

RAN: un programa informático para estimar patrones de respuesta aleatorios en tests objetivos

César Merino Soto y José López Pina*

Universidad de San Martín de Porres y * Universidad de Murcia

En la medición psicológica, especialmente de personalidad, el método lápiz-papel es usual entre psicólogos y profesiones aliadas. La puntuación obtenida por un examinado puede tener una validez cuestionable si respondió aleatoriamente, e identificar este problema ha sido abordado por diferentes métodos. La respuesta aleatoria tiende a ser una situación desafiante para el diagnóstico y la investigación en la medición de la personalidad y desempeño cognitivo evaluados mediante preguntas de tipo dicotómico. En este trabajo se presenta un programa informático que implementa el método de estimación puntual y por intervalos de confianza de Charter, para identificar si la puntuación obtenida por un examinado es probablemente aleatoria. Se analizan y discuten sus usos en la práctica profesional e investigación.

RAN: A computer program to estimate random response patterns in objective tests Computer program for random response. In the psychological measurement, especially in personality, the paper-pencil method is the usual device between the psychologist and allied professions. The score obtained by a person examined can have questionable validity if the responses were random. To identify this problem, different methods have been used. The random response is a challenging situation for the measurement research and diagnosis in cognitive performance and personality, when they have been measured by test of dichotomous type. In this study a computer program that implements the Charter's method of point estimate and confidence intervals is presented in order to identify if the score obtained can be considered random. We analyze and discuss this program uses in the research and professional practice.

Frederick y Crosby (2000) reconocieron varios estilos de respuesta relevantes para comprender y valorar la validez de los resultados obtenidos desde las pruebas de autoinforme. Así, desde un eje formado por el grado de esfuerzo y la motivación aplicado por el examinado para resolver los ítems de un instrumento de medida, Frederick y Crosby (2000) identificaron cuatro patrones de respuesta anómalas, pero tres han supuesto un desafío para el examinador al interpretar las puntuaciones. Una *respuesta inconsistente* sugiere variabilidad en el esfuerzo y motivación, con errores no esperados y pobre desempeño; un patrón de *respuestas suprimidas* proviene de una intención de simulación y de dar respuestas erróneas deliberadamente; finalmente, el patrón de *respuestas irrelevantes* identifica al examinado cuyas respuestas no guardan relación con el contenido de los ítems y no puede diferenciarse de un patrón del cuarto tipo, *respuestas aleatorias*, es decir aquellas respuestas en que el examinado responde con pobre motivación en la tarea y con bajo esfuerzo para representar su estatus en el constructo evaluado (Frederick y Crosby, 2000); tal patrón no es repli-

cable, y solo representan un proceso azaroso. Este último patrón es el objetivo del presente trabajo, y desde la cual se presentarán unos procedimientos estadísticos desarrollados por Charter (1994, 2000) para determinar si la puntuación obtenida por un examinado es diferente de la obtenida por azar y, derivar intervalos de confianza alrededor de una puntuación aleatoria esperada.

Woodard y Axelroad (1999) recomendaron que la interpretación de puntuaciones provenientes de ítems dicotómicos debiera incluir la evaluación de respuestas aleatorias, y consecuentemente aplicaron el método de Charter (1994, 2000) a la Escala de Depresión Geriátrica (Yesavage, 1986) para determinar la presencia de potenciales respuestas aleatorias dentro de los puntos de corte de identificación de la severidad de la depresión. Otros estudios aplicados también han usado este método (Charter, 1994; Charter y Lopez, 2003; Lopez y Charter, 2001), pero aparentemente no ha sido una metodología extendida a otros trabajos, aún cuando tiene un valioso potencial para aproximarse objetiva y cuantitativamente a la detección de un patrón aleatorio de respuestas. Comparado con la toma de decisiones basadas en la experiencia, los principios matemáticos demuestran ser sólidos en el establecimiento de puntos de corte o intervalos de confianza para sugerir la presencia de un patrón aleatorio de respuesta (Charter y Lopez, 2003), y son la base objetiva y fiable para la toma de decisiones.

En las evaluaciones de la personalidad, y dada su naturaleza explorativa, es común hallar indicadores de respuestas anómalas que sirven para obtener evidencias de la validez derivada de las pun-

tuciones; por ejemplo, las medidas verbales de personalidad para adultos (Butcher, Dahlstrom, Graham, Tellegen, y Kaemmer, 1989) y adolescentes del MMPI (Butcher et al., 1992), el *Inventario de Millon* (Millon, Davis, Millon y Grossman, 2006), y las medidas no verbales (por ejemplo, el *Cuestionario No Verbal de Personalidad*, (NPQ, Paunonen, Jackson y Ashton, 2004), incorporan subescalas para estimar la validez de las respuestas del examinado. Otras medidas clínicas más críticas, como las escalas de identificación del abuso infantil (Milner y Crouch, 1997), también poseen escalas de validez. Pero la efectividad de estos componentes es puesta a prueba en estudios empíricos; de este modo, las distinciones sobre estos estudios se pueden reconocer en sus aportaciones sobre datos normativos (p. e., Sánchez y Jiménez, 2003), puntos de corte y probabilidades asociadas a la detección de respuestas aleatorias (p. e., Lopez y Charter, 2001), y evaluaciones sobre la efectividad de las mismas (Charter y Lopez, 2002; Sánchez y Jiménez, 2003). Los diseños aplicados son generalmente experimentales y se analizan cuantitativamente con pruebas de hipótesis (Sánchez y Jiménez, 2003) o estimaciones de poder predictivo (Archer, Handel, Lynch y Elkins, 2002; Baer, Kroll, Rinaldo y Ballenger, 1999) variando la severidad de la presencia de respuestas aleatorias y el tipo de distorsión de respuesta (presentación favorable, presentación desfavorable, respuesta honesta y respuesta aleatoria).

Otro método es el uso de modelos estadísticos para determinar el estatus de un examinado respecto a sus respuestas, que no es nuevo pero tiende ser poco atendido por los usuarios. Estos métodos dan una estimación probabilística del grado de anomalía de respuesta. A menos que estos procedimientos se publiquen y describan en los manuales de pruebas, generalmente no llega al conocimiento del usuario aplicado o investigador debido que se difunden en revistas de compleja base matemática y estadística, que no son accesibles para el profesional de la Psicología. Se han desarrollado estimadores puntuales y límites de confianza para el grado de anormalidad de las diferencias entre una norma y un puntaje individual basados en distribuciones t (Crawford y Garthwaite, 2002, 2005; Crawford, Garthwaite y Howell, 1998), o aproximaciones normales a la distribución binomial (Charter, 1994; Charter y Lopez, 2003), pero aparentemente, no están tan difundidos como sería deseable. Otros métodos abordan el mismo problema pero diferente metodológicamente; por ejemplo, se tiene el comercialmente disponible *Indicador de Validez del Perfil* (Frederick y Crosby, 2000; Frederick, Sarfaty, Johnston, y Powel, 1994), pero requieren un largo proceso de validación y adaptación psicométrica y lingüística que no siempre accesible al profesional de la Psicología.

El objetivo del presente trabajo es describir un método para estimar la presencia de respuestas aleatorias dentro del conjunto de respuestas a un instrumento de tipo autoinforme. Este método fue propuesto por Charter (1994, 2000) y fue aplicado en varias ocasiones para examinar pruebas de personalidad (Lopez y Charter, 2001; Charter y Lopez, 2003) y de rendimiento cognitivo (Charter, 1994). Durante el recorrido de este documento, se presentarán las formulaciones estadísticas y ejemplos apropiados en varias situaciones de evaluación psicológica, pero se hará un énfasis en la interpretación juiciosa del coeficiente propuesto y sus intervalos de confianza. Finalmente, describiremos un programa informático de libre distribución.

Estimación Puntual por un Coeficiente Estandarizado

Al rastrear el inicio de este coeficiente, su primera exposición fue en una investigación de Charter (1994) para determinar si las

puntuaciones de pruebas de ritmo, percepción sonido-palabra y categorías de Seashore (todos de la *Batería Neuropsicológica Halstead-Reitan*; Reitan y Wolfson, 1993) son significativamente diferentes de un patrón de respuestas aleatorias; posteriormente, Charter (2000) amplió la exposición de la aproximación normal a la distribución binomial y presentó ecuaciones para la construcción de intervalos de confianza alrededor de la *respuesta aleatoria probable*. La distribución binomial es un modelo de probabilidad para variables discretas con dos categorías, por ejemplo, los puntajes de instrumentos de medida que usan ítems dicotómicos son una buena representación de ella (Sheskin, 2003). Más tarde, Charter y Lopez (2003) construyeron intervalos de confianza en los niveles del 95, 90 y 85% a las escalas de validez F, F Back y VRIN, todas provenientes del *MMPI-2* (Butcher et al., 1989).

Bajo el marco de la prueba de hipótesis, se probó la hipótesis nula de aleatoriedad en la puntuación del examinado; la hipótesis alternativa es la opuesta. En el cálculo del coeficiente, expresado como una puntuación típica z, se incluyen 1) la puntuación obtenida por el examinado que servirá para evaluar la aleatoriedad del patrón de respuestas, 2) el número de ítems y 3) la proporción aleatoria esperada (PAE). Este último concepto es la probabilidad de obtener una respuesta correcta o incorrecta en cada ítem. El cálculo se basa en la equiprobabilidad de las opciones de respuesta al ítem; por lo tanto, se calcula sencillamente como $1/n_{op}$, donde n_{op} es el número de opciones de respuesta en el ítem. Para ítems dicotómicos, el PAE es 0.5, es decir, 1/2. Para ítems de opción múltiple con cuatro alternativas, la PAE = 0.25 (1/4). Si se dispone de ítems con un número diferente de opciones de respuesta y agrupados en subtests, entonces ello produce también diferentes PAE; para tal situación se aplicaría el promedio ponderado. Por ejemplo, si disponemos de tres subtests, uno con 50 ítems de dos opciones (PAE = 0.5), otro de 20 ítems con cuatro opciones (PAE = 1/4 = 0.25) y el tercero formado por 25 ítems cinco opciones (PAE = 1/5 = 0.20), el promedio ponderado de los 95 ítems será

$$PAE_{pond} = \frac{[50(0.50) + 20(0.25) + 25(0.20)]}{95} = 0.368$$

La proporción aleatoria esperada (PAE) para los 95 ítems agrupados en un solo test será 0.368, y el coeficiente estandarizado para evaluar si una puntuación obtenida no se debe a un patrón aleatorio de respuestas es

$$z = \frac{\left(\frac{x}{k}\right) - PAE - \left(\frac{1}{2k}\right)}{\sqrt{\frac{PAE(1 - PAE)}{k}}} \quad (1)$$

donde

z = puntuación típica

x = puntuación obtenida por el examinado

k = número de ítems en el instrumento de medida.

PAE = Proporción Aleatoria Esperada

Esta fórmula es una razonable aproximación normal a la distribución binomial (Charter, 2000) siempre que K (PAE) o K (1-PAE) > 5. Siguiendo con el ejemplo anterior, un examinado que obtuvo una puntuación de 50 en el instrumento de 95 ítems ha obtenido la siguiente puntuación típica.

$$z = \frac{\left(\frac{50}{95}\right) - 0.368 - \left(\frac{1}{2(95)}\right)}{\sqrt{\frac{0.368(1 - 0.368)}{95}}} = 3.27$$

El coeficiente estandarizado es 3.27 y es estadísticamente significativo ($p < 0.001$), por lo tanto el examinador puede usar esta información estadística para decidir si la puntuación obtenida no proviene de un patrón aleatorio de respuesta y balancear la interpretación del puntaje. Este resultado es una estimación, e indica que es muy improbable que la puntuación provenga de tal patrón de respuesta aleatoria. Pero considerando esto técnicamente, es pertinente relacionar el número de ítems disponibles y la aproximación normal de la fórmula. En otras palabras, para lograr una más cercana aproximación normal a la distribución binomial, el valor de K (número de ítems) es clave; basándose en un estudio de simulación, Charter (2000) recomendó que un instrumento deba tener un mínimo número ítems con respecto al PAE (tabla 1).

K	> 30	> 50	> 80	> 200	> 600	> 1400
PAE	0.5	0.4 ó 0.6	0.3 ó 0.7	0.2 ó 0.8	0.1 ó 0.9	.05 ó 0.95

En otras palabras, teniendo una proporción aleatoria esperada (PAE) igual a 0.42, se sugiere que la prueba tenga más de 50 ítems para lograr una razonable aproximación normal usando el coeficiente estandarizado descrito en la ecuación 1.

Por otro lado, la ventaja de la aproximación normal a la distribución binomial derivada del coeficiente estandarizado afecta también a la construcción de intervalos de confianza para determinar si una puntuación es diferente de la respuesta aleatoria. Para construir intervalos de confianza, primero el profesional o investigador puede preguntarse cuál será la puntuación observada que no podría identificarse como derivado de una respuesta aleatoria, con un determinado nivel de confianza, por ejemplo, 95%. Otros niveles de significancia menos conservadores pueden también usarse (por ejemplo, 90% de confianza). Este es justamente el siguiente enfoque, la de construir intervalos de confianza para el rango probable de respuestas aleatorias.

Intervalos de confianza

Otro método recomendado, y que concuerda con las recomendaciones actuales en materia de publicación de resultados cuantitativos (Finch, Thomason y Cumming, 2002; Wilkinson y the Task Force on Statistical Inference, 1999), es la construcción de intervalos de confianza, que informa del rango probable de un parámetro en la población. Para nuestro objetivo, el parámetro de interés al cual se estimará el intervalo en un nivel de confianza particular será el puntaje aleatorio esperado. Por lo tanto, basándose en la aproximación normal a la distribución binomial (Charter, 2000), podemos construir uno de los elementos del intervalo de confianza para la respuesta aleatoria probable

$$A = z \sqrt{\frac{PAE(1 - PAE)}{k}} + \left(\frac{1}{2k}\right) \quad (2)$$

Los términos de la ecuación han sido definidos previamente; pero aquí, z es el valor crítico proveniente de la distribución normal estandarizada. El examinador puede elegir el nivel de confianza apropiado para sus propósitos: 99% ($z = 2.58$), 95% ($z = 1.96$), 90% ($z = 1.64$), ó 85% ($z = 1.44$). Los límites superiores e inferiores del intervalo son

$$LI = k(PAE - A) \quad (3)$$

$$LS = k(PAE + A) \quad (4)$$

Los límites centran el intervalo alrededor de la *puntuación aleatoria esperada*, PAES, que es igual a PAE multiplicado por K . Tomando nuevamente el ejemplo anterior, el puntaje aleatorio esperado para la prueba es PAES = (PAE) (K) = (0.36) (95) = 34. Luego, alrededor de este puntaje se construirá el intervalo de confianza que el examinador puede elegir en un nivel de 95%; entonces:

$$A = 1.96 \sqrt{\frac{0.368(1 - 0.368)}{95}} + \left(\frac{1}{2(95)}\right) = 0.102$$

$$\begin{aligned} LI &= 95 (0.368 - 0.102) = 25.24 \approx 25 \\ \in LS &= 95 (0.368 + 0.102) = 44.67 \approx 44 \end{aligned}$$

Charter (2000) recomendó redondear hacia arriba el límite inferior y hacia abajo el límite superior. Por lo tanto, el intervalo al 95% dentro del cual una respuesta se consideraría aleatoria se encuentra entre 25 y 44, y cuyo centro del intervalo es 34, lo que coincide con la estimación puntual proveniente del coeficiente estandarizado de nuestro ejemplo (50), y sugiere que este puntaje es mejor que el azar y se consideraría que la prueba no ha sido aleatoriamente respondida por el examinado, ya que esta fuera de este intervalo en el nivel de confianza prefijado. Además de probar la puntuación del examinado, el enfoque de intervalos de confianza informa del rango probable de puntuaciones aleatorias para que el investigador o clínico puedan contrastar otros puntajes.

Aplicaciones

Este método, y su programa informático acompañante (ver más abajo), puede jugar un papel importante en la revisión psicométrica de una prueba publicada, ya que la revisión puede incluir estos aportes complementarios o validar ciertos aspectos de del instrumento que sean consistentes con las hipótesis del examinador. Por ejemplo, el usuario puede calcular la respuesta esperada aleatoria y sus intervalos de confianza para los puntajes totales, ya que le dará información oportuna para el investigador o profesional que desea tomar decisiones con información técnica que las respalde. Finalmente, se podría poner a prueba los índices de validez de instrumentos clínicos publicados como en Charter y Lopez (2003) respecto al *MMPI*.

En el diseño de pruebas, este método presentado puede ayudar a determinar el número de ítems del formato final de una escala. Ya que el punto de corte obtenido por el método de Charter su-

giere una puntuación mínima (se deriva en un mínimo número de ítems), el constructor de una prueba determinará el grado en que es suficiente un determinado mínimo número de ítems para hacer menos probable la cantidad de respuesta aleatoria.

El clínico puede considerar que si los puntajes de un paciente son inconsistentes con ciertas hipótesis clínicas, una de las causas es la respuesta aleatoria que probablemente ha ocurrido; esta situación está relacionado directamente con la fiabilidad de los puntajes obtenidos; por lo tanto el clínico puede hallar razones para explorar los motivos de este patrón de respuestas o re-administrar el instrumento aplicado. Visto desde una perspectiva más amplia, la detección de la aleatoriedad de las respuestas puede jugar un papel clave en juzgar las evidencias de validez y fiabilidad de los instrumentos bajo análisis, incluso en la validez consecucional. De este modo, puntuaciones dentro o menores al intervalo de confianza para la respuesta aleatoria esperada pueden ser sometidos a exámenes adicionales como la consistencia interna y correlaciones en el grupo de examinados con estas puntuaciones; el investigador puede tener un apoyo para tomar mejores decisiones sobre si este grupo pueda ser eliminado o representa realmente el extremo posible más bajo de la distribución de las puntuaciones.

Hallar un resultado individual dentro de un intervalo que indique la probabilidad de respuesta aleatoria no debe ignorar los aspectos interpretativos. Más que una ciencia, el proceso interpretativo es un arte sensible a los aspectos que no reflejan las técnicas estadísticas, como son los factores relacionados con el examinado, motivación, salud, concentración, los cambios de administración de la prueba o aún los cambios de examinador (Feldt y Charter, 2003). Se ha puesto atención a los indicadores intrapersonales de los examinados al explorar el impacto del fingimiento y la respuesta aleatoria sobre la efectividad de los índices de respuestas válidas, correlaciones de los constructos medidos y propiedades psicométricas, pero se debería poner atención también en las propiedades estructurales de los instrumentos. Por ejemplo, la ubicación de los ítems dentro del instrumento (ítems agrupados vs. aleatoriamente distribuidos) son predictores parciales el grado de severidad del impacto del fingimiento y de la respuesta aleatoria (McFarland et al., 2002). En el mismo estudio de McFarland et al. (2002), las medidas de neuroticismo y conciencia (desde el modelo de lo Cinco Grandes de personalidad) fueron los más afectados. Por lo tanto, la distorsión de respuesta aporta una parte de la varianza de error.

Una diferencia que se debe alertar al usuario de este enfoque es entre un patrón de respuestas completamente aleatorio y otro *parcialmente aleatorio*, situación que ha sido explorada continuamente y con especial atención en el MMPI (Archer et al., 2002; Baer et al., 1999; Cramer, 1995). Y los métodos expuestos en nuestro trabajo no son sensibles a diferenciarlos. Aunque las respuestas completamente aleatorias en un protocolo tienden a ser más detectadas que las parcialmente presentes (Pinsoneault, 1999, 2006; Archer et al., 2002), el coeficiente no nos dice nada sobre cuál de estos problemas está presente, y el examinador es puesto a prueba para analizar cualitativamente el grado y la distribución de las respuestas aleatorias en el protocolo que está evaluando. En esta misma línea de discusión, es cierto que el poder discriminatorio de las técnicas cuantitativas presentadas aquí no alcanza a diferenciar estos tipos de respuestas anómalas ni los propuestos por Frederick y Crosby (2000), pero esta limitación de las inferencias que se pueden extraer de ellas podría ser una fortaleza. Creemos que discernir diferentes patrones de respues-

tas anómalas desde un solo índice cuantitativo puede promover su uso indiscriminado y desplazar a un plano secundario la incorporación del juicio del examinador ante un fenómeno complejo; por lo tanto, la estimación probabilística de la respuesta aleatoria puede ser indicador necesario aunque no suficiente para iniciar una investigación de tal respuesta en el examinado. Si hay el interés específico en, por ejemplo, diferenciar patrones de respuestas anómalas como los propuestos por Frederick y Crosby (2000), se puede recurrir al método evaluativo creado por ellos, pero cuya discusión está fuera de los límites del presente artículo.

De igual modo, el punto de corte generado de la aplicación del método que presentamos no es un dato absoluto para asegurar que un examinado ha respondido aleatoriamente (Charter, 2000). Otras explicaciones son igualmente posibles como estilos de respuesta particulares, fatiga o bajos niveles de alfabetización (Lopez y Charter, 2001), y que deben ser ponderados por el examinador. Como en muchas situaciones aplicadas y teóricas en la ciencia psicológica, la comunión de las técnicas y el juicio deben aliarse para tomar decisiones acertadas. Finalmente, las comparaciones interculturales entre puntajes de personalidad (por ejemplo, véase Scout y Wampa, 2000) pueden enriquecer la interpretación con índices estadísticos de validez de respuestas.

Es cierto que otros métodos más sofisticados y con validez de constructo aceptables pueden también ser utilizados (por ejemplo, las escalas de validez orientadas a la detección de respuestas aleatorias), pero pueden revelar un inestable éxito en diferentes instrumentos; y por otro lado, la sola nominación de *escala de aleatoriedad* no es un suficiente motivo para asignarle su efectividad pretendida. Por ejemplo, para un instrumento menos conocido, el *Inventario Jesness* (Jesness, 2000), su Escala de Aleatoriedad produjo bajas tasas de sensibilidad y especificidad que las escalas de Respuesta Variable y de Inconsistencia de Respuesta Variable incluídas en el mismo instrumento (Pinsoneault, 2006); y para un instrumento más conocido, el *Inventario Clínico Multiaxial de Millon (MCMI-III)*, sus índices de validez fueron pobremente eficientes en la detección de respuestas aleatorias (Charter y Lopez, 2002). Archer et al. (2002) también demostraron que las escalas F en la versión adolescente del *MMPI* son menos sensibles en condiciones de respuesta aleatoria parcial. Convergiendo con estos hallazgos, Lopez y Charter (2003) cuestionaron la utilidad clínica de la escala *TRIN del MMPI-2* para detectar confiablemente respuestas aleatorias. Aparentemente, las escalas de detección de patrones inválidos de respuesta son inestables y las investigaciones tienden a ser contradictorias.

Contar con métodos cuantitativos que amplifiquen la información sobre el examinado y que ayuden a tomar decisiones apropiadas, tiene consecuencias que superan el interés metodológico, pues si el clínico siempre pretende tomar decisiones adecuadas derivadas de los resultados de sus instrumentos, debería mantener un mínimo respaldo proveniente de la información disponible que tiene y su capacidad para integrarlos dentro de un marco de validez consecucional; y el método de Charter puede apuntar a este aspecto, y dar un aporte adicional a la validez consecucional de una práctica que caracteriza a los psicólogos, es decir, la evaluación psicológica.

Los ejemplos y la revisión de los métodos para determinar respuesta aleatoria aquí presentados, se han focalizado en instrumentos de personalidad, pero este no es el único contexto posible de aplicación del procedimiento delineado. El procedimiento puede

ser apropiadamente aplicado a cuestionarios conteniendo ítems dicotómicos, ya sea en aquellos temáticamente orientados hacia la medición de aspectos de la personalidad como los de desempeño máximo, por ejemplo instrumentos de aptitudes o logro académico. Es razonable suponer que en ambos tipos de instrumentos, la respuesta aleatoria puede ocurrir, y si la observación del examinador respalda esta suposición, la estimación de la probabilidad de tal respuesta en el examinado parece una práctica necesaria. Esta posición es consistente con las aplicaciones sugeridas por Charter (2000), aunque aparentemente hasta la fecha no ha sido aún muy utilizada en ambos ámbitos de evaluación psicológica. Aunque parece una limitación que los procedimientos de Charter se apliquen únicamente a instrumentos con ítems dicotómicos, estos tipos de medidas no son infrecuentes aún en la evaluación de personalidad. Las evaluaciones neuropsicológicas, de rendimiento académico, de inteligencia, de aptitudes y, en general, de máximo desempeño; y los instrumentos de personalidad que registran las respuestas binariamente (verdadero y falso), son las medidas en que idealmente pueden determinarse los estimadores puntuales y en los intervalos de confianza para la detección de respuestas aleatorias por el método de Charter (2000). No se conoce de un método que sea aplicable a puntajes provenientes de ítems politómico (de tipo Likert), y ello es más bien una vía de futuro desarrollo.

Finalmente dado que el todo es diferente a la suma de sus partes, creemos que la mejor estrategia sería combinar la evidencia psicométrica y el juicio clínico, tal como también lo sugieren Charter y Lopez (2002), para tomar decisiones bien informadas sobre la calidad del puntaje. Si el lector está interesado en aplicar los métodos aquí expuestos, el problema de los errores en el cálculo manual pueden estar presentes; por lo tanto, a continuación describimos un programa informático que facilitará estos cálculos.

Disponibilidad del programa

El programa está disponible en entorno Windows sin costo y puede solicitarse al primer autor de este trabajo mediante las direcciones de contacto. El programa se presenta como un ejecutable independiente y ha sido construido en el entorno Liberty-Basic; no obstante, este programa no requiere ninguna instalación de software especializado. El programa permite manejar archivos de respuestas correctas con 1's y 0's, donde 1 es acierto y 0 es fallo, o también permite corregir, con un conjunto de claves, un archivo que contiene las respuestas dadas por las personas a un conjunto de ítems. Los datos ausentes se contabilizan como fallos.

Si el instrumento está corregido, el programa necesita dos archivos:

1) Un archivo con una matriz rectangular (cada fila = una persona; y cada columna = un ítem. Las columnas deben estar separadas por comas ', '. Consultar el archivo de ejemplo "ex1.dat" que viene con el programa. Este archivo puede tener cualquier nombre, pero la extensión debe ser necesariamente "*.dat".

2) Un archivo con el número de opciones que tiene cada subtest. Por ejemplo, si un prueba está formado por tres subtests de 10 ítems, donde los 10 primeros ítems tienen 4 opciones, los 10 siguientes tienen 2 opciones y los 10 últimos tienen 2 opciones, entonces el archivo tiene 3 filas y dos columnas. El primer valor de

la fila corresponde al número de ítems en el subtest y el segundo valor corresponde al número de opciones de los ítems de ese subtest. Consultar el archivo "ex1.sbt". Este archivo puede tener cualquier nombre, pero la extensión debe ser necesariamente "*.sbt". Si se trata de una prueba completa, sin subtests, entonces este archivo tendrá una fila (número de ítems) y una columna (número de opciones).

Estos dos archivos deben crearse previamente a la puesta en marcha del programa con cualquier editor de textos o procesador de textos que permita escribir/leer los archivos en formato texto (ASCII).

Puesta en marcha del programa

El programa es autoejecutable desde el explorador de Windows o desde la ventana del MS-DOS. También se puede llevar el icono con la antorcha, que identifica al ejecutable del programa, al escritorio y lanzarlo desde allí.

Archivo con las respuestas correctas e incorrectas (1's y 0's)

Una vez activado el programa, aparece una pantalla que pide todos los datos del análisis que queremos realizar:

1) El nombre del archivo con los datos (DATA MATRIX) sin extensión.

2) El nombre del archivo de salida (OUTPUT FILE) sin extensión.

3) El tamaño muestral (SAMPLE SIZE).

4) La longitud del instrumento (TEST LENGTH) incluidos todos los subtests.

5) El número de subtests (NUMBER OF SUBTEST).

6) El nombre del archivo (sin extensión) con los subtests que incluye el instrumento completo (SUBTEST FILE NAME).

7) La puntuación típica, al nivel de confianza deseado (CONFIDENCE LEVEL (ST. SCORE)). Por ejemplo, si queremos trabajar al nivel de confianza del 90% escribiremos 1.64; al nivel de confianza del 95% escribiremos 1.96, y al nivel de confianza del 99%, escribiremos 2.56. La puntuación típica se escribirá sin signos de más o menos. Se puede emplear cualquier nivel de confianza, siempre que introduzcamos en esta casilla la puntuación típica correspondiente según la tabla de áreas bajo la curva normal. El programa no calcula esta puntuación típica, debemos buscarla manualmente en las tablas correspondientes. Si la puntuación típica que utilizamos no corresponde a ninguno de los tres niveles de confianza especificados anteriormente, el programa nos pedirá que introduzcamos el nivel de confianza correspondiente a esa puntuación típica.

8) El archivo con la clave de corrección de los ítems (KEY FILE). En este caso, como la matriz de entrada es de 1's y 0's, el archivo con las claves (KEY FILE) se dejará en blanco.

Archivo con las opciones de los ítems marcadas por las personas

Cuando el archivo con la matriz de datos no contenga los aciertos y fallos, sino las opciones seleccionadas por las personas, debemos utilizar la clave de corrección (KEY FILE). La clave para cada ítem se escribirá en un nuevo archivo que puede tener cualquier nombre pero su extensión debe ser necesariamente "*.key". Este archivo tiene tantas filas como ítems del test (contando todos

los subtests) y sólo una columna. En esta columna debe aparecer la opción correcta para cada ítem. Para el resto, operar como aparece en la sección anterior.

Archivo de salida

El programa produce un archivo de salida (*.out) que contiene el intervalo confidencial al nivel de confianza especificado, la proporción aleatoria esperada, un listado con el número de la perso-

na, la puntuación total y la puntuación típica correspondiente al patrón de respuestas obtenido, y el signo de (*) indica que el patrón de respuestas es no aleatorio.

La salida finaliza con un conteo del número de personas cuyo patrón de respuestas es no aleatorio y el número de personas cuyo patrón de respuestas es aleatorio. También calcula el coeficiente de fiabilidad KR-20 y la variante de Horst (1953) para la muestra completa, y para el grupo de personas cuyo patrón de respuestas no fue aleatorio.

Referencias

- Archer, R.P., Handel, R.W., Lynch, K.D., y Elkins, D.E. (2002). MMPI-A validity scale uses and limitations in detecting varying levels of random responding. *Journal of Personality Assessment*, 78, 417-431.
- Baer, T.A., Kroll, L.S., Rinaldo, J., y Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment*, 72, 308-320.
- Butcher, J., Williams, C., Graham, J., Archer, R., Tellegen, A., Ben-Porath, Y., y Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory Adolescent: Manual for administration and scoring*. Minneapolis: University of Minneapolis Press.
- Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., y Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minneapolis Press.
- Charter, R.A. (1994). Determining random responding for the Category, Speech-Sounds perception, and Seashore Rhythm tests. *Journal of Clinical and Experimental Neuropsychology*, 16, 744-748.
- Charter, R.A. (2000). Determining random responding to objective test. *Journal of Psychoeducational Assessment*, 18, 308-315.
- Charter, R.A., y López, M.N. (2002). Millon Clinical Multiaxial Inventory (MCMI-III): The inability of the validity conditions to detect random responders. *Journal of Clinical Psychology*, 58, 1615-1617.
- Charter, R.A., y López, M.N. (2003). MMPI-2: Confidence Intervals for random responding to the F, F Back, and VRIN scales. *Journal of Clinical Psychology*, 59, 985-990.
- Cramer, K.M. (1995). Comparing three new MMPI-2 randomness indices in a novel procedure for random profile derivation. *Journal of Personality Assessment*, 65(3), 514-520.
- Crawford, J.R., y Garthwaite, P.H. (2005). Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology*, 19, 318-331.
- Crawford, J.R., y Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, 40, 1196-1208.
- Crawford, J.R., Garthwaite, P.H., y Howell, D.C. (1998). Payne and Jones revisited: Estimating the test scores abnormality using a modified paired samples t-test. *Journal of Clinical and Experimental Neuropsychology*, 20, 898-905.
- Elkins, D.E., y Archer, R.P. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment*, 73, 407-421.
- Feldt, L.S., y Charter, R.A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological Methods*, 8, 102-109.
- Finch, S., Thomason, N., y Cumming, G. (2002). Past and future American Psychological Association guidelines for statistical practice. *Theory & Psychology*, 12, 825-853.
- Frederick, R.I., y Crosby, R.D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior*, 24, 59-82.
- Frederick, R.I., Sarfaty, S., Johnston, J., y Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology*, 8, 118-125.
- Horst, P. (1953) Correcting the Kuder-Richardson reliability formula for dispersion of item difficulties. *Psychological Bulletin*, 50, 371-374.
- Jesness, C.F. (2000) *The Jesness inventory manual* (Rev. Ed.). North Tonawanda, NY: Multi-Health Systems.
- López, M.N., y Charter, R.A. (2001). Random responding to the MMPI-2 F scale. *Psychological Reporting*, 88, 398.
- McFarland, L.A., Ryan, A.M., y Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78, 348-369.
- Millon, T., Davis, R., Millon, C., y Grossman, S., (1997). *The Millon Clinical Multiaxial Inventory-III, Third Edition (MCMI-III)*. Minneapolis, MN: National Computer Systems.
- Milner, J.S., y Crouch, J.L. (1997). Impact and detection of response distortions on parenting measures used to assess risk for child physical abuse. *Journal of Personality Assessment*, 69, 633-650.
- Paunonen, S.V., Jackson, D.N., y Ashton, M.C. (2004) *Nonverbal Personality Questionnaire*. Port Huron, MI: Sigma Assessment Systems, Inc.
- Pinsoneault, T.B. (1999). Efficacy of the Three Randomness Validity Scales for the Jesness Inventory. *Journal of Personality Assessment*, 73, 395-406.
- Pinsoneault, T.B. (2006). Updating the Jesness Inventory randomness validity scales for the Jesness Inventory-Revised. *Journal of Personality Assessment*, 86, 190-195.
- Reitan, R.M., y Wolfson, D. (1993). *The Halsted-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press
- Sánchez, G., y Jiménez, F. (2003). La Escala Superlativa S de Butcher y Han (1995): el fingimiento en la adaptación española del MMPI-2. *Revista de Psicología de la Pontificia Universidad Católica del Perú*, 21, 5-39.
- Sheskin, D.J. (2003). *Handbook of parametric and nonparametric statistical procedures* (3th ed.). Boca Raton, FL: Chapman & Hall.
- Scout, R.L., y Wampa, W.M. (2000). The MMPI-2 in Peru: A normative study. *Journal of Personality Assessment*, 74, 95-105.
- Wilkinson, L., y the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and Explanations. *American Psychologist*, 54, 594-604.
- Woodard, J.A., y Axelrod, B.N. (1999). Interpretative guidelines for neuropsychiatric measures with dichotomously scored items. *International Journal of Geriatric Psychiatry*, 14, 385-388.
- Yesavage, J.A. (1986). The use of self-rating depression scales in the elderly. En L.W. Poon (Ed.): *Handbook for clinical memory assessment of older adults*. Washington DC: American Psychological Association.