

Article

Cognitive Reflection and General Mental Ability as Predictors of Job Performance

Jesús F. Salgado *, Inmaculada Otero and Silvia Moscoso

Department of Political Science and Sociology, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain; inmaculada.otero@usc.es (I.O.); silvia.moscoso@usc.es (S.M.)

* Correspondence: jesus.salgado@usc.es; Tel.: +34-881-815-147

Received: 11 September 2019; Accepted: 12 November 2019; Published: 18 November 2019

Abstract: This paper presents a study with four independent samples on the validity of cognitive reflection (CR) for predicting job performance and academic outcomes. The results showed that CR was a valid predictor of academic outcomes (i.e., grades and exam marks) and job performance (i.e., assessment center ratings). The magnitude of the CR validity was similar to the validity of GMA. Moreover, we found that CR and GMA were moderately correlated, and multiple regression analyses showed that CR added a small amount of validity over GMA validity for predicting both types of performance. Finally, we discuss the implications of the findings for the theory and practice of predicting performance and organizational sustainability.

Keywords: organizational sustainability; sustainable employability; cognitive reflection; job performance; grade point average; incremental validity; general mental ability; dual-process theory

1. Introduction

Sustainability refers to concerns with the effects of human practices on the physical environment and how to conserve and recycle natural resources in an efficient way but, recently, the impact of organizations on the physical environment has also been a focus of interest, and some organizations are now called “green” organizations to reflect the extent of their effective management of organizational and individual environmental-related behaviors [1–4]. However, as Pfeffer [5] pointed out, we should also be concerned with the organizational effects on the social environment, because organizations and their managerial practices have profound impacts on human beings and social environments. To this purpose, Pfeffer [5] coined the term organizational sustainability (OS) to refer to how organizational activities and human management practices affect employee health, well-being, and performance sustainably. Sustainable performance at work means maximizing job performance, as well as worker health and well-being. As suggested by Di Fabio [6], sustainable performance is a necessary condition to foster positive, healthy organizations.

An important issue related to OS is sustainable employability (SE), which has been conceptualized as employees’ abilities to function adequately at work and in the labor market throughout their working lives [7,8]. OS and SE are two critical factors in the capacity of organizations to implement structures, processes, and dynamics that can ensure their development, progress, continuity, and endurance, as well as their capacity to renew and recycle; i.e., to be sustainable.

Both concepts, OS and SE, may be integrated within the framework of the psychology of sustainability and sustainable development (PSSD), defined by Di Fabio [2] and Di Fabio and Rosen [9] as the study of the psychological variables and processes to improve the quality of life of each human being with and in the environments (e.g., natural, personal, social, and organizational). The PSSD aims to enhance the sustainability of intra and interpersonal talents to promote effective and

sustainable well-being for individuals, groups, organizations, and societies from a psychological research perspective.

SE is a characteristic of the individual (i.e., actual and potential employees) that depends on personal factors (e.g., cognitive abilities, rationality, knowledge, personality, motivation, education, and so on), work factors (e.g., job complexity; role clarity; leadership practices), organizational factors (e.g., selection and assessment practices, culture), and societal factors (e.g., political and legal system). As conceptualized by van der Klink et al. [7], and Fleuren et al. [8], SE is a multidimensional/multifaceted construct that can be defined as the degree to which individuals are able to be employed throughout their entire working life. SE is the responsibility of both employees and employers to maintain the employee's ability to work, as the employee's capacity to perform efficiently depends on individual, work, and organizational factors. SE means that, throughout their working lives, workers can achieve tangible opportunities in the form of a set of capabilities. Research on the role of cognitive processes and factors (e.g., dual-processing, cognitive reflection) and on individual differences (e.g., cognitive abilities, personality) is critical for a comprehensive explanatory model of SE and its effects on employee job performance and well-being.

The challenges of the current labor market (e.g., innovation, competitiveness, and increased productivity) have shown that OS and SE have become increasingly relevant for both organizations as employers and individuals as employees [5]. Organizations must develop effective processes for hiring, developing, maintaining, and motivating employees to achieve organizational goals. Simultaneously, individuals need to provide organizations with abilities, knowledge, and skill for ensuring productivity and innovation at work.

Among the processes that help organizations to become sustainable, their entry processes (i.e., selection and assessment processes; socialization processes) are particularly important, because they allow organizations to improve performance and productivity, reduce turnover, insure working life conditions of quality, reduce absenteeism, and increase job satisfaction, fairness, and inclusiveness. Therefore, a fundamental requirement for effective OS and SE is that the selection processes allow the incorporation of potential employees with high performance. To this respect, sustainable personnel selection can be conceptualized as an organizational intervention that focuses on the individual having (a) right cognitive, emotional, and social abilities, personality, skills, knowledge, and competences to perform the job, (b) on individual high-performance and productivity aspects (e.g., task performance, citizenship behavior; innovation, knowledge sharing), and (c) on long-term effects of individual behavior at work. Organizations need highly innovative and productive workers to survive, and personnel selection can help to identify those workers. Hiring highly innovative, high-performance employees is of critical importance for building sustainable organizations. Selection processes are the first step for successful and sustainable organizational entry.

Depending on how they are managed, personnel selection practices can have positive or negative consequences on employees' life and individual job performance, and on the organizations as a whole. For example, sustainable selection practices can improve person-job fit and person-organization fit producing inclusive and discrimination-free workplaces [10]. "Bad" selection practices can create adverse impact, discrimination at work, unfairness, and exclusiveness.

A great deal of research has examined the validity of personnel selection procedures for predicting organizational criteria and, particularly, individual and organizational performance. Literally, thousands of studies have been conducted for estimating, among others, the relationships between performance and general cognitive ability, emotional intelligence, personality, interviews, assessment center, and work-sample tests. However, no studies tested the validity of cognitive reflection (CR) for predicting job performance and training success (e.g., academic outcomes). Moreover, another scarcely explored issue is the relationship of CR with general mental ability, when the two variables are used for predicting performance. A third untested issue is whether CR shows incremental validity over general mental validity for predicting job performance and training success.

This article aims to shed further light on four neglected (and critical) issues that current organizational psychology research has overlooked: (a) to review the conceptualization of CR and

the research conducted until now related to organizational settings; (b) to examine the validity of CR for predicting academic performance and job performance; (c) to estimate the relationship between CR and general mental ability; and (d) to determine the incremental validity of CR over GMA for predicting these two crucial outcomes for building a sustainable organization.

2. The Dual-Process Theory of Cognition and Cognitive Reflection

The view that the human mind operates through two types of highly differentiated cognitive processes has been proposed in several psychological disciplines, including cognitive, social, and personality psychology and neuroscience [11–19]. This consensus view of mind has been termed Dual-Process Theory (DPT) of cognition, and it is one of the most relevant approaches in the study of human rationality [20,21].

The DPT of cognition distinguishes between two processing systems, System 1 (S1) and System 2 (S2) [13,19,22]. On occasion, S1 is also called Type 1 processes, and S2 is called Type 2 processes. S1 is fast, impulsive, automatic, unconscious, is always active, and operates in parallel with no (or little) effort and without interfering with other Type 1 processes or with Type 2 processing [13,21,23]. S2 is slow, purposeful, reflective, and operates with effort and concentration. S2 is mainly serial, controlled, works linearly, and one of its most important functions is to override S1 [13,21,23]. S1 encompasses processes that imply affective (emotional) responses, related to unconscious experiences and learning, and with rules and principles that have been automatized (because they were introduced by S2). Furthermore, S1 tends to operate using heuristics, biases, and shortcomings that help it to function quickly and without effort [13,23]. However, S2 runs with great attention and concentration. It is capable of solving complex problems with high accuracy, but is computationally expensive. Because of these last characteristics, Toplak, West, and Stanovich [24] suggested that human beings are “cognitive misers” due to the tendency of our brain to avoid the cognitive cost implied in activating the S2. Therefore, we tend to process with S1, i.e., using cognitive shortcomings that require less effort and concentration, although this can cause less accurate and much more biased responses.

Concerning the DPT of cognition, Kahneman and Frederick [14] developed the concept of “cognitive reflection” (CR), which has been defined as the capacity of an individual to annulate (stop) the first impulsive response, frequently wrong, that our mind offers (typical of S1) and to activate the reflection mechanisms (i.e., S2) that allow us to find a response, make a decision, or carry out a specific behavior in a more reflexive and correct way [14,25].

Frederick [25] developed a test to assess the individual differences in the activation of the two systems of processing. This evaluative procedure is known as the Cognitive Reflection Test (CRT). The CRT consists of three apparently simple arithmetical problems that provoke an immediate and wrong response. The individual has to suppress this answer in favor of an alternative one, which is reflective, deliberative, and correct [20,26].

A relevant characteristic (and limitation) of Frederick’s CRT is that the distributions of scores show substantial degrees of kurtosis and skewness, and that the scores are not normally distributed (see, for instance, [25] Table 1, p. 29). These characteristics render problematic the use of typical bivariate and multivariate statistical analyses that assume bivariate and multivariate normality (e.g., correlation, ANOVA, multiple regression, structural equation modeling) and the estimation of some psychometric proprieties (e.g., internal consistency). Due to the non-normal distribution of the scores, some of the effect-size estimates obtained in the studies on the relationship between the CRT and other variables (e.g., intelligence) might not be robust findings. Currently, several additional tests can assess cognitive reflection, and they were developed for overcoming the psychometric limitations of the CRT. For instance, the CRT-10 [27] evaluates CR using 10 items for avoiding the large kurtosis and skewness found with the Frederick’s [25] CRT test. Other CR tests have been developed in the last few years, for instance, by Baron, Scott, Fincher, and Metz [28], Böckenholt [29], Borghans and Golsteyn [30], Finucane and Gullion [31], Grossman, van der Weele, and Andrijevich [32], Kinnunen and Windmann [33], Mata and Almeida [34], Mata, Fiedler, Ferreira, and Almeida [35], Mendonça [36], Primi, Morsanyi, Chiesi, Donati, and Hamilton [37], Reuben, Sapienza, and

Zingales [38], Shtulman and McCallum [39], Thomson and Oppenheimer [40], and Toplak et al. [24]. Many of these tests consisted of the three original items of Frederick [25] plus one, two, three, or four new items.

Despite that, the CRT has been used in numerous studies, which showed that high scores in this test are associated with very different aspects of everyday life. For instance, individuals scoring high in the CRT obtained better results in decision-making tasks, showed less risk aversion and greater patience of recompense return [25,41,42], they are capable of better interpreting the humor scenes [43], showed less tendency to religious beliefs and beliefs in paranormal issues [44–46], they achieved better strategic behaviors [47], and, in general, they used fewer heuristics and committed less cognitive biases in reasoning [13,48–50]. These results seem to indicate that higher scores in CRT could be associated with higher performance and productivity in several areas of our life (e.g., in the personnel field, academic field, and occupational field). This trait would imply the efficient management of cognitive resources and social resources to maximizing the outcomes. Therefore, the organizations could be interested in hiring people with this trait to help it function more sustainably.

3. Cognitive Reflection, Job Performance, Academic Performance, and General Mental Ability

Although the predictive capacity of Frederick's CRT has been examined in very different situations, and with various criteria, there are five issues in which research is scarce or non-existent. The first one refers to the prediction of academic outcomes. There are some studies on the relationship between the CRT and the academic outcomes (e.g., grade point average; GPA), but the findings are non-conclusive. For instance, Insler, Compton, and Schmitt [51] found a correlation coefficient of 0.31 ($n = 364$) between the CRT and GPA, and Toplak et al. [24] found a correlation coefficient of 0.23 ($n = 160$) between the CRT and the self-reported GPA. However, Corgnet, Hernán Gonzalez, and Mateo [52] did not find a relationship between the CRT and GPA ($r = 0.04$; $n = 264$). The second issue refers to the relationship between CR and job performance. To the best of our knowledge, this relationship has not been examined until now. On the one hand, academic performance (achievement) is the most relevant measure of university success and, at the same time, is a consistent predictor of job performance [53]. On the other hand, job performance is the primary individual dependent variable in work and organizational studies. Therefore, the first two goals of this research are to examine the predictive capacity of the CRT to predict academic performance and job performance, as these two criteria are particularly relevant from an applied point of view.

A third issue that deserves more research is the relationship between the CRT and general mental ability (GMA), and some specific cognitive abilities (e.g., numerical reasoning). Several studies have found that there is a moderate correlation between CRT and GMA (e.g., [25,42,54–56]). The magnitude of the correlation ranged from 0.17 to 0.64. Also, it has been found that the CRT is strongly correlated with the numerical ability (e.g., [31,42,54,55,57]) and that this ability has a role in the right solution of the arithmetical problems presented in the CRT items [58]. However, these studies have two limitations. The first one is related to the fact that the CRT score is not normally distributed and presents high skew and kurtosis. Consequently, the Pearson correlations can be severely affected by these biases. The second limitation is that the true correlation between the CRT and GMA was not estimated. It is well known that the observed correlations underestimate the true correlation due to the joint effect of two artifacts: the measurement error and the potential range restriction in both variables [59–61]. In other words, all of the observed coefficients reported until now have systematically underestimated the true correlation between CRT and GMA. Therefore, it seems necessary to examine the true relationship between the CRT and GMA after controlling for these artifactual errors. This would be the third research goal.

The fourth unexplored issue is about the capacity of GMA and CRT for predicting academic performance and job performance together has not been studied. Multiple primary studies and several meta-analyses have examined the validity of GMA for predicting job performance (e.g., [62–65]). The main conclusion is that GMA is the best predictor of job performance and training proficiency, and that the validity generalizes across organizations, jobs, and samples. The validity of GMA to predict academic performance has also been studied through meta-analysis, and the

findings have demonstrated the existence of a positive and consistent relationship between GMA and academic performance (e.g., [66–72]). A recent meta-analysis conducted by Salgado and Moscoso [73] found an average true correlation of the GMA of 0.44 and 0.62 to predict job performance and academic performance, respectively. We aim, as the fourth objective, to examine the multiple correlation of GMA and CRT to predict these two criteria and test whether the CRT shows incremental validity over and beyond GMA for predicting these criteria.

A fifth unexamined issue that must be mentioned is related to the fact that all the studies carried out with the CRT reported the observed correlation coefficients only. It is well-known that the observed correlation underestimates the true relationship between the two correlated variables. In other words, the observed correlation indicates how well one fallible measure predicts another fallible measure [59]. The fallibility of the measures is mainly due to three statistical artifacts: the predictor reliability, the criterion reliability, and the range restriction in the predictor [59–61,74,75]. However, if one wants to know how the true variances of the two variables are related (i.e., the true-score correlation between the two variables), it must be taken into account the reliability of the measures and the potential effects of range restriction. To the best of our knowledge, no published study corrected the observed correlations for measurement errors and range restriction. Therefore, all reported correlations of the relationship between CR and a second variable are an underestimation of the true magnitude of the relationship. This is important because the researchers seem to be interested in the true correlation rather than relations between fallible measures (e.g., [21,23,24,56]). Therefore, for modeling the relationships at the construct-level among CR, GMA, and performance, the true correlations (i.e., the observed correlation corrected for measurement errors and range restriction) should be used, and this will be the fifth objective of this research.

In summary, this research has five main objectives. The first two goals are to test the validity of CR to predict job performance and to predict academic performance by using a CR measure that overcomes the limitations of Frederick's CRT mentioned above. The third goal is to estimate the true-correlation between CR and GMA. The fourth goal is to determine the joint true-correlation capacity of GMA and CR for predicting job performance and academic performance, and the fifth goal is to know whether CR shows incremental validity over GMA validity for the prediction of these two criteria.

Based on the literature review and the findings mentioned above, we posit the following three hypotheses:

Hypothesis 1. *Cognitive reflection predicts job performance and academic performance.*

Hypothesis 2. *Cognitive reflection correlated moderately with GMA.*

Hypothesis 3. *Cognitive reflection shows incremental validity over GMA for the prediction of job performance and academic performance.*

4. Method

4.1. Samples

To test the hypotheses posited and fulfill the objectives of this research, we collected four independent samples. We describe the characteristics of the samples below.

Sample 1. 100 students from the University of Santiago de Compostela participated in the study. The average age was 22.31 years old ($SD = 2.93$), and 47% were women. All students were offered the chance to participate in the study as a voluntary activity for the development of their skills.

Sample 2. 318 students from the University of Santiago de Compostela belonging to different degrees took part. The average age was 21.50 years old ($SD = 3.42$), and 69.5% were women. All students participated voluntarily in the study.

Sample 3. The sample consisted of 130 students from the University of Santiago de Compostela, belonging to different degrees. The average age of the sample was 21.38 ($SD = 4.25$), and 69.23% were women.

Sample 4. The sample 4 consisted of 157 students from the Faculty of Business of the University of Santiago de Compostela. 68.2% were second-year students, and the rest were final-year students. The average age was 20.84 years old ($SD = 1.69$), and 49% were women.

4.2. Measures

Cognitive Reflection. In all samples, cognitive reflection was measured with a compound created with the 10 items of the Cognitive Reflection Test developed by Salgado [27] and the 3 items originally designed by Frederick [25]. This compound measure of CR will be called CRT-13 in this research. Each item consists of a small arithmetical problem that is solved by simple calculations. Participants had to answer each item by choosing, from among two answer choices, the answer they consider to be correct. One of the answer options would represent the impulsive and erroneous answer offered by our system 1; the other would represent the reflexive and correct answer offered by S2. The participants did not have a time limit to complete the test. The scores could range from 0 (no item answered correctly) to 13 (all items answered correctly). Therefore, high scores in this measure indicate greater cognitive reflection. The reliability of the CRT-13 ranged from 0.76 to 0.82 (average $r_{xx} = 0.79$; $n = 705$). Table 1 shows the CRT-13 reliability coefficient for each of the samples.

General Mental Ability. In samples 1, 2, and 3, GMA was measured with a Spanish version of the Wonderlic Personnel Test (WPT; [76]). The WPT consists of 50 items that should be answered in 12 min. The WPT score ranges from 0 (no items completed correctly) to 50 (all items completed correctly). High scores in the WPT indicate greater GMA. Table 1 shows the reliability coefficient for each of the three samples. The reliability ranged from 0.77 to 0.81 (average $r_{xx} = 0.80$; $n = 548$).

In Sample 4, Cattell's "g" Factor Test was used as a measure of GMA [77]. In this sample, the scale 3, form A, composed of 4 sub-tests, was applied. The participants had 12^{1/2} min to complete the test. The scores range from 0 (no items answered correctly) to 50 (all items answered correctly). The reliability was 0.67 ($n = 157$).

Job Performance. In Sample 1, the overall rating of an assessment center (AC) was used as a job performance measure. The AC consisted of four exercises: (1) an in-basket test; (2) a speech about some of the proposed topics and their subsequent exposure to the evaluators, and (3) two group discussions. Four judges evaluated the participants in six competencies considered relevant to perform a job. The interrater reliability was 0.84 for the overall score. The assessed competencies were: (a) work organization and planning, (b) argumentation of their proposals and interventions, (c) thoroughness in carrying out the tasks and the presentation, (d) oral communication, (e) contribution to teamwork, and (f) initiative in the group discussions. To carry out the evaluations, the judges used behavioral-anchored rating scales (BARS) from 1 to 5, for each of the competencies, where 1 indicated poor performance and 5 showed excellent performance. The overall rating of the AC performance consisted of the sum of the scores of the six competencies.

Academic Performance. The academic performance served as the criterion variable in samples 2, 3, and 4. Three different measures were used in these samples, as described below.

- (A) Grade Point Average (GPA). In sample 2, the average of the academic grades of each participant was used as a measure of academic performance. The grades were taken from the official transcript with the permission of the participants. High scores indicated better academic performance. Research has shown that the GPA has acceptable internal reliability and temporal stability [78,79]. In this sense, Salgado and Tauriz [79] developed an empirical distribution of GPA reliability and found an average reliability coefficient of 0.83. This coefficient is widely accepted and used and was the reliability coefficient used in sample 2 for this criterion measure.
- (B) University Entry Score (UES). The performance measure used in sample 3 was the university entry score. The Spanish university system requires the student to pass an official exam to be accepted into the university. The score on this exam determines the type of university degrees (e.g., Law, Economics, Engineering, and so on) that the individual can study. Each participant voluntarily provided a copy of the official document containing the UES. The UES ranges from

5 to 10 (scores below 5 means fail, and the student is not granted for the university degrees). The assumed reliability was 0.83.

- (C) Exam Mark. In sample 4, the criterion used was the mark achieved by the participants in the final exam of one of the subjects (e.g., marketing). The scores range from 0 to 10, where the higher scores indicated better performance, and 5 is the cutoff point of pass-fail. The raw scores were transformed in z-scores to control the differences in the subjects. The assumed reliability was 0.83.

4.3. Procedure

The testing of sample 1 was conducted in two sessions. The first session lasted 4 h and 30 min, where participants covered a GMA test and participated in an assessment center (AC). The second session took place a few weeks later, and the participants answered the CRT-13. In return for their participation, the subjects received an economic bonus of 10 euros.

The testing of samples 2, 3, and 4 was conducted in a single session. During the testing session, the participants answered a GMA test and the CRT measure. At the end of the session, the participants of sample 2 voluntarily provided a copy of the official transcript, and the participants of sample 3 provided a copy of their university entry score. In the case of sample 4, the grade on a marketing exam served as the criterion of academic performance. Each participant received 10 euros as compensation for their collaboration in the study.

4.4. Statistical Analyses

To correct for measurement error in the dependent and independent variables, we have used the formula to correct the observed correlation for attenuation (e.g., [60,61]). This formula requires knowing the reliability coefficient of the CRT-13 and the GMA test, and the reliability coefficient of job performance and academic performance.

Regarding range restriction (RR), we have used the formula to correct for direct RR developed by Schmidt and Hunter [61]. The formula of RR correction requires knowing the degree of homogeneity presented in the sample, indicated by u value. Table 1 reported the u -values for CRT and GMA in these samples.

Consequently, the artifacts considered here were the predictor reliability, the reliability of the performance measures, and the direct range restriction in the predictor variables. The three artifacts reduce the real size of the correlations, and their effects are cumulative [61]. True score correlation represents the correlation between CR and GMA in the absence of artifactual errors. Therefore, true score correlation is used for modeling the relationships between CR, GMA, and criteria at the construct-level [61,80].

Every sample was individually corrected for the three artifacts, to obtain the cumulative value of the true correlations between variables. According to Schmidt and Hunter [61], three phases were carried out to compute the meta-analysis. Firstly, every sample was corrected individually for the three artifacts. Then, the results of the samples were combined. Finally, the mean and variance of true effect size correlations were estimated.

To test the hypotheses, we carried out a series of multiple regression (MR) analyses. Methodologists advise researchers that measurement error and range restriction (when appropriate) should be controlled (corrected) when MR, structural equation modeling, and other multivariate techniques are used to analyze the relationships among dependent (criterion) and independent (predictors) variables (e.g., [60,61,81]). The basis for the recommendation is that the imperfect measurement of X and Y , and range restriction produce a violation of some fundamental assumptions of MR and, consequently, the parameters estimated are biased, for instance, reducing the squared multiple correlation and the size of the standardized regression weights (Betas). In other words, the researcher must be sure that the variables are perfectly reliable, and, therefore, the observed correlations must be corrected for measurement error in both variables X and Y [61,81].

Another critical issue in MR analysis is that the multiple correlation (R), the square multiple correlation (R^2), and the adjusted-square multiple correlation (adjusted R^2) are biased because MR

analysis capitalizes on chance. A more efficient estimate of the effect size is the squared population cross-validity coefficient (R^2_{cv}), which can be obtained by formulae. Monte Carlo examinations of the effectiveness of various R^2_{cv} formulas agreed that the one proposed by Browne [82] outperformed other formulae [83–85]. Therefore, we have used Browne's formula for estimating the squared population cross-validity.

5. Results

We conducted three sets of statistical analyses. The first one is the analysis of the correlation between GMA and CR with the performance measures. The second set is a series of quantitative syntheses (i.e., psychometric meta-analyses) conducted to establish the best estimate of the relationship between GMA and CR, GMA and the academic performance measures, and CR and the academic performance measures. The third set was two MR analyses conducted to estimate the joint capacity of GMA and CR for predicting job and academic performance, and to determine the degree of incremental validity of CR over GMA for predicting these two dependent variables.

Table 1 reports the results of the correlation analysis for the four samples. From left to right, the first three columns indicate the mean, standard deviation, and range restriction values of the variables. The next three columns report the correlation between the variables. The observed correlations appear below the diagonal, and the true correlations appear above the diagonal. The reliability of the measures appears between parentheses in the diagonal.

Table 1. Descriptive statistics, range restriction value, reliability, and correlations among the variables in the four samples.

Variable	\bar{X}	SD	u	CR	GMA	Performance
Sample 1 ($n = 100$)						
CR	8.68	3.54	1.03	(0.84)	0.65 **	0.37 **
GMA	25.69	6.01	1.05	0.55 **	(0.81)	0.44 **
JP	20.37	3.05	-	0.32 **	0.36 **	(0.84)
Sample 2 ($n = 318$)						
CR	8.33	3.41	0.99	(0.78)	0.46 **	0.24 **
GMA	22.18	5.00	0.88	0.33 **	(0.81)	0.28 **
GPA	6.93	0.86	-	0.19 **	0.20 **	(0.83) †
Sample 3 ($n = 130$)						
CR	7.46	3.17	0.92	(0.77)	0.68 **	0.54 **
GMA	23.23	5.17	0.91	0.50 **	(0.77)	0.68 **
UES	7.20	1.22	-	0.41 **	0.50 **	(0.83) †
Sample 4 ($n = 157$)						
CR	7.67	3.45	1.00	(0.81)	0.50 **	0.18 **
GMA	22.46	4.26	1.00	0.37 **	(0.67)	0.26 **
EM	0.00	1.00	-	0.15 *	0.19 **	(0.83) †

Note. The observed correlations appear below the diagonal and the true correlations appear above the diagonal. CR = cognitive reflection; GMA = general mental ability; JP = job performance; GPA = grade point average; UES = university ENTRY score; EM = exam mark; u = range restriction value. The coefficients reported between parentheses in the diagonal of the matrix are the reliability coefficients. † Taken from Salgado and Tauriz [79] and Salgado and Moscoso [73]. * $p < 0.05$; ** $p < 0.01$.

The results showed that CR and GMA were valid predictors of job and academic performance in the four samples. In the case of job performance, the observed correlation between CR and job performance was 0.32 ($p < 0.001$), which shows that CR is a relevant predictor of this performance measure. The observed validity of CR was 0.19 ($p < 0.001$) for predicting GPA, it was 0.41 ($p < 0.001$) for predicting the university entrance score, and it was 0.15 ($p < 0.10$) for predicting the exam mark. Therefore, the results supported Hypothesis 1.

Concerning the results for GMA, as can be seen, GMA predicted job performance very efficiently. The observed validity was 0.36 ($p < 0.001$), which is a similar value to the one found in the meta-analyses of the relationship between GMA and job performance for occupations of high level of job complexity (e.g., [63,73,86]. Regarding the educational performance criteria, the correlations of GMA were 0.20 ($p < 0.001$), 0.50 ($p < 0.001$), and 0.19 ($p < 0.02$) with GPA, UES, and EG, respectively.

The observed correlations between GMA and CR ranged from 0.33 to 0.55, and the weighted-sample average correlation was 0.40 ($p < 0.001$), which is a substantial correlation. When this correlation was corrected for predictor and criterion reliability, and range restriction in GMA, the true correlations ranged from 0.46 to 0.68, and the weighted-sample average true correlation was 0.56. Therefore, this empirical evidence supported Hypothesis 2.

As we have obtained the correlation between CR and GMA in four independent samples and the correlations of CR and GMA with education outcomes in three independent samples, it is possible to conduct three quantitative syntheses (i.e., meta-analyses), with correction for artifacts (i.e., sample size, criterion reliability, predictor reliability, and range restriction). The first quantitative synthesis is about the relationship between CR and GMA, the second one is about the relationships between GMA and the educational performance, and the third one is about the relationship between CR and the educational performance. To conduct these quantitative syntheses, we used the meta-analysis method developed by Schmidt and Hunter [61], implemented in the software program developed by Schmidt and Le [87]. The software program corrects the observed correlations for the reliability of predictor, criterion, and range restriction, and it corrects the observed variance for the following four artifactual errors, sample size, predictor reliability, criterion reliability, and range restriction.

Table 2 reports the results of these quantitative syntheses. From left to right, the first two columns reflect the number of independent samples that have been integrated into each meta-analysis and the total sample size. The next four columns show the observed validity, the standard deviation of the observed validity, the true correlation (ρ), and the standard deviation of the true correlation. The last three columns refer to the percentage of variance explained by artifactual errors, 90% credibility value, and the 95% confidence interval of ρ .

Table 2. Quantitative syntheses of the results of samples 1, 2, 3, and 4.

Relationship	<i>N</i>	<i>K</i>	r_{xy}	SD_r	ρ	SD_ρ	%VE	90%CV	95%CI $_\rho$
GMA-CR	705	4	0.40	0.056	0.56	0.000	100	0.56	0.48/0.64
GMA-AP	605	3	0.26	0.123	0.36	0.139	28	0.18	0.17/0.54
CR-AP	605	3	0.23	0.096	0.28	0.100	43	0.15	0.13/0.43

Note. GMA = general mental ability; CR = cognitive reflection; AP = academic performance; *N* = sample size; *K* = number of independent samples; r_{xy} = weighted-sample average observed correlation; SD_r = standard deviation of r_{xy} ; ρ = true correlation; SD_ρ = standard deviation of true correlation; %VE = percentage of variance in corrected correlations attributable to all artifacts; CV = credibility value; CI $_\rho$ = confidence interval of true correlation.

Regarding the relationship between GMA and CR, the observed validity was 0.40, and, once corrected for lack of reliability in the measure of the predictor and the criterion and for indirect restriction in the range of the predictor, the true correlation was 0.56. The percentage of variance explained by artifactual errors was 100%. The 90%CV was 0.56, indicating that the results are generalizable. The lower and upper limits of the 95% confidence interval for ρ were 0.48 and 0.64, which showed that the true correlation is highly significant.

The second meta-analysis shows the validity of GMA to predict performance. In this case, the observed validity weighted by sample size was 0.26, and the true correlation was 0.36. The percentage of variance explained by artifactual errors was 28%. The lower and upper limits of the 95% confidence intervals were both positive and, therefore, showed that the true correlation was statistically significant. The 90%CV was also positive and different from zero, indicating that GMA generalized its validity for predicting educational performance measures.

The third meta-analysis was about the validity of CR to predict performance. The observed validity was 0.23, and the true correlation was 0.28. The variance accounted for by the artifactual errors was 43%. The 90%CV was positive and different from zero, showing evidence of validity generalization. The lower and upper limits of the 95%CI showed that the true correlation was statistically significant.

Finally, in order to establish the joint capacity of GMA and CR to predict job and academic performance, and the incremental validity of CR over the GMA validity, we conducted two analyses of multiple regression (MR). The first MR analysis was performed using the correlations reported in Table 1. The second MR analysis was done with the true correlations found in the three meta-analyses reported in Table 2. Tables 3 and 4 report the results of these two MR analyses. The two tables indicate the multiple correlation (R), the explained variance (R^2), adjusted R^2 , statistical estimate of the cross-validated square multiple correlation (R^2_{cv}), betas for GMA and CR, and the incremental validity of CR.

As can be seen in Table 3, the joint effect of GMA and CR on job performance was 0.465, and the magnitude of the incremental validity of CR over GMA was 0.025. Therefore, R was larger than the correlation of GMA and CR with job performance considered individually. The examination of the Beta weights shows that they were positive and marginally significant ($p < 0.10$), which means that both variables contributed to the explanation of the job performance variance.

Table 3. Multiple regression analysis of the joint relationship of GMA and CR with job performance.

Variable	R	R^2	R^2_{adj}	R^2_{cv}	R_{cv}	F	p
GMA + CR	0.465	0.216	0.200	0.190	0.437	13.35	0.000
Δ CR over GMA	0.025						
Beta _{GMA}	0.340 ($p = 0.002$)						
Beta _{CR}	0.180 ($p = 0.10$)						

Note. GMA = general mental ability; CR = cognitive reflection; R^2_{adj} = adjusted squared multiple correlation; R^2_{cv} = statistical estimated of the population squared cross-validated multiple regression; R_{cv} = statistical estimated of the population cross-validated multiple regression.

The joint predictive effect of GMA and CR on academic performance appears in Table 4. The results are very similar to the one found for job performance. The multiple correlation was 0.372, and the incremental validity of CR over GMA was 0.012. Once again, R was larger than the correlation of GMA and CR with academic performance, respectively. Also, the Beta weights were positive and statistically significant, and, consequently, both variables contributed to the explanation of the academic performance variance. Therefore, as a whole, the results reported in Tables 3 and 4, concerning the incremental validity of CR over GMA, give support to the Hypotheses 3.

Table 4. Multiple regression analysis of the joint relationship of GMA and CR with academic performance.

Variable	R	R^2	R^2_{adj}	R^2_{cv}	R_{cv}	F	p
GMA + CR	0.372	0.139	0.136	0.132	0.363	48.41	0.000
Δ CR over GMA	0.012						
Beta _{GMA}	0.296 ($p = 0.000$)						
Beta _{CR}	0.114 ($p = 0.013$)						

Note. GMA = general mental ability; CR = cognitive reflection; R^2_{adj} = adjusted squared multiple correlation; R^2_{cv} = statistical estimated of the population squared cross-validated multiple regression; R_{cv} = statistical estimated of the population cross-validated multiple regression.

6. Discussion

Their selection processes are one of the factors that help organizations to become sustainable because they allow organizations to improve performance and productivity, hiring highly innovative and high-performance employees. Thousands of studies have been published for

estimating the ability of several individual differences to predict performance. However, the validity of cognitive reflection to predict performance has not been investigated. The purpose of this research was to extend the current knowledge on CR by examining: (1) the relationship of CR with job performance and academic outcomes, (2) the relationship of CR with GMA, and (3) whether CR added validity over GMA validity for predicting these two critical criteria.

This research has made several unique contributions to the literature on CR. The first unique contribution has been to show that CR is a relevant predictor of job performance and educational outcomes, and that the magnitude of the validity is similar or even more substantial than the validity of other well-studied predictors (e.g., cognitive abilities, personality dimensions; [88–90]). Moreover, in the case of academic performance, the findings showed that CR was a valid predictor for the three types of educational outcomes (i.e., GPA, UEM, and EG). According to our findings, the best estimate of the relationship between CR and job performance is 0.37, and the best estimate of the CR-academic performance relationship is 0.28. These findings have implications for both personnel selection processes and student university admission processes. Based on these findings, we suggest that CR can be used as a tool for making decisions in these two applied domains. Hiring people with high CRT scores, organizations could improve their performance and productivity, and hence, encourage organizational sustainability and sustainable employability. In addition, hiring individuals with high CRT scores might contribute to more harmonious psychological processes [91].

The second contribution has been to demonstrate that CR shared a substantial percentage of variance with GMA. The true correlation between the two constructs was 0.56, which means that they shared 31% of the variance. At the same time, this figure also suggests that CR and GMA are relatively independent constructs and, therefore, the contribution to the prediction of performance can be cumulated. This was, in fact, the third contribution of this research.

Multiple regression analyses pointed out that the two variables can be used together for a better prediction of job performance and academic performance. When the CR is included in a regression equation with GMA, the predictive validity is larger for both types of performance, which means that CR added validity over and beyond GMA. Nevertheless, although the incremental validity of CR is relevant, it must be pointed out that the magnitude of the incremental validity is small for both criteria (i.e., 0.025 in the case of job performance and 0.012 for academic performance). Although, indeed, the increase in predictive power of cognitive reflection over GMA has been small, it is necessary to point out that this result has important theoretical implications: (1) the result allows us to explain a greater amount of variance in performance. Part of the explained variance is not shared by the two measures, implying that cognitive reflection also contributes to explaining performance variance; and (2) the result might suggest that general mental capacity and cognitive reflection do not measure exactly the same cognitive constructs. In this sense, Stanovich [21] stated that intelligence tests are incomplete measures of cognitive functioning and fail to measure particular cognitive abilities associated with reflection. For their part, Pennycook and Ross [92] state that CRT, in addition to cognitive ability, indexes a certain degree of willingness or propensity to think analytically, so that “if someone is not willing to think analytically, they will not fully exercise their cognitive ability and will not correctly solve the problem. Naturally, the opposite is also true: if someone does not have enough cognitive capacity, it will not matter how much time and effort he is willing to spend thinking about the problem” [92]. This may explain why cognitive reflection and general mental capacity are highly related to each other and, simultaneously, cognitive reflection adds validation.

Two additional contributions of this research, the fourth and fifth contributions, although not unique, should be mentioned. The fourth contribution was that GMA, once again, showed to be the most relevant cognitive predictor of job performance and academic performance. This finding agrees with the results of previous meta-analyses of the relationship between GMA and performance (e.g., [73,86,93]). The fifth contribution was that GMA was shown to predict academic performance independently of the measure used to evaluate the educational outcomes, although the magnitude

of the correlation was not the same across the three measures of academic performance. Future studies should examine if the outcome measure is a moderator of the validity of GMA.

Limitations of the Study and Future Research

Like many studies, the current one has some limitations, and three should be specially noticed. First, the reliability of academic performance measures used in this study could not be estimated. To correct the correlations for measurement error, we use an empirical distribution of GPA reliability developed by Salgado and Tauriz [79]. Second, the results of meta-analyses should be taken with caution since they only integrate 4 samples, and the total sample size is 705 subjects. Third, the CR validity for predicting job performance was examined with a specific measure of occupational performance (i.e., assessment center), which is mainly related to task performance. Consequently, the current findings should not be generalized to other performance dimensions, such as citizenship performance, innovative performance, and counterproductive performance. Future studies should be conducted to determine whether CR is also a valid predictor for these performance dimensions.

7. Conclusions

In summary, this research has demonstrated that cognitive reflection is a valid predictor of job performance and academic performance, showing a degree of predictive validity similar to that of GMA and other frequently used cognitive and personality assessment procedures. Furthermore, despite the moderately large magnitude of the correlation between the two predictors, CR added validity to GMA to predict performance. These findings provide new information to achieve sustainable selection practices, with better person-job fit and person-organization fit and, as a result, facilitate more sustainable organizations.

Author Contributions: All authors listed have made a relevant and substantial contribution to this research. Formal analysis, J.F.S., I.O. and S.M.; Funding acquisition, J.F.S.; Investigation, J.F.S., I.O. and S.M.; Writing—original draft, J.F.S., I.O. and S.M.

Funding: This research was partially supported by Grant PSI2017-87603-P from the Spanish Ministry of Economics and Competitiveness.

Acknowledgments: The authors sincerely thank Pamela Alonso, Alexandra Martínez and Dámaris Cuadrado for their help in collecting data and reviewing the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. De Jonge, J.; Peeters, M.C.W. The vital worker: Towards sustainable performance at work. *Int. J. Environ. Res. Public Health* **2019**, *16*, 910. doi:10.3390/ijerph16060910.
2. Di Fabio, A. The psychology of sustainability and sustainable development for well-being in organizations. *Front. Psychol.* **2017**, *8*, 1534. doi: 10.3389/fpsyg.2017.01534.
3. Ones, D.S.; Wiernik, B.M.; Dilchert, S.; Klein, R.M. Multiple Domains and Categories of Employee Green Behaviors: More than Conservation. In *Research Handbook on Employee Pro-Environmental Behavior*; Wells, V., Gregory-Smith, D., Manika, D., Eds.; Edward Elgar Publishing: Cheltenham, UK; Northampton, MA, USA, 2018; pp. 13–38.
4. Wiernik, B.M.; Ones, D.S.; Dilchert, S.; Klein, R.M. Individual Antecedents of Pro-Environmental Behaviours: Implications for Employee Green Behaviours. In *Research Handbook on Employee Pro-Environmental Behavior*; Wells, V., Gregory-Smith, D., Manika, D., Eds.; Edward Elgar Publishing: Cheltenham, UK; Northampton, MA, USA, 2018; pp. 63–82.
5. Pfeffer, J. Building sustainable organizations: The human factor. *Acad. Manag. Perspect.* **2010**, *24*, 34–45. doi:10.5465/amp.24.1.34.
6. Di Fabio, A. Positive healthy organizations: Promoting well-being, meaningfulness, and sustainability in organizations. *Front. Psychol.* **2017**, *8*, 1938. doi:10.3389/fpsyg.2017.01938.

7. van der Klink, J.J.; Bültmann, U.; Burdorf, A.; Schaufeli, W.B.; Zijlstra, F.R.; Abma, F.I.; Brouwer, S.; van der Wilt, G.J. Sustainable employability—Definition, conceptualization, and implications: A perspective based on the capability approach. *Scand. J. Work Environ. Health* **2016**, *42*, 71–79. doi:10.5271/sjweh.3531.
8. Fleuren, B.P.; de Grip, A.; Jansen, N.W.; Kant, I.; Zijlstra, F.R. Critical reflections on the currently leading definition of sustainable employability. *Scand. J. Work Environ. Health* **2016**, *42*, 557–560. doi:10.5271/sjweh.3585.
9. Di Fabio, A.; Rosen, M.A. Opening the black box of psychological processes in the Science of Sustainable Development: A new frontier. *Eur. J. Sustain. Dev.* **2018**, *2*, 47. doi:10.20897/ejosdr/3933.
10. Salgado, J.F.; Moscoso, S.; García-Izquierdo, A.L.; Anderson, N.R. Inclusive and Discrimination-Free Personnel Selection. In *Shaping Inclusive Workplaces through Social Dialogue*; Arenas, A., Di Marco, D., Munduate, L., Euwema, M.C., Eds.; Springer: New York, NY, USA, 2017; pp. 103–119.
11. Chaiken, S.; Liberman, A.; Eagly, A.H. Heuristic and Systematic Information Processing within and beyond the Persuasion Context. In *Unintended Thought*; Uleman, J.S., Bargh, J.A., Eds.; Guilford: New York, NY, USA, 1989; pp. 212–252.
12. Evans, J.S.B.; Wason, P.C. Rationalization in a reasoning task. *Br. J. Psychol.* **1976**, *67*, 479–486. doi:10.1111/j.2044-8295.1976.tb01536.x.
13. Kahneman, D. *Thinking, Fast and Slow*; Farrar, Strauss, and Giroux: New York, NY, USA, 2011.
14. Kahneman, D.; Frederick, S. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In *Heuristics of Intuitive Judgment: Extensions and Applications*; Gilovich, T., Griffin, D., Kahneman, D., Eds.; Cambridge University Press: New York, NY, USA, 2002; pp. 49–81.
15. Kahneman, D.; Frederick, S. A Model of Heuristic Judgment. In *The Cambridge Handbook of Thinking and Reasoning*; Holyoak, K.J., Morrison, R.G., Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 267–293.
16. Metcalfe, J.; Mischel, W. A hit/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychol. Rev.* **1999**, *106*, 3–19.
17. Stanovich, K.E.; West, R.F. Variation in how we cope with uncertainty. *Am. J. Psychol.* **2002**, *115*, 127–132. doi:10.2307/1423679.
18. Strack, F.; Deutsch, R. Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* **2004**, *8*, 220–247. doi:10.1207/s15327957pspr0803_1.
19. Wason, P.C.; Evans, J.S.B. Dual processes in reasoning? *Cognition* **1975**, *3*, 141–154. doi:10.1016/0010-0277(74)90017-1.
20. Mastrogiorgio, A.; Petracca, E. Numerals as triggers of System 1 and System 2 in the ‘bat and ball’ problem. *Mind Soc.* **2014**, *13*, 135–148. doi:10.1007/s11299-014-0138-8.
21. Stanovich, K.E. *What Intelligence Tests Miss: The Psychology of Rational Thought*; Yale University Press: New Haven, CT, USA, 2009.
22. Stanovich, K.E. *Who Is Rational? Studies of Individual Differences in Reasoning*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1999.
23. Stanovich, K.E. *Rationality and the Reflective Mind*; Oxford University Press: New York, NY, USA, 2011.
24. Toplak, M.E.; West, R.F.; Stanovich, K.E. Assessing miserly information processing: An expansion of the cognitive reflection test. *Think. Reason.* **2014**, *20*, 147–168. doi:10.1080/13546783.2013.844729.
25. Frederick, S. Cognitive reflection and decision making. *J. Econ. Perspect.* **2005**, *19*, 25–42. doi:10.1257/089533005775196732.
26. Albaity, M.; Rahman, M.; Shahidul, I. Cognitive reflection test and behavioral biases in Malaysia. *Judgm. Decis. Mak.* **2014**, *9*, 149–151.
27. Salgado, J.F. *Cognitive Reflection Test (CRT-10)*; Technical Report; Unpublished work.
28. Baron, J.; Scott, S.; Fincher, K.; Metz, S.E. Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *J. Appl. Res. Mem. Cogn.* **2015**, *4*, 265–284. doi:10.1016/j.jarmac.2014.09.003.
29. Böckenholt, U. The cognitive-miser response model: Testing for intuitive and deliberate reasoning. *Psychometrika* **2012**, *77*, 388–399. doi:10.1007/s11336-012-9251-y.
30. Borghans, L.; Golsteyn, B.H. Default options and training participation. *Empir. Econ.* **2014**, *46*, 1417–1428. doi:10.1007/s00181-013-0717-2.
31. Finucane, M.L.; Gullion, C.M. Developing a tool for measuring the decision-making competence of older adults. *Psychol. Aging* **2010**, *25*, 271–288. doi:10.1037/a0019106.

32. Grossman, Z.; van der Weele, J.; Andrijevik, A. A Test of Dual-Process Reasoning in Charitable Giving; Working Paper; 2014. Available online: <https://escholarship.org/uc/item/4tm617f7> (accessed on 10 December 2018).
33. Kinnunen, S.P.; Windmann, S. Dual-processing altruism. *Front. Psychol.* **2013**, *4*, 193. doi:10.3389/fpsyg.2013.00193.
34. Mata, A.; Almeida, T. Using metacognitive cues to infer others' thinking. *Judgm. Decis. Mak.* **2014**, *9*, 349–359.
35. Mata, A.; Fiedler, K.; Ferreira, M.B.; Almeida, T. Reasoning about others' reasoning. *J. Exp. Soc. Psychol.* **2013**, *49*, 486–491. doi:10.1016/j.jesp.2013.01.010.
36. Mendonça, C.S.P. Cognitive Reflection Test Eteorias Dualistas no Raciocínio e nas Atitudes. Master's Thesis, Psychology Faculty, Porto, Portugal, 2012.
37. Primi, C.; Morsanyi, K.; Chiesi, F.; Donati, M.A.; Hamilton, J. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *J. Behav. Decis. Mak.* **2015**, *29*, 453–469. doi:10.1002/bdm.1883.
38. Reuben, E.; Sapienza, P.; Zingales, L. Procrastination and impatience. *J. Behav. Exp. Econ.* **2015**, *58*, 63–76. doi:10.1016/j.socec.2015.07.005.
39. Shtulman, A.; McCallum, K. Cognitive Reflection Predicts Science Understanding. In Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec City, QC, Canada, 23–26 July 2014; Bello, P., Guarini, M., McShane, M., Scassellati, B., Eds.; Cognitive Science Society: Austin, TX, USA, 2014; pp. 2937–2942.
40. Thomson, K.S.; Oppenheimer, D.M. Investigating an alternate form of the cognitive reflection test. *Judgm. Decis. Mak.* **2016**, *11*, 99–113.
41. Campitelli, G.; Labollita, M. Correlations of cognitive reflection with judgments and choices. *Judgm. Decis. Mak.* **2010**, *5*, 182–191.
42. Cokely, E.T.; Kelley, C.M. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgm. Decis. Mak.* **2009**, *4*, 20–33.
43. Ventis, L. Thinking fast and slow in the experience of humor. *Int. J. Humor Res.* **2015**, *28*, 351–373. doi:10.1515/humor-2015-0070.
44. Cheyne, J.A.; Pennycook, G. Sleep paralysis post-episode distress: Modeling potential effects of episode characteristics, general psychological distress, beliefs, and cognitive style. *Clin. Psychol. Sci.* **2013**, *1*, 135–148. doi:10.1177/2167702612466656.
45. Pennycook, G.; Cheyne, J.A.; Seli, P.; Koehler, D.J.; Fugelsang, J.A. Analytic cognitive style predicts religious and paranormal belief. *Cognition* **2012**, *123*, 335–346. doi:10.1016/j.cognition.2012.03.003.
46. Shenhav, A.; Rand, D.G.; Greene, J.D. Divine intuition: Cognitive style influences belief in God. *J. Exp. Psychol. Gen.* **2012**, *141*, 423–428. doi:10.1037/a0025391.
47. Brañas-Garza, P.; García-Muñoz, T.; González, R.H. Cognitive effort in the beauty contest game. *J. Econ. Behav. Organ.* **2012**, *83*, 254–260. doi:10.1016/j.jebo.2012.05.018.
48. Hoppe, E.I.; Kusterer, D.J. Behavioral biases and cognitive reflection. *Econ. Lett.* **2011**, *110*, 97–100. doi:10.1016/j.econlet.2010.11.015.
49. Moritz, B.; Siemsen, E.; Kremer, M. Judgmental forecasting: Cognitive reflection and decision speed. *Prod. Oper. Manag.* **2014**, *23*, 1146–1160. doi:10.1111/poms.12105.
50. Sirota, M.; Juanchich, M. Role of numeracy and cognitive reflection in bayesian reasoning with natural frequencies. *Stud. Psychol.* **2011**, *53*, 151–161.
51. Insler, M.; Compton, J.; Schmitt, P. The investment decisions of young adults under relaxed borrowing constraints. *J. Behav. Exp. Econ.* **2015**, *64*, 106–121. doi:10.1016/j.socec.2015.07.004.
52. Corgnet, B.; Hernán Gonzalez, R.; Mateo, R. Cognitive reflection and the diligent worker: An experimental study of millennials. *PLoS ONE* **2015**, *10*, e0141243. doi:10.1371/journal.pone.0141243.
53. Roth, P.L.; BeVier, C.A.; Switzer, F.S.; Schippman, J. Meta-analyzing the relationship between grades and job performance. *J. Appl. Psychol.* **1996**, *81*, 548–556. doi:10.1037/0021-9010.81.5.548.
54. Liberali, J.M.; Reyna, V.F.; Furlan, S.; Stein, L.M.; Pardo, S.T. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *J. Behav. Decis. Mak.* **2012**, *25*, 361–381. doi:10.1002/bdm.752.
55. Obrecht, N.A.; Chapman, G.B.; Gelman, R. An encounter frequency account of how experience affects likelihood estimation. *Mem. Cogn.* **2009**, *37*, 632–643. doi:10.3758/MC.37.5.632.

56. Toplak, M.E.; West, R.F.; Stanovich, K.E. The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Mem. Cogn.* **2011**, *39*, 1275–1289. doi:10.3758/s13421-011-0104-1.
57. Campitelli, G.; Gerrans, P. Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Mem. Cogn.* **2014**, *42*, 434–447. doi:10.3758/s13421-013-0367-9.
58. Pennycook, G.; Cheyne, J.A.; Koehler, D.J.; Fugelsang, J.A. Is the cognitive reflection test a measure of both reflection and intuition? *Behav. Res. Meth.* **2015**, *48*, 341–348. doi:10.3758/s13428-015-0576-1.
59. Guilford, J.P. *Psychometric Methods*; McGraw-Hill: New York, NY, USA, 1954.
60. MacKinnon, D.P. *Introduction to Statistical Mediation Analysis*; Lawrence Erlbaum: New York, NY, USA, 2008.
61. Schmidt, F.L.; Hunter, J.E. *Methods of Meta-Analysis*, 3rd ed.; Sage: Newbury Park, CA, USA, 2015.
62. Bertua, C.; Anderson, N.; Salgado, J.F. The predictive validity of cognitive ability tests: A UK meta-analysis. *J. Occup. Organ. Psychol.* **2005**, *78*, 387–409. doi:10.1348/096317905X26994.
63. Hunter, J.E.; Hunter, R.F. Validity and utility of alternative predictors of job performance. *Psychol. Bull.* **1984**, *96*, 72–98. doi:10.1037/0033-2909.96.1.72.
64. Pearlman, K.; Schmidt, F.L.; Hunter, J.E. Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *J. Appl. Psychol.* **1980**, *65*, 373–406. doi:10.1037/0021-9010.65.4.373.
65. Salgado, J.F.; Anderson, N.; Moscoso, S.; Bertua, C.; de Fruyt, F.; Rolland, J.P. A meta-analytic study of general mental ability validity for different Occupations in the European Community. *J. Appl. Psychol.* **2003**, *88*, 1068–1081. doi:10.1037/0021-9010.88.6.1068.
66. Kunzel, N.R.; Credé, M.; Thomas, L.L. The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Rev. Educ. Res.* **2005**, *75*, 63–82. doi:10.3102/00346543075001063.
67. Kuncel, N.R.; Hezlett, S.H.; Ones, D.S. A Comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychol. Bull.* **2001**, *127*, 162–181. doi:10.1037//Q033-2909.127.1.162.
68. Postlethwaite, B.E. Fluid Ability, Crystallized Ability, and Performance across Multiple Domains: A Meta-Analysis. Ph.D. Thesis, The University of Iowa, Iowa City, IA, USA, 2011. doi:10.17077/etd.zopi8wvs.
69. Richardson, M.; Abraham, C.; Bond, R. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychol. Bull.* **2012**, *138*, 353–387. doi:10.1037/a0026838.
70. Robbins, S.B.; Lauver, K.; Le, H.; Davis, D.; Langley, R.; Carlstrom, A. Do psychosocial and study skill factors predict college outcomes? A Meta-Analysis. *Psychol. Bull.* **2004**, *130*, 261–288. doi:10.1037/0033-2909.130.2.261.
71. Schmitt, N.; Gooding, R.Z.; Noe, R.A.; Kirsch, M. Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Pers. Psychol.* **1984**, *37*, 407–422. doi:10.1111/j.1744-6570.1984.tb00519.x.
72. Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **2007**, *35*, 401–426. doi:10.1016/j.intell.2006.09.004.
73. Salgado, J.F.; Moscoso, S. The validity of general mental ability for five performance criteria: Hunter and Hunter (1984) revisited. *Front. Psychol.* **2019**, in press.
74. Gulliksen, H. *Theory of Mental Tests*; Wiley: New York, NY, USA, 1950.
75. Lord, F.M.; Novick, R. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Oxford, UK, 1968.
76. Wonderlic & Associates. *Wonderlic Personnel Test and Scholastic Level Exam*; User's Manual; Wonderlic, Inc.: Libertyville, IL, USA, 1998.
77. Cattell, R.B.; Cattell, A.K.S. *Measuring Intelligence with the Culture-Fair Tests (Manual for Scale 2 and 3)*; TEA Ediciones: Champaign, IL, USA, 1986.
78. Bacon, D.R.; Bean, B. GPA in research studies: An invaluable but neglected opportunity. *J. Mark. Educ.* **2006**, *28*, 35–42. doi:10.1177/0273475305284638.
79. Salgado, J.F.; Tauriz, G. The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *Eur. J. Work Organ. Psychol.* **2014**, *23*, 3–30. doi:10.1080/1359432X.2012.716198.

80. Schmidt, F.L.; Shaffer, J.A.; Oh, I. Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Pers. Psychol.* **2008**, *61*, 827–868. doi:10.1111/j.1744-6570.2008.00132.x.
81. Fritz, M.S.; Kenny, D.A.; MacKinnon, D.P. The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivar. Behav. Res.* **2016**, *51*, 681–697. doi:10.1080/00273171.2016.1224154.
82. Browne, M.W. Predictive validity of a linear regression equation. *Brit. J. Math. Stat. Psychol.* **1975**, *28*, 79–87. doi:10.1111/j.2044-8317.1975.tb00550.x.
83. Cattin, P. Estimation of the predictive power of a regression model. *J. Appl. Psychol.* **1980**, *65*, 407–414. doi:10.1037/0021-9010.65.4.407.
84. Cattin, P. Note on the estimation of the squared cross-validated multiple correlation of the regression model. *Psychol. Bull.* **1980**, *87*, 63–65. doi:10.1037/0033-2909.87.1.63.
85. Yin, P.; Fan, X. Estimating R² Shrinkage in multiple regression: A comparison of different analytical methods. *J. Exp. Educ.* **2001**, *69*, 203–224. doi:10.1080/002200970109600656.
86. Salgado, J.F. Using Ability Tests in Selection. In *Handbook of the Psychology of Recruitment, Selection, and Retention*; Goldstein, H., Pulakos, E., Passmore, J., Semedo, C., Eds.; Wiley/Blackwell: Malden, MA, USA, 2017; pp. 115–150.
87. Schmidt, F.L.; Le, H. *Software for the Hunter-Schmidt Meta-Analysis Methods*; Department of Management and Organizations, University of Iowa: Iowa City, IA, USA, 2014.
88. Berges, A.; Fernández-del-Río, E.; Ramos-Villagrasa, P.J. The prediction of training proficiency in Firefighters: A study of predictive validity in Spain. *J. Work Organ. Psychol.* **2017**, *34*, 10–15. doi:10.5093/jwop2018a2.
89. Lado, M.; Alonso, P. The Five-Factor model and job performance in low complexity jobs: A quantitative synthesis. *J. Work Organ. Psychol.* **2017**, *33*, 175–182. doi:10.1016/j.rpto.2017.07.004.
90. Salgado, J.F. Moderator effects of job complexity on the validity of forced-choice personality inventories for predicting job performance. *J. Work Organ. Psychol.* **2017**, *33*, 229–239. doi:10.1016/j.rpto.2017.07.001.
91. Di Fabio, A.; Tsuda, A. The psychology of harmony and harmonization: Advancing the perspectives for the psychology of sustainability and sustainable development. *Sustainability* **2018**, *10*, 4726. doi:10.3390/su10124726.
92. Pennycook, G.; Ross, R.M. Commentary: Cognitive reflection vs. calculation in decision making. *Front. Psychol.* **2016**, *7*, 9. doi:10.3389/fpsyg.2016.00009.
93. Collins, J.M.; Schmidt, F.L.; Sanchez-Ku, M.; Thomas, L.; McDaniel, M.A.; Le, H. Can basic individual differences shed light on the construct meaning of assessment center evaluations? *Int. J. Select. Assess.* **2003**, *11*, 17–29. doi:10.1111/1468-2389.00223.

