

## Procedimientos gráficos para evaluar la consistencia intra-individual en el análisis factorial de ítems

Pere Joan Ferrando Piera y Fàbia Morales Vives  
Universidad Rovira i Virgili

A nivel aplicado puede resultar de gran utilidad la identificación de sujetos que no responden un test de forma consistente, especialmente en el ámbito clínico y en selección de personal. La identificación de estos patrones de respuesta evita la interpretación errónea de las puntuaciones totales de un individuo en el test. En este artículo se proponen dos procedimientos gráficos, basados en el modelo de análisis factorial lineal, que permiten establecer las posibles causas de la baja consistencia intra-individual en las puntuaciones de un sujeto, cuando su patrón de respuesta ha sido previamente detectado como inconsistente. Además, estos procedimientos muestran qué ítems han contribuido a la inconsistencia detectada. También se presentan varios ejemplos empíricos que muestran cómo interpretar los resultados obtenidos mediante estos procedimientos. Finalmente se discuten las implicaciones de los resultados en la medición de la personalidad.

*Graphical procedures for assessing person-fit in item factor analysis.* Flagging the individuals who did not answer consistently can be very useful in certain applied domains, especially in clinical and personnel selection areas. Identification of inconsistent patterns prevents erroneous interpretations of test scores. Two graphic procedures based on linear factor analysis are proposed in this paper. They allow the possible causes of low intra-individual consistency to be assessed once a pattern has been flagged as inconsistent. Moreover, these procedures allow us to identify the items that have contributed the most to the inconsistency. The procedures are illustrated with some empirical examples in personality. Lastly, implications of the results in the construction of personality measures are discussed.

El estudio de la consistencia de las respuestas de una persona a los ítems de un test tiene un claro interés aplicado, ya que un patrón de respuestas inconsistente indica que, posiblemente, no podrán hacerse inferencias válidas a partir de la puntuación obtenida por dicho individuo. En el dominio de personalidad, además, este tipo de consistencia es un tópico de interés teórico ya que el grado de consistencia se considera un indicador de la relevancia o centralidad que tiene el rasgo a medir en la estructura de personalidad del individuo (e.g. Tellegen, 1988). La generalización del uso de modelos de teoría de respuesta a los ítems (TRI) ha dado un impulso notable a la investigación del tema a partir de los años 80 (véase e.g. Meijer y Sijtsma, 1995, 2001). Además, el marco teórico de los modelos paramétricos de TRI permite una interpretación clara del concepto: la consistencia se evalúa en relación al modelo utilizado en la calibración de los ítems y a la puntuación estimada del individuo. En la terminología anglosajona este tipo de evaluación se denomina actualmente "person-fit". Aquí utilizaremos el término "consistencia intra-individual" (CII).

La evaluación de la CII es un proceso que idealmente debería proceder en dos etapas (Emons, Sijtsma y Meijer, 2004, 2005). En la primera se utilizaría un índice global, con valores referidos a una distribución teórica, para detectar a los sujetos potencialmente inconsistentes. En la segunda se utilizarían una variedad de procedimientos para evaluar: (a) el tipo de inconsistencia, (b) los ítems específicos en los que se producen las respuestas más inconsistentes, y (c) el posible impacto en la estimación del rasgo. Entre los procedimientos empleados en la segunda etapa cabe citar: los índices de CII a nivel de ítem, procedimientos gráficos, análisis de tendencias, e información cualitativa obtenida de entrevistas u observaciones.

Para los modelos TRI basados en respuestas binarias, existe actualmente una gran variedad de procedimientos e índices del tipo descrito arriba (véase e.g. Meijer y Sijtsma, 1995, 2001). En menor medida, existen también herramientas de este tipo para modelos de respuesta graduada, utilizados en la práctica para ítems con un número reducido de categorías. En los dominios de personalidad y medición de actitudes, sin embargo es bastante más común utilizar formatos de respuesta más continuos (Dawes, 1972, Ferrando, 2002, Hofstee, Ten Berge y Hendricks, 1998). En las aplicaciones, estos ítems se calibran habitualmente mediante el modelo de análisis factorial (AF) lineal, modelo que, en general resulta bastante apropiado (Ferrando, 2002). Sin embargo, la evaluación de la CII basada en el modelo AF no está tan avanzada como en el caso de los modelos TRI arriba citados. Recientemente

Ferrando (2007, 2009b) ha propuesto índices CII globales para este modelo, así como medidas a nivel de ítem e indicadores del grado de impacto sobre la estimación del rasgo. Sin embargo, los autores no tenemos noticia de que se hayan propuesto procedimientos gráficos detallados para evaluar la CII basados en el modelo AF lineal, y este es el objetivo del presente trabajo. Más específicamente, propondremos aquí dos procedimientos gráficos que permiten obtener información acerca del tipo de inconsistencia y sus posibles causas en patrones que han sido previamente detectados en la primera etapa mediante el uso de índices globales. Estos procedimientos están pensados para ítems con formato de respuesta graduada en 5 o más puntos, o con formatos más continuos, que son calibrados mediante el modelo lineal AF unidimensional.

*Revisión del Modelo y Planteamiento General*

Considérese un test formado por  $n$  ítems, con formato de respuesta (aproximadamente) continuo, que mide un rasgo  $\theta$ . El modelo lineal de respuesta para el individuo  $i$  que responde al ítem  $j$ , es:

$$X_{ij} = \mu_j + \lambda_j \theta_i + \varepsilon_{ij} \tag{1}$$

El modelo (1) es el modelo de un factor común de Spearman, que, en la literatura psicométrica, recibe el nombre de modelo para ítems congenéricos (Jöreskog, 1971, Mellenbergh, 1994). Por brevedad lo denominaremos en adelante modelo congenérico. En las aplicaciones aquí propuestas, las puntuaciones observadas  $X_{ij}$  estarán escaladas con valores entre 0 y 1, y  $\theta$  estará en escala estándar (media 0 y varianza 1). Para un nivel fijo en  $\theta$  las puntuaciones en los diferentes ítems se distribuyen independientemente (independencia local). La distribución condicional se supone normal, con media y varianza dadas por:

$$E(X_j | \theta) = \mu_j + \lambda_j \theta_i \quad ; \quad Var(X_j | \theta) = \sigma_{\varepsilon_{ij}}^2 \tag{2}$$

En la formulación (1), los parámetros correspondientes a cada ítem son: el intercepto  $\mu_j$ , la pendiente o carga factorial,  $\lambda_j$  y la varianza de error del ítem  $\sigma_{\varepsilon_{ij}}^2$ . Ferrando (2009a) ha propuesto recientemente la siguiente parametrización alternativa:

$$E(X_j | \theta) = 0.5 + \lambda_j (\theta_i - \beta_j) \tag{3}$$

con:

$$\beta_j = \frac{1 - 2\mu_j}{2\lambda_j} \tag{4}$$

En la forma (3) el modelo es más similar al modelo de dos parámetros de la TRI. Al igual que en este último, el parámetro  $\beta_j$  es el índice de dificultad, y se define como el nivel en el rasgo al que corresponde una puntuación esperada de 0,5 (el punto medio de la escala). El parámetro  $\lambda_j$  tiene la misma interpretación que en la formulación (1): es el índice de discriminación (Mellenbergh, 1994), y aquí consideraremos que sólo adopta valores positivos (es decir, que todos los ítems se puntúan en la misma dirección). En

los métodos que propondremos se hará uso de ambas formulaciones.

La estimación del modelo congenérico procede habitualmente en dos etapas. En la primera (calibración) se estiman los parámetros de los ítems y se evalúa el ajuste del modelo. En la segunda, los valores estimados se toman como fijos y conocidos y se estiman los niveles de los sujetos en el rasgo a medir (puntuaciones factoriales estimadas). En este trabajo supondremos que los ítems se han calibrado en una muestra grande y representativa que haga razonable el supuesto de parámetros conocidos. Y supondremos también que el ajuste del modelo es aceptable. En cuanto a la segunda etapa, consideraremos que, además de las puntuaciones estimadas, se han calculado también índices globales de CII para detectar a los sujetos potencialmente inconsistentes. Los métodos propuestos a continuación serán pues herramientas para analizar en detalle estos patrones detectados como inconsistentes y obtener información adicional.

*Superficie de respuesta de la persona*

Consideremos para un sujeto la puntuación factorial estimada en la segunda etapa del AF. Los métodos aquí propuestos son aplicables a cualquiera de los estimadores habituales de  $\theta$ . Sin embargo, los índices globales propuestos por Ferrando (2007) se basan en las puntuaciones factoriales de Bartlett y estas son las que consideraremos aquí preferentemente. Bajo el supuesto de normalidad condicional, la puntuación factorial de Bartlett es el estimador máximo verosímil del nivel del sujeto en el rasgo (Mellenbergh, 1994). De acuerdo con (2), la puntuación esperada para el sujeto  $i$  en el ítem  $j$  sería ahora:

$$E(X_{ij} | \hat{\theta}_i) = \mu_j + \lambda_j \hat{\theta}_i \tag{5}$$

donde  $\theta$  es la puntuación factorial estimada. Este valor se toma como fijo, y se considera una escala de “posición” ( $X_1$ ) donde se sitúan los  $n$  interceptos  $\mu_j$  ordenados de menor a mayor, y una escala de “discriminación” ( $X_2$ ) donde se sitúan los  $n$  pesos  $\lambda_j$  ordenados del mismo modo. Como función de  $\mu$  y  $\lambda$  los valores esperados en (5) definen un plano de regresión con pendiente de 1 sobre  $X_1$  y de  $\theta$  sobre  $X_2$ . Este plano puede interpretarse como el conjunto de respuestas esperadas para el sujeto  $i$  según el modelo (1) dadas las características de los ítems y el nivel estimado de dicho sujeto. Proponemos denominar a este plano como “superficie de respuesta de la persona” (SRP) esperada.

La representación propuesta es un gráfico tridimensional en el que los valores de  $\mu$  y de  $\lambda$  se sitúan en los ejes  $X_1$  y  $X_2$ . La propuesta consiste en mostrar simultáneamente la SRP esperada junto con las puntuaciones observadas en los  $n$  ítems representadas como puntos en el gráfico. La figura 1 muestra un gráfico de este tipo que se discutirá en la sección del ejemplo ilustrativo. En principio el gráfico permite una evaluación rápida del grado en que las respuestas observadas se ajustan a las esperadas, y permite identificar también aquellos ítems o conjuntos de ítems en los que se observan las respuestas más inconsistentes.

La representación gráfica propuesta en esta sección se basa en la formulación (1) del modelo congenérico. Esto se hace por dos razones. En primer lugar,  $\mu$  y  $\lambda$  en (1) son ‘a priori’ independientes. En cambio,  $\beta$ , y  $\lambda$  no lo son (véase ecuación 4). En segundo

lugar, el rango de valores de  $\mu$  y  $\lambda$  es más similar (entre 0 y 1) y esto da lugar a una representación visual más clara.

#### Curva de respuesta de la persona

En el caso de que todos los ítems tuviesen el mismo poder discriminativo ( $\lambda_j = \lambda$ ), el modelo congenérico (1) daría lugar al modelo tau-equivalente (Jöreskog, 1971). La formulación (3) muestra con claridad que, en este caso, el modelo AF se convertiría en el equivalente lineal del modelo TRI de un parámetro. Puesto que  $\lambda$  es constante, la representación tridimensional propuesta en la sección anterior podría reducirse ahora a una representación bidimensional en la que las puntuaciones fuesen tan sólo función de las posiciones de los ítems. En forma más general, este tipo de representación fue inicialmente propuesta por Trabin y Weiss (1983) con el nombre de “curva de respuesta de la persona” (CRP) que es el que utilizaremos aquí. Es una metodología que ha recibido bastante atención en el estudio de la CII (Emons, Sijtsma y Meijer, 2004, 2005). Sin embargo, hasta donde tenemos noticia, su uso se ha limitado a ítems binarios.

En la práctica no cabe esperar en general que un test de personalidad o actitud se ajuste al modelo tau-equivalente. Sin embargo, la experiencia con este tipo de ítems sugiere que la variabilidad de los índices de discriminación es relativamente reducida (Ferrando, 2009a, Levy, 1973). En el modelo tau-equivalente, si se situaran las  $n$  posiciones de los ítems ordenadas en el eje de abscisas, entonces las puntuaciones esperadas definirían una recta. En el modelo congenérico, y teniendo en cuenta la observación anterior, se esperaría que las puntuaciones formasen una nube de puntos relativamente compacta que sería bien ajustada por una recta de regresión. Definimos entonces la CRP esperada (CRP-E) como la recta de regresión de mejor ajuste (en los términos habituales de mínimos cuadrados) a la nube de puntos esperados.

En el gráfico bidimensional que se está describiendo, consideramos que la formulación (3) permite aquí una mejor interpretación que la formulación (1) propuesta en la anterior sección. Si se representan las dificultades  $\beta$  en el eje de abscisas, entonces se puede localizar en el gráfico el nivel estimado del sujeto como el valor en el eje de las  $\beta$  que corresponde a una puntuación esperada de 0,5. Además, la pendiente de la recta refleja la capacidad discriminativa media de los ítems y esto, como se discute abajo, permite obtener información adicional de interés.

En conjunto, la propuesta CRP consiste en un gráfico con las puntuaciones en el eje de ordenadas y las dificultades  $\beta$  en el eje de abscisas. En este gráfico se representa la recta CRP-E y, conjuntamente, las respuestas observadas en los  $n$  ítems representadas como puntos. Si el número de puntos es suficiente, se puede ajustar además una curva de regresión a los puntos observados. Esta curva sería la CRP observada (CRP-O). En el caso de un patrón detectado como inconsistente, la tendencia de las respuestas observadas no tiene por qué ser lineal y, por tanto, la forma de la CRP-O es desconocida. Para ajustar la línea de regresión a los puntos observados (es decir, obtener la CRP-O), se propone utilizar un procedimiento de regresión no paramétrico que no fuerce a priori ninguna forma específica de curva. Procedimientos de este tipo se encuentran implementados en los paquetes estadísticos más habituales. En nuestros trabajos utilizamos con buenos resultados un procedimiento de suavizado con función gaussiana. El estimador correspondiente al punto de evaluación  $\beta$  se obtiene como

$$\hat{m}(\beta) = \frac{\sum_{j=1}^n K\left(\frac{\beta - \beta_j}{h}\right) \hat{X}_{ij}}{\sum_{j=1}^n K\left(\frac{\beta - \beta_j}{h}\right)} \quad (6)$$

donde  $K(x)$  es la función de suavizado (en este caso gaussiana), no-negativa, continua y simétrica, que asigna valores máximos cerca del punto 0,0, y que decrece al alejarse de dicho punto en cualquier dirección. El parámetro  $h$  (amplitud de banda) lo selecciona el usuario y controla el grado de suavizado. La idea general del procedimiento es obtener una estimación no-paramétrica de la CRP-O mediante una media ponderada obtenida en cada punto de evaluación. El procedimiento se explica con detalle en Härdle (1990).

Al igual que en el gráfico SRP, las comparaciones entre la CRP-E y la nube de puntos observados pueden dar una idea general de las discrepancias entre el patrón de respuestas observado en este individuo, y el patrón predicho por el modelo. Si el test no es tau-equivalente, se perderá algo de información con respecto al gráfico SRP, pero la visualización será más simple. Por otra parte, las comparaciones entre la CRP-E y la CRP-O pueden indicar tendencias que, a su vez, pueden ser útiles para detectar formas particulares de inconsistencia. Por ejemplo, si la inconsistencia fuese debida a cierta tendencia a responder al azar, entonces se esperaría que la CRP-O tuviese menor pendiente que la CRP-E. Y en el caso de respuestas totalmente al azar, la CRP-O se esperaría que fuese plana. Sin embargo, para que la comparación entre curvas sea precisa e interpretable es necesario disponer de un número substancial de ítems (quizás un mínimo de 35 ítems). Con conjuntos mayores se podrían establecer bandas de confianza en torno a una de las dos curvas, lo que permitiría no sólo un análisis general de tendencias sino también un análisis local de discrepancias entre ambas curvas.

#### Ejemplo ilustrativo

La utilidad de los procedimientos propuestos se ilustra en esta sección mediante un test de personalidad adaptado y administrado por los autores. Se trata de una escala de Extraversión de 35 ítems que utilizan un formato de respuesta graduada en 5 puntos. A fin de controlar potenciales efectos de aquiescencia, 10 de los ítems miden en la dirección de Introversión y los 25 restantes en la dirección de Extraversión. El cuestionario se administró a un grupo de 479 estudiantes universitarios que respondieron voluntariamente. De acuerdo con los requisitos descritos antes, las respuestas se reescalaron para que tomaran valores entre 0 y 1, y, para el análisis, se puntuaron todas en la dirección de Extraversión.

En la fase de calibración se ajustó el modelo congenérico (1) mediante mínimos cuadrados simples tal como se encuentran implementados en la versión 8.3 de LISREL (Jöreskog y Sörbom, 1996). El ajuste del modelo era aceptable: los residuales se distribuían simétricamente en torno a una mediana de cero y, una vez estandarizados, la mayor parte de ellos estaban dentro del intervalo (-3,+3), con una raíz media cuadrática de 0,05. El índice de bondad de ajuste gamma-GFI tenía un valor de 0,93.

En la segunda fase, los parámetros de los ítems se consideraron fijos y, para cada sujeto, se calcularon la puntuación factorial estimada de Bartlett y el índice global de CII  $lcz$ , el cual es una es-

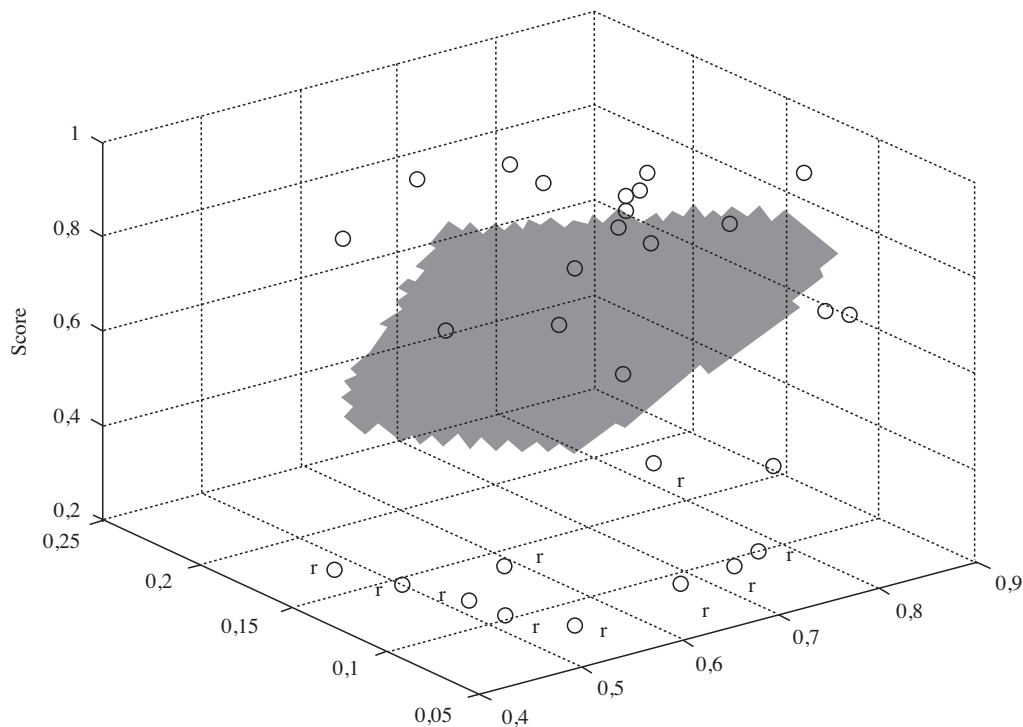


Figura 1. Gráfico SRP correspondiente al sujeto con instrucciones de respuesta aquiescente

tandarización del índice cuadrático  $lco$ , ambos propuestos por Ferrando (2007):

$$lco = \sum_{j=1}^n \frac{(X_{ij} - \mu_j - \lambda_j \hat{\theta}_i)^2}{\sigma_{\epsilon j}^2} ; lcz = \frac{lco - E(lco)}{\sqrt{Va(lco)}} \quad (7)$$

donde  $\theta$  es la puntuación factorial de Barlett. El índice refiere (aproximadamente) a la distribución normal estándar. Valores altos y positivos (digamos por encima de 2) sugieren inconsistencia en el patrón de respuestas evaluado.

Los dos primeros ejemplos corresponden a participantes que recibieron instrucciones especiales destinadas a provocar un tipo específico de respuesta inconsistente. Se trata, por tanto, de ver hasta qué punto los gráficos propuestos permiten recuperar el tipo específico de inconsistencia. En el primer ejemplo se intentó provocar un sesgo de respuesta aquiescente. El participante recibió instrucciones de responder sinceramente a los ítems en lo posible, pero también, al mismo tiempo, de mostrar acuerdo con el contenido de los mismos cualquiera que fuese la dirección del enunciado. El valor del índice  $lcz$  para este sujeto fue de 3,12, valor que lo detectaría como potencialmente inconsistente. El gráfico SRP se muestra en la figura 1.

El gráfico en la figura 1 sugiere que el participante siguió bastante fielmente las instrucciones. La mayor parte de las respuestas no se alejan demasiado del plano de respuesta esperado. Sin embargo, hay un conjunto de 10 ítems, localizados mayoritariamente en la parte inferior izquierda del gráfico, cuyas respuestas parecen claramente inconsistentes. Estos ítems que se han marcado con una *r* (revertidos) son, precisamente, los

que medían en la dirección opuesta. De acuerdo con su nivel estimado, este sujeto no debería haber estado de acuerdo con ellos. Lo estuvo por la tendencia a responder de forma aquiescente.

En el segundo ejemplo se ilustra la utilidad del procedimiento CRP. Al segundo participante se le pidió que respondiera al cuestionario totalmente al azar. El valor del índice  $lcz$  para este sujeto fue de 6,09, un valor muy extremo y que sugeriría un elevado grado de inconsistencia. El gráfico CRP se muestra en la figura 2. La recta con trazado grueso es la PRC-E. La curva con trazado fino es la PRC-O obtenida mediante suavizado.

En la figura 2 se aprecia claramente una gran dispersión de todos los valores observados en general, sin ninguna tendencia definida. La CRP-O resume bien este aspecto: es mucho más plana que la CRP-E, sin tendencia clara ni creciente ni decreciente. Es el gráfico que cabría esperar en este tipo de inconsistencia.

El último ejemplo pretende mostrar la utilidad de la CRP en una situación opuesta a los dos anteriores. Cuando se estudian los valores del índice global de CII, generalmente sólo se tienen en cuenta los valores extremos en la dirección de inconsistencia (valores altos y positivos en el caso de  $lcz$ ). Sin embargo, los valores que indican una extrema consistencia también merecen ser inspeccionados. Este sería el caso del participante nº 18, con un valor de  $lcz$  de -3,18. El resultado sugiere que este sujeto responde de forma más determinística de lo que cabría esperar desde el modelo. El gráfico CRP se muestra en la figura 3.

Como se ha comentado antes, la pendiente de la CRP-E reflejaría la capacidad discriminativa promedio de los ítems. La pen-

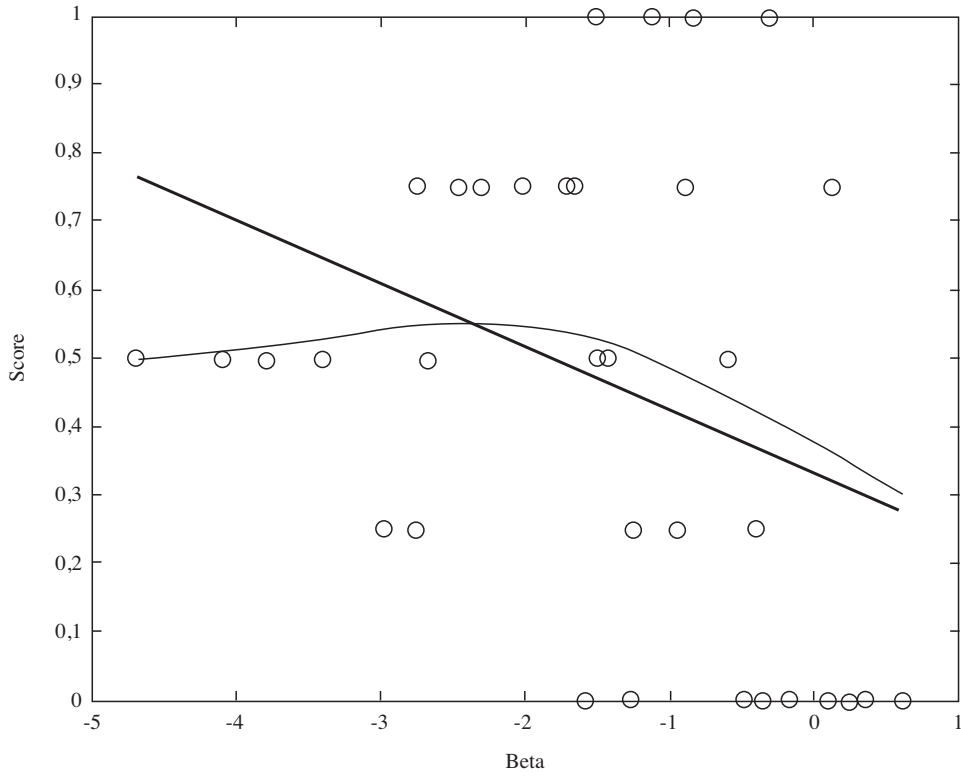


Figura 2. Gráfico CRP correspondiente al sujeto con instrucciones de respuesta al azar

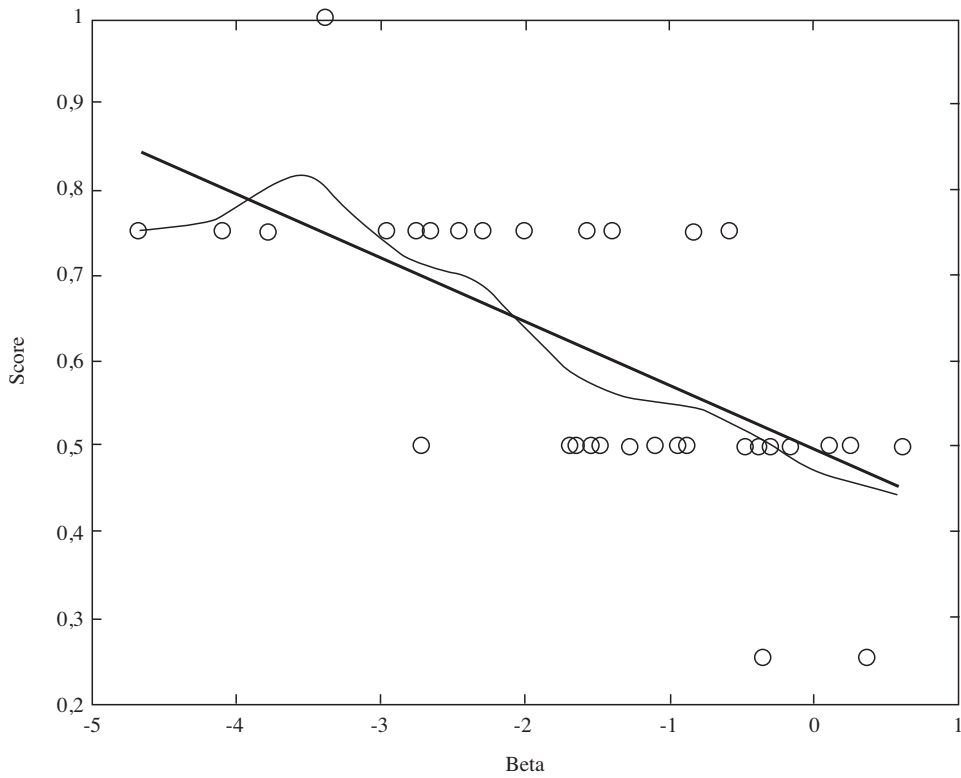


Figura 3. Gráfico CRP correspondiente al sujeto nº 18

diente del la CRP-O en el sujeto 18 es mayor que la de la CRP-E lo que sugiere que responde en forma más consistente de lo que cabría esperar dada la consistencia promedio del test. En teoría de la personalidad esto sugeriría que la estructura del rasgo de Extraversión está muy bien organizada en este sujeto. Sin embargo, la inspección de los puntos indica también que esta sobreconsistencia se obtiene, al menos en parte, gracias al uso restringido de la escala de respuesta. El sujeto sólo utiliza 4 puntos de respuesta adyacentes de los 5 disponibles (no utiliza el extremo inferior en ningún caso). Y sólo utiliza el extremo superior en una ocasión.

### Discusión y conclusiones

Para que una puntuación en un test sea interpretable y permita llegar a inferencias correctas (toma de decisiones, relación con otras variables, etc.) debe basarse en un patrón de respuestas consistente (e.g. Rojas, González, Padilla y Pérez, 2000; Hambleton, 1990). Este requisito está cada vez más asumido en las aplicaciones basadas en los modelos más comunes de TRI. Sin embargo, está casi totalmente descuidado en aquellas basadas en el AF. En efecto, una vez calibrado el test mediante AF y evaluado (en el mejor de los casos) su ajuste global, las puntuaciones factoriales se interpretan y utilizan generalmente sin ningún tipo de verificación. Esta situación afecta principalmente a las aplicaciones en medición de actitudes y personalidad, que son las que más utilizan el modelo lineal.

En este trabajo se presentan dos procedimientos gráficos basados en el modelo unidimensional de análisis factorial lineal, que complementan los resultados obtenidos con índices globales de consistencia intra-individual. Estos procedimientos, pensados para ítems aproximadamente continuos, permiten establecer las posibles causas de los patrones inconsistentes de respuesta detectados previamente. En ambos casos, la utilización de las técnicas propuestas requiere en primer lugar la calibración de los ítems y la evaluación global del ajuste del modelo. Si el ajuste es adecuado, se toman los parámetros de los ítems como fijos y, para cada sujeto, se calculan las puntuaciones factoriales estimadas y un índice de CII. Finalmente, una vez identificado los sujetos potencialmente inconsistentes, se procede a la representación gráfica de las puntuaciones observadas en cada ítem y de las puntuaciones factoriales estimadas de uno de estos sujetos.

En el primer procedimiento propuesto, la representación gráfica tridimensional de los valores esperados del sujeto en cada ítem, como función de los interceptos y de los pesos, forman un plano de regresión denominado Superficie de Respuesta de la Persona. Observando la posición en el gráfico de las puntuaciones observadas, en relación a dicha superficie, es posible determinar qué ítems son los responsables de la baja consistencia intra-individual detectada mediante el índice global de CII. El ejemplo empírico proporcionado en este artículo ilustra cómo, este procedimiento gráfico permite detectar un patrón inconsistente debido a la aquiescencia. Como era de esperar en un patrón de este tipo, las respuestas observadas en aquellos ítems formulados en sentido inverso son las que están más alejadas de la Superficie de Respuesta de la Persona.

El segundo procedimiento parte del modelo de ítems tau-equivalentes, según el cual todos los ítems del test se caracterizan por tener el mismo poder discriminativo. Los ítems de personalidad y de actitudes suelen presentar valores similares en

cuanto a su discriminación, por lo que el uso de este modelo puede funcionar como una razonable aproximación. La representación gráfica en este caso es bidimensional, y las puntuaciones estimadas de los sujetos estarán en función de las posiciones de los ítems, pero no de su nivel de discriminación. En este procedimiento se propone utilizar una formulación en la que la posición del ítem se define en forma similar a los modelos básicos de TRI. Con respecto al primer procedimiento, éste es más simple, y permite obtener una CRP-O que puede compararse con la CRP-E, lo que, a su vez, permite detectar tendencias de respuesta que son la causa de la inconsistencia. Así, en un patrón consistente, ambas curvas mostrarán la misma tendencia y estarán próximas. En cambio, una CRP-O aproximadamente horizontal indicaría que el sujeto tiende a responder siempre de la misma manera, independientemente del nivel de dificultad del ítem. En este artículo se presenta un ejemplo empírico, correspondiente a un sujeto que recibió instrucciones de responder al azar. La CRP-O, correspondiente al patrón de respuestas observadas de este individuo, es muy diferente de la CRP-E, constituyendo en gran medida una línea aproximadamente horizontal.

Pero no solamente la inconsistencia intra-individual debe ser considerada y evaluada. El hecho de que una persona presente un patrón de respuestas especialmente consistente también puede resultar una valiosa fuente de información, porque podría indicar que el rasgo medido está altamente estructurado en el sujeto. Por lo tanto, podría observarse en personas que tienden a responder de forma muy consistente en las diversas situaciones en las que el rasgo está implicado. Con el fin de ilustrar esta posibilidad, se proporciona un último ejemplo empírico, correspondiente a un sujeto con un valor CII *lcz* que sugiere una consistencia intra-individual especialmente elevada. Sin embargo, el gráfico muestra que este patrón de respuestas también se puede obtener cuando el sujeto se limita a responder utilizando un número limitado de categorías de respuesta. Por lo tanto, es necesario descartar esta posibilidad antes de concluir que el comportamiento del sujeto es muy consistente a través de las situaciones.

Los procedimientos desarrollados en este trabajo pueden resultar de utilidad tanto en investigación como a nivel aplicado, en diversas áreas de la Psicología. Por ejemplo, en el caso de que los ítems de un test se refieran a diversas situaciones en las que se puede manifestar el rasgo evaluado, estos procedimientos gráficos permitirían determinar si una persona presenta patrones de respuesta inconsistentes en situaciones concretas. De hecho, diversos autores han reconocido la importancia de conocer hasta qué punto es consistente la conducta de una persona, a través de diversas situaciones. Por ejemplo, según Deschamps y Beauvois (1996), las personas explicamos las conductas ajenas en función de diversas variables, y una de ellas es la consistencia intra-individual de la persona que presenta la conducta en cuestión. A nivel educativo, por otra parte, también hay autores que han señalado la consistencia intra-individual como variable a considerar. Rodríguez, Cabanach, Piñeiro, Valle, Núñez y González-Pienda (2001), por ejemplo, señalan que la orientación de algunos alumnos hacia metas concretas (metas de aprendizaje, metas de aproximación a la tarea, metas en defensa del yo, etc.) dependería de la situación académica, mientras que en otros alumnos las orientaciones hacia las metas serían consistentes a través de las situaciones académicas.

## Referencias

- Dawes, R.M. (1972). *Fundamentals of attitude measurement*. New York: Wiley.
- Deschamps, J.C., y Beauvois, J.L. (1996). Atribuciones intergrupos. En J.F. Morales y D. Páez (Ed.): *Estereotipos, Discriminación y Relaciones entre grupos* (pp. 91-112). Madrid: McGraw Hill.
- Emons, W.H.M., Sijtsma, K., y Meijer, R.R. (2004). Testing hypothesis about the person-response-function in person-fit analysis. *Multivariate Behavioral Research*, 39, 1-35.
- Emons, W.H.M., Sijtsma, K., y Meijer, R.R. (2005). Global, local and graphical person-fit analysis using person-response functions. *Psychological Methods*, 10, 101-119.
- Ferrando, P.J. (2002). Theoretical and empirical comparisons between two models for continuous item responses. *Multivariate Behavioral Research*, 37, 521-542.
- Ferrando, P.J. (2007). Factor-analytic procedures for assessing response pattern scalability. *Multivariate Behavioral Research*, 42, 481-508.
- Ferrando, P.J. (2009a). Difficulty, discrimination and information indices in the linear factor analysis model for continuous responses. *Applied Psychological Measurement*, 33, 9-24.
- Ferrando, P.J. (2009b). Some statistics for assessing person-fit based on continuous-response models. *Applied Psychological Measurement* (en prensa).
- Hambleton, R.K. (1990). Item response theory: Introduction and bibliography. *Psicothema*, 2, 97-107.
- Härdle, W. (1990). *Applied nonparametric regression*. London: Chapman & Hall.
- Hofstee, W.K.B., Ten Berge, J.M.F., y Hendricks, A.A.J. (1998). How to score questionnaires. *Personality and Individual Differences*, 25, 897-910.
- Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Jöreskog, K.G., y Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software.
- Levy, P. (1973). On the relation between test theory and psychology. En P. Kline (Ed.): *New approaches in psychological measurement* (pp. 1-42). New York: Wiley.
- Meijer, R.R., y Sijtsma, K. (1995). Detection of aberrant item scores patterns: A review and new developments. *Applied Measurement in Education*, 8, 261-272.
- Meijer, R.R., y Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25, 107-135.
- Mellenbergh, G.J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, 29(3), 223-237.
- Rodríguez, S., Cabanach, R., Piñeiro, I., Valle, A., Núñez, J.C., y González-Pienda, J.A. (2001). Metas de aproximación, metas de evitación y múltiples metas académicas. *Psicothema*, 13, 546-550.
- Rojas, A.J., González, A., Padilla, J.L., y Pérez, C. (2000). Comparación de estrategias de ajuste de los datos al modelo de crédito parcial. *Psicothema*, 12, 296-302.
- Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality*, 56, 622-663.
- Trabin, T.E., y Weiss, D.J. (1983). The person response curve fit of individuals to item response theory models. En D.J. Weiss (Ed.): *New horizons in testing* (pp. 83-108). New York: Academic Press.