Two tasks for predicting guessing tendencies

Can we predict risk-taking behavior?
Two behavioral tests for predicting guessing tendencies in a multiple choice test

Running head: Two behavioral tests for assessing risk-propensity

Víctor J. Rubio(1), José M. Hernández(1), Flor Zaldivar(2), Mª Oliva Márquez(1), and José Santacreu(1)

(1)Universidad Autónoma de Madrid (Spain)
(2) Universidad de Almería (Spain)

Abstract
Objective behavioral personality measures are an alternative to traditional self-report personality assessment and overcome some of the difficulties traditional assessment shows. This paper presents two studies in which two measures of risk propensity are validated. In Study 1, 1,325 applicants for an ab initio air traffic control training program were assessed using the Roulette Test (RT) and the Crossing the Street Test (CtST). Convergent as well as criterion validity were tested using the guessing tendency demonstrated by participants in a multiple choice test. Once the effect of general cognitive ability was leveled out, results showed the predictive capacity of both tests. Study 2 replicated the previous study. 127 3rd year Psychology students were assessed using the RT. Results showed that the test had an even higher predictive validity of the Guessing Tendency (GT) shown in a multiple choice test. Usefulness for personnel recruitment processes and other assessment contexts are emphasized as are future trends of objective personality assessment.

Keywords: risk propensity, risk-taking behavior, objective personality assessment, behaviorally-based assessment, guessing tendency.
As any observer can assert, individuals clearly differ in risk-taking behavior. Moreover, it is assumed that human beings, at least some of them (Soane & Cmiel, 2005), demonstrate consistent risk-taking tendencies (Levin, Hart, Weller & Harshman, 2007; Meertens & Lion, 2008; Zuckerman, 1979, 1994). In fact this idea of people being consistent in their risk-taking behavior and showing individual differences in risk propensity (RP) has important implications for practical insights (Nicholson, Soane, Fenton-O’Creevy & Willman, 2005). For instance, many personnel recruitment processes have included the assessment of applicants’ risk tendencies (e.g. air traffic controllers) as a relevant variable and also carry out managerial placement programs in risk sensitive industries or sectors (e.g. investment advisers, control-room operators of nuclear power plants).

However, the same consensus is not found when considering how to measure risk propensity (see Harrison, Young, Butow, Salkeld, & Solomon, 2005, for a review of RP measures). For instance, several measures for more general personality traits supposedly related to RP (Extraversion, Psychoticism, Neuroticism) have been employed. This would be the case of the Eysenck Personality Questionnaire (EPQ) (Eysenck & Eysenck, 1998). Other instruments measure specific traits or facets such as Sensation Seeking, Impulsivity or Boredom Susceptibility. The Sensation Seeking Scale (SSS, Zuckerman, Eysenck & Eysenck, 1978) is one of the most frequently used. There are also everyday life situation tests in which simulated situations of uncertainty are proposed to participants that are intended to elicit their RP, such as Kogan & Wallach’s (1964) Choice Dilemma Questionnaire (CDQ). Other sorts of measures that have been employed to assess risk-taking tendency have been specific self-reports with a more or less wider domain scope, such as the Risk Propensity Scale (Nicholson, et al., 2004), which assesses recreational, health, career, financial, safety, and social risks; or the Domain Specific Risk Attitude Scale, DOSPERT (Weber, Bleis & Betz, 2002), assessing financial, health/safety, recreational, ethical and social risk domains. These instruments usually assess how often people engage in various risky behaviors or, alternatively, how they would behave in different situations (DOSPERT also includes how risky the situation is perceived in order to discriminate between risk perception and attitudes toward risk; see Weber et al., 2002).

The main concern of the measures cited above regards the fact that all of them are self-report based instruments. Self-reports referring to how people behave, what they say, think or feel (Q-data in Cattell’s terminology –see Cattell & Kline, 1977) are sensitive to response distortions due to self-presentation biases, acquiescence or faking (Robie, Born & Schmit, 2001). Cattell himself emphasized the need to complement self-report based information with data from the observation of individuals’ behavior when facing tasks (T-data in Cattell’s terminology).

The use of objective behavioral measures for assessing RP

Interestingly, in recent times a T-data alternative to examining personality factors of risky behavior has been produced. Three behavioral measures of risk propensity can be found in the recent scientific literature. Firstly, the Iowa Gambling Task (IGT, Bechara, Damasio, Damasio, & Anderson, 1994) is a computer-based task in which people have to select cards from one out of four decks in order to get as much money as they can; when the card is selected, the amount won is indicated. The decks differ in the amount of immediate gain versus the possibility of penalty with larger rewards and losses in decks A and B than in decks C and D. The optimum results are achieved avoiding decks A, B and selecting cards from decks C and D. The IGT is not a pure RP measure. Instead, it is a measure of decision-making impairment usually present in ventro-medial prefrontal cortex damaged patients. In fact, the expected values of the four decks are not the same. Thus, it is expected that impaired people will choose cards from the “advantageous” decks.

A second behavioral measure is the Balloon Analogue Risk Task (BART, Lejuez, et al., 2002) in which examinees pump up an on-screen balloon with the aim of making it as large as possible without causing it to explode. The balloons have different explosion probabilities from the first to the 128th pump; as long as it doesn’t explode, the larger the...
Two tasks for predicting guessing tendencies

balloon is pumped, the greater the gain is. The BART can be considered a “pure” measure of RP as long as it clearly presents a risky situation in which alternatives have different (uncertain) probabilities and rewards.

These two measures have demonstrated a relationship with some risk-taking behaviors such as substance abuse, alcohol consumption, smoking, etc. (Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005; Bechara, 2003; Skeel, Neudecker, Pilarski, & Pytlak, 2007; Skeel, Pilarski, Pytlak, & Neudecker, 2008). Even though most of the outcomes used have been self-reported risk-taking behaviors and no high correlations are presented, it is interesting to note that they usually predict different aspects to traditional questionnaire-based instruments. In fact, it has been suggested that questionnaires account for significant levels of overall risk behavior measures though behavioral measures account for the variance of specific subtypes of risky behaviors (Skeel et al., 2007; see Weiner, 2005 for a review on the differences between self-report and performance-based measures). This result supports the idea of Q-data and T-data capturing different parts of the construct.

A third behavioral measure of RP is the Betting Dice Test (BDT) (Arend, Botella, Contreras, Hernández & Santacreu, 2003). It consists of betting on one out of four alternatives which can result when throwing two dice. The higher the probability of occurrence, the lower the reward associated. Contrary to the BART, with the expected values of the alternatives changing along the test, the expected values of the BDT options are the same. It is therefore assumed that subjects who bet on the more probable but less rewarded alternative are making a more conservative election than those who choose the less probable though more rewarded bet. Moreover, no feedback is given after each bet. Thus, individuals’ elections are free from the previous results obtained. The BDT assumes that risk propensity is a stable way of behaving that is idiosyncratic, historically configured, and is demonstrated in interaction with a risky situation in which two or more response alternatives are involved. That is, the election should not be context-specific. The test has developed from a risk-return trade-off framework: potential return rises with an increase in risk; low levels of risk are associated with low potential returns, whereas high levels of risk are associated with high potential returns.

The BDT has shown satisfactory internal consistency as well as promising criterion validity (see Arend, et al, 2003; Rubio, Santacreu & Hernández, 2004, 2006; Santacreu, Rubio, & Hernández, 2006). For instance, the BDT has shown significant regression coefficients in the prediction of success in passing the ab initio Air Traffic Control (ATC) training course which include, among others, an outcome regarding the safety distance maintained between airplanes in the simulator (Santacreu et al., 2006).

Nevertheless, there is an important limitation of these behavioral measures: their lack of convergent validity. That is the case for the IGT and the BART (Aklin et al., 2005; Skeel et al., 2007, 2008) and it is also the result obtained by the authors between the BART and the BDT in research currently in progress (Rubio, Márquez, Narváez & Santacreu, 2008). Thus, the need for new behavioral measures to investigate convergent validity of objective testing arises. In fact, the lack of convergent validity of objective behavioral measures is, at least partially, related to the controversy about domain specificity of risk-taking behavior versus consistency of risk preferences. If risk-taking behavior is mainly domain specific, convergence is expected only between instruments which assess the same risk-domain. Moreover, it is expected that behavioral measures’ scores will be context determined to a greater extent than self-report based instruments, particularly those including items with non specific situations. This is due to the fact that questionnaire items usually present equivalent wording (Zuckerman, 1979) as well as statements which would be applied to a large range of situations. This would also contribute to explaining the lack of convergent validity between self-reports and behaviorally-based measures. Contrary to that, if RP is a consolidated behavioral pattern, it is expected that RP will be shown in such situations where there is no “correct response” to cope with the task. We refer to “correct response” as taking into account the
Two tasks for predicting guessing tendencies

fact that a situation might impose that just one and only one response is acceptable or valid. When abilities such as mathematical skills or mental reasoning are assessed, there is only one correct response. However, when personality is assessed there is no correct response: there should not be a unique response maximizing the efficacy of solving the task. This lies at the basis of the faking phenomenon. Personality questionnaires would be filled out according to what the examinee thinks the assessor expects. Objective behaviorally-based assessment must involve task designs in which response efficiency is independent of the behavioral tendency. Therefore, if risk propensity is a consolidated behavioral pattern as a result of previous interactions with risky situations, and assuming that people have had the chance in their interactive history of setting this consolidated pattern, it is expected that different assessment methods using situations in which several alternatives have the same objective utility should share at least in part their variance.

On this line, two new tests have been developed. The measures presented are objective tests in Cattell’s sense (Cattell & Kline, 1977). Assessment procedures do not provide any feedback on the performance of individuals. Feedback on previous results has shown an important influence on following risk-taking behavior (Romanus & Gärling, 1999; Sullivan & Kida, 1995). Furthermore, assessment aims are masked: subjects are not informed about the objectives of the measurement. Rather, they are encouraged to perform their best in order to obtain as many points as possible. This strategy reduces biases as well as increases ecological validity (Hundleby, 1973).

The first test is the Roulette Test (RT). The RT is equivalent to the BDT but using a roulette game scenario in which people have to bet on one out of four possibilities: thirty numbers (from 1 to 30, \( p = 30/36 = .83 \)), fifteen numbers (even numbers from 2 to 30, \( p = 15/36 = .42 \)), six numbers (from 31 to 36, \( p = 6/36 = .17 \)), and a straight bet (number 17, \( p = 1/36 = .03 \)). Individuals are told that each one is associated with different amounts of reward: 1, 2, 5, and 30 points, respectively. They are encouraged to get as many points as they can. Participants had ten 20-second trials. It was understood they gathered points from trial to trial. In each one, participants had to make their bets but they were not informed of the number of trials, nor of the results of their bets. Each trial finished with a standard message such as ‘OK. And now what is your bet?’ At the end, the message ‘The task is finished’ was shown. The risk value of each alternative is defined as the inverse of the probability of getting it correct. In order to avoid the overweight the option ‘Straight bet’ produces, a natural logarithmic transformation has been used as follows:

\[
\text{Task 1 (roulette) risk score} = \frac{10}{\sum_{i=1}^{10} \ln(1/p_i)} \quad [1]
\]

It is considered that the higher the score, the greater the risk assumed. Previous research (see Arend et al, 2003) has shown satisfactory internal consistency (Cronbach’s \( \alpha = .83 \)) and moderate temporal stability (one-year test-retest \( r = .43 \)).

The second test is the “Crossing the Street” test (CtST), which consists of deciding where a pedestrian representing the examinee should cross from one sidewalk to another in order to reach a Chemist located on the other side of the road at the left side of the screen. The aim is that the pedestrian should get to the chemist as quickly as possible but without causing a car crash or being knocked down. The starting position of the pedestrian is also the left side of the screen and it is moved by clicking over the arrow keys displayed. In the beginning, cars run from the left to the right coming out from a tunnel and the arrows are not active. After 10 seconds, the road is “veiled”. Thus, no car can be seen. Participants then have to decide where they will cross paying attention to the fact that the closer to the left corner of the screen they decide to cross, the less they are able to see if a car is coming. However, the further away they are from the left corner, the more time is spent crossing the street. It is assumed that participants are greater risk-takers the closer to the left side of the screen they decide to cross. Subjects had 10 identical trials and they were not informed about any sort of accident they could have
produced nor about the number of trials. The risk score is equal to the average of the distance ($d$) from the left side of the screen in a horizontal axis computed in pixels, as follows:

$$\text{Task 2 (crossing the street) risk score} = \frac{\sum_{i=1}^{10} d_i}{10} \times (-1)$$

[2]

The measure has shown high levels of internal consistency (Cronbach’s $\alpha = .96$) (Rubio et al., 2006).

Using guessing tendency as a risk propensity validity criterion

The present paper analyses the convergent and criterion validity of two objective behavioral measures of RP using the guessing tendency (GT) shown by participants on a multiple choice test as a criterion. GT has been used as validity criterion because, firstly, natural setting criteria are needed instead of the commonplace convergent correlation with some other (usually self-report) instrument supposedly measuring the same construct. Secondly, GT is a more direct measure of RP behavior compared to frequently used others which are much more multi-determined, such as driving offenses or smoking behavior. Indeed, guessing behavior has been considered in itself as a measure of risk-taking behavior (e.g. Slakter, 1969; Ziller, 1957a). Thirdly, because guessing behavior in a multiple choice test under certain circumstances (when wrong answers are penalized) clearly involves risk-taking behavior (Ben-Shakhar & Sinai, 1991; Slakter, 1969; Ziller, 1957a, 1957b): the objective is to obtain as many points as possible, there are several alternatives (omitting or guessing) associated to different outcomes (obtaining the point or not) with different probabilities each (1.0 versus 1/3). Nevertheless, GT is probably a multi-determined phenomenon also affected by other aspects beyond RP. It is assumed that if an examinee knows the right response to a particular item he/she will answer correctly. But if he/she does not know the answer, he/she would either skip that question or guess the answer at random. However, examinees who have partial information about an item do not respond at random, nor do those who wrongly believe they know the correct answer. In these situations wrong answers cannot be equally attractive to the examinee. Thus, factors such as impulsivity, extraversion and anxiety would play a role in guessing behavior (Avila & Torrubia, 2004; Matters & Burnett, 2003). In accordance with this, it is expected that RP will not account for 100% of the variance of guessing tendency.

Cognitive ability (CA) has also been associated with GT (von Schrader & Ansley, 2006). Indeed, several authors have pointed out that guessing is the wiser behavior when coping with a multiple-choice test (Slakter, Koehler, Hampton & Grennell, 1971). That is, a “rational examinee” should guess if he/she is aiming to maximize his/her expected score. Thus, for checking criterion validity of RP measures, general CA should be considered.

In order to test the convergent and criterion validity of the objective measures using the GT shown in a multiple choice test, two different studies have been carried out. In Study 1, extremely motivated applicants for a highly complex job were assessed using RT and CTST as well as a multiple choice test as a part of their selection process. Study 2 used graduate students who were assessed in RT and a multiple choice test exam was conducted in order to check the robustness of our prior result as well as to improve the criterion measure.

Study 1

Method

Participants. 1,325 Spanish university graduates, 743 males (M age = 28.7 years; SD = 4.1) and 582 females (M age = 28.3 years; SD = 3.9). All participants were applicants for an ab initio air traffic control (ATC) training program. The job of an Air Traffic Controller (ATCo) is highly valued in terms of prestige, as well as earnings. Moreover, in the recruitment process, only 300 people were to be selected. Thus, it can be expected that all participants were highly motivated to do their best.
Two tasks for predicting guessing tendencies

Measures. Two different computer-based behavioral tests were used for assessing RP: the Roulette Test and the “Crossing the Street” Test described above.

A 20-item test (four alternatives each) for the assessment of ATC basic knowledge, such as aerodynamics, meteorology, cartography, ATC regulations, etc., was also used. The test was designed by the Spanish National Agency in Charge of Air Navigation (AENA). Instructions about completing and scoring were given before starting the test. Applicants were informed that they should not answer a question if they were not sure which was the right option since a 1/3 of the errors would be deducted from the total score. Test score was the sum of the correct responses minus 1/3 of the errors.

A computer adaptive test for assessing reasoning, the TRASI (Rubio & Santacreu, 2003), was also applied in order to control the effects of general cognitive ability on the guessing behavior. TRASI consists of a bank of 98 items calibrated using an IRT three-parameter (3p) model (Birnbaum, 1968).

Procedure. Participants were assessed over a period of eight weeks. All of them completed the two behavioral tests, the TRASI and the ATC basic knowledge test, in a computer-based format as part of the recruitment process. Firstly, Guessing Tendency was estimated using an IRT 2p model (Birnbaum, 1968). For this, correct responses had been ruled out and errors and omitted responses were coded as 1 and 0, respectively. Thus, the latent trait $\theta$ represents the tendency to omit and it was computed for each subject using the BILOG-W (Mislevy & Bock, 1990). Subsequently, $\theta$ was inverted in order to represent the GT. Secondly, subjects with ten or more errors/omitted responses were selected. Otherwise, the GT standard error of estimation would be too high (higher than .65). 514 participants fulfilled this requirement (249 females and 265 males). Thirdly, the accuracy of the GT estimation was checked through the correlation between $\theta$ and Ziller’s $G_1$ (1957a, 1957b), which is generally used for the measurement of the tendency to omit. $G_1$ is based on the ratio of the estimated number of guesses to the estimated number of items the examinee does not know. It ranges from 1 (maximum GT) to 0 (no item guessed). The advantage of this index is that it controls the true ability of the examinee, as follows:

$$G_1 = \frac{W_2}{W_2 + O_1}$$  \[3\]  
where $O_1$ is the total number of items omitted, $W_2$ is the estimated number of guesses:

$$W_2 = W_1 + W_1/(K - 1)$$  \[4\]  
$W_1$ is the number of wrong answers and $K$ is the number of alternative answers per question.

The correlation between $\theta$ and $G_1$ was $r = .99$ which confirms the accuracy of the estimated GT.

Fourthly, in order to control the influence of intelligence on guessing behavior, a linear regression analysis was carried out, using the TRASI scores as the predictive, and guessing behavior as the dependent, variable. Standardized residuals were then saved.

Fifthly, Correlations between RT and CtST scores as well as both tasks and the Guessing Tendency were computed. Subsequently, a multiple regression analysis was carried out in order to establish the power of the two behavioral tests in predicting the GT of individuals. Giving attention to the fact that some other variables have demonstrated a role in risk-taking behavior, such as gender (see the Byrnes, Miller & Schafer, 1999’s meta-analysis) and age (see Elander, West, & French, 1993’s review), both were also included as predictors in the analysis. The SPSS v15.0 statistical package was used for the calculations.

Results

The internal consistencies obtained with the current sample were Cronbach’s $\alpha_{RT} = .82$ and $\alpha_{CtST} = .93$. Table 1 shows the descriptive data of the two tests and the GT.
Two tasks for predicting guessing tendencies

The correlation between the two behavioral tests was \( r = .51, p < .0001 \). The correlations between the two behavioral test scores and the Guessing Tendency were not high but were significant: \( r_{RT,GT} = .28, p < .0001; r_{CtST,GT} = .27, p < .0001 \).

The previous linear regression for checking the effects of intelligence on guessing behavior showed that cognitive ability accounted for only 9% of the variance of the GT. A new regression analysis using the standardized residuals showed that 16% of the variance of GT was accounted for by the four variables included in the model. Nevertheless, only RT and CtST showed significant regression coefficients (see Table 2).

However, the distribution of the guessing tendency showed a very skewed shape. It was due to the fact that the knowledge test was too easy for most of the participants. In order to improve the measure of the dependent variable using a more balanced distribution, a second study has been carried out employing a more difficult test.

Study 2

Method

Participants. 127 3rd-year Psychology degree students (22 males and 105 females; \( M \text{ age} = 22.9; SD = 4.4 \)) volunteered to participate in the study in exchange for credits in a Psychological Assessment course.

Measures. Due to time restrictions, only the Roulette Test which was used in the previous study was also used here as a measure of RP. A 50-item four-choice test for assessing the contents of the Psychological Assessment course was used for assessing GT. Examinees were informed that the score for this test would be the sum of the correct responses minus 1/3 of the errors. Students needed to obtain at least 20 points to pass the exam. The TRASI was also used as a measure of cognitive ability.

Procedure. At the beginning of the course, each student’s general mental ability was assessed for teaching purposes. At the end of the semester, those who wanted to participate were assessed, in groups of 20, on the RT before the exam date. In order to increase motivation and ensure their commitment to the task, students were informed that they were able to obtain credits for their course according to their performance in the computer task (though all of them finally obtained the credits). Two weeks later, they were tested in order to pass the course. Taking into account that the procedure, based on an IRT-2p model proposed in the previous study, requires large samples and that the correlation between the \( \theta \) obtained and Ziller’s \( G_1 \) (1957a, 1957b) was almost perfect, in this case the \( G_1 \) was used to estimate GT. Individuals with more than 25 correct answers were ruled out in order to make a robust estimation of their GT using a similar criterion as in Study 1. 95 students remained for further calculations.

Subsequently, standardized regression residuals of guessing behavior using intelligence as a predictor were saved in order to rule out the influence of cognitive ability on guessing behavior. Then, a multiple regression analysis using RT, age and gender as predictors of the standardized residuals of \( G_1 \) as the criterion variable was computed.

Results

The internal consistency obtained with the current sample was Cronbach’s \( \alpha_{RT} = .72 \). Table 3 shows the descriptive data of the variables of Study 2.

Correlation between RT and GT was \( r = .48 (p < .0001) \). The previous linear regression of intelligence on guessing behavior showed that cognitive ability accounted for only 12.7% of the variance of the GT. The multiple regression between predictors and GT as a criterion amounted to \( R^2 = .25 \). Only RT had a significant effect (see Table 4).

General Discussion
Two tasks for predicting guessing tendencies

Previous research has suggested that objective behaviorally-based measures may be used to measure RP. The present paper attempted to test the criterion as well as the convergent validity of two behavioral measures of risk propensity, the Roulette Test and the Crossing the Street Test, designed from Cattell's perspective, using a real-life situation: guessing in a performance test.

Regarding the criterion validity, both Study 1 and 2 demonstrated that the Roulette Test and the Crossing the Street Test have a good capacity to predict the Guessing Tendency of individuals when responding on a multiple choice test. Even though GT is a multi-determined phenomenon also affected by other factors such as extraversion, anxiety and impulsivity, objective tests have shown a satisfactory predictive power. Furthermore, general cognitive ability, which has also been related to GT (von Schrader & Ansley, 2006), was previously leveled out, and the percentage of the GT variance explained could be directly attributed to the risk tests: 16% and 25%, respectively. Particularly the second study has raised the correlation between the behavioral test and the criterion to .48.

Consequently, it can be argued that both tests, the RT and the CtST, are useful as measures of RP and are able to predict the behavior of people in a real-life situation. These results support those obtained using other behavioral tests and different risk-taking behaviors (Bechara, 2003; Aklin et al., 2005; Santacreu et al., 2006; Skeel et al., 2007). In fact, it has been pointed out that self-report personality measures and behaviorally-based measures of risk-taking might each have valuable unique inputs in the prediction of risky behaviors. For instance, Skeel et al. (2007) have found that questionnaire personality variables accounted for significant proportions of variance within overall risk behaviors, while the BART and the IGT accounted for a significant proportion of unique variance beyond that accounted for by self-report in the prediction of rebellious behaviors (a subtype of risky behaviors which are not significantly predicted by questionnaire factors). In the end, the low correlations between these kinds of measures reflect the differences between what Cattell called T-data and Q-data. In other words, the differences between what people do and what people say about what they do. Thus, a complete assessment of personality variables should require both types of information in order to improve the predictions of future behavior. As Skeel et al (2007) noted, the use of behaviorally-based measures would provide a number of advantages in addition to more traditional personality measures. They may add utility by discriminating between the different factors involved in risky behaviors (e.g. environmental triggers).

Finally, these instruments which are (1) not based on the individuals' verbal statement and, (2) keep the assessment aim masked would be very useful in such settings (forensic, personnel recruitment) in which risk-taking behavior is a relevant variable and where possible faking due to self-presentation biases would distort the results of the assessment (e.g. forensic and personnel recruitment).

Nevertheless, objective risk propensity behaviorally-based tests have also shown a lack of correlation among each other (Aklin et al., 2005; Rubio et al, 2008; Skeel et al., 2007, 2008). Another objective of the present paper has been to design new measures to investigate convergent validity. According to the results obtained, it cannot be affirmed that both tests are measuring exactly the same facet of the same construct. However, the results do not support the idea of risk propensity as a domain specific construct either. Contrary to this, a significant correlation of about .5 has been found between both tests although each one refers to a different domain. Even so, they were both developed as risky situations in which different alternatives have pay-offs and probabilities of occurrence that are inversely proportional, and higher convergent validity was expected to be found. The results obtained would be due to several differences between the tests. Firstly, RT presents a game of luck, whilst CtST presents a natural setting closer to real situations individuals have to face. Each one clearly regards a different risk domain. But each one would also elicit different anticipated emotions. The CtST induces an “emergency context” (the examinee needing to cross the street in order to reach the chemist) which would evoke negative experiences. As several authors have pointed out (Cooper, Flanagan,
Talley & Micheas, 2006), positive emotional experience is less consequential for risk taking behavior than negative experience due to negative emotions having strong motivational consequences. Secondly, even though neither task includes losses, the CtST represents a situation in which the losses are implicit: the examinee is looking for a chemist and would be knocked down when crossing the street. Thus, it is possible the CtST leaks a framing situation of losses instead of the RT. Thirdly, the CtST scenario would be perceived as a controllable situation. Contrary to this, the RT context is an uncontrollable one. In the context of risk acceptability, people tend to prefer controllable over uncontrollable risks (Klein & Kunda, 1994). Moreover, individuals tend to underestimate risks that are under their control (Weinstein, 1984). The possible effects of these factors are in opposite directions and would be non-linear. Thus, correlations between both tests are not as high as expected. This would also explain why RT is a better predictor than CtST in Study 1. Once the role of a general cognitive ability has been leveled out, residuals of guessing tendency are closer to the uncontrollable context proposed by the RT than the apparently more controllable situation faced in the CtST.

Two limitations should be mentioned. Firstly, it should be noted that the criterion used in Study 1 was a test with a very skewed distribution. This reduced variability might limit the accuracy of the prediction. When a longer and more difficult test was used, as in Study 2, predictions were improved. On the same line, data showed that samples from Study 1 and from Study 2 were quite different in average RP as well as the role of general cognitive ability. Our judgment is, however, that this serves to increase the robustness of the results. The second limitation we would like to consider is related to the fact that no traditional self-report personality measure was used in combination with the behavioral measures. Future studies should compare the results of predicting individuals' risk-taking behavior in natural settings using both types of assessment procedure.

In summary, this methodology, based on individuals' performance rather than their self-report, opens up a promising scenario. According to our results, these behavioral measures (T-data in Cattell's terminology –Cattell & Kline, 1977) may be generalized to natural settings.
Two tasks for predicting guessing tendencies

References


Two tasks for predicting guessing tendencies

Table 1. Descriptive statistics (between brackets, Standard Error of Skewness and Kurtosis) of Roulette Test (RT), Crossing the Street Test (CtST) and Guessing Tendency (GT) for the whole sample (N = 1,325) and for the subsample of those participants who were able to guess (with less than 10 correct responses in the multiple-choice test; N = 514).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>.87</td>
<td>.94</td>
<td>.51</td>
<td>.53</td>
<td>1.15</td>
<td>1.35</td>
<td>2.95</td>
<td>3.77</td>
</tr>
<tr>
<td>CtST</td>
<td>333.68</td>
<td>296.63</td>
<td>169.90</td>
<td>158.32</td>
<td>.44</td>
<td>-.64</td>
<td>-.65</td>
<td>.15</td>
</tr>
<tr>
<td>GT</td>
<td>.34</td>
<td>-.72</td>
<td>1.10</td>
<td>.66</td>
<td>-.61</td>
<td>-.23</td>
<td>.33</td>
<td>-.58</td>
</tr>
</tbody>
</table>
Two tasks for predicting guessing tendencies

Table 2
Study 1. Summary of Multiple regression analysis for variables predicting Guessing Tendency (N = 514)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>.40</td>
<td>.12</td>
<td>.19**</td>
</tr>
<tr>
<td>CtST</td>
<td>.01</td>
<td>.01</td>
<td>.13*</td>
</tr>
<tr>
<td>Age</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
</tr>
<tr>
<td>Gender</td>
<td>-.01</td>
<td>.14</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Note. RT: Roulette Test; CtST: Crossing the Street Test; GT: Guessing Tendency; $B$: regression coefficient; $SE B$: Standard error of $B$; $\beta$: Standardized regression coefficient. $R^2 = .16$

* $p < .05$; ** $p < .01$
Table 3
Study 2. Descriptive statistics (between brackets, Standard Error of Skewness and Kurtosis) of Roulette Test (RT) and Guessing Tendency (GT) for the whole sample (N = 127) and for the reduced subsample of those participants who were able to guess (with less than 25 correct responses in the multiple-choice test; N = 95).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
<th>Whole sample</th>
<th>Reduced subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>1.29</td>
<td>1.35</td>
<td>.96</td>
<td>.55</td>
<td>.96</td>
<td>1.43</td>
<td>.88</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.21)</td>
<td>(.30)</td>
<td>(.43)</td>
<td>(.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>.49</td>
<td>.56</td>
<td>.20</td>
<td>.20</td>
<td>.41</td>
<td>.16</td>
<td>-.02</td>
<td>-.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.23)</td>
<td>(.32)</td>
<td>(.45)</td>
<td>(.63)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Study 2. Summary of Multiple regression analysis for variables predicting Guessing Tendency (N = 95)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>1.00</td>
<td>.20</td>
<td>.47***</td>
</tr>
<tr>
<td>Age</td>
<td>.04</td>
<td>.02</td>
<td>.16</td>
</tr>
<tr>
<td>Gender</td>
<td>.10</td>
<td>.25</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note. RT: Roulette Test; GT: Guessing Tendency; B: regression coefficient; SE B: Standard error of B; β: Standardized regression coefficient. $R^2 = .25$  
* p < .05; ** p < .01; ***p < .001