Análisis y Modificación de Conducta, 2004, Vol. 30, Nº 134

CONSISTENCY PARADOX REVISITED. A PROCEDURE FOR ESTIMATING INTRAINDIVIDUAL CONSISTENCY INDEPENDENTLY FROM THE MEASURE

José-Manuel Hernández[1]
Víctor J. Rubio
José Santacreu
Javier Revuelta

UNIVERSIDAD AUTÓNOMA DE MADRID

RESUMEN

Este trabajo ofrece un procedimiento estadístico capaz de determinar el grado de consistencia intraindividual, independientemente de la medida utilizada. Se sugiere una adaptación del estadístico π^* (Rudas et asl., 1994). Se han llevado a cabo tres estudios para poner a prueba la conveniencia del estadístico π^* y la proporción de sujetos que actúan de forma consistente. Los resultados demuestran que la adecuación del estadístico propuesto, así como el porcentaje de individuos consistentes depende de si los ítems son equivalentes o no, y el número de alternativas de respuesta que poseen.

Palabras clave: ***CONSISTENCIA INTRAINDIVIDUAL, ESTADÍSTICA DE LA CONSISTENCIA.***

[1] Correspondence to the author: Dr. José Manuel Hernández. Faculty of Psychology. University Autonoma of Madrid. 28049 Madrid (Spain). E-mail: josemanuel.hernandez@uam.es. Web site: www.uam.es/psimad.

SUMMARY

Mischel (1968) argued against the idea of a general consistency of human beings. The present paper aims to design a statistical procedure able to determine the degree of intraindividual consistency independently of the measure used. For that, an adaptation of the statistic π^* (Rudas et al., 1994) is suggested. Using objective tests, three studies have been carried out for testing the suitability of the π^* statistic and the proportion of subjects that act consistently. Results have shown the availability of the statistic proposed as well as that the percentage of consistent individuals depends on whether test items can be assumed as equivalents or not, and the number of response alternatives they contained.

Key words: ***INTRAINDIVIDUAL CONSISTENCY, STATISTICS FOR CONSISTENCY.***

The notion of intra-individual consistency lies at the basis of all personality psychology. Ever since Mischel (1968) formulated his theoretical and methodological criticism of the concept of consistency sustained by trait psychologists, various authors have debated the different conceptions of consistency (Mischel and Peake, 1982, 1983; Epstein, 1983a, 1983b, 1984; Bem, 1983; Funder, 1983a, 1983b; Funder and Ozer, 1983; Ozer, 1986; Kenrick and Funder, 1988), chiefly throughout the 1980s.

The debate originated with the conception defended by trait psychology that there exists a universal structure of human personality which is expressed in the ontological assumption underlying the proposition: *All human beings are consistent*. This conception supposes that, since people are consistent in some of their behaviors, we can identify the same kind of consistent behaviors corresponding to the major dimensions of personality in all subjects.

Given this, a good instrument for assessing a personality trait is one made up of elements that can detect the consistency of the individuals irrespective of the magnitude of the trait variable in each of them. Consequently, the process of constructing an assessment instrument that aims to measure personality traits demands the elimination of those items that do not contribute to increasing the internal consistency of the test. This is because it is supposed that those items are not measuring the dimension or trait being explored. *That the individuals may not be consistent* is not doubted. This is why, in spite of the evident fact that the sources of variation in the scores of a test are two: a) the items of the test and b) the individuals tested, methodological and analytical procedures have not been implemented to allot separate values to intra-individual consistency and to the consistency of the elements that make up the test.

We propose a procedure that allows estimating the proportion of consistent and inconsistent subjects and the probability of the consistency of each subject independently of the subjects' level in the variable measured and independently of that variable's items being equivalent. The procedure is an adaptation to the psychometric context of the statistic proposed by Rudas, Clogg and Lindsay (1994) for the analysis of contingency tables. Two strategies are used to verify the model's goodness of fit. On the one hand, the statistic π^* . On the other, the likelihood ratio statistic ($G^{2)}$ is used to asess the fit.

The objectives of the study are the following:

1.To estimate the percentage of subjects that act *consistently* in a task requiring conscientiousness (described in Hernández, Sánchez-Balmisa, Madrid and Santacreu, 2003).

2.To assess the *equivalence* of the various tasks or items that comprise the test.

METHOD

A sample of 428 participants carried out 15 items of a Methodicalness Test (an aspect of Conscientiousness) with a Cronbach's coefficient of 0.76 (Hernández et al, 2003). The task consists in identifying and marking an object (a type of tree) in a matrix containing that object mixed with other objects (other types of tree). Since the configuration of each matrix is different, each item is morphologically different (the type and the position of the tree to be identified varies) although functionally the same, because they can be identified by using the same behavioral strategy. The score is based on the order in which the objects are identified. The range of scores of each item corresponds to an ordinal scale from 0 to 7 where the 0 score corresponds to no conscientiousness and the score of 7 to maximum conscientiousness. Because of the small number of subjects, only the first 7 items were analyzed. Furthermore, the responses were dichotomized using the value 3.5 as the cut-off point. This way the number of possible response patterns is $2^7 = 128$, which allows assessing the model's goodness of fit. (Figure 1 is an example of an item).

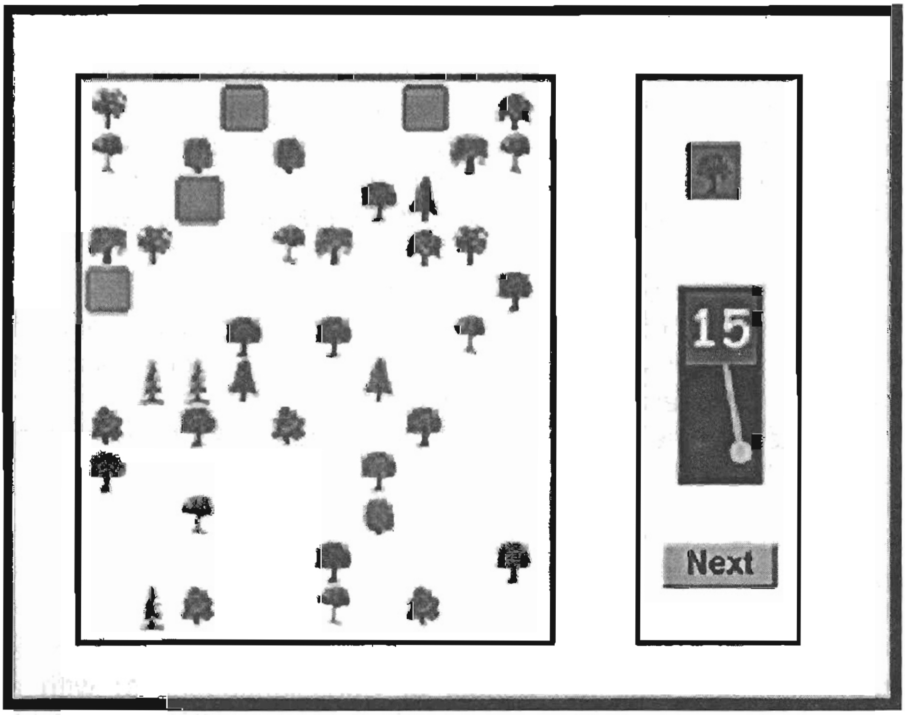

Figure 1.- Example of an item of the Methodicalness Test

Two *2pl-* based models were tested. Parameter restrictions were imposed to assess both objectives. Model 1 assumes that the items are not equivalent (parameters *a* and *b* can be different for each item) and that the subjects are consistent (*q* equal). Model 2 assumes that the items are equivalent (equal parameters). Model 2 is nested in model 1. This allows testing each model's goodness of fit by means of the likelihood ratio statistic (G^2).

RESULTS

Table 1 contains the outcome of the estimation of the two models, which is used to assess the goodness of fit. The second column is the logarithm of function log *h(**x**)* evaluated in the maximum likelihood estimator. The third column contains the number of free parameters.

Table 1.- Goodness of fit for models 1 (non equivalent items) and 2 (equivalent items) (Sample size = 428)

Model	*log h(**Y**)*	*p.l.*
1	-1506.7	14
2	-2061.8	2

The value of G^2 for model 1 is 92.1 with 113 degrees of freedom (p-value equals to 0.925). Therefore it fits the data well. Comparing model 1 with model 2, the value of G^2 is 1110.2, which, with 12 degrees of freedom, is significant at 99%. The lack of fit of model 2 means that the items are not equivalent.

Table 2 contains the estimated values of the parameters together with the standard errors. As item-difficulty index, this model generally uses the *q* value for which the probability of a correct response is 0.5, which coincides with *–b/a*. The difficulty appears in the right-hand column of table 2. The mean difficulty is –1.04.

Table 2.- Item parameter estimates, item difficulty (-b/a) and proportion of subjects out of model (δ') (Sample size = 428)

	Non equivalents				Equivalents				
Item	*a*	S_a	*b*	S_b	*a*	S_a	*b*	S_b	*-b/a*
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				
π^*	0.436				0.978				
π^*_L	0.270				0.944				

The estimated value of for model 1 is $\pi^* = 0.436$ and the lower limit of the interval is $\pi^*_L = 0.270$. The results for model 2 are $\pi^* = 0.978$ and $\pi^*_L = 0.944$, respectively. That is, in spite of the model's goodness of fit, it is estimated that somewhat more than 40% of the subjects fall outside of it. This result may be due to the small sample size and the numerous response patterns with low or null frequencies.

DISCUSSION

As to the equivalence hypothesis, the result is clear. The items are not equivalent. The various items can not be considered tasks that are equivalent to each other in spite of having tried to construct equal tasks, except in the situation of the targets. 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, whether it is due to the lack of consistency or to something else. The low percentage of consistent subjects can be explained by the small sample size, which can cause an overestimation of π^*.

In conclusion, it is possible to test both hypotheses at the same time but this requires large samples. With tests that have more than 10 items, which is common in psychological assessment, or with items that have more than two categories of response, the number of subjects should be of the order of several thousands.

REFERENCES

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 basics 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 controversie 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 fron 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.