

## LA PARADOJA DE LA CONSISTENCIA VUELTA A REVISAR. UN PROCEDIMIENTO PARA LA ESTIMACIÓN DE LA CONSISTENCIA INTRASUJETOS INDEPENDIENTE DE LA MEDIDA <sup>1 2</sup>

José Manuel Hernández<sup>3</sup>

Víctor J. Rubio

José Santacreu

Javier Revuelta

UNIVERSIDAD AUTÓNOMA DE MADRID

### RESUMEN

*El presente trabajo retoma el clásico debate sobre la consistencia en la Psicología de la Personalidad. El origen del debate parte de la idea de una estructura universal de la personalidad que es defendida por los psicólogos del rasgo. Esta idea se expresa en la asunción ontológica que implica que todos los seres humanos son consistentes. Así, aunque dos son las fuentes de variación en las puntuaciones de los tests: los individuos y los ítems, no se han desarrollado procedimientos metodológicos ni analíticos que asignen diferentes valores a la consistencia intraindividual y a la consistencia de los ítems.*

*Este trabajo propone un procedimiento que permite estimar la proporción de sujetos consistentes e inconsistentes de manera independiente a la medida utilizada en una amplia muestra. El procedimiento es una adaptación al contexto de la psicometría del estadístico  $\pi^*$  propuesto por Rudas, Clogg y Lindsay (1994) para*

---

<sup>1</sup> Esta investigación ha sido parcialmente financiada por el proyecto AENA-UAM / 785002

<sup>2</sup> Poster presentado a la 12th European Conference on Personality, Groningen, The Netherlands, Julio 2004

<sup>3</sup> Correspondencia: Dr. José Manuel Hernández. Facultad de Psicología. Universidad Autónoma de Madrid. 29049 Madrid (Spain). e-mail: josemanuel.hernandez@uam.es. Web site: <http://www.uam.es/psimad>.

*el análisis de tablas de contingencia. En esta primera aproximación, los datos permiten sostener que aproximadamente un 60% de los sujetos se comportan de forma consistente. Con respecto al otro 40%, el método no permite determinar cual es la causa del desajuste al modelo, si la falta de consistencia u otra.*

## SUMMARY

*The present paper revisits the consistency debate in Personality Psychology. The origin of the debate comes from the idea that there is a universal human personality structure, which is defended by trait psychologists. This idea is expressed in the ontological assumption that underlies the proposal that all the human beings are consistent. Then, even though it is obvious that there are two sources of variance in the tests scores, items and individuals, no methodological nor analytical procedures have been developed in order to assign different values to intraindividual consistency and item consistency.*

*This paper proposes a procedure that allows estimating the proportion of consistent and inconsistent subjects independently of the measure used. The procedure is an adaptation to the psychometric context of the statistic  $\pi^*$  proposed by Rudas, Clogg and Lindsay (1994) for the analysis of contingency tables. In this first approach the data allow sustaining that approximately 60% of the participants behave consistently. With respect to the other 40%, the method does not allow determining why the model does not fit, wheter it is due to the lack of consistency or to something else.*

## INTRODUCCIÓN

La noción de consistencia intraindividual está a la base de toda la Psicología de la Personalidad. Desde que Mischel (1968) formulase su crítica teórica y metodológica al concepto de consistencia

que sostienen los psicólogos del rasgo, varios autores han protagonizado un debate (Mischel y Peake, 1982, 1983; Epstein, 1983a, 1983b, 1984; Bem, 1983; Funder, 1983a, 1983b; Funder y Ozer, 1983; Ozer, 1986; Kenrick y Funder, 1988), principalmente a lo largo de la década de los 80 del siglo pasado, en torno a los distintas concepciones de consistencia.

El origen de tal debate se sitúa en la concepción, defendida por la psicología del rasgo, de que existe una estructura universal de la personalidad humana que se expresa en la asunción ontológica subyacente a la proposición: *Todos los seres humanos son consistentes*. Esta concepción supone que dado que las personas son consistentes en algunos de sus comportamientos podemos identificar en todos los sujetos la misma clase de comportamientos consistentes que se corresponderían con las grandes dimensiones de la personalidad.

Dado esto, un buen instrumento de evaluación de un rasgo de personalidad es aquél que está compuesto por elementos que son capaces de detectar la consistencia de los individuos con independencia de la magnitud de la variable-rasgo en cada uno de ellos. Por ello, el proceso de construcción de un instrumento de evaluación que pretende medir los rasgos de personalidad exige eliminar aquellos ítems que no contribuyan a incrementar la consistencia interna del test bajo el supuesto de que no están midiendo la dimensión o rasgo a explorar, sin poner en duda *que los individuos puedan no ser consistentes*. Por esta razón, aunque es evidente que las fuentes de variación de las puntuaciones de un test son dos: a) los ítems del test y b) los individuos evaluados, no se han desarrollado procedimientos metodológicos y analíticos que sirvan para otorgar valores separados de consistencia intraindividual y de consistencia de los elementos que constituyen el test.

Pretendiendo llenar el vacío metodológico, existente en la teoría de la personalidad, el objetivo de este artículo es la puesta a prueba de un procedimiento estadístico para, dadas las puntuaciones de un sujeto en una prueba objetiva de evaluación compuesta por ítems morfológica y funcionalmente iguales, poder determinar el grado de consistencia intraindividual que muestra con independencia del instrumento de medida utilizado.

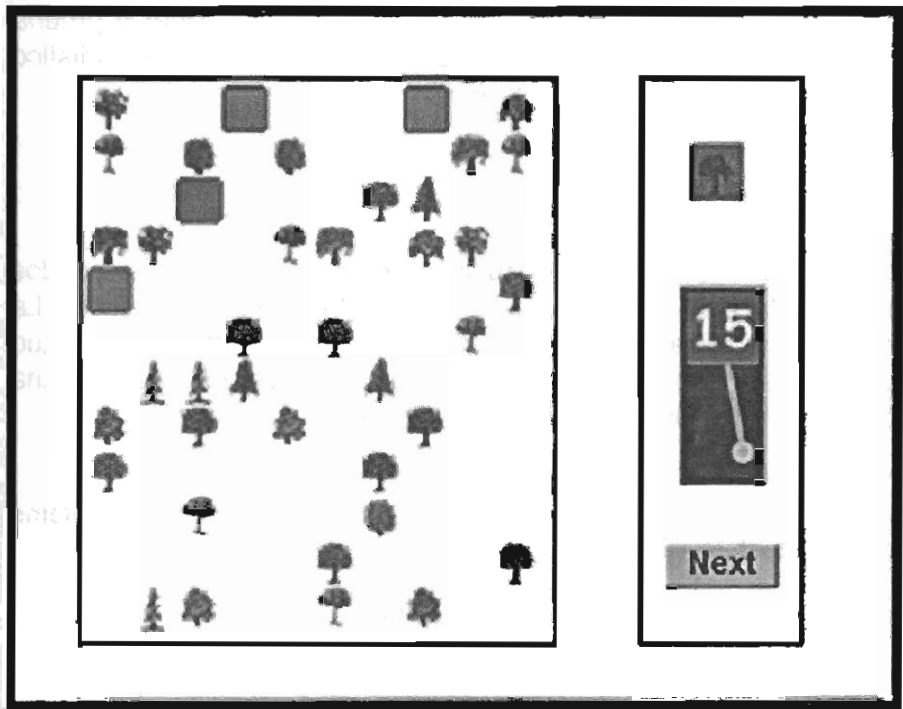
Para resolver la cuestión planteada proponemos un procedimiento que permita estimar la proporción de sujetos consistentes e inconsistentes y la probabilidad de la consistencia de cada sujeto, con independencia del nivel de los sujetos en la variable medida y de que los ítems de la misma sean equivalentes. El procedimiento es una adaptación al contexto de la psicometría del estadístico  $\pi^*$  propuesto por Rudas, Clogg y Lindsay (1994) para el análisis de tablas de contingencia.

Los objetivos del estudio son los siguientes:

1. Estimar el porcentaje de sujetos que actúan de forma *consistente* en una tarea de minuciosidad (descrita en Hernández, Sánchez-Balmisa, Madrid y Santacreu, 2003).
2. Evaluar la *equivalencia* de las diferentes tareas o ítems que componen el test.

## MÉTODO

Se dispone de una muestra de 428 participantes que han respondido a 15 ítems de un test de minuciosidad con un valor del coeficiente de Cronbach de 0.76 (Hernández et al, 2003). La tarea consiste en identificar y marcar un objeto (en concreto, un árbol) en una matriz donde se refleja ese objeto entremezclado con otros objetos (distintos tipos de árbol). Dado que la configuración de cada matriz es distinta, cada ítem es morfológicamente diferente (varía el tipo y la posición del árbol a identificar) aunque funcionalmente igual ya que se pueden identificar utilizando la misma estrategia comportamental. La puntuación se establece con base en el orden en el cual se han identificado los objetos. El rango de puntuaciones de cada ítem corresponde a una escala ordinal de 0 a 7 donde la puntuación 0 corresponde a nula minuciosidad y la puntuación 7 a máxima minuciosidad. Debido al reducido número de sujetos se han analizado únicamente los 7 primeros ítems, además, las respuestas se han dicotomizado utilizando el valor 3,5 como punto de corte. De esta forma, el número de posibles patrones de respuesta es  $2^7 = 128$ , que permite evaluar la bondad de ajuste del modelo. (Véase un ejemplo de ítem en la figura 1)



**Figura 1.- Ejemplo de matriz de la prueba de evaluación del estilo interactivo ejecutar una tarea de manera ordenada, organizada y siguiendo un patrón sistemático. Los cuadrados de la matriz corresponden a celdas del target pulsadas por el sujeto**

Se han puesto a prueba dos modelos basados en el *2pl*, al que imponen restricciones en los parámetros para evaluar los dos objetivos.

- El modelo 1 asume que los ítems son no equivalentes (los parámetros  $a$  y  $b$  pueden ser diferentes para cada ítem) y los sujetos consistentes ( $\theta$  igual).

- El modelo 2 asume que los ítems son equivalentes (parámetros iguales).

El modelo 2 está anidado en el 1. Esto permite poner a prueba la bondad de ajuste relativa de cada modelo mediante el estadístico de la razón de verosimilitudes ( $G^2$ ).

## RESULTADOS

La tabla 1 contiene el resultado de la estimación de los dos modelos, la cual se utiliza para evaluar la bondad de ajuste. La segunda columna es el logaritmo de la función de verosimilitud evaluada en el estimador máximo verosímil. La tercera columna contiene el número de parámetros libres.

**Tabla 1.- Estimación de los modelos (1= Ítems no equivalentes; 2= Ítems equivalentes). N=428**

Modelo	$\log h(\mathbf{Y})$	<i>p.l.</i>
1	-1506.7	14
2	-2061.8	2

El valor de  $G^2$  para el modelo 1 es 92.1, con 113 grados de libertad, obteniéndose  $p= 0.925$ , por lo que su ajuste a los datos es bueno. El valor de  $G^2$  al comparar el modelo 1 con el modelo 2 es 1110.2 que, con 12 grados de libertad, resulta significativo al 99%. La falta de ajuste del modelo 2 significa que los ítems no son equivalentes.

La tabla 2 contiene los valores estimados de los parámetros junto con los errores típicos. Como índice de dificultad del ítem suele utilizarse con este modelo el valor de  $\Theta$  para el cual la probabilidad de acierto es 0,5, lo cual coincide con  $-b/a$ . La dificultad aparece en la columna derecha de la tabla 2. La media de las dificultades es -1,04.

Tabla 2.- Valores estimados de los parámetros y dificultad de los ítems ( $p'$ =Proporción de sujetos inconsistentes; N=428)

Ítem	No equivalentes			Equivalentes			$-b/a$		
	$a$	$S_a$	$b$	$S_b$	$a$	$S_a$		$b$	
1	0.912	0.099	-0.296	0.088	0.400	0.030	0.165	0.027	0.325
2	0.970	0.136	-2.256	0.156	0.400	0.030	0.165	0.027	2.326
3	0.497	0.052	2.618	0.050	0.400	0.030	0.165	0.027	-5.268
4	1.204	0.086	0.0186	0.081	0.400	0.030	0.165	0.027	-0.015
5	0.401	0.054	1.973	0.052	0.400	0.030	0.165	0.027	-4.920
6	1.840	0.111	-1.509	0.137	0.400	0.030	0.165	0.027	0.820
7	1.002	0.076	0.542	0.070	0.400	0.030	0.165	0.027	-0.541
$G^2$	92.1			1202.2					
$gl$	113			125					
$p$	0.995			0.990					
$\pi^*$	0.436			0.978					
$\pi_L^*$	0.270			0.944					

El valor estimado de  $\pi^*$  para el modelo 1 ha sido 0.436 y el límite inferior del intervalo es  $\pi_L^* = 0.270$ . Para el modelo 2 los resultados son  $\pi^* = 0.978$  y  $\pi_L^* = 0.944$ . Es decir, a pesar del buen ajuste del modelo, se estima que algo más del 40% de los sujetos están fuera del mismo. Este resultado puede deberse al reducido tamaño muestral y los muchos patrones de respuesta con frecuencia baja o nula.

## DISCUSIÓN

Con respecto a la hipótesis de equivalencia el resultado es claro, los ítems no son equivalentes. Los distintos ítems no pueden considerarse tareas equivalentes entre sí a pesar de que se han tratado de construir iguales excepto en la situación de los targets.

Los datos permiten sostener que aproximadamente un 60% de los sujetos se comportan de forma consistente. Con respecto al otro 40%, el método no permite determinar cual es la causa del desajuste al modelo, si la falta de consistencia u otra. Una explicación del bajo porcentaje de sujetos consistentes es el reducido tamaño muestral, que puede producir una sobrestimación de  $\pi^*$ .

En conclusión, es posible evaluar las dos hipótesis simultáneamente pero se necesitan muestras de gran tamaño. Si se utilizan tests de más de 10 ítems, lo que suele ser habitual en la evaluación psicológica, o ítems con más de dos categorías de respuesta el número de sujetos debería ser del orden de varios miles.

## REFERENCIAS

- Bem, D.J. & Funder, D.C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 85, 485-501.
- Epstein, S. (1983a). The stability of confusion: A reply to Mischel and Peake. *Psychological Review*, 90, 179-184.
- Epstein, S. (1983b). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51, 360-392.



- Epstein, S.** (1984). The stability of behavior across time and situations. In R.A. Zucker, J. Aronoff & A.I. Rabin (Eds.). *Personality and prediction of behavior* (pp. 209-268). New York: Academic Press.
- Funder, D.C.** (1983a). Three issues in predicting more of the people: A reply to Mischel and Peake. *Psychological Review*, *90*, 283-289.
- Funder, D.C.** (1983b). The consistency controversy and the accuracy of personality judgement. *Journal of Personality*, *5*, 346-359.
- Funder, D.C. & Ozer, D.J.** (1983). Behavior as a function of situation. *Journal of Personality and Social Psychology*, *44*, 107-112.
- Hernández, J.M., Sánchez-Balmisa, C., Madrid, B. & Santacreu, J.** (2003). La evaluación objetiva de la minuciosidad. Diseño de una prueba conductual. [The objective assessment of conscientiousness. The design of a behavioral test] *Análisis y Modificación de Conducta*, *29*, 457-479.
- Kenrick, D.T. & Funder, D.C.** (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*, 23-34.
- Mischel, W.** (1968) *Personality and assessment*. New York: Wiley.
- Mischel, W. & Peake, P.K.** (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, *89*, 730-755.
- Mischel, W. & Peake, P.K.** (1983). Some facets of consistency: replies to Epstein, Funder y Bem. *Psychological Review*, *90*, 394-402.
- Ozer, D.J.** (1986). *Consistency in personality: A methodological framework*. Berlin: Springer.
- Rudas, T., Clogg, C. C. & Lindsay, B. G.** (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society, series B*, *56*, 623-639.