THE OBJECTIVE ASSESSMENT OF PERSONALITY. AN ALTERNATIVE TO SELF-REPORT BASED ASSESSMENT[1]

Víctor J. Rubio[2]
José Santacreu
José-Manuel Hernández

UNIVERSITY AUTÓNOMA OF MADRID

RESUMEN

El objetivo de este trasbajo consiste en presenter una alternative a las medidas tradicionales de personalidad basada en los datos T de Cattell. Una evaluación objetiva de la personalidad no debería evaluar lo que las personas dicen que hacen usualmente, sino más bien lo que hacen. En otras palabras, utilizando una metodología en la que las puntuaciones no estén basadas en la auto-evaluación de los individuos, ni en la necesidad de los evaluadores para interpretar los datos o para modificarlos. Además, utilizando tareas que exijan poca habilidad, lo que significa que todo participante está homologado respecto a este criterio. De igual modo, los procedimientos de evaluación no ofrecen feedback sobre la actuación

2 Correspondence to the author: Dr. Víctor J. Rubio. Faculty of Psychology. University Autonoma of Madrid. 28049 Madrid (Spain). E-mail: victor.rubio@uam.es Web site: www.uam.es/psimasd.

de los individuos. Sin feedback, los participantes están, en parte,
libres de las influencias del contexto de modo que no
puedenaprender la respuesta correcta necesaria para obtener
resultados positivos. Los propósitos de la evaluación deberían ser
enmascarados. Esto significa que los sujetos no son informados
sobre los objetivos de las tareas, más bien se les anima a rendir
al máximom para obtener tantos puntos como puedan en una tarea
fácil. Esta estrategia reduce los sesgos involuntarios y la distorsión
de respuesta intencional que podría estar presente en las evalua-
ciones tradicionales de personalidad. Este trabajo presenta tres
pruebas diseñadas desde esta perspectiva, que evalúan coopera-
ción, metodicidad y asunciòn de riesgo, así como los resultados
en términos de fiabilidad y validez de las medidas obtenidas.

Palabras clave: *EVALUACIÓN OBJETIVA DE LA PERSONALIDAD, EVA-
LUACIÓN DE COOPERACIÓN, METODICIDAD Y ASUNCIÓN DE RIESGO. FIA-
BILIDAD Y VALIDEZ EN MEDIDAS OBJETIVAS DE PERSONALIDAD.*

SUMMARY

The objective of this paper is to present an alternative to traditional
personality measures based on Cattell's T-data. An objective
assessment of personality should not assess what people say they
usually do. Rather, what they really do must be assessed. In other
words, using methodology where the scores are not based on the
self-assessment of individuals nor the assessor's need to make any
interpretation of or changes to the data. Moreover, using tasks that
require little skill, which means that every applicant is evenly matched
in this respect. Likewise, the assessment procedures do not provide
any feedback on the performance of individuals. Without feedback,
participants are, in part, free from context influences, so cannot
learn the correct response needed to obtain positive results.
Assessment aims should be masked. This means that subjects are
not informed about the objectives of the tasks. Rather, they are
encouraged to do their best in order to obtain as many points as
they can in an easy task. This strategy reduces involuntary bias
and intentional response distortion which might be present in
traditional personality assessments.

This paper presents three tests designed from that perspective, which assess Cooperation, Methodicalness and Risk Taking, as well as the results in terms of reliability and validity of the measurements that have been obtained.

The notion of intra-individual consistency lies at the basis of all personality psychology. As regards trait psychology, there exists a universal structure of human personality. Traditional personality assessment tries to reveal this universal structure through the identification of those verbal statements that tend to be grouped into the dimensions which make up this personality structure. Whether or not a verbal statement belongs to the dimension is established by the co-variation with others belonging to this category across a broad sample of subjects.

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. Thus, it is supposed, on the one hand, that there is a direct connection between what people say and how they behave, and, on the other hand, that there is a stable way of behaving, independently of environmental circumstances.

However, there is a lack of satisfaction about traditional self-report measures of personality. This lack of satisfaction is due to factors such as involuntary biases (e.g. acquiescence), voluntary response distortion (e.g. social desirability, faking) (Amelang, Schafer & Yousfi, 2002) and the limitations in predicting future behavior (Schmidt & Hunter, 2004). The three of them point out that none of the above mentioned suppositions can be firmly maintained.

Regarding the first one, the correlations between behavior and its verbal description are far from being high (Skinner & Howarth, 1975). It should be noted that people usually try to be consistent in their

statements. Thus, verbal descriptions are typically *congruent.* However, that does not imply verbal statements should coincide to what people really do. Moreover, if a person is aware he/she has been inconsistent, that person will try to justify or reformulate the statement. Consequently, consistency failures are due to, a) contingencies associated to those statements which result in social desirability or faking (e.g. when an applicant for a job is asked about whether or not he/she considers him/herself as an adaptive person), or, b) language inaccuracy, omissions and mistakes (e.g. when someone says he/she drinks alcohol frequently meaning once a week or daily). Those elements can affect the predictability of that measure.

As regards the second one, the stable way to answer regardless of the environmental circumstances, from our point of view, is Personality an historical notion which should not be considered irrespective of the contextual influences that have been shaping it. Furthermore, we think that personality or what we prefer to call *interactive style* is a *tendency for individuals to behave in a certain way in a specific situation.* This brings us to consider the situation and the relationship that the person has with the situation as a necessary element in proper personality assessment

Lastly, traditional self-report personality assessment has shown its usefulness in classifying individuals (particularly improved since the spread of the Big Five Model) but, as was mentioned, there are severe limitations in predicting future behavior. These limitations could be due to the above factors, but also because of the limited focus on the behavioral sequence questionnaire items usually have. Consider, for example, a person who answers, "I always have difficulty in striking up a conversation with strangers" on a typical introversion scale. From the point of view of personality, knowing how this result is manifested, is what is relevant. So, it is not the same if this person "avoids" interaction with strangers or if he/she simply does not have the opportunity to mix with them more than sporadically or if, without being able to avoid the situation, this individual has difficulties. The self-report assessment format precludes making these type of distinctions.

When the nomothetic outlook is superseded by further objectives such predicting the future behavior of an individual, the usefulness

of those instruments is clearly below expectations (Rees & Metcalf, 2003). In the end, it is not recommended to assess behavior using verbal statements. Actually, it should be consider as a paradox, the fact that no one will probably be able to accept the assessment of aptitudes by asking questions, as traditional personality assessment does, instead of observing how people behave in situations in which the aptitude should be shown.

As an alternative to that self-reported based measurement of personality, an objective personality measurement is proposed in which what people really do is assessed rather than what they say they usually do. In other words, it is considered the assessment of people's learning history crystallized through behavioral tendencies (which can be measured using T-data using Cattell's terminology) instead of this other which is synthesized from verbal statements (Q-data in Cattell's terminology).

An objective measure of personality means that the scores are not based on the self-assessment of individuals nor does the assessor need to make any interpretation of or changes to the data. Moreover, tasks require few skills, which means that every applicant is evenly matched in this respect. Likewise, assessment procedures do not provide any feedback on the performance of individuals. Without feedback, participants are, in part, free from context influences, so they cannot learn the *correct response* needed to obtain positive results. Assessment aims should be masked. This means that subjects are not informed about the objectives of the tasks. Rather, they are encouraged to do their best in order to obtain as many points as they can in an easy task. This strategy reduces involuntary biases and intentional response distortion (Robie, Born & Schmit, 2001) which might be present in traditional personality assessments (Hough, Eaton, Dunnette, Kamp & McCloy, 1990). This is particularly relevant for personnel selection processes where employment decisions might be affected (Christiansen, Goffin, Johnston & Rothstein, 1994). It also increases ecological validity (see Hundleby, 1973, Hernández, Santacreu & Rubio, 1999 for a review of objective personality assessment).

The present paper shows an example of an objective assessment of personality. Which is the use of three tests which assess Cooperation, Methodicalness, and Risk Taking Behavior.

THE OBJECTIVE TESTS

Cooperation Test

The Cooperation Test tries to assess the interactive style, *cooperation with a competitor* which is an aspect of pro-social behavior and a common scenario human beings have to cope with.

The test we have designed consists of a computer-based task in which the screen is divided into two panels, the upper panel corresponding to the candidate and the lower panel supposedly corresponding to another candidate present in the same room (a virtual partner) (see Figure 1). On the left side of each panel there is a frame with the participant's puzzle pieces randomly distributed inside, and on the right an empty frame. Each piece has to be moved to its correct place in the empty frame by tracking it with the mouse and releasing it in the right-hand frame. Participants can move their own and their partner's puzzle pieces, and the partner can also move pieces in both puzzles. When a piece is moved to the wrong place the error is indicated by a sound, and the piece is moved back automatically to its original place. When the piece is moved correctly, a different sound is made, and the piece remains in its place. Participants can see how the partner moves his/her pieces. There is a button labeled "continue", which can be used to skip the puzzle if the subject has finished his/her own puzzle and does not wish to wait for his/her partner. Participants are told that the task consists of completing the puzzle quickly, gaining as many points as possible in the shortest time. However, subject's and partner's scores were linked, in an attempt to make the participant understand that cooperation could be beneficial. Specifically, the participant was informed about the following rules for gaining points:

• If the participant moves a piece correctly in his/her own puzzle, +1 point for him/her.

• If the participant moves a piece correctly in his/her partner's puzzle, +1 point for him/her and for the partner.

• If the participant moves a piece to a wrong place in his/her own puzzle, -1 point.

• If the participant moved a piece to a wrong place in his/her partner's puzzle, -1 point for him/her and for the partner.

The display will show you two puzzles (yours and your partner's) with the pieces equally mixed up at the left side.

Timer

Your puzzle

you 00:02 Continue →

your partner
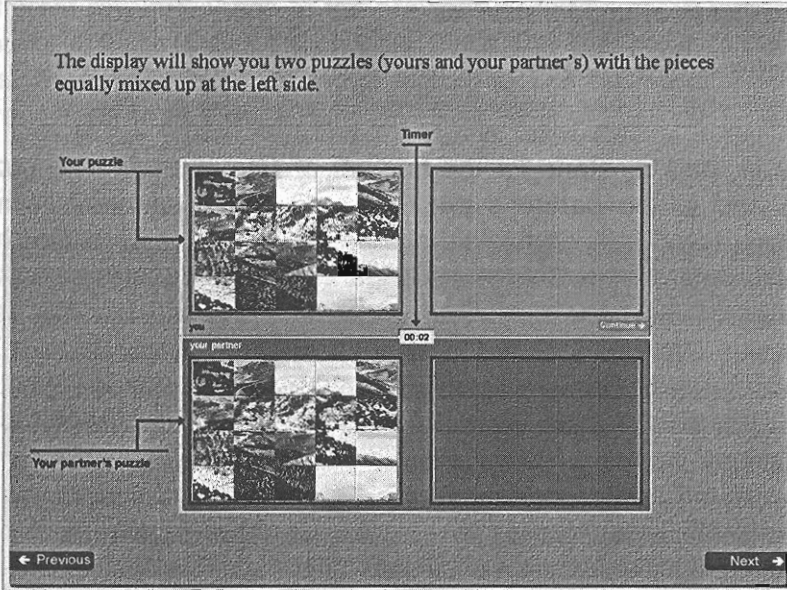
Your partner's puzzle

← Previous Next →

Figure 1.- An example of an instructions screen of the Cooperation Test

This scoring system served to foster the participant's predisposition to cooperate while suggesting that a good performance in one's own puzzle also leads to gaining points. The virtual partner was programmed to cooperate with the subject at a few fixed moments throughout the task if the participant did no cooperate at all. However, once the participant cooperates, the virtual partner essentially mimicks the participant's behavior.

The *index of cooperation* is,

where H_A and H_E stand for the number of *Hits* in one's own puzzle (*Ego*) and in the partner's puzzle (*Alter*). As is obvious from its structure, the *C* index reflects the degree to which participants share their hits with the partner. A participant that never cooperates with the partner will have C = 0. A systematic cooperator, that is, a participant that always shares hits with the partner, will have a 0.50 value. Intermediate values reflect different degrees of cooperation during the task, rather than a systematic way of facing it. Values significantly higher than 0.50 would be considered as reflecting a "perverse behavior", because it would indicate that the participant is checking the correct place by using the partner's panel.

Previous research has shown satisfactory internal consistency (Cronbach's a = 0.940) as well as temporal stability (one-year test-restest r = 0.404).

Methodicalness Test

Methodicalness is understand as an aspect of conscientiousness and is defined as an interactive style which consists of performing a task in an ordered, organized and according to a systematic pattern. The test designed consists of a computer task in which individuals should find all the icons in a 10-row x 12-column matrix that match the target (a specific kind of tree) intermingled with other icons (different kind of trees) and clicking them with the mouse as fast as possible (see Figure 2). The test is made up of 15 20-second trials and each matrix includes 14 target icons (distributed in seven rows and columns, 2 per row and 2 per column) and another 58 non-target icons, which means 60% of the cells filled. The methodicalness score is computed in accordance to the trajectory followed by the subject in his/her clickings. More precisely, one point is given if a subject correctly chooses an icon which is in the same row or column as the previous click. No point is given if the following icon is chosen from

A

a different row or column to the previous one. The same strategy is pursued until the end of the choices. Thus, the final score is the sum of these points, which can range from 0 to 7 in each trial. The total score ranges from 0 to 105.
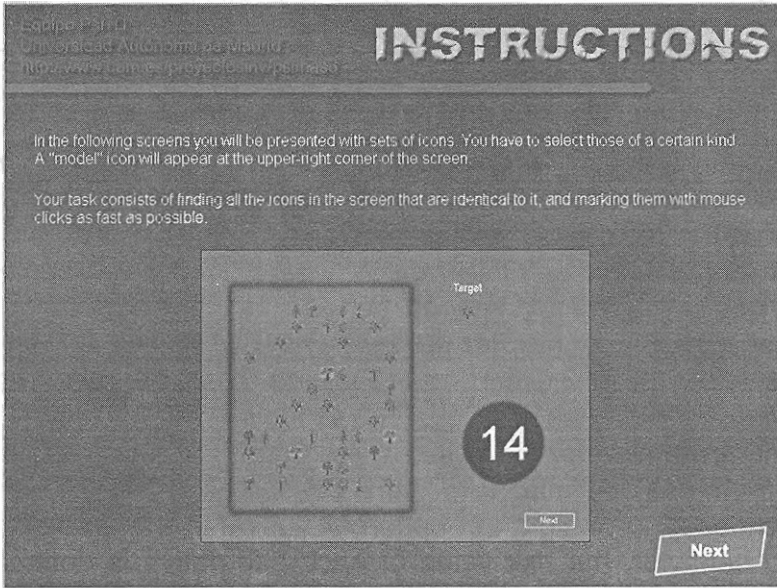

Figure 2.- An example of an instructions screen of the Methodicalness Test

Previous research (Hernández, Sánchez-Balmisa, Madrid & Santacreu, 2003) has shown quite acceptable reliability: Cronbach's $a = 0.760$; split-half realiability $r = 0.700$; one-year test-retest $r = 0.503$.

Risk taking Test

The Risk Taking Test has been designed as a Betting Dice task. In this case, people have to estimate the result of the sum of two dice. They have to bet on one of four options: *More than 4* (p = 30/ 36 = 0.83), More than 7 (p = 15/36 = 0.42), *More than 9* (p = 6/36 = 0.17), and a straight bet *Number 12* (p = 1/36 = 0.03). Individuals are told that each one is associated to different amounts of reward: 1 point, 2 points, 5 points, and 30 points, respectively (see Figure 3). They are encouraged to get as many points as they can. Subjects had 10 20-second-trials. It is assumed they are gathering points trial to trial. In each one subjects have to make their bets but are not informed of the number of trials during instructions, or about the results of their bets. Each trial finishes with a standard message such as *OK. And now what is your bet.* If someone does not bet in the 20 second period the trial consists of, the message *You have not bet on anything. You have not won any points* appears. In the end, the message *The task has finished* is shown. As can be verified, the expected values of the four options are the same, so it is assumed that subjects who bet on thirty numbers are making a more conservative bet than those who choose the straight bet option. As usual in the study of risk taking behavior, the risk value of each alternative is defined as the inverse of the probability of getting it right. In this task, the subject's risk-taking score is calculated as the average of the natural logarithm of the inverse of the probabilities of his or her choices ($1/p_{More\_than\_4}$ = 1.2; $1/p_{More\_than\_7}$ = 2.4; $1/p_{More\_than\_9}$ = 6; $1/p_{Number\_12}$ = 36, respectively) as follows:

$$R \text{ score} = \frac{\sum_{10}^{l=1} \ln(1/p_i)}{10}$$

Using natural logarithms gives a closer index of the risk behavior subjects are assuming when they are opting for one alternative out

of the others (see Arend, Botella, Contreras, Hernández & Santacreu, 2003). Previous research (see Sante & Santacreu, 2001) has shown satisfactory internal consistency (Cronbach's $a = .800$) and temporal stability (one-year test-retest $r = .600$).
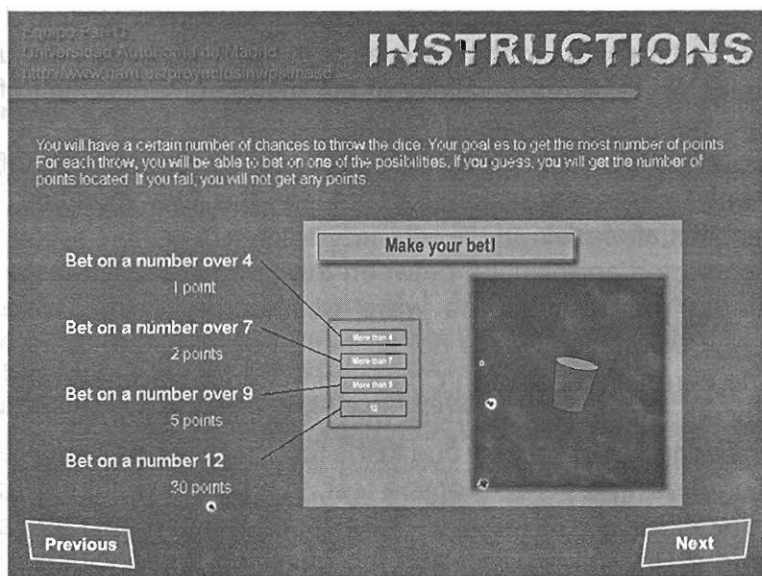

Figure 3.- An example of an instructions screen
of the Risk Taking Test

VALIDATION STUDY

A predictive validation study has been carried out using a sample of 4,966 university graduates, applicants for an *ab initio* air traffic control (ATC) training program. Applicants were assessed using a test battery that included, among others, the Cooperation Test, the Methodicalness Test, and the Risk-Taking Test. They were also

assessed in dimensions such as Spatial Ability, Attention, Reasoning, Verbal Comprehension, Emotional Adjustment, Extroversion, Normlessness, Persistence, Tolerance to Frustration, and Perception of Control. 47 applicants were selected for the ATC training program which consisted of five different modules: Basic Knowledge, Area Control Unit, Approach Control Unit, Aerodrome Control Unit, and No-ATS unit. The training program scores of the five modules were used as a criteria for testing the validity of the measurements. First of all, the range restriction corrected R was reasonably high for the five modules (adjusted R^2 were 0.198, 0.577, 0.715, 0.764, and 0.650, respectively). The regression ANOVA for the five modules was significant for all cases.

The corrected correlations between the three objective tests and the training program scores were satisfactory (see Table 1). Expectedly, correlations were not so high. For instance, there were almost no correlation between Cooperation and Methodicalness and Basic Knowledge. However, there were a significant correlation with the Risk Taking Test, probably due to the fact that this basic knowledge is measured using a multiple option test. In a previous study (Rubio, Hernández & Santacreu, submitted), we found a significant correlation between our Risk taking Test and the guessing tendency shown in a multiple option test which is also negatively related to the success on this test. Moreover, as regards the Cooperation Test, the higher correlation is found with the Area Control Unit in which working with a partner is fundamental.

Table 1.- Corrected correlations between the tests and the ATC training course scores

Training course units	Betting Dice Test	Crossing the Street Test
BASIC KNOWLEDGE	-.229	.200
AREA CONTROL UNIT	-.295	.082
APPROACH CONTROL UNIT	-.165	.403
AERODROME CONTROL UNIT	-.089	.193
NO-ATS UNIT	-.078	.153

The results show satisfactory reliability as well as validity of the tests and they support the feasibility of the objective assessment of personality.

REFERENCES

Amelang, M., Schafer, A. & Yousi, S. (2002). Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of the fake good instructions. *Psychologisque Beitrage*, 44, 24-41.

Arend, I., Botella, J., Contreras, M.J., Hernández, J.M. & Santacreu, J. (2003). A betting dice test to study the interactive style of risk-taking behavior. *The Psychological Record*, 53, 217-230.

Christiansen, N.D., Goffin, R.D., Johnston, N.G. & Rothstein, M.G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47, 847-860.

Hernández, J.M., Santacreu, J. & Rubio, V.J. (1999). Evaluación de la personalidad: una alternativa teórico-metodológica. *Escritos de Psicología, 3*, 20-28.

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. & McCloy, (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

Hundleby, J.D. (1973). The measurement of personality by objective tests. In P. Kline (Ed.), *New approaches in psychological measurements.* N. York: Wiley.

Rees, C.J. & Metcalfe, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology, 18, 156-165.*

Robie, C., Born, M.P. & Schmit, M.J. (2001). Personal and situational determinants of personality responses: A partial reanalysis and reinterpretation of the Schmit et al. (1995) data. *Journal of Business & Psychology, 16, 101-117.*

Rubio, V.J., Hernández, J.M. & Santacreu (submitted). *Can we predict Risk-Taking behavior? Two tasks for predicting guessing tendencies in a multiple option test.* Personality and Individual Differences.

Sante, L. & Santacreu, J. (2001). La eficiacia (o la suerte) como moduladora en la evaluación del estilo interactivo tendencia al riesgo. *Acta Comportamentalia*, 9, 463-486.

Schmidt, F.L. & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162-173.

Skinner, N.S.F. & Howarth, E. (1975). Cross-media independence of questionnaire and objective test personality factors. *Multivariate Behavioral Research*, 8, 23-40.