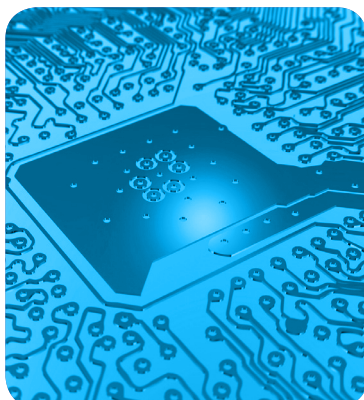
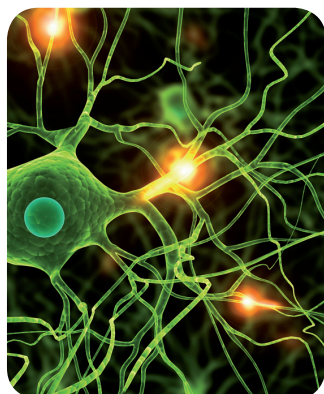


# MÁSTERES de la UAM

Facultad de Psicología /11-12

Máster en Metodología  
de las Ciencias del  
Comportamiento  
y de la Salud



**Efecto de la  
amplitud desigual  
de las categorías so-  
bre la fiabilidad y la  
capacidad de pre-  
dicción en la escala  
de ansiedad del PAI**

*William Fernando  
Durán*



## Índice

Resumen .....	3
Introducción .....	4
Método .....	10
Participantes .....	10
Instrumentos.....	11
PAI.....	11
Procedimiento .....	11
<i>Predicción de un criterio externo</i> .....	24
<i>Simulación de datos</i> .....	26
Discusión y Conclusiones .....	30
Referencias .....	34
Anexos.....	36
Anexo 1. Descripción de las escalas del PAI.....	36
Anexo 2. Sintaxis de Matlab utilizada para la simulación .....	38
Anexo 3. Sintaxis de Multilog.....	46
Anexo 4 Archivo Ejecutable.bat .....	47

### **Resumen**

Los ítems que conforman el PAI (Personality Assessment Inventory) presentan una estructura poco común que no ha recibido investigación psicométrica. En estos, se utiliza una escala Likert de 4 puntos en donde la primera categoría representa la opción “falso” y las siguientes son graduaciones de la respuesta a “verdadero”. Se exploró la posibilidad de que las categorías utilizadas en estos ítems ocuparan amplitudes asimétricas del continuo psicológico, lo cual podría tener implicaciones en su funcionamiento. Se estudió el efecto de la amplitud desigual de la primera categoría en ítems Likert de 4 puntos sobre la fiabilidad y la capacidad de predicción a partir de puntuaciones con el Modelo de Respuesta Graduada y puntuaciones clásicas; para esto, se trabajó con una muestra española de 3057 casos y se complementó con técnicas de simulación. Se mostró que la distinta amplitud de las categorías tiene un impacto en la fiabilidad y en la capacidad de predecir variables criterio asociadas. El Modelo de Respuesta Graduada parece ser más robusto que la puntuación clásica, pero la diferencia en precisión es apenas de algunas centésimas.

## Introducción

La elaboración de ítems para medir variables clínicas de personalidad ha tenido un largo proceso de desarrollo, en el cual han surgido múltiples alternativas. Quizá el formato de ítems más reconocido para su medición sea el dicotómico (Si-No), en donde el evaluado responde a un reactivo, señalando si la situación planteada le ocurre en algún contexto vital o no. Este tipo de ítems ha sido criticado ya que obliga al participante a tomar de manera forzosa uno de los dos extremos perdiendo la información cuantitativa intermedia sobre la intensidad en que considera que el contenido del ítem lo representa (Nunnally, 1979). El impulso más fuerte para este tipo de ítems se dio por su uso en el MMPI, un clásico instrumento de personalidad clínica desarrollado en 1930 que se convirtió durante años en un estándar de la evaluación.

Como alternativa se han propuesto ítems con escalas Likert, comúnmente de 4 puntos para evitar una respuesta intermedia por parte del evaluado; este es el caso del PAI (Personality Assessment Inventory) un instrumento desarrollado en Estados Unidos como alternativa a los métodos clásicos, que buscó introducir las mejoras metodológicas y conceptuales que se han dado en el entorno clínico, además de tener en cuenta las necesidades del profesional encargado de la intervención (Morey, 1991). Es un instrumento bastante completo compuesto por 22 escalas, de las cuales 11 evalúan variables clínicas, 4 son escalas de validez, 5 evalúan variables relacionadas con el tratamiento y 2 tienen en cuenta los elementos de relación interpersonal. (Ver Anexo 1 con la lista de las escalas del PAI)

Los ítems Likert según fueron propuestos en 1929 y se basan, según Barbero (2007) en cinco principios:

1. Es posible estudiar una dimensión a partir de un conjunto de enunciados que operen como reactivos para los sujetos
2. Los evaluados pueden situarse en la variable desde el punto más favorable al desfavorable. La variación de las respuestas será debida a diferencias individuales de los sujetos.

3. La valoración de los sujetos en la variable no supone una distribución uniforme sobre el continuo, sino su posición favorable o desfavorable sobre el conjunto estudiado.
4. La suma de las curvas características de los ítems es una función monotónica y aproximadamente lineal respecto a la actitud medida
5. Los elementos que componen la escala deben estar midiendo una única dimensión. Se trata por lo tanto de una escala unidimensional (Lopez Pérez 1985; en Barbero 2007).

El método de construcción de ítems Likert consiste en crear un conjunto de enunciados relacionados con el constructo que se desea medir, presentarlos con una escala de respuesta graduada y obtener la puntuación total de cada sujeto como suma o promedio de los valores de las alternativas que ha escogido. Las alternativas graduadas llevan un entero asociado que refleja su posición en el grupo de opciones de respuesta (Meliá Navarro, 1991)

No hay un pronunciamiento tajante sobre la relación que deben guardar los puntos utilizados en la escala graduada, aunque de manera clásica se podría hablar que hay un principio implícito que señala que estos guarden una relación progresiva constante. (Wakita, Ueshima, & Noguchi, 2012). Un ejemplo es el siguiente es un ítem extraído de Abad, Olea, Ponsoda y García (2011):

Las drogas pueden realmente resolver problemas de uno mismo

- Muy en desacuerdo
- Bastante en desacuerdo
- Neutral
- Bastante de acuerdo
- Muy de acuerdo

En el anterior ítem se observa una progresión constante en la dimensión (actitud hacia las drogas), de manera que el continuo psicológico podría dividirse aproximadamente en cinco categorías simétricas.

En el PAI se construyeron ítems que buscaban evaluar a partir del mismo reactivo la intensidad y la ocurrencia (frecuencia de ocurrencia) de la situación señalada para el examinado además de presentar categorías que se alejan de una progresión constante. Por lo tanto con el mismo ítem se intentó extraer aspectos cuantitativos potenciales de la variabilidad de la respuesta sobre su frecuencia y su intensidad con un tipo de formato novedoso. La escala de respuesta graduada utilizada en el PAI es:

- Falso
- Ligeramente verdadero
- Bastante verdadero
- Completamente verdadero

El primer punto o categoría graduada estaría señalando la “No ocurrencia” de la información señalada por el reactivo, mientras que los tres restantes conformarían una gran categoría que indica la “Si ocurrencia”; por lo tanto, a la hora de evaluar si el contenido de un ítem ocurre o encuentra un correlato en la experiencia vital del participante, se estaría en realidad utilizando un ítem dicotómico (No-Si). Pero si se ubicaran los puntos de este ítem en un continuo psicológico la primera categoría de respuesta debería ocupar un espacio mayor en comparación con las tres siguientes, ya que las demás son apenas graduaciones de la “si ocurrencia”.

La intensidad estaría evaluada por la graduación que tiene la respuesta de “Si ocurrencia” en tres niveles, de este modo, a la vez que el sujeto responde que la afirmación si ocurre en su contexto vital, también está indicando que tanto le sucede.

Las causas por las cuales Morey (1991) plantea este tipo de ítems para el PAI son las siguientes:

En Primer lugar, reconoce que al usar una escala de cuatro puntos se captura más variabilidad en cada uno de los ítems que al usar una dicotómica, lo cual permite mejorar la fiabilidad de la escala con pocos ítems. Una de sus preocupaciones en el desarrollo del instrumentó fue la extensión que debía tener, ya que los test de personalidad clínica en el mercado eran bastante extensos lo que dificultaba la aplicación; al usar escalas Likert de 4 puntos se podría tener una mayor fiabilidad con menor número de ítems

Segundo, la escala de cuatro puntos eliminaría la categoría central lo cual lleva consigo importantes ventajas psicométricas y señala además que esta permitiría dicotomizar si se deseara en algún momento, algo que se hizo para el desarrollo de la versión de screening del PAI llamada PAS (Morey, 1997). Este es un elemento que despierta ciertas dudas ya que podría haber algún tipo de contradicción entre su filosofía de creación de ítems con asimetría en la primera categoría. Si se realizara la dicotomización siguiendo su planteamiento se debería poner el punto de corte en el primer punto.

Tercero, los puntos de anclaje Verdadero-Falso, son familiares para muchos evaluados y facilitan el entendimiento por personas de bajo nivel educativo. Y finalmente en el cuarto argumento señala que este tipo de puntos subsume un número de potenciales aspectos cuantitativos incluyendo frecuencia e intensidad.

Esta forma de los ítems pondría en duda el supuesto explícito 5 de los ítems Likert clásicos (unidimensionalidad), debido a que un mismo ítem estaría midiendo tanto la ocurrencia como la intensidad de los enunciados por parte del evaluado y además, las categorías o puntos graduados no siguen una clara progresión constante, ya que el espacio en el rasgo que tomaría el primer punto (opción “falso” debería ser mayor que el ocupado por los otros tres (graduaciones de la opción “Si ocurre”). En los ítems del PAI se espera que exista asimetría entre los rangos ocupados por cada categoría a lo largo del continuo psicológico evaluado. Para el presente estudio se entenderá asimetría como el espacio o rango distinto ocupado por la primera categoría graduada de un ítem, en comparación con las demás. La asimetría puede estar presente en cualquiera de las categorías que conformen un ítem Likert, pero para este estudio se pondrá la atención en la primera, ya que es la condición que caracteriza a los ítems del PAI.

Morey (2007) indica que este tipo de ítems se adecuan más al examinado y le facilitan su respuesta, pero es necesario explorar las consecuencias metodológicas que puede traer este cambio de filosofía en la construcción de los ítems.

Una posible consecuencia de romper el criterio implícito de progresión constante es que las categorías no pueden ser puntuadas como enteros fijos asociados a cada una, ya que el espacio en el continuo psicológico que ocuparían no sería el mismo. Este es un punto de

vital importancia para el PAI ya que la forma de calificación propuesta por el autor es la sumatoria de respuestas, lo cual plantea un problema debido a la falta de claridad en cuanto a si la asimetría de las categorías afecta o no el modo de puntuación.

Para evaluar el efecto de la asimetría en la primera categoría puede ser útil el modelo de respuesta graduada (MRG) de Samejima (1969, 1972, 1997; en Abad, Ponsoda y Revuelta, 2006), el cual hace parte de la familia de modelos de Teoría de Respuesta al Ítem. En éste la aportación a la theta estimada ( $\theta$ ) de la categoría elegida no es un escalar constante, sino que depende de las propiedades psicométricas del ítem. En teoría, si en verdad los ítems tienen asimetría entre sus categorías, el modelo de respuesta graduada sería más robusto para estimar las puntuaciones en este tipo de ítems que la puntuación clásica, además permitiría una mayor fiabilidad.

El MRG está diseñado para trabajar con ítems de respuestas ordenadas las cuales siguen un sentido teórico en el rasgo que se desea medir; se asume que las opciones pueden ordenarse por el nivel de rasgo que implica elegir las. El modelo propuesto permite estimar K-1 curvas características operantes de un ítem de K categorías, en donde cada una representa la probabilidad de elegir dicha categoría o alguna superior (Abad, Olea, Ponsoda y García, 2011). El modelo permite estimar un parámetro b para cada una de las curvas características, el cual indicaría el nivel de rasgo asociado a la respuesta a cada categoría y además un parámetro a, que señala la capacidad de discriminación del ítem.

En la literatura pueden encontrarse varios estudios que realizan comparaciones de los modelos TRI para ítems de respuesta máxima con las puntuaciones clásicas (Fan, 1998; Macdonal y Paunonen, 2009; Barbero, Prieto, Suarez y San Luís, 2001) pero es más escasa para los ítems de respuesta típica. Entre los que se reportan Dumenci y Achembach (2008) estudiaron las diferencias que podrían surgir al utilizar distintas maneras de puntuar escalas de depresión, ansiedad y agresión. Compararon las puntuaciones obtenidas por medio de modelos TRI (Crédito parcial y Modelo de respuesta graduada), puntuación clásica (sumatoria de respuestas) puntuaciones factoriales y análisis de componentes principales. Igualmente consideran que las puntuaciones pueden agruparse en dos tipos: aquellas que son sensibles a la distribución ordinal de las categorías de los ítems (en donde se ubicaran los Modelos de TRI) y los que no son sensibles a esta distribución. Los resultados de su



trabajo les permiten concluir que la elección entre estos métodos es inconsecuente ya que las diferencias entre estas son insignificantes.

En otro estudio, Asún y Zúñiga (2008) comparan el Modelo de respuesta graduada con la puntuación la Teoría clásica de los Test, en relación a las propiedades de los ítems y de las puntuaciones de los sujetos concluyendo que hay una enorme similitud entre ambos. Sostienen que el MRG permite ganar conocimiento sobre el funcionamiento de los ítems que la teoría clásica no muestra, entre otras cosas permite conocer la información que aporta cada una de las categorías de respuesta ayudando a planear las futuras aplicaciones. Con base en los resultados señala también que las categorías intermedias de la escala comúnmente son menos informativas y con cuatro alternativas parece suficiente para cubrir el espectro del rango medido.

En el trabajo de Xu y Stone (2012) se usan el método de Monte Carlo para comparar la estimación de un rasgo y la capacidad de predicción de una variable externa por el MRG y por la puntuación clásica. Se realizaron 500 replicas manipulando el tamaño del efecto, la extensión de la escala y la correlación entre el rasgo y la variable externa. En el estudio se llega a la conclusión que ambas formas de puntuación son comparables, de hecho la puntuación clásica al parecer es más robusta cuando se trabaja con pocos ítems (en su estudio fueron 10) y un bajo tamaño muestral (250 casos).

El PAI es un instrumento que ha ganado espacio en el campo aplicado durante los últimos años, ya existe una validación y tipificación a una muestra española (Ortiz-Tallo, Santamaría, Cardenal y Sánchez, 2012) y varios estudios que ratifican su calidad psicométrica en diferentes poblaciones. Siefert, Sinclair, Kehl-Fie, & Blais (2009) evidenciaron que el PAI mantiene la consistencia interna en sus escalas y subescalas cuando se aplica a pacientes psiquiátricos con valores entre 0,55 y 0,90 y media de 0,78. La menor consistencia interna se encuentra en la sub escala de nivel de actividad perteneciente a la escala de manía. La correlación ítem-escala tiene un rango de 0,27 a 0,68 con una media de 0,49. Karlin et al. (2005) estudian las propiedades psicométricas del instrumento en pacientes con dolor crónico; reportan que la media en consistencia interna de las escalas es de 0,78 siendo este muy similar al valor reportado por Morey (1991) en el manual para una muestra clínica (0,79). La media de la correlación ítem-escala es de 0,20 también un

valor similar al reportado en el manual (0,24). Corsica, Azarbad, McGill, Wool, & Hood (2010) encuentran valores similares en candidatos a tratamiento bariátrico, los valores de fiabilidad interna de las escalas varían entre 0,53 y 0,91 con una media de 0,78. Las sub escalas tienen fiabilidad interna entre 0.49 y 0.84 con valor medio de 0.72. El valor medio de la correlación ítem-test es de 0,23 con valores entre 0,12 y 0,47.

Si bien se reconocen las propiedades psicométricas del instrumento no hay estudios que profundicen en el funcionamiento de estos ítems tan particulares que lo conforman. En el presente estudio se buscó explorar cuales son las consecuencias metodológicas del uso de ítems con asimetría en la primera categoría graduada en relación con su fiabilidad y su capacidad predictiva. Se espera que la asimetría afecte la fiabilidad de las puntuaciones clásicas y TRI (puntuación  $\theta$ ) de manera distinta, ya que el MRG debería ser más robusto y ajustarse de una mejor manera a las categorías con espacios distintos ya que tiene en cuenta la distribución de estas en el rasgo medido; aunque según los antecedentes, no se espera que la diferencia entre ambas sea amplia. También se espera que el MRG prediga de un modo más preciso algún criterio asociado al nivel de habilidad real de los sujetos.

Para poner a prueba el efecto de la asimetría en la primera categoría de la escala sobre la fiabilidad y la capacidad predictiva, se trabajó con los datos de tipificación y aplicación del PAI en España y con técnicas de simulación de datos. Para generar la condición de asimetría se desplazó los parámetros  $b$  del MRG hacia la parte derecha del continuo psicológico, de tal manera que todo el espacio anterior a  $b_1$  sea para la categoría asimétrica.

## Método

### Participantes

Se trabajó con los datos de 3426 casos de aplicación del PAI provenientes de toda España, de los cuales se descartaron 369 por contar con elevados valores en las escalas de Inconsistencia e Infrecuencia, las cuales están diseñadas para detectar posibles casos de respuesta al azar o patrones de respuesta dudosos. Se eliminaron todos los casos con puntuaciones en alguna de estas escalas superior a 75 en puntuación  $T$  (Media=50,

Desviación=10). Este valor es el recomendado en el manual de la prueba a partir de los análisis realizados con la muestra de tipificación española.

Por el amplio número de la muestra con el que se contó fue posible descartar también cualquier caso que tuviera por lo menos un valor perdido. La muestra final estuvo conformada por 3057 personas, 60,6% mujeres, con edad media de 31,66 años y en un rango entre 18 y 84 años. De esta muestra, 1907 fueron casos normales (sin diagnóstico o sospecha de problemas psicológicos) y 1150 clínicos (en consulta o tratamiento o con historial psicopatológico).

## Instrumentos

### *PAI*

El Personality Assessment Inventory es un instrumento de medición de variables clínicas compuesto por 22 escalas de las cuales 4 son de validez, 11 son escalas clínicas, 5 evalúan consideraciones para el tratamiento y 2 exploran elementos de relación interpersonal. El presente estudio se centró el análisis en la escala de Ansiedad, compuesta por 24 ítems Likert de 4 puntos, la cual hace parte de las escalas clínicas del test.

## Procedimiento

Los datos fueron proporcionados por TEA Ediciones S.A, empresa que realizó la validación de la prueba al español y se encargó además de la tipificación de la misma a la población española. En la muestra que conforma el presente estudio se trabajó con los datos de la muestra original de tipificación y se añadió algunos casos nuevos reunidos por la empresa a través de su plataforma en Internet (Ortiz-Tallo, Santamaría, Cardenal y Sánchez, 2012).

Con el conjunto de datos se realizó, en primera instancia, un análisis de unidimensionalidad para 18 escalas del PAI, excluyendo las escalas de validez. Se realizó un conjunto de análisis factoriales confirmatorios con el programa Mplus utilizando como método de estimación RWLS por tratarse de datos politómicos.

A partir de los resultados se decidió realizar las siguientes fases de análisis con la escala de Ansiedad, ya que cumplía con la unidimensionalidad, mantenía valores adecuados de asimetría y curtosis, tenía el mayor número de ítems y además pertenecía a las escalas clínicas. Este último elemento le dio un interés adicional ya que las demás escalas que alcanzaban la unidimensionalidad indagaban por aspectos relacionados con el tratamiento, lo cual es un suplemento informativo para el profesional pero no constituyen el corazón de un test aplicado a la clínica.

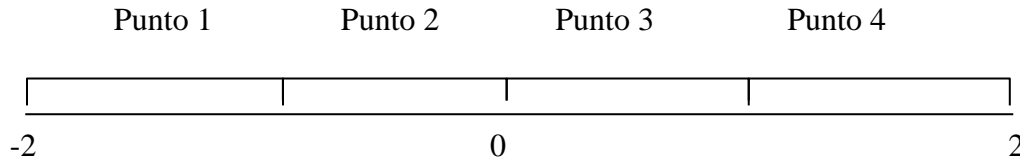
Se aplicó el modelo de respuesta graduada (MRG) a los datos de la escala de Ansiedad para explorar la posible asimetría en los puntos de la escala Likert utilizados en el PAI. Con este fin, se utilizó el programa Multilog. Se utilizó como método de estimación la Máxima Verosimilitud Marginal. Con los valores de los parámetros estimados, se estudió el ajuste al modelo utilizando para este fin la macro de Excel Modfit (Drasgow, F., Levine, M. V. 2001) y el programa R. Se calcularon los valores de doublets para la muestra total pero, debido a la sensibilidad de la prueba ante tamaños de la muestra grandes, surgieron desajustes en todos los ítems, por lo cual se seleccionaron sub-muestras de 300, 900, 1500, 2100 y 3000 para realizar el análisis. No se identificó un desajuste fuerte en ninguno de los 24 ítems que conforman la escala con el análisis de las sub-muestras.

Con los parámetros de cada ítem, se estimaron además los valores de  $\theta$  para cada sujeto nuevamente con Multilog. Tomando los parámetros, se calcularon las distancias entre los puntos de la escala Likert de dos maneras distintas:

La primera, utilizó la mediana de los parámetros  $b$  para encontrar tres puntos representativos de toda la escala; se tomó como dominio del rasgo, los valores comprendidos entre -2 a 2 por considerarse suficientemente representativo de todo el continuo y se ubicaron las medianas calculadas en esa recta. Para este análisis, se utilizó la mediana y no la media, ya que los parámetros del ítem 8 fueron bastante irregulares, lo cual podría sesgar el promedio debido a la sensibilidad que tiene a los valores extremos.

Con los valores de la mediana ubicados en el continuo de -2 a 2, se calculó la distancia entre los puntos de la siguiente forma:

El primer punto tomó toda el área que se encontraba antes de éste hasta -2, el segundo punto tomó el área entre la segunda y la primera mediana; el tercer punto, se quedó con el espacio entre la tercera y la cuarta mediana y finalmente el cuarto punto, tomó el complemento hasta 2.



**Figura 1.** Distribución de las categorías en un ítem

La segunda manera en que se calcularon las distancias siguió la misma estrategia de ubicación de puntos en el continuo del rasgo de -2 a 2 pero se hallaron las diferencias para cada ítem y éstas fueron promediadas para obtener valores globales de distancia en la escala. Se eliminó de este procedimiento los ítems 8, 14 y 15 por tener valores en el parámetro  $b_1$  inferiores a -2. Ambas maneras de estudiar las distancias llegaron a conclusiones similares sobre el comportamiento de los puntos de la escala.

Luego de identificar la asimetría en los puntos de los ítems del PAI, se exploró el efecto de ésta a la hora de predecir un criterio externo. El instrumento fue diseñado para proporcionar información sobre variables clínicas que permitieran apoyar los diagnósticos realizados por profesionales, de lo cual se sigue que la manera en que se califica las escalas del instrumento, puede afectar las conclusiones a las que llegue un profesional; por esta razón se decidió comparar el modo de calificación clásico de la prueba (sumatoria de respuestas dadas por los participantes en cada ítem) con las puntuaciones  $\theta$  del modelo de respuesta graduada. Una fortaleza de los modelos de TRI frente a la Teoría Clásica de los Test radica en permitir que las puntuaciones dadas a los sujetos no sean constantes sino que puedan estar en función de las propiedades psicométricas de cada ítem, por lo cual se exploró la posibilidad que la asimetría encontrada tuviera un efecto en las puntuaciones obtenidas de los sujetos y en la predicción que realizaban de una variable externa.

Durante la recolección de datos, se preguntó a los participantes si tenían o tuvieron algún diagnóstico clínico, lo cual permitió identificar grupos de casos con diagnósticos previamente identificados. Se encontraron personas con trastornos de ansiedad, trastornos

del estado de ánimo, trastornos psicóticos, comportamiento antisocial, intentos de suicidio, adicciones, trastornos adaptativos, trastornos orgánicos, somatomorfos, de alimentación, entre otros. Para el presente estudio se evaluó la capacidad predictiva de las puntuaciones clásicas de la escala de Ansiedad y las puntuaciones  $\theta$  del modelo de respuesta graduada para clasificar correctamente aquellos casos previamente diagnosticados con trastorno de ansiedad. Se utilizó la curva COR y la regresión logística binaria como estrategias de análisis y se trabajó con el programa SPSS 18. El número de personas diagnosticadas con problemas de Ansiedad fue de 288, un valor bastante pequeño en comparación con la muestra total, por lo que se decidió seleccionar de manera aleatoria otros 288 casos entre aquellos que no tuvieran este diagnóstico. Estos dos grupos se unieron en una misma base de datos para efectuar la regresión logística.

Como estrategia de profundización de los resultados, se recurrió a la simulación de datos a partir del programa Matlab; la calibración de respuestas simuladas se realizó con Multilog (los programas estuvieron encadenados por un archivo de extensión .bat). La idea fue estudiar con mayor precisión las consecuencias que puede tener la asimetría entre los puntos de un ítem Likert sobre la fiabilidad y la predicción de un criterio externo. El proceso que se siguió en la simulación fue el siguiente:

1. Generación aleatoria de valores de una distribución normal que representaban los sujetos del estudio (este valor fue llamado Z real).
2. Simulación de tres variables correlacionadas con la Z real en 0.3, 0.5 y 0.7 respectivamente.
3. Aplicación de los ítems, siempre con parámetros conocidos, a los sujetos simulados para obtener los patrones de respuesta
4. Cálculo de las puntuaciones clásicas en la escala a partir de la sumatoria de las respuestas simuladas.
5. Exporte de las respuestas simuladas a Multilog por medio de un fichero .bat.
6. Calibración de los parámetros y las puntuaciones  $\theta$ .
7. Recuperación de los archivos .par y .sco que contienen los valores de las parámetros y las puntuaciones  $\theta$  desde Matlab

Este proceso se llevó a cabo para cada una de las réplicas realizadas. (Los Anexos 2, 3 y 4 permiten observar la sintaxis de MATLAB utilizada para la simulación, las sintaxis de Multilog y el ejecutable .bar).

La fiabilidad se definió como la capacidad que tenían ambas formas de puntuar (clásica y  $\theta$  estimada) para recuperar los valores de Z real, y se obtuvo mediante la correlación de Pearson entre las Z reales y los niveles de rasgo estimados a partir de la TCT y de la TRI. La asimetría se definió como la amplitud ocupada por el primer punto de la escala en relación con cada uno de los demás puntos, de tal manera que se amplió progresivamente el espacio del punto uno, manteniendo simetría entre el punto dos y tres y entre el tres y el cuatro. En otras palabras cada uno de los niveles de asimetría desplazó los valores de los parámetros b de los ítems hacia la derecha del dominio del rasgo.

Las variables que se simularon correlacionadas con la Z real, sirvieron como criterio externo que permitió estudiar cuál de las dos formas de puntuación (clásica y  $\theta$ ) reproducía de mejor manera la correlación simulada. Por lo tanto, de esta manera, se analizó el efecto de la asimetría al momento de predecir un criterio. Para la generación de estas variables se tuvo en cuenta el  $R^2$  de cada valor fijado en la correlación y su complemento.

$$1 = R^2_{coef} + R^2_{error}$$

Por lo cual, para reproducir la correlación de 0.3 se encontraron los valores de  $R^2$  del coeficiente (en este caso 0.09) y del error (siendo este el complemento 0.91). A estos valores se les calculó la raíz cuadrada la cual fue utilizada en la ecuación de simulación. Por lo tanto, para la correlación de 0.3, 0.5 y 0.7 se utilizaron respectivamente las siguientes ecuaciones:

$$Y_1 = 0,3.Z_{real} + 0,9539.x_1$$

$$Y_2 = 0,5.Z_{real} + 0,8660.x_2$$

$$Y_3 = 0,7.Z_{real} + 0,7141.x_3$$

En cada ecuación siempre se utilizó hasta el cuarto decimal para el coeficiente de error. Z real estuvo definida por el vector de valores generados como nivel de rasgo para

los sujetos y  $x$  representó el error de predicción o esa parte que no se logra explicar con el modelo. Los valores que adoptó  $x$  fueron números aleatorios generados a partir de una distribución normal (0,1).

Se simularon en total 3000 sujetos (generados a partir de valores aleatorios de la distribución normal de media 0 y desviación 1) con 1000 réplicas para cuatro niveles de asimetría diferentes en cuatro condiciones distintas, con lo cual en total se realizaron 16 simulaciones. Los niveles de asimetría fueron:

1. Parámetros estimados: con los parámetros calibrados de los ítems del PAI en la escala de Ansiedad, se simularon las respuestas dadas por 3000 sujetos a 1000 réplicas.
2. Simetría: se buscó simular la situación de simetría entre los puntos. Todos los procesos de simulación se desplazaron en el rango entre -2 y 2; en este nivel se fijó como parámetros  $b$  (-1, 0 y 1), de tal manera que los puntos tuvieron una unidad del dominio manteniendo la simetría.
3. Asimetría 1: se fijó como valores de los parámetros  $b$  (0, 0.7 y 1.4), de tal manera que el espacio desde -2 hasta 0 fue dado al primer punto de la escala.
4. Asimetría 2: se fijaron como valores de  $b$  (1, 1.33 y 1.66) dejando para el primer punto de la escala el espacio entre -2 y 1.

En todos los niveles de asimetría se mantuvo como parámetros  $a$  los valores identificados en la calibración, de tal manera que éste fuera constante para cada nivel.

Las cuatro condiciones que se simularon fueron:

1. Condición 1: Se simuló la situación en la cual cada uno de los niveles de asimetría conformaba un test diferente aplicado a una muestra de sujetos diferente, de tal manera que los valores de  $Z$  real cambiaban para cada uno de los niveles.
2. Condición 2: Se buscó simular la aplicación de test distintos a los mismos sujetos, por lo tanto los valores de  $Z$  real fueron iguales para cada uno de los niveles.
3. Condición 3: Se estudió el efecto de la asimetría cuando se ve reducida la fiabilidad del test, para lo cual se seleccionaron 12 ítems aleatorios de los 24 que



conforman la escala. Esto se realiza bajo el conocimiento que la extensión del test está en clara relación con la fiabilidad.

4. Condición 4: nuevamente se redujo la extensión del test a 6 ítems seleccionados de manera aleatoria para disminuir un poco más la fiabilidad y ver el comportamiento de la asimetría en esta situación.

Se planeó que MATLAB generara ficheros con extensión .dat que contuvieran las salidas del proceso y pudieran ser llevadas a SPSS 18 para analizarlas.

En la fase de análisis se estudió el sesgo de cada simulación y la recuperación correcta de los parámetros del modelo de respuesta graduada, correlacionándolos con los valores estimados. Lo anterior se realizó para cada una de las réplicas y se promedió los valores de sesgo y de correlación.

### Resultados

La aplicación de un modelo de TRI requiere la comprobación de supuestos que hagan válida su aplicación a una muestra determinada, por lo cual el primer paso fue la corroboración de la unidimensionalidad por medio de análisis factorial confirmatorio. En la tabla 1 se presentan los estadísticos descriptivos y los índices de ajuste a la unidimensionalidad para cada una de las escalas del PAI.

**Tabla 1.** Descriptivos e índices de ajuste de las escalas del PAI

Escala	N ítems	Media	Desv	Asimetría	Curtosis	Índices de ajuste		
						CFI	TLI	RMSEA
SOMATIZACIÓN	24	16,15	12,414	1,278	1,460	0,798	0,95	0,088
ANSIEDAD	24	27,57	14,259	,631	-,245	0,844	0,969	0,085
TRAS REL ANSIEDAD	24	25,95	11,492	,554	-,057	0,795	0,905	0,104
DEPRESIÓN	24	20,65	13,270	,952	,489	0,648	0,907	0,142
MANÍA	24	23,35	9,606	,562	,282	0,563	0,718	0,113
PARANOIA	24	24,29	10,302	,750	,839	0,675	0,878	0,105
ESQUIZOFRENIA	24	17,90	10,081	,953	1,294	0,462	0,739	0,145
RASGOS LIMITES	24	25,23	11,733	,670	,166	0,623	0,838	0,114
RASGOS ANTISOCIALES	24	15,88	9,788	1,311	2,448	0,721	0,902	0,07
AGRESIÓN	18	16,47	8,977	,820	,690	0,775	0,915	0,103
ALCOHOL	12	4,94	5,952	2,166	5,522	0,918	0,967	0,099
DROGAS	12	5,99	6,828	1,755	3,191	0,928	0,967	0,091
IDEAS SUICIDAS	12	4,96	7,178	2,111	4,395	0,952	0,989	0,093

ESTRÉS	8	7,92	4,852	,791	,227	0,879	0,879	0,154
F APOYO SOCIAL	8	6,70	4,334	,704	,212	0,893	0,906	0,126
R. TRATAMIENTO	8	12,66	5,250	-,265	-,625	0,922	0,944	0,104
DOMINANCIA	12	19,60	5,368	-,125	-,091	0,827	0,852	0,122
AFABILIDAD	12	21,00	5,515	-,258	,100	0,693	0,776	0,14

Tomando como criterios los valores de ajuste, no habría en primera instancia una escala que cumpliera de manera ideal con los criterios establecidos para cada uno de los índices, pero si hay algunas de ellas que merecen un análisis más minucioso. Las escalas de Alcohol, Drogas, Ideación suicida y rechazo al tratamiento presentan valores de CFI y TLI aceptables pero valores muy elevados en RMSEA; en cambio, la escala de Ansiedad presenta valores aceptables para TLI y RMSEA. Con estas escalas que parecen mostrar indicios de unidimensional, se decidió aplicar un análisis confirmatorio para explorar el posible efecto del tamaño de la muestra, por lo cual se aplicó el análisis a 3 muestras de 300 casos seleccionados al azar y se promedió los índices de cada ocasión. Los valores pueden observarse en la tabla 2.

**Tabla 2.** Índices de ajuste de sub muestras

	CFI	TLI	RMSEA
ANSIEDAD	0,92	0,97	0,08
ALCOHOL	0,95	0,97	0,11
DROGAS	0,97	0,98	0,08
IDEAS SUICIDAS	0,98	0,99	0,08
R. TRATAMIENTO	0,95	0,97	0,08

La escala de Alcohol no muestra un valor aceptable en RMSEA, con lo que fue descartada como posible candidata de análisis. Entre las demás escalas se acepta que existe unidimensionalidad pero se tomó la decisión de trabajar con la escala de Ansiedad, ya que cumplía con los criterios pese a tener un número mayor de ítems y no mostraba una asimetría fuerte como si sucede en las escalas de Drogas y en la de Ideas Suicidas. La escala de Rechazo al tratamiento fue descartada por considerarse un constructo propio del PAI que funciona como índice de orientación para el profesional a la hora de adoptar estrategias de intervención con el paciente, pero no es en esencia una característica

psicopatológica. De este modo, se concluyó que lo más adecuado sería trabajar con la escala de Ansiedad para la calibración con el modelo de respuesta graduada.

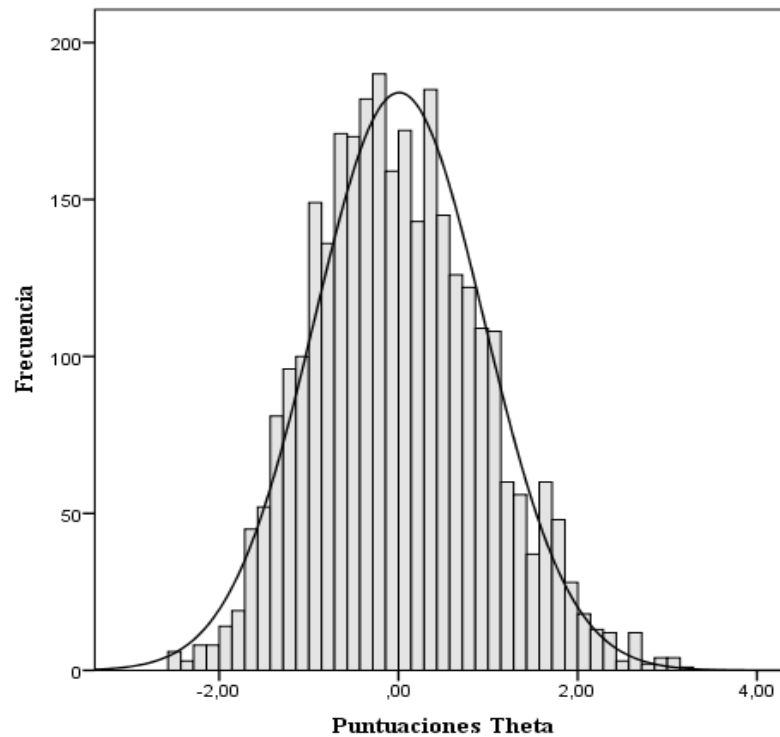
Como un último filtro se utilizó el método de análisis paralelo con 300 casos usando la correlación policórica y se concluye de igual manera que los ítems se ajustan a una estructura unidimensional.

Con la escala seleccionada, se procedió a realizar la calibración de los ítems en MULTILOG, utilizando estimación Máximo Verosímil Marginal, llegando a los valores de los parámetros que se observan en la tabla 3.

**Tabla 3.** Parámetros de los Ítems de la escala de Ansiedad

	A	B1	B2	B3
Ítem 1	2,0522	-0,4375	0,6708	1,5627
Ítem 2	1,9491	-0,0610	0,9318	1,7499
Ítem 3	3,2491	0,0210	0,8723	1,5473
Ítem 4	1,3948	-1,1527	0,0857	1,2888
Ítem 5	1,5922	-1,5712	-0,3290	0,9581
Ítem 6	2,1365	0,7758	1,3523	1,8927
Ítem 7	0,9877	-1,6915	0,1025	1,7675
Ítem 8	0,5263	-4,6471	-2,3568	0,4290
Ítem 9	2,2623	-0,5452	0,4350	1,2248
Ítem 10	2,1660	-0,2661	0,8043	1,5224
Ítem 11	1,7485	-0,0944	0,9446	1,7436
Ítem 12	1,5297	0,3731	1,4513	2,2724
Ítem 13	1,3695	-0,4929	0,8670	1,9985
Ítem 14	0,5765	-2,8047	-0,4234	2,0090
Ítem 15	1,0798	-2,0811	-0,7526	0,7136
Ítem 16	1,4758	-1,6341	-0,7293	0,3827
Ítem 17	1,9583	-0,6103	0,6785	1,7016
Ítem 18	2,2385	0,7586	1,5250	2,1466
Ítem 19	1,5075	0,1316	1,0573	1,9287
Ítem 20	1,2772	-0,5088	0,6399	1,7008
Ítem 21	1,7472	0,0839	0,9077	1,6343
Ítem 22	0,6877	1,6072	3,0497	4,2660
Ítem 23	1,4299	-1,7492	-0,5441	0,7887
Ítem 24	0,7593	-1,1378	0,8210	2,6975

No se encontró ningún problema de estimación en la aplicación del programa. Se terminó la estimación en el ciclo 62 con un cambio interciclo de 0.00089. También, con el mismo programa, se estimaron los valores de  $\theta$  para los sujetos, los cuales tuvieron una media de 0,0081 y una desviación de 0.946.



**Figura 2.** Distribución de Theta

Un elemento importante para tener en cuenta al aplicar la TRI radica en evaluar el ajuste del modelo a los datos, para lo cual se utilizó el programa R comparando los ítems por parejas a modo de doublets (Ali, Drasgow, Williams, 2011) y la Macro de Excel Modfit.

Al aplicar el análisis a la muestra completa de 3057 casos, se encontró que todos los ítems mostraban diferencias significativas en al menos una pareja indicando desajuste, algo que parecía contradictorio en relación a las gráficas de ajuste que se obtuvieron con Modfit, por lo que se pensó explorar si el desajuste mostrado en los valores de  $\chi^2$  y de  $\chi^2/\text{gl}$  podría deberse a un efecto del tamaño muestral, con lo cual se estudió el ajuste con 300, 900, 1500, 2100 y 3000 casos. En la tabla 4, se puede observar la frecuencia y el porcentaje de

parejas de ítems que fueron significativos tomando el valor de  $\chi^2$  en cada una de los tamaños de muestra analizados.

Tabla 4. Análisis de ajuste al modelo

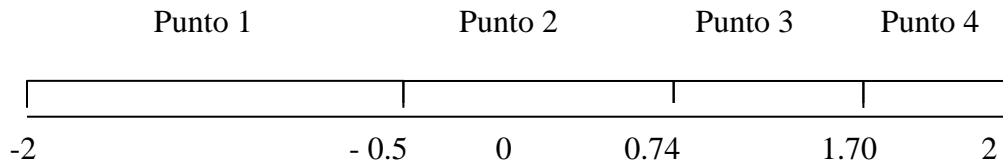
Tamaño de muestra	Frecuencia	Porcentaje
300	0	0%
900	8	2,90%
1500	25	9,06%
2100	52	18,84%
3000	123	44,57%

El porcentaje de parejas de ítems significativos aumenta con el incremento del tamaño de la muestra. El ítem 8 fue el que mayor desajuste mostró pero solo cuando se realizó el análisis con 2100 casos; al ver las gráficas no se evidenció un desajuste fuerte ni en este ni en otros por lo cual se concluyó que el modelo de TRI utilizado ajustaba de manera admisible a los datos.

El objetivo de llevar a cabo la calibración de los ítems con el modelo de respuesta graduada fue evaluar el comportamiento de los puntos de la escala Likert usada en el PAI, de tal manera que se pudiera explorar si en verdad la distancia entre el primer punto y el segundo (la cual representa la opción “Falso”) era mayor que las demás. En otras palabras, se buscó comprobar que existía asimetría en la distancia entre puntos. Con este propósito, se calculó la mediana en los parámetros b para los 24 ítems como indicador de cada punto. Si en verdad la distancia entre los dos primeros es mayor que las restantes, el dominio del primer punto debería ser mayor que el de los demás y las curvas estarían desplazadas hacia los niveles altos del rasgo evaluado. Por el contrario, si hay simetría, los dominios de cada opción serán relativamente similares. Con el objetivo de limitar el dominio posible de habilidad, se tomó en cuenta los valores entre -2 y 2.

Se utilizó como estadístico la mediana y no la media de los parámetros b, ya que el ítem 8 toma valores extremos que podrían distorsionar los promedios; ante esta situación, la mediana es menos sensible a la presencia de valores atípicos. La mediana para el punto b1 fue -0.50, para b2, 0.74 y para b3, 1.70, de tal manera que el área anterior al nivel de rasgo -0,50 correspondería a los valores en que es probable responder “falso” a los ítems de la

escala de Ansiedad del PAI; el espacio entre  $-0.50$  y  $0.74$ , señalaría el dominio de habilidad en que es más probable responder al segundo punto de la escala; la distancia entre  $0.74$  y  $1.70$  indicaría el tercer punto y el dominio superior a  $1.70$  representaría al cuarto punto de la escala Likert (Figura 3).



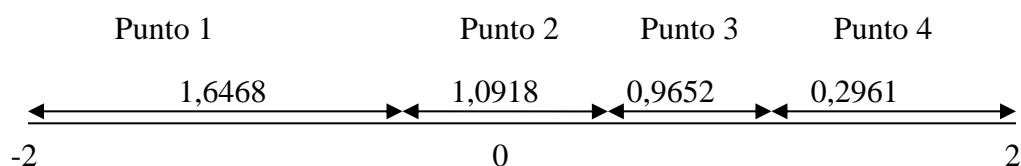
**Figura 3.** Distancia entre puntos primer método

Tomando como indicador las distancias entre los puntos, se puede señalar que la dimensión del punto 1 es de 1,5 unidades del continuo psicológico; el punto 2 tiene 1,24; el punto 3 tiene 0,96 unidades y mientras que el último cuenta con 0,30 unidades.

Se utilizó un segundo modo alterno al anterior para calcular las diferencias entre los puntos. Tomando nuevamente el dominio entre  $-2$  y  $2$  para el rasgo, se calcularon las distancias entre puntos para cada uno de los ítems, tomando en cada caso los valores de los parámetros  $b$  de igual manera como se hizo anteriormente. Con las distancias identificadas, se promedió para tener valores representativos de toda la escala. Los ítems 8, 14 y 15 se excluyeron de este análisis, ya que sus parámetros estaban por debajo de  $-2$ .

**Tabla 5.** Distancia entre los puntos de la escala Likert

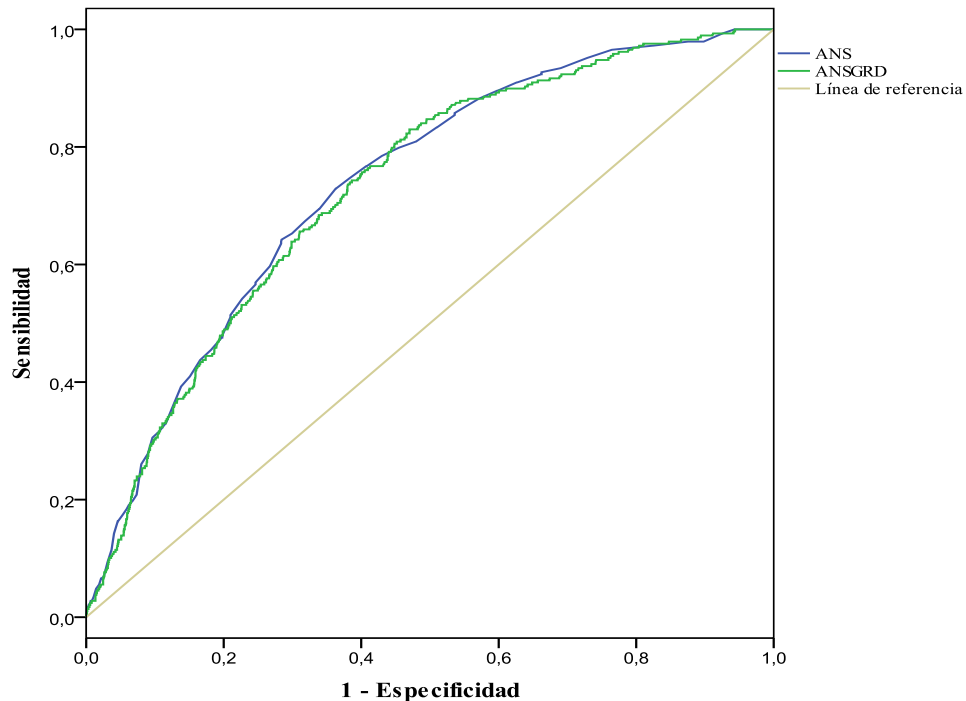
Ítem	Distancia Punto 1	Distancia Punto 2	Distancia Punto 3	Distancia Punto 4
Ítem 1	-1,5625	-1,1082	-0,8920	0,4373
Ítem 2	-1,9390	-0,9928	-0,8181	0,2501
Ítem 3	-2,0210	-0,8513	-0,6750	0,4527
Ítem 4	-0,8473	-1,2384	-1,2032	0,7112
Ítem 5	-0,4288	-1,2422	-1,2872	1,0419
Ítem 6	-2,7758	-0,5765	-0,5405	0,1073
Ítem 7	-0,3086	-1,7939	-1,6651	0,2325
Ítem 9	-1,4548	-0,9803	-0,7897	0,7752
Ítem 10	-1,7339	-1,0704	-0,7181	0,4776
Ítem 11	-1,9056	-1,0390	-0,7990	0,2564
Ítem 12	-2,3731	-1,0782	-0,8211	-0,2724
Ítem 13	-1,5071	-1,3599	-1,1315	0,0015
Ítem 16	-0,3659	-0,9048	-1,1120	1,6173
Ítem 17	-1,3897	-1,2888	-1,0230	0,2984
Ítem 18	-2,7586	-0,7665	-0,6216	-0,1466
Ítem 19	-2,1316	-0,9257	-0,8714	0,0713
Ítem 20	-1,4912	-1,1487	-1,0608	0,2992
Ítem 21	-2,0839	-0,8238	-0,7267	0,3657
Ítem 22	-3,6072	-1,4425	-1,2163	-2,2660
Ítem 23	-0,2508	-1,2051	-1,3328	1,2113
Ítem 24	-0,86216	-1,95881	-1,87653	-0,6975
Promedio	-1,6468	-1,0918	-0,9652	0,2961

**Figura 4.** Distancia entre puntos segundo método

Las distancias que se calcularon anteriormente por el primer método expuesto son muy similares a las que se encuentran con el segundo. Con estos datos parece confirmarse la asimetría entre los puntos de la escala de Ansiedad: el punto 1 ocupa un mayor espacio en el continuo, los puntos dos y tres tienen distancias similares entre ellos y el punto 4 es el que menor espacio ocupa.

*Predicción de un criterio externo*

Luego de identificar la presencia de asimetría entre los puntos, se exploró las implicaciones que esto tendría a la hora de predecir un criterio. El modelo de respuesta graduada permite que los valores de  $\theta$  estimados para los sujetos sean sensibles a las distancias entre los puntos, algo que no sucede en el método de calificación en donde se otorga un valor constante a cada punto. Por lo tanto se consideró relevante estudiar el efecto de la asimetría a la hora de predecir una variable externa. En los datos de la muestra española del PAI, se preguntó por la presencia de alguna patología o problema psicológico diagnosticado en los participantes de manera externa a la calificación de la prueba, con lo cual se contó con 288 pacientes diagnosticados con problemas de Ansiedad. A partir de esto se estudió la capacidad diagnóstica de la escala calificada de dos maneras, por el método clásico y por los valores de  $\theta$ , para clasificar de manera correcta a los participantes, que se sabe de antemano, tenían este cuadro clínico. La figura 5 muestra el análisis de la curva COR aplicado a estos valores.



**Figura 5.** Curva COR Puntuación clásica y Theta



El área bajo la curva para los puntajes con el modelo clásico es de 0,735 y para los valores de  $\theta$  es de 0,730. La diferencia entre los valores es bastante pequeña, por lo cual puede indicarse que ambos modos de puntuación son similares; el grado de asimetría que se encuentra en los ítems de Ansiedad no parece afectar la clasificación realizada.

Por medio de regresión logística binaria, se estimó el porcentaje de clasificación correcta para ambas maneras de calificar la escala. Como la muestra era tan grande en comparación con el sub grupo de pacientes diagnosticados con ansiedad (muestra total=3057 y muestra de diagnosticados=288), el análisis se aplicó a una sub-muestra de 576 casos compuesta por los pacientes con diagnóstico y 288 casos aleatorios seleccionados de los casos sin diagnóstico.

**Tabla 6.** Porcentaje de clasificación correcta MRG

Respuesta Graduada			
Observado	Pronosticado		% correcto
	Sin diagnostico	Con diagnóstico	
Sin diagnóstico	199	89	69,1
Con diagnóstico	91	197	68,4
% global			68,8

**Tabla 7.** Porcentaje de clasificación correcta Puntuación clásica

Puntuación Clásica			
Observado	Pronosticado		% correcto
	Sin diagnóstico	Con diagnóstico	
Sin diagnóstico	202	86	70,1
Con diagnóstico	94	194	67,4
% global			68,8

El porcentaje de clasificación correcto para ambas calificaciones es de 68,8%. Hay una insignificante diferencia en sensibilidad y especificidad; al parecer las puntuaciones con el modelo de respuesta graduada tienen un punto más de sensibilidad, mientras que la puntuación clásica tiene un punto más de especificidad, pero la diferencia es tan pequeña que no merece mayor mención. También los valores de R cuadrado de Negelkerke son similares, 0,288 para el modelo de respuesta graduada y 0,246 para la puntuación clásica.

### *Simulación de datos*

Para clarificar los resultados encontrados en los análisis de la muestra española del PAI, se recurrió a la simulación de datos. Se estudió el efecto de la asimetría en los puntos de una escala Likert variando el espacio ocupado por el primer punto para conocer su efecto sobre la fiabilidad (recuperación de valores originales de rasgo) y en la predicción de tres criterios externos (variables simuladas con correlaciones de 0,3 0,5 y 0,7). Se revisó cuidadosamente el sesgo de cada una de las simulaciones para evaluar la calidad de proceso realizado, la tabla 8 resume los valores de sesgo de cada uno de los parámetros estimados a lo largo de todas las condiciones y niveles del estudio.

**Tabla 8.** Medición del sesgo en las réplicas

Condición 1				
	Estimados	Simetría	Asimetría 1	Asimetría 2
Sesgo A	-0,0009	0,0061	-0,0816	-0,0035
Sesgo B1	-0,0126	0,0069	-0,0488	-0,0056
Sesgo B2	-0,0089	0,001	-0,0786	0,002
Sesgo B3	-0,0044	0,0057	-0,0987	0,0018
Condición 2				
	Estimados	Simetría	Asimetría 1	Asimetría 2
Sesgo A	-0,0008	0,0052	0,0006	0,0061
Sesgo B1	-0,002	0,0057	0,012	0,0061
Sesgo B2	-0,005	0,0003	0,0101	0,0062
Sesgo B3	0,0016	-0,0054	0,0085	0,006
Condición 3				
	Estimados	Simetría	Asimetría 1	Asimetría 2
Sesgo A	-0,003	0,0005	-0,0033	-0,0048
Sesgo B1	0,02638	0,0001	0,0032	0,0012
Sesgo B2	0,0295	0,0014	0,0033	-0,0021
Sesgo B3	0,044	0,0026	0,0034	-0,0062
Condición 4				
	Estimados	Simetría	Asimetría 1	Asimetría 2
Sesgo A	-0,0013	-0,002	-0,002	-0,0042
Sesgo B1	0,0016	-0,0024	0,0061	-0,0029
Sesgo B2	-0,0013	-0,0012	0,0011	-0,0074
Sesgo B3	-0,0026	-0,0031	-0,005	-0,0117

Los valores de sesgo son bastante pequeños, por lo cual se puede indicar que el proceso de simulación de datos fue adecuado. La generación de datos y las estimaciones que se realizaron no muestran sesgo, algo que le da fuerza a las conclusiones que se desprendan de los análisis. También se estudió la recuperación de los parámetros originales de la simulación, se muestran en la Tabla 9 los resultados obtenidos con los parámetros del nivel de asimetría 1, en el que se usaron como valores de inicio los parámetros estimados de la escala de Ansiedad.

**Tabla 9.** Recuperación de parámetros

	Parámetros Estimados			
	Condición 1	Condición 2	Condición 3	Condición 4
A	.9961	.9929	.9946	.9936
B1	.9982	.9965	.9988	.9986
B2	.9984	.9969	.9986	.9990
B3	.9955	.9911	.9967	.9963

Luego de revisar la calidad del proceso de simulación, se evaluó la fiabilidad y la capacidad de predicción de las puntuaciones clásica y  $\theta$  por medio de la correlación. La tabla 10 resume los valores promedio de correlación para las replicas de cada uno de los niveles y condiciones.

En cada una de las tablas se observa el valor de correlación de la Z real con las variables simuladas como criterio, este dato se incluyó en las tablas para confirmar que efectivamente se logró simular de manera correcta la relación de esta con su criterio. Los valores muestran que el proceso de simulación fue exitoso y logró valores muy precisos de relación con el criterio, en algunos de ellos exactos y en otros con diferencia de algunas centésimas. Este valor es el punto de comparación para los valores medios de correlación de las puntuaciones clásicas y de  $\theta$ .

La columna de fiabilidad indica cual es la relación que guarda cada una de las formas de puntuación con la Z real, lo cual explica por qué la relación de esta consigo misma sea 1. Al trabajar con los parámetros estimados en los ítems del PAI y con el nivel de simetría la fiabilidad es alta, siendo levemente superior en la simulación simétrica, pero esta decae cuando se introducen los niveles asimétricos. En el primer grado de asimetría simulado la

fiabilidad cae, pero en el nivel de mayor asimetría el descenso es muy marcado. En todos los niveles la puntuación  $\theta$  se muestra más fiable y más robusta que la puntuación clásica, al parecer la asimetría afecta de manera distinta las puntuaciones.

También la calidad de la predicción del criterio simulado se ve reducida con la introducción de la asimetría. Por el contrario la simulación simétrica mantiene la mejor predicción de los tres criterios tanto para puntuación clásica como para  $\theta$ . Se observa además que la correlación media es menor cuando se predice criterios que guardan una relación más fuerte con la Z real, en todos los casos el mayor descenso se encuentra en el criterio correlacionado al 0,7.

La puntuación  $\theta$  nuevamente es más robusta para predecir los tres criterios, pese a que se ve afectada por la asimetría su descenso no es tan fuerte como sí sucede en la puntuación clásica.

Las condiciones 1 y 2 tienen resultados similares, se encuentra en ambas el mismo patrón de resultados los valores son muy cercanos para todos los niveles en Fiabilidad y en relación con el criterio externo, al parecer no hay diferencias entre estas dos maneras de simulación.

Cuando se reduce la Fiabilidad por vía de reducción de los ítems (condición con 12 ítems) se observa un fuerte descenso de esta en los niveles de asimetría y una disminución leve en el nivel simétrico y en el de parámetros estimados. De nuevo la asimetría afecta más a la puntuación clásica que la  $\theta$ , a tal grado que en asimetría 2 la Fiabilidad se reduce a 0,8186. Sin embargo, la caída que tiene la puntuación  $\theta$  también es importante, llegando a 0,8583.

Las dos puntuaciones también reducen marcadamente la capacidad predictiva de los criterios simulados, algo mucho más evidente para la puntuación clásica que en el nivel de mayor asimetría tiene una correlación de 0,5733 para el criterio simulado como 0,7.

**Tabla 10.** Medidas de Fiabilidad y correlación con los criterios simulados

Z real distinta																
	Estimados				Simetría				Asimetría 1				Asimetría 2			
	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad
Z real	0,2996	0,499	0,7001	1	0,2996	0,499	0,7001	1	0,3001	0,5	0,6998	1	0,2997	0,4991	0,7001	1
Clásica	0,283	0,4719	0,6621	0,9533	0,2857	0,476	0,6679	0,9537	0,2793	0,465	0,6505	0,9297	0,2575	0,4289	0,6022	0,8598
$\theta$	0,2887	0,4813	0,6752	0,9642	0,2905	0,4841	0,6792	0,97	0,2878	0,4794	0,6707	0,9583	0,2732	0,4551	0,6389	0,9122
Z real igual																
	Estimados				Simetría				Asimetría 1				Asimetría 2			
	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad
Z real	0,3009	0,4998	0,7004	1	0,2997	0,2988	0,6996	1	0,2995	0,5004	0,6995	1	0,2988	0,4991	0,6999	1
Clásica	0,285	0,4723	0,6622	0,9455	0,2858	0,4752	0,6674	0,9537	0,2785	0,4651	0,6504	0,9297	0,2567	0,4298	0,6017	0,8599
$\theta$	0,2906	0,4817	0,6754	0,9642	0,2908	0,4836	0,6789	0,9701	0,2871	0,4796	0,6704	0,9584	0,2727	0,4559	0,6383	0,9121
12 Ítems																
	Estimados				Simetría				Asimetría 1				Asimetría 2			
	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad
Z real	0,3002	0,5001	0,6999	1	0,2999	0,4999	0,7002	1	0,2997	0,4999	0,7	1	0,2997	0,5001	0,6999	1
Clásica	0,2736	0,456	0,6379	0,9116	0,275	0,4586	0,6429	0,9175	0,2677	0,4461	0,6248	0,8927	0,2453	0,4093	0,5733	0,8186
$\theta$	0,2793	0,4652	0,6507	0,9298	0,2813	0,4691	0,6574	0,9385	0,2764	0,4605	0,6449	0,9213	0,2574	0,429	0,601	0,8583
6 Ítems																
	Estimados				Simetría				Asimetría 1				Asimetría 2			
	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad	0,3	0,5	0,7	Fiabilidad
Z real	0,301	0,5001	0,6996	1	0,2998	0,5007	0,6998	1	0,3004	0,4992	0,6996	1	0,2998	0,5	0,7	1
Clásica	0,2401	0,3988	0,5573	0,7968	0,2443	0,4074	0,5693	0,8130	0,2373	0,3933	0,5517	0,7881	0,2131	0,3564	0,4982	0,7117
$\theta$	0,2483	0,4126	0,5765	0,8241	0,2576	0,4294	0,6002	0,8573	0,251	0,416	0,5832	0,8336	0,2258	0,3783	0,5291	0,7554

Las tendencias que marcan los resultados de la condición 3 se agravan para la condición 4; al reducir la fiabilidad trabajando con 6 ítems el nivel de simetría se muestra como el más robusto en términos de fiabilidad y en predicción de los criterios para ambas formas de puntuación, seguido por el nivel en que utiliza los parámetros estimados en el PAI, pero en los niveles de asimetría 1 y 2 el efecto es muy fuerte, la fiabilidad cae y los criterios no son predichos de manera precisa. Aunque también confirma que la puntuación  $\theta$  es más robusta que la puntuación clásica, en el mayor grado de exigencia al que se llevó las puntuaciones ambas fallan en predecir los criterios

### Discusión y Conclusiones

El propósito de este estudio fue explorar las consecuencias metodológicas que tiene el uso de ítems Likert con asimetría en la primera categoría graduada sobre la fiabilidad y la predicción de criterios. Los análisis realizados permiten señalar que sí hay un efecto de la asimetría sobre ambos elementos; existe una relación negativa entre el grado de asimetría de los ítems con la fiabilidad y su capacidad para predecir criterios externos.

En todas las condiciones simuladas el nivel denominado de simetría entre puntos se mostró como el más robusto en fiabilidad y capacidad de predicción, seguido de cerca por la condición que utilizó los parámetros estimados en la escala de Ansiedad. En la condición de simetría, además, se evidenció la menor diferencia entre la puntuación clásica y la puntuación  $\theta$ , por lo cual, se puede pensar que la suma de puntuaciones necesita ítems con baja asimetría entre categorías para funcionar de manera adecuada. Las puntuaciones  $\theta$  fueron sistemáticamente robustas en comparación con las clásicas, en ninguna ocasión se mostró que el patrón se invirtiera, lo cual apoya la idea que los modelos de TRI son sensibles a elementos que la sumatoria simple de respuestas no puede detectar.

Tomando como base estos resultados es adecuado cuestionarse si vale la pena cambiar la manera como actualmente se puntúa la escala por valores estimados a partir de un modelo de respuesta graduada, los cuales al parecer son más sólidos en su funcionamiento. Pero la diferencia de los resultados entre ambas puntuaciones no es suficientemente amplia para apoyar una decisión como esta; en el presente estudio se llega

a la misma conclusión que Dumenci y Achenbach (2008) y Xu y Stone (2012) quienes aseguran que la diferencia entre ambas es mínima; es verdad que hay una ganancia en fiabilidad y en capacidad de predicción, pero estas se encuentran en un rango entre una y cuatro centésimas. En términos prácticos no parece razonable optar por la aplicación de un modelo TRI por una ganancia tan minúscula, menos aún si se tiene en cuenta las dificultades técnicas que se tuvieron para lograr calibrar los ítems de la escala; para hacerlo sería necesario reelaborar completamente el test, para que sus 22 escalas se ajusten a criterios de unidimensionalidad.

Tampoco se logró demostrar con los datos de la muestra Española del PAI que se alcance un mayor porcentaje de clasificación correcta con la puntuación  $\theta$  que con la clásica, ambas alcanzan el mismo número de aciertos totales, lo cual deja en desventaja práctica a la TRI, ya que al parecer con la sumatoria de respuestas se alcanzarían similares resultados.

Siguiendo con esta línea de argumentación debe recordarse que los ítems del PAI están pensados para evaluar ocurrencia e intensidad de los reactivos, lo cual pone en tela de juicio la válida aplicación de un modelo de TRI ya que no es del todo claro que se respete el supuesto de unidimensionalidad. Quizá a esto se deba las dificultades que se tuvieron para seleccionar una escala del PAI para el presente estudio; la escala de Ansiedad cumple con el ajuste a la unidimensionalidad pero en las demás podría haber un efecto del estilo de los ítems. Para siguientes investigaciones podría trabajarse con modelos de TRI multidimensional con el objetivo de evaluar las dos dimensiones implícitas en los ítems y evaluar su funcionamiento.

También genera algunas dudas el hecho que la intensidad y la ocurrencia sean medidas con ítems de nivel de medida distinto, mientras que la primera sería evaluada con una medida ordinal (Ligeramente verdadero, bastante verdadero, completamente verdadero), la segunda lo haría con una medida nominal (Si ocurre- No ocurre). Esta propiedad de los ítems debe ser estudiada a profundidad ya que no es claro cual es su alcance y sus consecuencias.

Quizá el resultado de mayor interés que se desprende de este estudio sea la identificación de la asimetría como un elemento moderador de la capacidad predictiva y la fiabilidad de una escala. Este si constituye un peligro a tener en cuenta en el desarrollo de ítems de respuesta graduada ya que la excesiva relación asimétrica puede llevar a problemas importantes en la puntuación. En la escala de Ansiedad los ítems 2, 3, 6, 11 y 21 tiene parámetros similares a los que en este trabajo se han denominado Asimetría 1, en donde el valor del parámetro  $b_1$  es cercano a 0, por lo tanto pueden estar contribuyendo a un funcionamiento inadecuado de la escala, por otro lado, el ítem 22 sería un ejemplo de Asimetría 2, el máximo nivel que se reprodujo con las simulaciones. Lo que no es del todo claro y no puede conocerse a partir de los resultados actuales es el porcentaje de ítems que pueden tolerarse en la escala con problemas de asimetría, puede suceder que estos seis ítems identificados que representan el 25% de la escala no estén afectando la medición final, mucho mas cuando los ítems 4, 5, 7, 16, 23 y 24 serían casos de simetría. Es algo que debe resolverse de manera empírica y que se desprende como una posible investigación futura a partir de los resultados presentes.

Una de las limitaciones del trabajo realizado puede estar en el rango del dominio psicológico escogido para efectuar las simulaciones, se consideró que tomar los valores entre -2 y 2 era bastante representativo de la situación, pero es necesario replicar el estudio con un rango más amplio. Los resultados actuales se limitan al funcionamiento que puedan tener los ítems que se muevan en el rango delimitado, pero no pueden generalizarse a valores por fuera de este. Además de lo anterior, al fijar este rango de trabajo se introdujo de manera implícita una variable extraña que no fue controlada ni se identificó a tiempo, esta es la reducción gradual del espacio ocupado por los puntos 2, 3 y 4. En la planificación de la simulación se buscó ampliar el rango del punto 1, pero no se tuvo en cuenta que a la vez se estaba reduciendo el espacio de los otros puntos del ítem. Hay un claro efecto de reducción de fiabilidad y capacidad predictiva en el presente estudio, pero no es claro que los resultados obtenidos se deban de manera única a la asimetría, podría haber un peso importante en la reducción gradual que tuvieron los demás puntos de la escala, ya que en el nivel de Asimetría 2, los parámetros  $b_2$  y  $b_3$  estuvieron separados tan solo por 0,33 unidades del continuo psicológico, lo cual parece poco y puede llevar a pensar que estas se fundieron en una sola. Lo adecuado para la investigación pudo haber sido mantener para



todos los niveles y condiciones distancias iguales en los puntos 2, 3 y 4 de manera que al aumentar el espacio del punto 1, no se diera un ningún cambio en el rango que estos ocuparan.

Los ítems del PAI representan un cambio en la construcción de ítems para instrumentos de personalidad clínica que pueden llevarse a otros espacios de aplicación, pero es indispensable que se continúe con la investigación de sus propiedades psicométricas, sus fortalezas y debilidades y los modelos psicométricos adecuados para describir su funcionamiento. En el estudio realizado se trabajó a partir del Modelo de Respuesta Graduada, pero la aplicación de otros modelos como el nominal o el de crédito parcial pueden abrir nuevas alternativas para su comprensión.

## Referencias

- Abad, F.J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.
- Abad, F.J. y Revuelta, J. (2006). *Modelos Politómicos de Respuesta al Ítem*. Cuadernos de Estadística. Madrid: Muralla.
- Asún, R. Zúñiga, C. (2008). Ventajas de los Modelos Politómicos de Teoría de Respuesta al Ítem en la Medición de Actitudes Sociales, El Análisis de un Caso. *PSYKHE*, 17(2), 103-115.
- Barbero, M. I. (2007). *Métodos de Elaboración de Escalas*. Unidad didáctica. Madrid: UNED.
- Barbero, M. I., Prieto, P., Suárez, J. y San Luis, C. (2001). Relaciones empíricas entre los estadísticos de la teoría clásica de los test y los de la teoría de respuesta a los ítems. *Psicothema*, 13(2) 324-329.
- Corsica, J. a, Azarbad, L., McGill, K., Wool, L., & Hood, M. (2010). The Personality Assessment Inventory: clinical utility, psychometric properties, and normative data for bariatric surgery candidates. *Obesity surgery*, 20(6), 722–31.
- Dumenci, L. y Achenbach, T. (2008). Effects of Estimation Methods on Making Trait-Level Inferences From Ordered Categorical Items for Assessing Psychopathology. *Psychological Assessment*, 20(1) 55–62.
- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3) 357-381.
- Karlin, B. E., Creech, S. K., Grimes, J. S., Clark, T. S., Meagher, M. W., & Morey, L. C. (2005). The Personality Assessment Inventory with chronic pain patients: Psychometric properties and clinical utility. *Journal of clinical psychology*, 61(12), 1571–85.

- Macdonald, P., Paunonen, S. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement*. 62(6) 921-943.
- Morey, L (1991). *The personality Assessment Inventory Professional Manual*. Odessa. New York: Psychological Assessment Resources.
- Morey, L. (2007). *Personality assessment inventory PAI*. New York: Psychological Assessment Resources, Inc.
- Nunnally, J. (1979). *Psychometric theory* (2da ed). New York: Mcgraw-Hil.
- Morey, L. (1997) *Personality Assessment Screener*. Odessa, Fl: Psychological Assessment Resources.
- Ortiz-Tallo, M., Santamaría, P., Cardenal, V. y Sánchez, Ma. P. (2011). *Inventario de Evaluación de la Personalidad. Adaptación española*. Manual Técnico. Madrid: TEA Ediciones. S.A.
- Ortiz-Tallo, M., Santamaría, P., Cardenal, V. y Sánchez, Ma. P. (2011). *Inventario de Evaluación de la Personalidad. Adaptación española*. Manual de Aplicación, Corrección e Interpretación. Madrid : TEA Ediciones. S.A.
- Siefert, C. J., Sinclair, S. J., Kehl-Fie, K. a, & Blais, M. a. (2009). An item-level psychometric analysis of the personality assessment inventory: clinical scales in a psychiatric inpatient unit. *Assessment*, 16(4), 373–83.
- Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological Distance Between Categories in the Likert Scale: Comparing Different Numbers of Options. *Educational and Psychological Measurement*, 72(4), 533–546.
- Xu, T. y Stone, C. (2012). Using IRT Trait Estimates Versus Summated Scores in Predicting Outcome. *Educational and Psychological Measurement*. 72(3) 453 –468.

## Anexos

### Anexo 1. Descripción de las escalas del PAI

Escala	Descripción
Escalas de validez	
Inconsistencia	Indica si el evaluado responde de forma inconsistente a lo largo de la prueba. Para ello tiene en cuenta su respuesta a pares de ítems alternamente relacionados entre sí
Infrecuencia	Indica si el evaluado ha respondido poco cuidadosamente, al azar o de una forma muy peculiar. Los ítems son neutrales con respecto a psicopatología y resultan muy poco frecuentes tanto en personas normales como en pacientes clínicos
Impresión negativa	Indica si el evaluado está presentando una imagen de sí mismo extremadamente desfavorable que pueda sugerir exageración de síntomas
Impresión positiva	Indica si el evaluado está presentando una imagen de sí mismo muy favorable y evita admitir pequeños defectos
Escalas clínicas	
Quejas somáticas	Evalúa las preocupaciones relacionadas con la salud física y las quejas somáticas habitualmente presentes en los tratamientos de somatización o de conversión.
Ansiedad	Evalúa las manifestaciones y los signos observables de ansiedad con especial énfasis en la evaluación de sus distintas modalidades (cognitiva, emocional y fisiológica).
Trastornos relacionados con la ansiedad	Evalúa los síntomas y conductas relacionadas con trastornos específicos de la ansiedad, específicamente fobias estrés postraumático y síntomas obsesivos compulsivos.
Depresión	Evalúa las manifestaciones y síntomas de los trastornos depresivos.
Manía	Evalúa los síntomas afectivos, cognitivos y conductuales de la manía y la hipomanía.
Paranoia	Evalúa los síntomas de los trastornos paranoides y las características más estables de la personalidad paranoide.
Esquizofrenia	Evalúa los síntomas clave del amplio espectro de los trastornos esquizofrénicos.
Rasgos límites	Evalúa los atributos indicativos de la personalidad límite, entre los que se incluyen las relaciones interpersonales inestables y fluctuantes, la impulsividad, la inestabilidad, la labilidad emocional y la ira incontrolada.
Rasgos antisociales	Evalúa el historial de actos ilegales y problemas con la autoridad, así como el egocentrismo, la falta de empatía y lealtad, la inestabilidad y la búsqueda de sensaciones.
Problemas con el alcohol	Evalúa las consecuencias negativas del consumo de alcohol y los rasgos indicativos de su dependencia.
Problemas con las drogas	Evalúa las consecuencias negativas del uso de drogas y los rasgos indicativos de su dependencia.

Escalas relacionadas con el tratamiento	
Agresión	Evalúa las características y actitudes relacionadas con la ira, la asertividad, la hostilidad y la agresión.
Ideaciones suicidas	Evalúa la ideación suicida cubriendo un amplio rango que va desde la desesperanza hasta los pensamientos y planes concretos de suicidio.
Estrés	Mide el impacto de circunstancias o situaciones estresantes recientes en la vida del sujeto.
Falta de apoyo social	Evalúa la falta de apoyo social percibido teniendo en cuenta el nivel y la calidad del apoyo disponible.
Rechazo al tratamiento	Evalúa aquellos atributos y actitudes que indican una falta de interés y motivación para hacer cambios personales psicológicos o emocionales.
Escalas de relación interpersonal	
Dominancia	Evalúa el grado en que una persona es controladora e independiente en sus relaciones personales. Las puntuaciones altas están asociadas a un estilo dominante, mientras que las puntuaciones bajas reflejan un estilo sumiso.
Afabilidad	Evalúa el grado en que una persona se interesa por aquellas relaciones personales empáticas y de apoyo. Las puntuaciones altas están asociadas a un estilo afable, sociable y las puntuaciones bajas a un estilo frío e indiferente.

**Anexo 2.** Sintaxis de Matlab utilizada para la simulación

```

clear all;
muestras=1000;
sujetos=3000;
nitems=12;
ncatego=4;
fidsal1= fopen('salida1.dat', 'wt');
fclose(fidsal1);
fidpar1= fopen('PAR1.dat', 'wt');
fclose(fidpar1);
fidcor1= fopen('crit1.dat', 'wt');
fclose(fidcor1);
fidsal2= fopen('salida2.dat', 'wt');
fclose(fidsal2);
fidpar2= fopen('PAR2.dat', 'wt');
fclose(fidpar2);
fidcor2= fopen('crit2.dat', 'wt');
fclose(fidcor2);
fidsal3= fopen('salida3.dat', 'wt');
fclose(fidsal3);
fidpar3= fopen('PAR3.dat', 'wt');
fclose(fidpar3);
fidcor3= fopen('crit3.dat', 'wt');
fclose(fidcor3);
fidsal4= fopen('salida4.dat', 'wt');
fclose(fidsal4);
fidpar4= fopen('PAR4.dat', 'wt');
fclose(fidpar4);
fidcor4= fopen('crit4.dat', 'wt');
fclose(fidcor4);
for h=1:muestras

```

\*\*\*\*\*DEFINICIÓN DE Z REAL Y VALORES DE LA DISTRIBUCIÓN UNIFORME PARA GENERAR PATRONES DE RESPUESTA\*\*\*\*\*

```

Z=randn(sujetos,1);
a=rand(sujetos,nitems);
s11=0.3*Z+0.953939*randn(3000,1);

```

\*\*\*\*\*SIMULACIÓN DE VARIABLES CRITERIO CORRELACIONADAS A 0.3, 0.5 Y 0.7

```

COR1=corr2(Z,s11)
s12=0.5*Z+0.866*randn(3000,1);
COR2=corr2(Z,s12)
s13=0.7*Z+0.71414*randn(3000,1);
COR3=corr2(Z,s13)

```

\*\*\*\*\*PARÁMETROS DE LOS ÍTEMS, aa INDICA LOS VALORES DEL PARÁMETRO a Y bbLOS VALORES DE LOS PARÁMETROS b ESTIMADOS CON MULTILOG. EN ESTE CASO SE OBSERVAN LOS VALORES DE LOS 24 ÍTEMS DE LA ESCALA DE ANSIEDAD, PARA GENERAR LA SIMULACIÓN CON 12 O 6 SE SELECCIONAN AL AZAR LOS ÍTEMS Y SE INTRODUCEN SUS PARÁMETROS EN ESTE PUNTO \*\*\*\*\*

```
aa = [2.05215    1.94911 3.24906 1.39481 1.59216 2.13648 0.98768 0.52629 2.26231 2.16601 1.74845
1.52967 1.36948 0.57646 1.07976 1.4758  1.9583  2.23845 1.50748 1.27719 1.74716 0.6877  1.42993
0.75933];
```

```
bb1 =[-0.43746, 0.67075, 1.56274;
-0.06103, 0.93175, 1.74989;
0.02099, 0.87227, 1.54729;
-1.15271, 0.08565, 1.28884;
-1.57122,-0.32904, 0.95813;
0.77575, 1.35225, 1.8927;
-1.69145, 0.10246, 1.76752;
-4.64706,-2.35678, 0.42895;
-0.54524, 0.43504, 1.22477;
-0.2661, 0.8043, 1.52237;
-0.09437, 0.9446, 1.74357;
0.37305, 1.45128, 2.27235;
-0.4929, 0.86698, 1.99847;
-2.80474,-0.42338, 2.00902;
-2.08114, -0.75258, 0.71363;
-1.63411, -0.7293, 0.3827;
-0.61028, 0.67854, 1.70158;
0.75857, 1.52504, 2.14659;
0.13162, 1.05733, 1.92871;
-0.50876, 0.63993, 1.70076;
0.08388, 0.90767, 1.63434;
1.60723, 3.0497, 4.266;
-1.7492, -0.54409, 0.78873;
-1.13784, 0.82097, 2.6975];
```

\*\*\*\*\*DEFINICIÓN DE MATRICES, BUCLES\*\*\*\*\*

```
pa1=zeros(nitems,(ncatego+1));
p1=zeros(nitems,ncatego);
t1=zeros(nitems,(ncatego-1));
r1=zeros(sujetos,nitems);
for i=1:sujetos;
for j=1:nitems;
for k=1:(ncatego+1);
    if k==1 ;
        pa1(j,k)=1;
    elseif k==(ncatego+1);
        pa1(j,k)=0;
    else
        pa1(j,k)=1/(1+(exp((-1*aa(j))*(Z(i)-bb1(j,k-1)))));
    end
end
for k=1:(ncatego);
    p1(j,k)=pa1(j,k)-pa1(j,k+1);
end
for k=1:(ncatego);
    if k==1;
        t1(j,k)=p1(j,k);
    else
        t1(j,k)=t1(j,k-1)+p1(j,k);
    end
end
for k=1:(ncatego);
```

```

    if a(i,j) < t1(j,k);
        s1=a(i,j);
        m1=t1(j,k);
        r1(i,j)=k;
        break
    end
end
end
end

*****GENERACIÓN DEL ARCHIVO CON LOS PATRONES DE RESPUESTA PARA QUE SEA
RETOMADO POR MULTILOG*****
fid1 = fopen('ANS.dat', 'wt');
for i=1:sujetos;
    fprintf(fid1, '%5d', (i));
    for j=1:nitems;
        fprintf(fid1, '%1d', r1(i,j));
    end
    fprintf(fid1, '\n', r1(i,j));
end
fclose(fid1);

*****LLAMADA AL ARCHIVO DE EXTENCIÓN.BAT, ESTE EJECUTA MULTILOG REALIZANDO
LA CALIBRACIÓN DE LOS PARÁMETROS Y LA ESTIMACIÓN DE THETA*****

system('sim.bat');

*****CÁLCULO DE LA PUNTUACIÓN CLÁSICA*****
PC1=sum(r1,2);

*****SE IMPORTA EL ARCHIVO .SCO CON LAS PUNTUACIONES THETA. PARA EVITAR QUE LA
SIGUIENTE RÉPLICA SE GUARDE ENCIMA DE LA ANTERIOR, CON EL EJECUTABLE . BAT SE
CAMBIÓ EL NOMBRE Y LA EXTENCIÓN DEL ARCHIVO DE SALIDA DE MULTILOG ANS1.SCO
POR THETA.DAT. ESTE ÚLTIMO ES EL QUE SE IMPORTA*****

TH1=load('THETA.dat');

*****GENERACIÓN DE ARCHIVO DE SALIDA CON LA Z REAL, LA PUNTUACIÓN THETA Y
CLÁSICA*****
salida1=[Z PC1 TH1];
fidsal1= fopen('salida1.dat', 'at');
for i=1:sujetos;
    fprintf(fidsal1, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h,
salida1(i,1),salida1(i,2),salida1(i,3),salida1(i,4),salida1(i,5));
end
fclose(fidsal1);
crit1=[Z s11 s12 s13];
fidcor1= fopen('crit1.dat', 'at');
for i=1:sujetos;
    fprintf(fidcor1, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h, i, crit1(i,1),crit1(i,2),crit1(i,3),crit1(i,4));
end
fclose(fidcor1);

*****SE IMPORTA EL ARCHIVO .PAR CON LOS PARÁMETROS. TAMBIÉN CON EL EJECUTABLE
.BAT SE CAMBIO EL ARCHIVO ANS.PAR POR PARAM.DAT*****

```



```

PR1= importdata('PARAM.dat');
fidpar1= fopen('PAR1.dat', 'at');
for j=1:nitems;
    fprintf(fidpar1, '%4d %8.4f %8.4f %8.4f %8.4f\n', h, PR1(j,1),PR1(j,2),PR1(j,3),PR1(j,4));
end
fclose(fidpar1);

*****INICIO DE LA SIMULACIÓN PARA EL NIVEL SIMÉTRICO (PARÁMETROS b CON VALORES
-1, 0, 1, LA ESTRUCTURA DEL PROGRAMA ES IGUAL A LA ANTERIOR*****
s21=0.3*Z+0.953939*randn(3000,1);
COR21=corr2(Z,s21)
s22=0.5*Z+0.866*randn(3000,1);
COR22=corr2(Z,s22)
s23=0.7*Z+0.71414*randn(3000,1);
COR23=corr2(Z,s23)

*****SE INTRODUCE EN ESTE PUNTO LOS VALORES DE LOS PARÁMETROS b. LOS VALORES
DE a SON CONSTANTES Y EL PROGRAMA TOMARÁ LOS QUE SE INTRODUCEN AL
PRINCIPIO*****
bb2 =[-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;
-1,0,1;];
pa2=zeros(nitems,(ncatego+1));
p2=zeros(nitems,ncatego);
t2=zeros(nitems,(ncatego-1));
r2=zeros(sujetos,nitems);
for i=1:sujetos;
for j=1:nitems;
for k=1:(ncatego+1);
    if k==1 ;
        pa2(j,k)=1;
    elseif k==(ncatego+1);
        pa2(j,k)=0;
    else
        pa2(j,k)=1/(1+(exp((-1*aa(j))*(Z(i)-bb2(j,k-1)))));
    end
end
for k=1:(ncatego);
    p2(j,k)=pa2(j,k)-pa2(j,k+1);
end
for k=1:(ncatego);
    if k==1;
        t2(j,k)=p2(j,k);
    else
        t2(j,k)=t2(j,k-1)+p2(j,k);
    end
end

```

```

end
for k=1:(ncatego);
    if a(i,j) < t2(j,k);
        s2=a(i,j);
        m2=t2(j,k);
        r2(i,j)=k;
        break
    end
end
end
end
end
fid2 = fopen('ANS.dat', 'wt');
for i=1:sujetos;
    fprintf(fid2, '%5d', (i));
    for j=1:nitems;
        fprintf(fid2, '%1d', r2(i,j));
    end
    fprintf(fid2, '\n', r2(i,j));
end
fclose(fid2);
system('sim.bat');
PC2=sum(r2,2);
TH2=load('THETA.dat');
salida2=[Z PC2 TH2];
fidsal2= fopen('salida2.dat', 'at');
for i=1:sujetos;
    fprintf(fidsal2, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h,
    salida2(i,1),salida2(i,2),salida2(i,3),salida2(i,4),salida2(i,5));
end

fclose(fidsal2);
crit2=[Z s21 s22 s23];
fidcor2= fopen('crit2.dat', 'at');
for i=1:sujetos;
    fprintf(fidcor2, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h, i, crit2(i,1),crit2(i,2),crit2(i,3),crit2(i,4));
end
fclose(fidcor2);
PR2= importdata('PARAM.dat');
fidpar2= fopen('PAR2.dat', 'at');
for j=1:nitems;
    fprintf(fidpar2, '%4d %8.4f %8.4f %8.4f %8.4f\n', h, PR2(j,1),PR2(j,2),PR2(j,3),PR2(j,4));
end
fclose(fidpar2);

*****INICIO DE LA SIMULACIÓN PARA ASIMETRÍA 1, CON VALORES DEL PARÁMETRO b (0,
0,7 y 1,4). LA ESTRUCTURA DEL PROGRAMA ES IGUAL A LAS ANTERIORES*****

s31=0.3*Z+0.953939*randn(3000,1);
COR31=corr2(Z,s31)
s32=0.5*Z+0.866*randn(3000,1);
COR32=corr2(Z,s32)
s33=0.7*Z+0.71414*randn(3000,1);
COR33=corr2(Z,s33)
bb3 =[0,0.7,1.4;
    0,0.7,1.4;
    0,0.7,1.4;

```

```

0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4;
0,0.7,1.4];
pa3=zeros(nitems,(ncatego+1));
p3=zeros(nitems,ncatego);
t3=zeros(nitems,(ncatego-1));
r3=zeros(sujetos,nitems);
for i=1:sujetos;
for j=1:nitems;
for k=1:(ncatego+1);
    if k==1 ;
        pa3(j,k)=1;
    elseif k==(ncatego+1);
        pa3(j,k)=0;
    else
        pa3(j,k)=1/(1+(exp((-1*aa(j))*(Z(i)-bb3(j,k-1))))));
    end
end
for k=1:(ncatego);
    p3(j,k)=pa3(j,k)-pa3(j,k+1);
end
for k=1:(ncatego);
    if k==1;
        t3(j,k)=p3(j,k);
    else
        t3(j,k)=t3(j,k-1)+p3(j,k);
    end
end
for k=1:(ncatego);
    if a(i,j) < t3(j,k);
        s3=a(i,j);
        m3=t3(j,k);
        r3(i,j)=k;
        break
    end
end
end
end
end
fid3 = fopen('ANS.dat', 'wt');
for i=1:sujetos;
fprintf(fid3, '%5d', (i));
for j=1:nitems;
fprintf(fid3, '%1d', r3(i,j));
end
fprintf(fid3, '\n', r3(i,j));
end

fclose(fid3);
system('sim.bat');
PC3=sum(r3,2);

```

```

TH3=load('THETA.dat');
salida3=[Z PC3 TH3];
fidsal3= fopen('salida3.dat', 'at');
for i=1:sujetos;
    fprintf(fidsal3, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h,
salida3(i,1),salida3(i,2),salida3(i,3),salida3(i,4),salida3(i,5));
end
fclose(fidsal3);
crit3=[Z s31 s32 s33];
fidcor3= fopen('crit3.dat', 'at');
for i=1:sujetos;
    fprintf(fidcor3, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h, i, crit3(i,1),crit3(i,2),crit3(i,3),crit3(i,4));
end
fclose(fidcor3);
PR3= importdata('PARAM.dat');
fidpar3= fopen('PAR3.dat', 'at');
for j=1:nitems;
    fprintf(fidpar3, '%4d %8.4f %8.4f %8.4f %8.4f\n', h, PR3(j,1),PR3(j,2),PR3(j,3),PR3(j,4));
end
fclose(fidpar3);

```

\*\*\*\*INICIO DE LA SIMULACIÓN PARA EL NIVEL DE ASIMETRÍA 2, CON PARÁMTEROS b (1, 1,3333 Y 1,6666). EL PROGRAMA TIENE LA MISMA ESTRUCTURA ANTERIOR\*\*\*\*

```

s41=0.3*Z+0.953939*randn(3000,1);
COR41=corr2(Z,s41)
s42=0.5*Z+0.866*randn(3000,1);
COR42=corr2(Z,s42)
s43=0.7*Z+0.71414*randn(3000,1);
COR43=corr2(Z,s43)
bb4 =[1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666;
1,1.3333,1.6666];
pa4=zeros(nitems,(ncatego+1));
p4=zeros(nitems,ncatego);
t4=zeros(nitems,(ncatego-1));
r4=zeros(sujetos,nitems);
for i=1:sujetos;
for j=1:nitems;
for k=1:(ncatego+1);
    if k==1 ;
        pa4(j,k)=1;
    elseif k==(ncatego+1);
        pa4(j,k)=0;
    else
        pa4(j,k)=1/(1+(exp((-1*aa(j))*(Z(i)-bb4(j,k-1))))));
    end
end
end

```

```

end
for k=1:(ncatego);
    p4(j,k)=pa4(j,k)-pa4(j,k+1);
end
for k=1:(ncatego);
    if k==1;
        t4(j,k)=p4(j,k);
    else
        t4(j,k)=t4(j,k-1)+p4(j,k);
    end
end
for k=1:(ncatego);
    if a(i,j) < t4(j,k);
        s4=a(i,j);
        m4=t4(j,k);
        r4(i,j)=k;
        break
    end
end
end
end
end
fid4 = fopen('ANS.dat', 'wt');
for i=1:sujetos;
    fprintf(fid4, '%5d', (i));
    for j=1:nitems;
        fprintf(fid4, '%1d', r4(i,j));
    end
    fprintf(fid4, '\n', r4(i,j));
end
fclose(fid4);
system('sim.bat');
PC4=sum(r4,2);
TH4=load('THETA.dat');
salida4=[Z PC4 TH4];
fidsal4= fopen('salida4.dat', 'at');
for i=1:sujetos;
    fprintf(fidsal4, '%4d      %8.4f      %8.4f      %8.4f      %8.4f      %8.4f\n',      h,
salida4(i,1),salida4(i,2),salida4(i,3),salida4(i,4),salida4(i,5));
end
fclose(fidsal4);
crit4=[Z s41 s42 s43];
fidcor4= fopen('crit4.dat', 'at');
for i=1:sujetos;
    fprintf(fidcor4, '%4d %8.4f %8.4f %8.4f %8.4f %8.4f\n', h, i, crit4(i,1),crit4(i,2),crit4(i,3),crit4(i,4));
end
fclose(fidcor4);
PR4= importdata('PARAM.dat');
fidpar4= fopen('PAR4.dat', 'at');
for j=1:nitems;
    fprintf(fidpar4, '%4d %8.4f %8.4f %8.4f %8.4f\n', h, PR4(j,1),PR4(j,2),PR4(j,3),PR4(j,4));
end
end
fclose(fidpar4);
end

```

**Anexo 3. Sintaxis de Multilog**

Se utilizaron dos sintaxis, la primera utiliza el comando RANDOM para la calibración y guarda el archivo.par

MULTILOG for Windows 7.00.2327.2

```
>PROBLEM RANDOM,
  INDIVIDUAL,
  DATA = 'ANS.DAT',
  NITEMS = 24,
  NGROUPS = 1,
  NEXAMINEES = 3000,
  NCHARS = 5;
>TEST ALL,
  GRADED,
  NC = (4(0)24);
>ESTIMATES NCYCLES=1000;
>SAVE;
>END ;
4
1234
11111111111111111111111111111111
22222222222222222222222222222222
33333333333333333333333333333333
44444444444444444444444444444444
(5A1,24A1)
```

La segunda sintaxis retoma el .par creado y calcula las puntuaciones theta generando el archivo .sco

MULTILOG for Windows 7.00.2327.2

```
>PROBLEM SCORE,
  INDIVIDUAL,
  DATA = 'ANS.DAT',
  NITEMS = 24,
  NGROUPS = 1,
  NEXAMINEES = 3000,
  NCHARS = 5;
>TEST ALL,
  GRADED,
  NC = (4(0)24);
>ESTIMATES NCYCLES=1000;
>START ALL, PARAM = 'ANS.PAR';
>SAVE;
>END ;
4
1234
11111111111111111111111111111111
22222222222222222222222222222222
33333333333333333333333333333333
44444444444444444444444444444444
(5A1,24A1)
```

#### **Anexo 4** Archivo Ejecutable.bat

```
@echo off
cd
del THETA.dat
del PARAM.dat
call MLG ANS
del ANS.out
call MLG ANS1
rename ANS1.SCO THETA.DAT
rename ANS.PAR PARAM.dat
exit
```