ojiva.

De la exposición hecha hasta ahora, cabe notar que, cuanto mayor sea el parámetro que hemos denominado «a», mayor será la rapidez con que se incrementa la probabilidad de acierto al aumentar el nivel en el rasgo latente. Como es sabido, «a» es el índice de discriminación en este modelo y, en forma conceptual, lo podríamos definir como la pendiente de la CCI entre los puntos de inflexión de dicha curva (Lord y Novick, 1968).

En suma, bajo la restricción de que la variable latente tenga una distribución normal, las relaciones entre el índice de discriminación correspondiente a un ítem en el modelo normal de dos parámetros «a» y la saturación factorial de dicho ítem «β», se encuentra determinada totalmente por la expresión (4).

Considerando ahora que el modelo logístico de dos parámetros constituye una aproximación al modelo normal, es inmediato que las relaciones hasta ahora expuestas podrán generalizarse al modelo logístico en la medida en que este último se aproxime al modelo normal. Como es sabido, los valores de probabilidad que proporcionan ambos modelos difieren uniformemente en menos de 0.01 (Birnbaum, 1968, p. 399).

Las relaciones, aparentemente bastante inmediatas, entre ambos modelos inducen a pensar que una posible vía para estimar el índice «a» en un modelo de TRI sería a partir de las saturaciones factoriales. De hecho, como se tratará de mostrar más adelante, si se dispusiese de las puntuaciones en las supuestas variables continuas de respuesta, bastaría ajustar un modelo de un factor común a la matriz de correlación entre dichas variables para obtener las saturaciones. Como es bien sabido la expresión general para estimar las saturaciones en este modelo sería:

$$(6) \qquad r_{kj} = \beta_k \, \beta_j$$

Donde «r_{kj}» sería la correlación habitual entre los continuos «k»,«j», siendo «β_k» y «β_j» las saturaciones de «k» y «j» respectivamente en el factor común. Esta correlación es desconocida, ya que sólo se dispone de las puntuaciones binarias. Su estimador máximo-verosímil, bajo el supuesto de normalidad bivariante es la correlación tetracórica (Elderton, 1953), de forma que, si se pretenden estimar las saturaciones, deberá factorizarse la matriz de r-tetracóricas. Tal análisis presenta una serie de problemas bien conocidos que no trataremos aquí. En este trabajo, basta decir que, si se utiliza directamente el coeficiente producto-momento (phi), en lugar de r-tetracórica, el valor de correlación se verá atenuado en la medida en que las proporciones de acierto-error para cada par de ítems sean distintas y, como consecuencia de (6) se obtendrán estimadores de saturación también atenuados.

Otro punto de partida para tratar de estimar «a» basándose en la equivalencia mostrada en (4) sería el siguiente. Asumiendo que la variable latente de respuesta fuese conocida, entonces la correlación entre los continuos «zy» y «θ» (es decir, «β») se estimaría mediante el coeficiente de correlación biserial «r_b» entre las puntuaciones dicotomizadas y la variable latente «θ». Nótese entonces que en (6), una correlación tetracórica «r_{kj}» se obtiene como el producto de dos coeficientes biseriales.

En la literatura psicométrica, es frecuente, en virtud de esta relación, sustituir la saturación factorial «β» por la correlación biserial entre el ítem y el total de la escala en (4), sin embargo esta equivalencia no es tan directa. Nótese de nuevo que «β» equivale a la correlación biserial entre el ítem y el rasgo latente, lo que no es lo mismo que la correlación biserial entre el ítem y el total de la escala, (que es tal como se computa habitualmente el índice de discriminación en TCT). En primer lugar, el total de la escala estima con error el nivel en el rasgo latente. En segundo lugar, aunque la puntuación verdadera fuese conocida, si bien es cierto que dicha puntuación verdadera y el rasgo latente tienen una relación uno a uno (Birnbaum, 1968; Lord, 1980), también lo es que la relación entre ambos no es lineal (Birnbaum, 1968). En suma, como adecuadamente señala Urry (1974) la correlación ítem-test depende del test, mientras que la correlación entre el ítem y el rasgo latente es generalizable.

De todo lo anterior se deriva que, en caso de sustituir la saturación factorial por el índice de discriminación biserial en (4), la relación con «a» ya no será igual sino tan sólo aproximadamente igual. La aproximación, por otra parte, dependerá de la relación que exista entre las saturaciones factoriales y el coeficiente «r_b», relación que se trata en el siguiente apartado.

Cabe decir, por último, que la estimación del índice de discriminación en el modelo de dos parámetros basada en el análisis factorial de la matriz de correlaciones tetracóricas o en los coeficientes biseriales ítem-total, suele recibir el nombre de estimación heurística o aproximada. Urry (1974, 1976) es quizás el principal proponente de este enfoque, habiendo desarrollado el programa ANCILLES basado en las equivalencias (3) y (4).

Saturaciones factoriales y correlación ítem-total

Las relaciones existentes entre el coeficiente biserial ítem-total y las saturaciones factoriales, fueron tratadas en un conciso y claro trabajo de Henrysson (1962). Este autor se basó en las saturaciones obtenidas al aplicar el método centroide, caso en que las relaciones se hacen muy directas, aunque, como tratará de mostrarse, no tanto como concluye el citado autor.

Como es bien sabido, la expresión para estimar la saturación factorial de una variable «k» en el centroide (p. ej. Comrey, 1985) es la siguiente:

$$(7) \qquad a_k = \frac{\sum_j^n r_{kj}}{\sqrt{\sum_k \sum_j r_{kj}}}$$

Considérese ahora «n» ítems tipificados y el total de la escala que se obtendrá mediante la suma simple de estos ítems. Nótese que el numerador de (7) es la suma de los elementos de la columna «k» de la matriz de correlaciones inter-ítem, por tanto, dicho numerador es la covarianza entre el ítem «k» y el total de la escala. Por otra parte, el denominador de (7) es la raíz de la suma de todos los elementos de la matriz de correlación inter-ítem y, por tanto, es la desviación típica del total de la escala. Dado que «k» se considera tipificado, el cociente entre la covarianza ítem total y la desviación típica del total es el habitual coeficiente producto-momento ítem-total. Si el ítem es binario, cabe concluir que la saturación estimada por el centroide equivale exactamente al coeficiente biserial-puntual ítem-total.

Sin embargo, cabe hacer aquí una observación. Cuando se calcula la correlación ítem total a partir de la matriz de correlación inter-ítem, los elementos de la diagonal principal de esta matriz son unos. En cambio, cuando se estima la saturación centroide, en la diagonal principal se colocan estimadores de comunalidad. Por tanto, los dos indicadores tan sólo serán aproximadamente iguales; por supuesto, más similares cuanto mayores sean las comunalidades.

En el estado actual de desarrollo del AF, la exposición anterior sólo puede tener el interés didáctico que posee el método centroide. Si pretenden buscarse las relaciones con métodos más sofisticados, tales relaciones dejan de ser tan directas.

Considérense dos métodos factoriales relativamente simples y ampliamente utilizados, como pueden ser el de ejes factoriales (PAF) y el de residuo mínimo (MINRES). En ambos casos los vectores de saturación se obtienen bajo el criterio de minimizar la suma cuadrática de las diferencias entre los elementos de la matriz de correlación observada y los correspondientes elementos de la matriz reproducida por el modelo. La única diferencia entre ambos métodos es la de que en PAF se minimizan tanto las diferencias no diagonales como las diagonales, mientras que en el método de residual mínimo se minimizan tan solo los elementos no diagonales.

Teniendo en cuenta todos los elementos residuales (como hace PAF) puede derivarse la expresión siguiente para estimar la saturación de una variable «k» en el caso de un solo factor:

$$(8) \qquad a_k = \frac{\sum_{j=1}^{n} r_{kj}\, a_j}{\sum_{j}^{n} a_j^2}$$

Esta expresión guarda cierta semejanza con (7). Aunque no creo que pueda obtenerse una equivalencia algebraica más simple entre los estimadores centroide y PAF, conceptualmente puede ser de ayuda una interpretación geométrica.

Si se representan las variables y el factor como vectores unitarios en el espacio generado por los sujetos, el factor centroide se sitúa sobre el eje que hace mínima la suma de las distancias en valor absoluto entre cada variable y dicho eje, mientras que el factor PAF se sitúa sobre el eje que hace mínima la suma de las distancias al cuadrado. Claramente las saturaciones (proyecciones) deberán ser bastante similares en ambos métodos.

En suma, puede decirse que, para un ítem determinado, la correlación ítem-total será similar a la saturación centroide de dicho ítem, en la medida en la que las comunalidades de todos los ítems tiendan a uno. En el caso de un sólo factor común, esto implica que, cuanto más altas sean las correlaciones ítem total, más altas serán también las saturaciones y más se parecerán ambos indicadores. Esta relación se mantendrá también, aunque no en forma tan directa, para métodos factoriales más sofisticados debido a que la solución centroide es una aproximación a las estimaciones que proporcionan dichos métodos.

Por último debe hacerse notar que en el apartado anterior se han analizado las equivalencias con el estimador biserial, mientras que en este apartado se ha tratado el coeficiente directo producto-momento o biserial puntual «r_{bp}». La equivalencia entre ambos indicadores es bien conocida y viene dada por la expresión:

$$(9) \qquad r_b = r_{bp} \frac{\sqrt{pq}}{y}$$

Donde «p» y «q» son, respectivamente, las proporciones de acierto y error en el reactivo, siendo «y» el valor de la función de densidad normal (ordenada) en el punto de división p–q.

Un ejemplo numérico

Para ilustrar las relaciones hasta ahora expuestas se llevó a cabo un trabajo de simulación siguiendo el procedimiento que se detallará seguidamente. Debe advertirse que tal estudio es meramente un ejemplo que permita poner de relieve las relaciones antes expuestas. Sin embargo, al final del mismo se expondrán los resultados en función de algunos de los principales estudios de simulación a gran escala que han tratado el tema.

En primer lugar se creó un vector de saturaciones «a» de 5 ítems en un factor común. Seguidamente se reprodujo la matriz de correlación inter-ítem «R» mediante el teorema de Thurstone:

$$R = a\,a' + U^2$$

A continuación, mediante el programa MULTINOR elaborado en nuestro laboratorio, se generaron 5.000 observaciones (5.000 sujetos × 5 ítems) distribuidas en forma normal multivariante y con matriz de correlación R. Las puntuaciones totales para cada sujeto se obtuvieron mediante la suma simple de los 5 ítems. Se calculó la correlación entre cada ítem y el total de la escala, así como el índice de discriminación esperado en caso de dicotomización mediante la expresión (4). El análisis de los datos originales se presenta en la Tabla 1*a*.

Tras este análisis inicial los datos fueron dicotomizados mediante el procedimiento RECODE de SPSS-PC. Todos ellos se dicotomizaron en el mismo punto ($z_c = 0.25$) para evitar efectos de atenuación en las correlaciones. Sobre los datos dicotomizados se estimaron los siguientes indicadores: *a*) saturación en el centroide; *b*) saturación en PAF; *c*) correlación biserial ítem-total; *d*) correlación biserial-puntual ítem-total y *e*) índice de discriminación en el modelo logístico de dos parámetros.

Dado que el AF centroide no se encuentra disponible en ninguno de los paquetes actuales, se escribió un sencillo programa en Fortran-77 para obtener las saturaciones, basado en el procedimiento habitual descrito en Comrey (1985). El análisis PAF se llevó a cabo mediante el programa FACTOR de SPSS-PC, utilizando como estimadores iniciales de comunalidad los valores absolutos máximos de correlación y llevando a cabo 30 iteraciones de refactorización. La correlación biserial ítem-total se obtuvo mediante la rutina BIS del paquete científico SSP. La correlación biserial puntual se obtuvo mediante el programa RELIABILITY de SPSS-PC. Por último, las estimaciones del modelo de dos parámetros se obtuvieron mediante el programa MULTILOG.

Los resultados de este segundo análisis se presentan en la Tabla 1*b*.

TABLA 1. ANÁLISIS DE LOS DATOS GENERADOS MEDIANTE SIMULACIÓN

a) Análisis con datos originales

Ítem	Saturación	r ítem-total	Valor esperado i. discriminación
1	0.1	0.40	0.110
2	0.3	0.53	0.314
3	0.5	0.64	0.577
4	0.7	0.72	0.980
5	0.9	0.78	2.064

b) Análisis con datos dicotomizados

Ítem	Saturación Centroide	Saturación PAF	r bis.	r bis.p	i. discriminación
1	0.082	0.082	0.5187	0.4096	0.106
2	0.236	0.235	0.6381	0.5037	0.310
3	0.399	0.397	0.7419	0.5857	0.580
4	0.568	0.571	0.8199	0.6487	1.020
5	0.783	0.783	0.8780	0.7020	2.210

Dado que el ejemplo es puramente ilustrativo, algunos de los resultados son totalmente previsibles. Así, la similitud entre los índices de discriminación estimados por MULTILOG y los que se derivan de la solución original, tan sólo informa de que el algoritmo utilizado por el programa tiene un funcionamiento correcto, siendo capaz de recuperar razonablemente bien las soluciones originales. Para un estudio detallado de la capacidad de recuperación de este programa, puede consultarse Stone (1992).

En la misma forma, en una situación en que no existen sesgos en las distribuciones de los ítems y en la que las saturaciones son relativamente claras, también es de esperar que las soluciones centroide y PAF sean prácticamente idénticas.

De cara a la aplicación en análisis de ítems, puede ser más interesante la observación general de que todos los índices de la tabla están dando la misma información y que tienen una perfecta relación de rango. Con los datos originales, además, es posible pasar de un índice a otro mediante las transformaciones comentadas en el trabajo. Con los datos dicotomizados la relación es mucho más aproximada, pero, como se ha dicho, el ordenamiento de los ítems según su capacidad de discriminación será el mismo con cualquiera de los indicadores.

A nivel más específico puede ser interesante comentar que en el caso de datos dicotomizados, se producen notables divergencias entre las saturaciones factoriales y los índices biseriales-puntuales cuando las saturaciones son bajas, pero a medida que éstas aumentan ambos indicadores tienden a dar resultados similares. Este fenómeno ya había sido comentado pero puede ahora generalizarse más para explicar los resultados del ejemplo. Si se observa la expresión (7) podrá derivarse que, a medida que aumente el número de ítems, la saturación y la correlación ítem total tenderán a hacerse más similares si los demás factores permanecen constantes. Esto es así ya que, al añadir un nuevo ítem a los «n» existentes, se añade un valor de comunalidad y 2(n−1) valores de correlación. Cuantos más ítems, por tanto, menor es la importancia relativa de las comunalidades respecto a la suma de las columnas de R (numerador) y de todos los elementos de R (denominador).

Las discrepancias observadas en el ejemplo se explican, por tanto, por el hecho de ser muy pocos los ítems y muy bajas algunas comunalidades. En un caso real, si el test es razonablemente largo y las saturaciones tienden a ser altas, cabe esperar que las correlaciones ítem total y las saturaciones factoriales tiendan a ser muy parecidas.

Los resultados obtenidos en este pequeño estudio están en congruencia con la evidencia empírica disponible. En sus estudios de simulación iniciales para el desarrollo del program ANCILLES, Urry (1974, 1976) y Jensema (1976) advirtieron que las estimaciones respecto al índice de discriminación se hacían aceptables cuando el test era largo, cuando el coeficiente de fiabilidad era alto (es decir, las saturaciones factoriales altas) y cuando el tamaño muestral era grande. Obviamente este último punto no interfiere en el estudio actual. Respecto a los dos primeros puntos, los autores citados llegaron a la conclusión de que se producían estimaciones razonables en test de más de 80 ítems con fiabilidades superiores a .90.

Estudios de simulación posteriores han sido realizados por McKingley y Reckase (1980) y por Swaminathan y Guiford (1983) siendo este último el trabajo más completo y el más citado en la literatura sobre el tema. Comparando las estimaciones producidas mediante el método aproximado y mediante la estimación máximo-verosímil, concluyeron que se producían estimaciones razonables del índice de discriminación usando el primer método cuando la longitud del test era de 80 o más ítems y cuando el tamaño muestral superaba los 1000 casos, siendo mucho más crítica la longitud del test que el tamaño muestral (como se ha visto en el estudio aquí presentado).

Algunas consideraciones de tipo aplicado

De la exposición anterior pueden derivarse algunas consideraciones de posible interés de cara a la construcción de escalas psicométricas y al análisis de reactivos. En líneas generales, quizás la idea general que este trabajo intenta mostrar es la de que no es necesario hacer una elección categórica entre los distintos modelos teóricos que subyacen al análisis de ítems, sino que, por el contrario, los procedimientos de ellos derivados pueden utilizarse en forma efectiva en sucesivas etapas.

El proceso de análisis o de construcción de una escala compuesta por reactivos binarios puede iniciarse utilizando los indicadores de correlación ítem-total basados en el modelo clásico, tanto el biserial como el biserial-puntual. La inspección de estos índices permite eliminar o revisar reactivos inapropiados, así como sugerir cuál será el modelo de TRI más apropiado para calibrar los ítems. Si los valores de correlación son bastante similares en todos los casos, esta uniformidad sugiere la utilización del modelo más parsimonioso en el que todos los índices de discriminación son iguales (modelo de Rasch). Valores discrepantes, por el contrario, apuntan la necesidad de utilizar modelos más flexibles. Por supuesto, si el psicómetra tiene en mente construir un test basado en el modelo más parsimonioso, en esta etapa deberá seleccionar ítems que ajusten lo mejor posible a la condición exigida de igualdad en las correlaciones ítem-total.

El AF utilizado en una segunda etapa ofrece interesante información adicional a la obtenida anteriormente. Como hemos visto, el ajuste al modelo de un factor común nos dará, para cada ítem, valores de saturación bastante similares a las correlaciones ítem total; más similares cuantos más ítems y cuanto mayores las comunalidades. Adicionalmente la inspección de los valores residuales tras el ajuste del modelo nos permitirá evaluar si el conjunto preseleccionado de reactivos se comporta en forma razonablemente unidimensional.

Asimismo, en el caso de construcción antes comentado, el modelo AF nos permite evaluar si los ítems resultan apropiados para una calibración basada en el modelo de igual discriminación. Para ello, mediante algún programa como LISREL o COSAN, puede llevarse a cabo un AF confirmatorio con la restricción de que todas las saturaciones son iguales y verificar estadísticamente el ajuste de tal modelo. Mediante comparación de los datos de bondad de ajuste puede llevarse a cabo un segundo proceso de selección que conduzca finalmente a un conjunto de reactivos apto para ser calibrado mediante el modelo de TRI elegido.

REFERENCIAS

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. & Novick, M.R., *Statistical theories of mental tests scores*. Massachusetts: Addison-Wesley.
Comrey, A.L. (1985). *Manual de Análisis Factorial*. Madrid: Cátedra.
Elderton, W.P. (1953). *Frequency curves and correlation*. Washington: Harren Press.
Henrysson, S. (1962). The relation between factor loadings and biserial correlations in item analysis. *Psychometrika, 27, 4*, 419-424.
Jensema, C.S. (1976). A simple technique for estimating latent trait mental test parameters. *Educational and Psychological Measurement, 36*, 705-715.
Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: LEA.
Lord, F.M. y Novick, M.R. (1968). *Statistical theories of mental tests scores*. Massachusetts: Addison-Wesley.
McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale: LEA.
McKingley, R.L. y Reckase, M.D. (1980). *A comparison of the ANCILLES and LOGIST parameter estima-*

tion procedures for the three parameter logistic model using goodness of fit as a criterion. Research Report 80-2. Columbia: University of Missouri.

Swaminathan, H. & Guiford, J.A. (1983). Estimation of parameters in the three parameter latent trait model. In Weiss, D. (Ed.), *New horizons in testing*. New York: Academic Press.

Stone, C.A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of Multilog. *Applied Psychological Measurement, 16*, 1, 1-16.

Takane, Y. & De Leeuw, J. (1987). On the relationship between Item Response Theory and factor analysis of discretized variables. *Psychometrika, 52*, 3, 393-408.

Urry, V.W. (1974). Approximations to item parameters of mental test models and their uses. *Educational and Psychological Measurement, 34*, 253-269.

Urry, V.W. (1976). *Ancillary estimators for the item parameters of mental tests*. Washington D.C.: Personnel Research and Development Center.