Maximum Effort May Not Be Required for Valid Intelligence Test Score Interpretations

Gilles E. Gignac, Asher Bartulovich, & Emilee Salleo

University of Western Australia

(A version of which to be published in Intelligence; Accepted May, 2019)

Author Note

Correspondence should be addressed to Gilles E. Gignac, School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia, 6009, Australia. E-mail: gilles.gignac@uwa.edu.au; thanks to Clare Starbuck for assistance with data collection.
Abstract

Intelligence tests are assumed to require maximal effort on the part of the examinee. However, the degree to which undergraduate first-year psychology volunteers, a commonly used source of participants in low-stakes research, may be motivated to complete a battery of intelligence tests has not yet been tested. Furthermore, the assumption implies that the association between test-taking motivation and intelligence test performance is linear – an assumption untested, to date. Consequently, we administered a battery of five intelligence subtests to a sample of 219 undergraduate volunteers within the first 30 minutes of a low-stakes research setting. We also administered a reading comprehension test near the end of the testing session (55 minutes). Self-reported test-taking motivation was measured on three occasions: at the beginning (as a trait), immediately after the battery of five intelligence tests (state 1), and immediately after the reading comprehension test (state 2). Six percent of the sample was considered potentially insufficiently motivated to complete the intelligence testing, and 13% insufficiently motivated to complete the later administered reading comprehension test. Furthermore, test-taking motivation correlated positively with general intelligence test performance ($r \approx .20$). However, the effect was non-linear such that the positive association resided entirely between the low to moderate levels of test-taking motivation. While simultaneously acknowledging the exploratory nature of this investigation, it was concluded that a moderate level of test-taking effort may be all that is necessary to produce intelligence test scores that are valid. However, in low-stakes research settings, cognitive ability testing that exceeds 25 to 30 minutes may be inadvisable, as test-taking motivational levels decrease to a degree that may be concerning for a non-negligible portion of the sample.

*Keywords: intelligence, motivation, validity, testing*
Maximum Effort May Not Be Required for Valid Intelligence Test Score Interpretations

**Introduction**

With respect to tests of maximum performance, such as intelligence tests, Cronbach (1990, p. 38) stated that examinees must be sufficiently motivated “…to earn the best score [they] can.” Furthermore, Cronbach (1990, p. 38) recommended that the goal of the maximum performance tester was to “…bring out [people’s] best possible performance (within the rules), and this means that [examinees] must want to do well…”. Correspondingly, Revelle (1993) stated that when researchers study human performance, it is assumed that “…subjects are alert and optimally motivated. It is also assumed that the experimenter’s task at hand is by far the most important thing the subject has to do at that time” (p. 352). However, individual differences in test-taking motivation, and positive correlations between test-taking motivation and cognitive ability test performance, have been reported (e.g., Schiel, 1996; Sundre & Kitsantas 2004). Furthermore, intelligence predictive validity coefficients have been contended to be mediated to a substantial degree by test-taking motivation (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011). However, to date, there is little test-taking motivation research with commonly used first-year undergraduate student samples requested to complete a moderately sized battery of intelligence tests during a low stakes testing session. Additionally, Gignac and Wong (2018) reported a curvilinear association between a behavioral measure of test-taking persistence and intelligence test performance, suggesting the possibility that the positive association between test-taking motivation and intelligence resides principally between the low to moderate levels of test-taking motivation.

Consequently, the purpose of this investigation was to measure individual differences in self-reported test-taking motivation, in order to estimate the percentage of first-year
undergraduate psychology student volunteers who may be considered insufficiently motivated to complete a moderately sized battery of intelligence tests. Furthermore, we were interested in comparing their state test-taking motivation with their trait test-taking motivation, in addition to the possibility that their state test-taking motivation may decline during a 60 minute testing session. Finally, we tested the hypothesis that the association between self-reported test-taking motivation and intelligence test performance is nonlinear.\footnote{We acknowledge that a nonlinear association may be monotonic, i.e., increasing or decreasing at a different rate, or non-monotonic, i.e., change in direction from positive to negative or negative to positive (Yitzhaki & Schechtman, 2012). For the purposes of this investigation, we use the term nonlinear to imply a monotonic association. However, we revisit this issue in the discussion.}

**Test-Taking Motivation in Undergraduate Research Volunteers**

Several investigations have found that approximately two thirds of studies published across a variety of peer-reviewed journals in psychology employed undergraduate university students as participants (Arnett, 2008; Wintre, North and Sugar, 2001). It may be presumed that, in most cases, the students participate as part of a course requirement and/or for partial course credit (e.g., Estrada, Román, Abad, & Colom, 2017; Kretzschmar, Neubert, Wustenberg, & Greiff, 2016; Ren, Wang, Sun, Deng, & Schweizer, 2018). As there are little, if any, consequences to performing well or poorly on tests in such scenarios, the testing situation may be described as low-stakes (Shrock & Coscarelli, 2008). Consequently, whether first-year undergraduate psychology research volunteers are motivated sufficiently to exert the assumed
maximum effort required to produce valid cognitive ability test scores (Cronbach, 1960; Roid & Barram, 2004; Revelle, 1993) may be questioned.

To date, a substantial portion of self-reported test-taking motivation and cognitive ability testing research has been conducted with primary and high-school students requested to complete academic achievement type tests (e.g., Hopfenbeck & Kjaernsli, 2016; Schiel, 1996). Such research may not be especially relevant, here, as the students were typically required to complete the testing as part of a school or government initiative. Additionally, completing a mass academic achievement test may not be comparable to individual intelligence testing in a laboratory setting. Nonetheless, we note that the mass academic achievement testing research suggests that as many as 30% of the students reported applying far less than maximum effort, when completing the academic achievement tests (Hopfenbeck & Kjaernsli, 2016; Schiel, 1996; Thelk et al., 2009).

Other possibly relevant research is that conducted with first-year undergraduate psychology students requested to complete relatively comprehensive neuropsychological test batteries, many of which include several cognitive ability tests in addition to validity test performance indicators (a.k.a., tests of malingering). For example, An, Zakzanis, and Joordens (2012) administered a comprehensive cognitive ability testing battery (e.g., Letter-Number Sequencing; Advanced Progressive Matrices), in addition to three tests designed to detect malingering (e.g., Dot Counting Test; Boone, Lu & Herzberg, 2002) to a sample of 36 first-year undergraduate psychology student volunteers (for extra course credit). An et al. (2012) considered performance on the malingering tests as indicative of test-taking motivation, as all undergraduate university students (i.e., young and neurologically healthy) should perform beyond a very poor level on the tests of malingering (i.e., > 5th percentile). On the basis of their
analysis, An et al. (2012) reported that 55% of the participants were identified as having engaged in a poor effort on at least one of the three malingering tests. Consequently, An et al. (2012) interpreted the results to suggest that using first year psychology student test scores in research may be problematic. Several other neuropsychological testing investigations have applied procedures similar to An et al. (2012) and found lower, but still potentially concerning, levels of implausibly poor performance from a non-negligible percentage of their first-year psychology undergraduate research volunteers (i.e., 8% to 18%; DeRight & Jorgensen, 2015; Ross et al., 2016; An, Kaploun, Erdodi & Abeare, 2017).

Furthermore, in the area of experimental cognition, Hawkins et al. (2013) administered an adaptive binary decision making task to a sample of 64 first-year psychology student volunteers in a low-stakes setting. The task involved hundreds of trials that required as much as 50 minutes to complete in total. Hawkins et al. (2013) also administered a short-survey that included the following 7-point Likert scale item: ‘How much effort did you put in throughout the task?’; 1 = lowest; 7 = highest. Hawkins et al. (2013) reported a mean of 4.41 (SD = 1.60) for the self-reported effort item. Thus, assuming a value less than 4.5 on the 7-point scale may be considered indicative of appreciably less than maximum effort (and assuming a relatively normal distribution of scores), then perhaps as many as one half of the undergraduate volunteers may be regarded as to have self-reported an insufficient amount of motivation to complete the cognitive task in a valid manner.

Although insightful, the research reviewed above may be regarded as limited in at least two ways. First, the amount of time required to complete the neuropsychological testing was estimated by us to be between 90 and 180 minutes, which may be considered rather substantial, in comparison to more typical differential psychology investigations (i.e., often involve less than
30 minutes of intelligence testing). Thus, it remains to be determined whether undergraduate university volunteers apply a respectable amount of effort during the completion of relatively brief low-stakes intelligence testing.

Secondly, most of the research reviewed above tended not to report whether the volunteers were screened for English language proficiency. If they were not, the increasing internationalisation of higher education (Larsen, 2016) suggests that a non-negligible percentage of the participants (10% +) may not have spoken English as a first language, which may have compromised their test performance, as well as possibly their motivation, on the English-based cognitive ability tests. Consequently, the estimates of insufficiently motivated examinees reported above may be overestimates.

Recently, Gignac (2018) conducted an investigation with first-year undergraduate volunteers (extra course credit) who were screened for English as a first language. A relatively brief intelligence test battery (five subtests; ≈ 25 minutes), in addition to a self-report inventory of test-taking motivation was administered. Specifically, Gignac (2018) administered the Student Opinion Scale, a 10-item inventory scored on a 5-point Likert scale (strongly disagree to strongly agree) designed to measure two moderately correlated dimensions of test-taking motivation: Importance and Effort. Although Gignac’s (2018) investigation included an experimentally manipulated financial incentive, for the condition in which no financial incentive was offered (N = 51), the self-reported test-taking Importance and test-taking Effort means were estimated at 3.82 and 3.40, respectively. If a mean item response of 3.5 or less may be regarded as appreciably less than maximum effort to complete the cognitive ability tests, then approximately 50% of Gignac’s (2018) sample may be considered to have been insufficiently motivated. However, a limitation with Gignac (2018) was that the sample size was only 51 in the
non-incentivised condition. Consequently, a substantial amount of confidence cannot be placed in the results. Another limitation with Gignac (2018) is that test-taking motivation was measured only from the perspective of a state.

**State versus Trait Test-Taking Motivation**

Broadly speaking, motivation is a construct that represents the psychological processes that produce the initiation, direction, intensity, and persistence of goal-directed behaviour (Mitchell, 1982). By comparison, a trait is a particular tendency to behave in the same way across situations (Allport, 1927; Hertzog & Nesselroade, 1987). Well-known traits in the area of personality include extraversion, neuroticism, sensation seeking, perfectionism, and harm avoidance, for example (Matthews, Deary & Whiteman, 2003). There is an impressive amount of evidence to support the notion that there are substantial inter-individual differences in these traits, and that these inter-individual differences are relatively stable across situations (e.g., Rantanen, Metasapelto, Feldt, Pulkkinen, & Kokko, 2007).

In contrast to a trait, the state representation of an attribute corresponds to the fluctuations across situations and contexts (Hertzog & Nesselroade, 1987; Rosenberg, 1998). The distinction between a state and a trait permeates several areas of psychology. For example, one of the most well-known state-trait distinctions is state and trait anxiety (Oatley, Keltner, & Jenkins, 2006; Spielberger, 1985). In the area of motivation, specifically, Brophy (1986) proposed the distinction between state motivation to learn (motivation to learn material in a particular situation

---

2 In a conference presentation, Gignac (2017) reported undergraduate university sample \((N = 173)\) low-stakes testing data for which the mean on the Student Opinion Scale test-taking Effort subscale was 3.87.
and time) and trait motivation to learn (an enduring disposition to learn material across situations). Correspondingly, Humphreys and Revelle (1984) distinguished between state achievement (motivation) and trait achievement (motive). State levels of an attribute have been reported to fluctuate by as much as a full standard deviation (Evans et al., 1988). Furthermore, the correlations between corresponding state and trait dimensions tend to range between .40 and .70, i.e., not so large as to suggest redundancy (e.g., Cohen, et al., 1995; Hong, 1998; Oei, Evans, & Crook, 1990).

Theoretically, trait test-taking motivation may be discussed within the context of intrinsic and extrinsic motivation (Deci & Ryan, 1985). That is, it may be suggested that, for some people, taking an intelligence test may be intrinsically motivating. Thus, when completing an intelligence test, it is reasonable to hypothesize that some people would be “…moved to act for the fun or challenge entailed, rather than because of external prods, pressures, or rewards” (Ryan & Deci, 2000, p. 56). Furthermore, like most traits in psychology, it would be expected that there would be individual differences with respect to trait test-taking motivation.

To date, only a small amount of state-trait related empirical research in the area of test-taking motivation has been conducted. For example, in a sample of German university student volunteers who completed a matrix reasoning type test, Freund and Holling (2011) found that self-reported interest (state) correlated positively ($r = .37$) with matrix reasoning test performance. Additionally, Freund and Holling (2017) found that trait conscientiousness correlated positively ($r = .17$) with individual difference in state interest, suggesting the possibility that some trait test-taking motivation variance may underlie some of the state test-taking motivation measurement responses.
In another investigation with a large sample of German ninth graders required to complete national academic achievement testing, Penk, Pöhlmann, and Roppelt (2014) investigated the influence of self-concept in mathematics (trait) and situational (state) test-taking motivation (invested effort and usefulness of the test) on the mathematics portion of the achievement testing. In a multiple regression model that included all three predictors, 11% of the variance in mathematics achievement was accounted for. Furthermore, the semi-partial correlations corresponded to approximately .13, .08 and .02 for self-concept, invested effort, and usefulness of the test scales, respectively.\(^3\) Thus, perhaps surprisingly, self-concept was a larger unique predictor of academic mathematics achievement test performance than state level test-taking motivation. Such a result is surprising, as the state representation of a test-taking motivational construct may be expected to relate more strongly with test performance scores obtained in that situation. A similar effect with similar measures was reported by Penk and Richter (2017) in another sample of German ninth graders who completed academic achievement testing in a low-stakes setting.

Although insightful, arguably, the construct of self-concept is not the same as trait test-taking motivation. For example, the self-concept measure included in the Penk and colleagues work included the following item, ‘I have always been good at math,’ which is more akin to an item that might be found in a multi-item measure of self-estimated intelligence (e.g., Gignac, Stough, & Loukomitis, 2004), rather than a clear indicator of test-taking motivation. Thus, although a few test-taking motivation questionnaires have been developed (or combined) to

\(^3\) We estimated the semi-partial correlations on the basis of the multiple regression t-values for each predictor’s unstandardized beta-weight.
incorporate some elements of both trait and state test-taking motivation (Arvey, Strickland, Drauden, & Martin, 1990; Eklöf, 2006; Freund & Holling, 2001; Penk et al., 2014; Rheinberg, Vollmeyer, & Burns, 2001), none, to our knowledge, measure dimensions such as test-taking importance and test-taking effort in a clearly comparable manner across the trait and state perspectives.

The administration of more directly comparable trait and state test-taking motivation questionnaires would facilitate answering two types of questions. First, the degree to which undergraduate research volunteers are motivated to complete intelligence testing in a low-stakes setting could be compared against their typical test taking-motivation. Thus, in the context of completing intelligence tests in a low-stakes research setting, the observation of a statistically significantly lower mean level of state test-taking motivation, in comparison to trait-level test-taking motivation, would suggest that the undergraduate university students may be, on average, relatively less motivated to complete intelligence tests in low-stakes research-settings. Secondly, as state test-taking motivation is more situationally proximal to the administered tests, state test-taking motivation should relate more substantially to test performance, in comparison to a trait representation of test-taking motivation. Such a pattern of correlations would help support the distinction between trait and state test-taking motivation.

To test such hypotheses, the items from a state only questionnaire of test-taking motivation, such as the Student Opinion Scale, could be adapted to measure test-taking motivation from a comparable trait perspective. For example, item 1 within the Student Opinion scale reads, ‘Doing well on these tests was important to me.’ A trait modification of such an item could read, ‘Doing well on tests is important to me.’ As another example, item 2 within the Student Opinion Scale reads, ‘I engaged in a good effort throughout these tests.’ A trait
modification of such an item could read, ‘I engage in a good effort when tested.’ Such slight modifications to all 10 of the Student Opinion Scale’s items (see Appendix A) would allow for more direct comparisons between trait test-taking motivation and state test-taking motivation.

**Changes in Test-Taking Motivation in Research Volunteers**

Another limitation with much of the existing test-taking motivation research is that test-taking motivation is typically measured only once within the testing session. The measurement of test-taking motivation across a testing session could help evaluate the possibility that test-taking motivation may decrease over time - a realistic possibility, as most volunteers in a low-stakes setting may not view the research as sufficiently important to maintain a significant effort for an extended period of time.

To date, a small amount of empirical research suggests that test-taking motivation does decrease across a testing session. For example, in large scale study with ninth grade students ($N = 42,298$), Penk and Richter (2017) measured self-reported test-taking motivation (test-taking motivation scale from Eklöf, 2001) three times over the course of a two hour testing session, during which the students completed academic achievement testing (mathematical and scientific literacy), as part of a large scale government initiative. Penk and Richter (2017) found that self-reported test-taking effort decreased by approximately $d \approx .30$ within the first hour of testing, and then by approximately $d \approx .20$ between the completion of the first and second-hour of testing. Additionally, self-reported test-taking importance also evidenced a statistically significant decrease; however, the decrease was observed primarily within the first hour ($d \approx .35$).

In a behavioral response time study with first-year university students ($N = 506$), Wise (2006) reported a negative correlation ($r = .69$) between item position (earlier to later administered items) and response time fidelity (absence of rapid response guessing) for a 79-item
information retrieval/reading comprehension test. Although all incoming students were required to complete the testing, Wise (2006) considered the test testing low-stakes, as the scores were not made known to the students, nor were they used for selection type purposes. Similar results have been reported by others (e.g., Horst, 2010; Wise et al., 2009). However, it remains to be determined whether first-year undergraduate university student volunteers report changes in test-taking motivation across a typical low-stakes research setting in a laboratory that involves approximately 60 minutes of total testing, i.e., arguably, a typical research scenario in differential psychology.

**Correlations with Test Performance**

Concerns about the validity of intelligence test scores rest, in part, upon the observation of individual differences in test-taking motivation (Duckworth et al., 2011; Wise & Kong, 2005), as well as empirical research that has reported consistent, positive correlations ($r \approx .20$ to .35) between individual differences in test-taking motivation and cognitive ability-type test scores (e.g., Hopfenbeck & Kjaermsli, 2016; Schiel, 1996; Sundre & Kitsantas 2004). However, relatively little of the research, to date, has been conducted with undergraduate university volunteers in a low-stakes setting. For example, although Chan et al. (1997) reported a positive correlation ($r \approx .30$) between self-reported test-taking motivation and IQ test performance in university volunteers, the participants were told that those who performed in the top 40% would be paid a sum of money. Borghans, Meijers, and Ter Weel, (2008) employed a similar procedure with undergraduate university volunteers requested to complete IQ test items. Consequently, it is doubtful that such a testing scenarios may be described as entirely low-stakes.

A rare exception is Gignac (2018) who administered the Student Opinion Scale to first-year undergraduate university students, along with a battery of five intelligence tests. Gignac
(2018) reported correlations of .28 and .37 between test-taking Importance/Effort and intelligence test performance, respectively. However, the sample in the condition for which there were no financial incentives was small (i.e., \( N = 51 \)). Thus, further research with a larger sample size would be useful.

Another limitation with the self-reported test-taking motivation test performance correlational research is that the association has been assumed to be linear. Recently, Gignac and Wong (2018) administered the Anagram Persistence Task (APT) and a battery of intelligence tests to a sample of undergraduate student volunteers in a low-stakes setting. The APT includes several essentially impossible anagrams, and the amount of time people spend attempting to complete the essentially impossible anagrams is considered a measure of test-taking persistence. On the basis of a curvilinear multiple regression, Gignac and Wong (2018) found a statistically significant curvilinear effect between test-taking persistence and overall intelligence test performance. Specifically, a positive association was observed from lower to moderate levels of persistence, in combination with a substantially decelerating association beyond a relatively moderate level of persistence. To our knowledge, the possibility that self-reported test-taking motivation may be associated with test performance in a curvilinear fashion has not yet been investigated. Importantly, the observation of a curvilinear association, such that the positive association dissipates beyond a moderate level of test-taking motivation, would suggest that a maximum amount of effort may not be required to produce an essentially valid intelligence test score - a widely held assumption in the psychometric literature (Cronbach, 1960; Roid & Barram, 2004; Revelle, 1993).
Summary and Purpose

Test-taking motivation has been argued to be a potential confound in the measurement of intelligence (Duckworth et al., 2011). However, to date, there has been relatively little research on the levels of test-taking motivation of first-year undergraduate volunteers – a commonly used source of participants in psychological research (Arnett, 2008; Wintre, North and Sugar, 2001). Additionally, the association between self-reported test-taking motivation and test performance has, thus far, only been assumed to be linear, rather than specifically examined to be linear.

Consequently, the purpose of this investigation was to estimate the degree to which first-year psychology undergraduates are motivated to complete intelligence testing in a low-stakes setting, in comparison to their trait level of test-taking motivation. Additionally, the degree to which test-taking motivation may change across an approximate one hour testing session was examined. Finally, the association between self-reported test-taking motivation and intelligence test performance was estimated from both a linear and curvilinear (quadratic) perspective.

Method

Sample

The original sample included 228 first-year undergraduate psychology students (English as a first language) from a large university in Australia. However, ten of the participants were excluded from the sample, as we did not have information on their age. Thus, the final sample was \( N = 219 \) (71.1% female). The mean age of the sample was 19.87 (\( SD = 4.60 \); range 17 to 55). All of the participants were volunteers, and they received extra course credit in return for their participation.
Measures

**Test-taking motivation.** Test-taking motivation was measured with the state and trait versions of the Student Opinion Scale (SOS). The state version of the Student Opinion Scale (Sundre, 1999; Thelk, et al., 2009) consists of 10 items designed to measure two dimensions of test-taking motivation: Importance (5-items) and Effort (5-items). Each item is responded to on a 5-point Likert-scale: 1 = Strongly Disagree; 2= Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree. As described in the introduction above, the trait version of the Student Opinion Scale was created for the purposes of this investigation (see Appendix A). For this sample, the reliabilities were: trait-Importance = .59; trait-Effort = .72; state-Importance time 1 = .67; state-Effort time 1 = .75; state-Importance time 2 = .83; state-Effort time 2 = .84.

**Intelligence.** Intelligence was measured across four dimensions of cognitive ability: processing speed, crystallised intelligence, fluid intelligence, and memory span.

*Processing speed* was measured with the four subtests (Numbers, Letters, Numbers/Letters, and Letters/Numbers) within the Connections battery (Salthouse et al., 2000). These tests are very similar to the well-known Trails A/Trails B tests (Reitan, 1958). On the basis of the correlations between the four processing speed tests, the internal consistency reliability of the overall processing speed composite scores was estimated at .72. *Crystallized intelligence* was measured with the Advanced Vocabulary Test (AVT; Gignac, Shankaralingam, Walker, & Kilpatrick, 2016), which is a 21-item multiple choice test comprised of relatively difficult words to define, by selecting the most appropriate of the five alternatives ($\alpha = .61$ this sample). *Fluid intelligence* was measured with the odd items from the Advanced Progressive Matrices (APM; Raven, 1998; $\alpha = .67$ this sample). Participants were given 10
minutes to complete the APM short form. Finally, memory span was measured with a slight adaptation of the Letter–Number Sequencing Test (LNS; Wechsler, 2008; \( \alpha = .79 \) this sample).

**Procedure**

The testing battery was administered as a part of a combined investigation relevant to differential psychology (i.e., test-taking motivation; attractiveness of intelligence). The entire testing protocol consisted of the following. First, after providing informed consent, the participants completed some basic demographic questions (age, gender). Next, they completed the trait version of the Student Opinion Scale (1 minute), after which they completed the following three surveys that were not relevant to this investigation: Buss’s ranking questionnaire (1.5 minutes; Buss & Barnes, 1986), a percentile-based attraction survey (8 minutes; Gignac & Starbuck, 2018), and a nine-item self-report intelligence questionnaire (1 minute; Gignac et al., 2004). All of these questionnaires were completed on a computer (via Qualtrics). Next, the participants interacted with the experimenter, in order to complete the Letter–Number Sequencing test (4-5 minutes) and the Connections test (2 minutes). Then, the participants were returned to the computer to complete the following two intelligence tests: the Advanced Progressive Matrices (10 minutes) and the Advanced Vocabulary Test (4 minutes). With the intelligence testing complete, the participants were administered the state version of the Student Opinion Scale (1 minute). They then completed the Balanced Inventory of Desirable responding (5 minutes; Paulhus, 1991) and the HEXACO-60-PI (7 minutes; Ashton & Lee, 2009) at the computer. Next, the participants watched a 3 minute video: half of the participants watched a motivational video and the other half watched a video of people walking in a naturalistic setting.
(participants were allocated randomly). Then, the participants read a short introduction on statistical variance (4 minutes; see Supplementary Materials, ‘Testing Differences in Variability Statistically’), after which the participants were tested on their reading comprehension with a short test on statistical variance (7 items; 2.5 minutes; see Supplementary Materials, ‘Reading Comprehension Test Questions’). As the reading comprehension test was associated with internal consistency reliability of only .22, it was not considered for analysis in this investigation. Finally, the participants completed the state Student Opinion Scale for a second time, immediately after the reading comprehension test. An experimenter was in the testing room at all times. All testing was completed individually and took approximately 55 minutes.

To reduce the chances of demand characteristics, the participants were not made aware that the purpose of the investigation was to determine the degree to which undergraduate university students are motivated to complete intelligence type tests. Instead, the participants were simply informed the following: “The purpose of the investigation is to understand why individuals vary in their performance on various tests. Of course, ability would be expected to play a part. However, a number of other factors would be expected to play a part, as well. For example, your perceived confidence at completing such tasks successfully, as well as your interest in completing them.”

\[^4\] The two videos were not found to have a differential impact on any of the scores. Specifically, state-Importance: \(t(216) = .60, p = .547\), Cohen’s \(d = .08\); state-Effort: \(t(216) = .04, p = .788\), Cohen’s \(d = \); reading comprehension test: \(t(216) = -.22, p = .828\), Cohen’s \(d = -.03\).
Data Analysis

In order to examine the levels of test-taking motivation reported by the sample, we calculated the means and standard deviations for the trait, state time 1, and state time 2 conditions for both the test-taking Importance and Effort subscales. Next, in order to test the hypothesis that, on average, the participants were potentially minimally acceptably motivated to complete the intelligence testing, we conducted a series of one-sample $t$-tests with a specified test value as the null expectation for the Student Opinion Scale subscales. Naturally, any specified numerical null test value would be, to some degree, arbitrary. We note that analyses of this nature with 5-point Likert scales tend to use a null test value of 3.0 for the one-sample $t$-test (i.e., neutral; e.g., Foster, 2001; Hoffmann & Fischer, 2012; Tetterton & Brodsky, 2007). However, we regarded a mean item response from an examinee equal to approximately 3.0 on the Student Opinion Scale as unlikely to be sufficiently motivated, as a mean item response of 3.0 corresponds to ‘neutral’ on the scale. Consequently, we chose a more conservative null test value of 3.5, as it was somewhere between ‘neutral’ and ‘agree’ on the 5-point Likert scale. We acknowledge that this somewhat arbitrary null specification renders this investigation as exploratory in nature. Additionally, for further insights, we reported the percentage of participants who rated their test-taking motivation less than 3.0 on the Importance and Effort subscales, as people who scored less than 3.0 were considered more clearly insufficiently motivated to complete the testing.

Next, in order to test the null hypothesis of equal test-taking motivation means across the trait/state conditions (trait, state-time 1, state-time 2), and across the Importance and Effort dimensions, we conducted a 3 x 2 factorial repeated measures ANOVA, with time as a within-
subjects factor (three levels: trait; state 1; state 2) and the two test-taking motivation sub-dimensions as a within-subjects factor (two levels; Importance and Effort).

The correlations between the variables included in this investigation were estimated with Pearson correlations. We created general intelligence component scores on the basis of a principal components analysis with the extraction of a single component, as component analysis, unlike common factor analysis, does not suffer from factor indeterminacy issues (Velicer & Jackson, 1990). In order to test the hypothesis that state test-taking motivation would correlate more substantially with general intelligence test performance than trait test-taking motivation, tests of the difference between dependent and overlapping (one variable in common) correlations were conducted with the Lee and Preacher web utility (Lee & Preacher, 2003; Steiger, 1980; Williams, 1959).

In order to test the hypothesis that self-reported test-taking motivation was associated with general intelligence test performance in a curvilinear fashion, two hierarchical multiple regressions were performed (for Importance and Effort, separately). Specifically, the test-taking motivation dimension was entered at step one and the corresponding squared variable (quadratic function) was entered at step two. A statistically significant change in $R^2$ between step one and step two was considered evidence for a curvilinear effect. Although we did not expect the data to be substantially non-normal, where possible, we nonetheless used bootstrapping as the estimation procedure to test the hypotheses for statistical significance (2000 resamples; bias-corrected accelerated confidence intervals), as bootstrapping has been found to be an all-around robust estimation approach (Chernick, 2008).

The data were analysed with IBM SPSS (Version 24). The data have been uploaded to the Open Science Framework and are available under the following URL: https://osf.io/jsnbx/
Results

An examination of the distribution of scores uncovered a low to moderate level of negative skewness for the test-taking motivation scores (≈ -.20 to -.50). However, no outliers were identified on the basis of the inter-quartile range rule with a 3.0 multiplier (Hoaglin & Iglewicz, 1987). Furthermore, the general intelligence test scores (IQg) were essentially normally distributed (skew = -.16).

Test-Taking Motivation: Means and Change

As can be seen in Table 1, the trait-Importance ($M = 4.19$) and trait-Effort ($M = 3.83$) means were statistically significant larger than 3.5, on the basis of one-sample $t$-tests. Thus, as a group, the participants were considered to have a stable disposition toward considering testing important, as well as to apply a potentially minimally acceptable effort at completing those tests. By contrast, on average, the participants did not consider the intelligence testing at time 1 to be important (state-Importance $T_1$, $M = 3.39$; $< 3.50$, $p = .008$). However, on average, they reported making a statistically significantly greater effort than the null expectation of 3.5 (state-Effort $T_1$, $M = 3.82$; $d = .53$, $p = .008$; see Table 1). Furthermore, 74.4% reported a state-Effort $T_1$ mean greater than 3.5. By contrast, 18.7% and 5.9% of the examinees reported state-Importance $T_1$ and state-Effort $T_1$ item means less than 3.0, respectively (see Figure 1 for violin plots with means).

With respect to the reading comprehension test administered later in the experimental protocol, on average, the participants continued to view the cognitive ability testing as relatively unimportant (i.e., state-Importance $T_2$, $M = 3.17$; i.e., $< 3.50$, $p < .001$). However, on average, the participants, nonetheless, continued to self-report a minimally acceptable level of test-taking effort for the second stage of cognitive ability testing (state Effort $T_2$, $M = 3.63$), on the basis of
a one-sample $t$-test ($d = .19, p = .008$; see Table 1). Furthermore, a total of 35.2% and 13.2% of the examinees reported state-Importance T2 and state-Effort T2 means less than 3.0, respectively.

Next, we examined the test-taking means from the perspective of change. The assumption of sphericity was violated for both the condition (trait, state T1 and state T2) main effect ($\varepsilon = .84, p < .001$) and the condition by subtest-type (Importance and Effort) interaction ($\varepsilon = .90, p < .001$). Consequently, the Huynh-Feldt adjusted statistics were consulted. With respect to differences between the test-taking means across the three conditions and two subtests, a factorial repeated measures ANOVA rejected the null hypothesis of equal main effect condition means, $F(1.69, 367.41) = 124.09, p < .001$, partial $\eta^2 = .363$, as well as equal main effect subtest-type means, $F(1, 218) = 31.03, p < .001$, partial $\eta^2 = .125$. However, a statistically significant interaction was identified, $F(1.79, 390.89) = 175.10, p < .001$, partial $\eta^2 = .445$. As can be seen in Figure 2, a classic disordinal interaction characterised the test-taking motivation means from trait to state time 1. Specifically, the magnitude of the decrease in test-taking Importance ($d = 1.54, t = 17.86, p < .001$) was more substantial than the decrease in test-taking Effort ($d = .01, t = 0.16, p = .218$). Stated alternatively, the participants applied an effort at completing the intelligence testing roughly equal to their trait disposition to making an effort at test-taking, whereas there was a substantial drop in rated test-taking Importance. By comparison, there was no interaction associated test-taking motivation Importance and Effort means leading from time 1 to time 2, $F(1, 218) = .58, p = .448$, partial $\eta^2 = .003$. Instead, both test-taking Importance and test-taking Effort decreased by essentially the same magnitude (state-Importance: $d = .33, t = 5.85, p < .001$; state-Effort: $d = .31, t = 5.05, p < .001$).
Test-Taking Motivation and IQ Test Performance

As can be seen in Table 2, at time 1, both state-Importance T1 ($r = .18; 95\%CI: .03/.32$) and state-Effort T1 ($r = .20; 95\%CI: .09/.31$) correlated positively with general intelligence test performance. By contrast, trait-Importance and trait-Effort did not correlate with general intelligence test performance significantly. The test of the difference between the corresponding trait/state correlations was found to be significant only for Effort (Importance, $\Delta r = .13, p = .124$; Effort, $\Delta r = .30, p < .001$).

Next, a series of two curvilinear multiple regressions were performed, by regressing general intelligence test performance onto test-taking motivation (Importance and Importance$^2$ in addition to Effort and Effort$^2$, separately). As can be seen in Table 3, both state-Importance T1 and state-Effort T1 were associated with statistically significant curvilinear effects. Specifically, across both models, approximately 6% of the total variance in general intelligence test performance was accounted for by the curvilinear models (linear + quadratic effects combined). Furthermore, the amount of unique variance attributed to the quadratic effect was 2.9% and 1.7% for the state-Importance and state-Effort dimensions, respectively, on the basis of the respective squared semi-partial correlations (i.e., $-.17^2$ and $-.13^2$; see Table 2).$^5$

As can be seen in Figure 3, for both state-Importance T1 and state-Effort T1, the curvilinear effect was consistent with an essentially inverted U-shaped effect, with a steeper,

---

$^5$ A multiple regression was also performed with all four variables (Importance/Importance$^2$ and Effort/Effort$^2$) entered into the equation. However, none of the beta-weights were found to be significant statistically (e.g., state-Importance T1, $\beta = .17, p = .065$), although the model accounted for 6.2% of the variance in IQg scores, $F(4, 214) = 3.53, p = .008$. 


positive slope between low and moderate levels of test-taking Importance and Effort, in addition to a slight, negative slope from elevated to very elevated levels of test-taking state-Importance and state-Effort. Based on a supplementary analysis that divided the sample into relatively low and relatively high test-taking Importance (low = 1 to 2.99; high = 3.0 to 5.0), the correlation between test-taking Importance and intelligence test performance was $r = .24$ for the low Importance group ($N = 120$) and $r = -.02$ for the high Importance group ($N = 99$). Additionally, based on another supplementary analysis that divided the sample into relatively low and relatively high test-taking Effort (low = 1 to 3.49; high = 3.5 to 5.0), the correlation between test-taking Effort and intelligence test performance was $r = .24$ for the low Effort group ($N = 120$) and $r = -.02$ for the high Effort group ($N = 99$). Thus, the results supported the hypothesis of a curvilinear effect on intelligence test performance for both test-taking state-Importance and test-taking state-Effort.

**Discussion**

We found that, while the volunteers did not, on average, consider the intelligence testing important, they did, on average, report that they applied a minimally respectable amount of effort completing the tests. However, we found that their test-taking motivation reduced statistically significantly from cognitive ability testing earlier in the session to later in the session. Finally, we found the association between test-taking motivation and intelligence test performance was curvilinear, such that the positive association diminished beyond a self-reported test-taking motivation level of approximately 3.5 on the Student Opinion Scale.

**Degree of Test-Taking Motivation**

With respect to trait-based test-taking motivation, we found that the undergraduate university student sample reported a disposition toward viewing test taking as important. They
also tended to apply a non-negligible effort to complete the tests, as the trait-Importance and trait-Effort scale means were statistically significantly greater than the null specification of 3.5. Such results may be considered unremarkable, as the university students would be expected to have taken academic testing seriously in their academic career, in order to become selected into a respected university (i.e., above average conscientiousness; Poropat, 2009). Correspondingly, a non-negligible percentage of university students tend to report substantial levels of test anxiety (Gerwing, Rash, Gerwing, Bramble, & Landine, 2015), likely because they view the outcomes of testing as important. Thus, our trait test-taking motivation results are consistent with this literature.

More insightfully, and a novel aspect to this investigation, the trait test-taking scores helped contextualise the self-reported state levels of self-reported test-taking motivation. Specifically, in this investigation, the majority of the undergraduate volunteer sample (55%) did not view the intelligence testing as important, which may be considered congruent with the fact that there were no direct consequences to their performance (i.e., low-stakes). Correspondingly, the standardized difference between the trait-Importance and state-Importance means \((d = 1.55)\) may be described as large from an effect size perspective (Cohen, 1992).

Despite the fact that most of the examinees did not view the intelligence testing as important, an appreciable majority (74.4%) self-reported a state-Effort item mean score greater than 3.50, and only 5.9% of the sample reported a state-Effort item mean score of less than 3.0. Correspondingly, the undergraduate volunteers’ test-taking state-Effort at time 1 was not found to be statistically significantly lower than their reported trait-Effort mean. Such results imply that test-taking importance may not be a necessary condition for the application of test-taking effort, which is contrary to the expectancy-value theory of test-taking motivation (Wigfield & Eccles,
Importantly, it has been shown that trait conscientiousness correlates positively with test-taking Effort, as measured by the Student Opinion Scale (Demars, Bashkov, & Socha, 2013). Additionally, the observation in this investigation of significant, positive correlations between trait-Importance/state-Importance and between trait-Effort/state-Effort also supports the notion that test-taking motivation may be influenced by trait variance. Consequently, the incorporation of trait-theory into the expectancy-value model of test-taking motivation may be required to provide a more complete account of individual differences in test-taking motivation. In this investigation, however, neither trait-Importance nor trait-Effort correlated significantly with test performance. Thus, any effect of trait variance on state variance may be entirely indirect (i.e., via its influence on state-motivation).

In comparison to Gignac’s (2018) sample of undergraduate volunteers (N ≈ 50) who completed a similar battery of intelligence tests, our sample self-reported a greater level of test-taking Effort (3.82 vs. 3.40). Given our larger sample size (N = 219), we place greater confidence in the results reported in this investigation. Thelk et al. (2009) also reported a test-taking Effort mean of approximately 3.40 in a large sample of undergraduate students (N = 3111), however, the examinees were essentially required to take the cognitive ability/achievement testing as part of a university wide initiative. Thus, research scenarios that are not viewed as voluntary by the examinee (e.g., attendance taken as part of a class that includes the testing) may yield an appreciably larger percentage of examinees with invalid cognitive ability/achievement test scores, if maximum test-taking motivation is required.

As hypothesized, the degree of self-reported test-taking Effort decreased across the testing session. Specifically, test-taking Effort dropped from 3.82 to 3.63 (d = .29), which is between a small and medium sized effect (Cohen, 1992). Thus, our results are consistent with
Penk and Richter (2017) who observed a similarly sized decrease in test-taking Effort, amongst a group of ninth grade students required to complete a battery of academic achievement tests. Furthermore, in our investigation, the percentage of volunteers who self-reported test-taking Effort less than 3.0 increased non-negligibly (i.e., 6% to 13%) from the intelligence testing (first 25-30 minutes) to the reading comprehension testing (last 50 to 55 minutes). Given the first phase of the cognitive ability testing in this investigation was conducted within the first 30 minutes, and the second phase at approximately the 50 minute mark, it may be suggested that the validity of cognitive ability test scores may be compromised beyond the 30 minute testing mark in low-stakes settings. Correspondingly, previously reported large percentages of undergraduate volunteers (20%+) who have been found to have failed one or more neuropsychology testing validity indicators (e.g., An et al., 2012) may be due to the significant testing time required to complete those batteries (often 2 to 3 hours). Thus, for research purposes that require general intelligence tests scores, a maximum of 20 to 30 minutes of total cognitive ability testing time may be considered advisable.

It is possible that cognitive ability tests could be developed in such a way as to increase engagement on the part of the research volunteer. Based on Lumsden et al.’s (2016) review, gamified versions of cognitive ability tests (e.g., Whack-a-Mole) can correlate respectably (≈ .60) with their non-gamelike counterparts (e.g., executive function tests such as Inhibition). Furthermore, people who complete gamified cognitive ability tests have been found to report greater levels of test-taking motivation, in comparison to completing the non-gamelike counterparts (Lumsden et al., 2016). However, meaningful effects may be essentially circumscribed to samples of people with difficulties of attention (e.g. children with ADHD; Dovis, van der Oord, Wiers & Prins, 2012). Importantly, based on an experiment with 127
healthy first-year psychology students, Hawkins et al. (2013) found that a gamelike version of an
adaptive binary decision making task increased self-reported enjoyment with the testing \((d = .43)\), however, there was no concomitant change in test performance \((d \approx .00)\). Such a result is
consistent with Gignac (2018) who, in another experiment, found that a financial incentive
increased self-reported test-taking effort, but not intelligence test performance. Thus, participant
enjoyment and/or motivation with testing may be amenable to increase. However, there does not
appear to be a corresponding change in test performance, at least not in healthy adult volunteers.

**Correlation with IQ Test Scores**

Consistent with several other investigations with students who completed academic
achievement tests (Hopfenbeck & Kjaernsli, 2016; Schiel, 1996; Sundre & Kitsantas 2004), as
well as Gignac (2018) with undergraduate university psychology volunteers, we found a
moderately sized \((\approx .25)\) positive correlation between self-reported test-taking motivation
(Importance and Effort) and intelligence test performance. On the basis of such a positive
association between test-taking motivation and intelligence, some researchers have contended
that cognitive ability-type test scores are confounded appreciably by individual differences in
test-taking motivation (Duckworth et al., 2011; Wise, 2009). Logically, such a contention
implies that test-taking motivation influences test performance, which, at least superficially,
seems plausible. If such a contention were established, researchers and clinical practitioners who
use intelligence tests in their low-stakes assessment work would be encouraged to take test-
taking motivation into consideration.

However, in this investigation, the nature of the association between test-taking
motivation and intelligence test performance was found to be curvilinear, such that the effect was
essentially restricted exclusively to the low to moderate (neutral) levels of test-taking motivation.
Such a result is consistent with Gignac and Wong (2018) who also found a similar curvilinear effect between test-taking persistence, measured behaviourally, and general intelligence performance. Thus, beyond a moderate level of test-taking motivation, the prospects of increasing maximal cognitive ability test performance via enhancements of test-taking motivation may be limited. In simple terms, the difference between ‘giving it a shot’ and trying one’s absolute best may be negligible, at least with respect to cognitive ability test performance. The establishment of such a position with further empirical work would challenge the long held notion that maximum performance tests require maximal effort (Cronbach, 1960; Roid & Barram, 2004; Revelle, 1993). Instead, a moderate effort may suffice. It is acknowledged that the development of theoretical work would be valuable to help support such an effect. Currently, there does not appear to be an existing theory of motivation to explain the phenomenon, unless test-taking effort is confounded with arousal (see Humphreys & Revelle, 1984).

We also note that the curvilinear effect may not be monotonic (see Jacoby, 2000). Instead, on the basis of a loess regression (uniform weighting; window = 50%), there was some evidence to suggest that the association was non-monotonic, with a trend toward decreasing intelligence test performance at very elevated levels of self-reported test-taking effort (see Figure S1 in the supplementary materials). With a much larger sample size, the possibility of a non-monotonic, curvilinear effect may be investigated with a more sophisticated procedure (e.g., Hoeffding analysis). If such a non-monotonic, curvilinear effect were established, it would suggest that both effort and arousal may have similar non-monotonic effects on performance (see Humphreys & Revelle, 1994).

Although it may seem surprising to suggest that a cognitive ability test may not require a maximal effort to yield an essentially valid score, other putatively considered maximal
performance tests may also only require a moderate effort to yield an essentially valid score. For example, Martin, Thompson, Keegan, Ball, and Rattray (2015) failed to observe a statistically significant effect of moderate mental fatigue (manipulated experimentally) on performance across a series of anaerobic exercises (e.g., leg extensions; see also van Cutsem et al., 2017). We also note that on the basis of an experimental investigation with a monetary incentive, Gignac (2018) found that test-taking motivation could be increased somewhat in first-year undergraduate psychology student volunteers, however, a corresponding statistically significant increase in cognitive ability test performance was not observed, a null finding consistent with the experimental research in industrial organisational psychology (Bonner & Sprinkle, 2000). Such a pattern of results supports the possibility that the direction of causality may lead from cognitive ability to test-taking motivation: a plausible suggestion, given the findings in the area of test-taking anxiety and intelligence test performance (Sommer & Arendasy, 2014, 2015; Wicherts & Scholten, 2010).

**Practical Implications**

Taking the results of this investigation and those of Gignac (2018), as well as the neuropsychological testing research with validity index tests (e.g., DeRight & Jorgensen, 2015; Ross et al., 2016; An, Kaploun, Erdodi & Abeare, 2017), it may be suggested that between 5% and 12% of undergraduate university volunteers likely provide an effort insufficient to yield valid cognitive ability test scores. Simulation research has shown that 10% of invalid responses in a sample of data is sufficient to seriously distort the results (Woods, 2006). Thus, even though the mean change in test-taking Effort was not alarming from time 1 to time 2 in this investigation (i.e., the mean at time 2 remained above 3.5), arguably, the more than doubling of the percentage of people (i.e., 6% to 13%) who scored less than 3.0 on the test-taking Effort subscale should
probably be considered a serious concern. Consequently, test-taking motivation should probably be taken into consideration in low-stakes intelligence research, in order to identify those clearly insufficiently motivated to complete the testing in a valid manner.

Any demarcation criterion specification for sufficient motivation will be, to some degree, arbitrary. However, taking into consideration the nature of the nonlinear association between test-taking motivation and intelligence test performance uncovered in this investigation, as well as common sense (i.e., someone should not be less than neutral about the amount of motivation they applied) a value somewhere between 3.0 and 3.5 on a 5-point Likert scale (e.g., Student Opinion Scale) may be considered a candidate criterion. Of course, more research is required to confirm the nonlinear, non-monotonic association between test-taking motivation and intelligence test performance reported in this investigation. Thus, to an appreciable degree, the results of this investigation should be regarded as exploratory. If the effect were to be established, in practice, a person’s intelligence score could be adjusted on the basis of their motivation. Alternatively, a clearly unmotivated case could be potentially excluded from the data. However, discarding data for any reason needs to be considered seriously, as the implications can have complicated consequences to the nature of the results (Rios, Guo, Mao & Liu, 2017).

**Limitations**

The sample size was broadly adequate for the types of analyses carried out in this investigation, as a sample size of 219 is associated with power of .91 to detect a correlation of .20 as significant ($p < .05$). However, with a larger sample size (> 500), the possibility of testing the curvilinear effects within a latent variable modeling framework would be advantageous, particularly considering that measurement error is more substantial near the ends of a continuum.
of psychometric scores (Dudek, 1979). Moreover, it remains a possibility that the decelerating rate of the positive slope may be a statistical artefact due to very low levels of internal consistency reliability at the high end of the general intelligence test scores. Additionally, a larger sample size would afford testing some interesting cross-lagged panel models and/or latent growth curve models to investigate the process of test-taking motivation across time.

A larger sample size from a more representative sample may also provide a clearer estimation of the curvilinear effect, as university students may be regarded to possess relatively elevated levels of conscientiousness, a known positive correlate with test-taking motivation (DeMars et al., 2013). That is, relatively few undergraduate volunteers scored less than 3.0 on the test-taking motivation Effort subscale. Thus, the prospect of obtaining larger numbers of people with self-reported test-taking Effort scores less than 3.0 would help fill the gap in the lower-left quadrant of the scatter plot in Figure 3 (panel B). Unfortunately, volunteers with very little motivation to complete intelligence testing may not turn up to a testing session at all, thus, sample sizes in the thousands may be required to help uncover precisely the nature of the association between test-taking motivation and intelligence test performance. Correspondingly, it should be acknowledged that these results may not generalise to people with less experience taking tests (e.g., typical high-school or primary school students), as the current sample was based on undergraduate university students from a selective university (i.e., students with substantial experience taking tests).

Additionally, the trait version of the Student Opinion Scale we created and administered in this investigation did not specifically direct the examinees to consider a broad context when responding to the items. Thus, it is possible that some of the student volunteers considered their academic testing experience, when responding to the items, while others may have considered
both their academic testing experience, as well as their research volunteer experience. Such possible individual differences in consideration may have contributed to the relatively low Student Opinion Scale trait-based score internal consistency reliability (especially trait-Importance), as well as the relatively low observed score correlations ($r \approx .20$ to $.30$) between the trait and state scores. Finally, we note that our results may not extend to purely online data collection efforts, as our data were collected in a laboratory.

**Conclusions**

It has long been assumed that examinees must provide their maximum effort when completing cognitive ability tests, in order to allow for valid interpretations of the scores (Cronbach, 1960; Roid & Barram, 2004; Revelle, 1993). However, the establishment of a curvilinear association between test-taking effort and intelligence test performance, as reported in this investigation, may undermine such a long-held assumption.
References


Borghans, L., Meijers, H., & Ter Weel, B. (2008). The role of noncognitive skills in explaining


determinants of cognitive performance. In A. D. Baddeley & A. D. Weiskrantz (Eds.),
Attention: Selection, awareness, and control: A tribute to Donald Broadbent (pp. 346-


on aggregated-scores: To filter unmotivated examinees or not?. International Journal of
Testing, 17(1), 74-104.

(SB5) Assessment. Hoboken, NJ: Wiley & Sons

General Psychology, 2, 247-270.

Performance invalidity base rates among healthy undergraduate research
participants. Archives of Clinical Neuropsychology, 31(1), 97-104.

Effects of aging on efficiency of task switching in a variant of the trail making test.

Schiel, J. (1996). Student Effort and Performance on a Measure of Postsecondary

Technical and legal guidelines for corporate training (3rd ed.). San Francisco, CA: John
Wiley & Sons, Inc.


Table 1

*Test-Taking Motivation Descriptives and One-Sample t-tests (Null Mean 3.5)*

<table>
<thead>
<tr>
<th></th>
<th>Descriptives</th>
<th>One-Sample t-Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Importance</td>
<td>Effort</td>
</tr>
<tr>
<td></td>
<td>$M$ (SD)</td>
<td>$M$ (SD)</td>
</tr>
<tr>
<td>Trait</td>
<td>4.19 (.45)</td>
<td>3.83 (.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Time 1</td>
<td>3.39 (.58)</td>
<td>3.82 (.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Time 2</td>
<td>3.17 (.77)</td>
<td>3.63 (.69)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $N = 219$; $d =$ Cohen’s $d$; the one-sample $t$-tests were conducted with a null value of 3.50.
Table 2

*Descriptive Statistics and Pearson Correlations (95% Confidence Intervals Above Main Diagonal)*

<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trait-Importance</td>
<td>-</td>
<td>.18/.42</td>
<td>.07/.35</td>
<td>-.10/.17</td>
<td>.01/.27</td>
<td>-.07/.17</td>
<td>-.06/.16</td>
<td>4.19</td>
<td>.45</td>
<td>-.46</td>
</tr>
<tr>
<td>2. Trait-Effort</td>
<td>.31</td>
<td>-</td>
<td>.07/.35</td>
<td>.13/.44</td>
<td>-.04/.28</td>
<td>.06/.37</td>
<td>-.21/.02</td>
<td>3.83</td>
<td>.53</td>
<td>-.28</td>
</tr>
<tr>
<td>4. State-Effort - Time 1</td>
<td>.03</td>
<td>.29</td>
<td>.38</td>
<td>-</td>
<td>.19/.44</td>
<td>.49/.71</td>
<td>.08/.32</td>
<td>3.82</td>
<td>.60</td>
<td>-.38</td>
</tr>
<tr>
<td>5. State-Importance - Time 2</td>
<td>.15</td>
<td>.12</td>
<td>.67</td>
<td>.32</td>
<td>-</td>
<td>.50/.70</td>
<td>-.01/.27</td>
<td>3.17</td>
<td>.77</td>
<td>-.32</td>
</tr>
<tr>
<td>6. State-Effort - Time 2</td>
<td>.05</td>
<td>.22</td>
<td>.37</td>
<td>.61</td>
<td>.61</td>
<td>-</td>
<td>.07/.36</td>
<td>3.63</td>
<td>.69</td>
<td>-.73</td>
</tr>
<tr>
<td>7. IQg</td>
<td>.05</td>
<td>-.10</td>
<td>.18</td>
<td>.20</td>
<td>.13</td>
<td>.21</td>
<td>-</td>
<td>.00</td>
<td>1.00</td>
<td>-.16</td>
</tr>
</tbody>
</table>

*Note. N = 219; correlations in bold were significant statistically via bootstrapping (p < .05); IQg = general intelligence test performance.*
Table 3

*Curvilinear Multiple Regressions: Test-Taking Motivation (Importance and Effort) Predicting General Intelligence Test Performance*

<table>
<thead>
<tr>
<th></th>
<th>(b) (95% CI)</th>
<th>(p)</th>
<th>(\beta)</th>
<th>(r)semi-partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>2.47 (.48/4.00)</td>
<td>.004</td>
<td>1.43</td>
<td>.19</td>
</tr>
<tr>
<td>Importance(^2)</td>
<td>-.33 (-.58/-0.02)</td>
<td>.011</td>
<td>-1.27</td>
<td>-.17</td>
</tr>
<tr>
<td>Effort</td>
<td>2.17 (.65/4.36)</td>
<td>.001</td>
<td>1.29</td>
<td>.15</td>
</tr>
<tr>
<td>Effort(^2)</td>
<td>-.25 (-.47/-0.08)</td>
<td>.018</td>
<td>-1.10</td>
<td>-.13</td>
</tr>
</tbody>
</table>

\(\alpha = -4.52; R^2 = .059; \text{Model} \, F(2, 216) = 6.78, \, p = .001\)

\(\alpha = -4.59; R^2 = .055; \text{Model} \, F(2, 216) = 6.31, \, p = .002\)

*Note.* \(N = 219\); dependent variable = IQg; \(\alpha\) = intercept; \(b\) = unstandardized beta-weight; \(\beta\) = standardized beta-weight; the Importance\(^2\) and Effort\(^2\) terms in the models represent the quadratic curvilinear effect; confidence intervals and \(p\)-values estimated via bootstrapping.
Figure 1. Violin plots with means (dark grey rectangles)
Figure 2. Plot of test-taking motivation means (error-bars are 95% confidence intervals); Trait = trait test-taking motivation; State 1 = state test-taking motivation (time 1); State 2 = test-taking motivation (time 2); the dotted line represents the test-taking Importance within-subjects factor; the solid line represents the test-taking Effort within-subjects factor.
Figure 3. Scatter plots depicting the association between test-taking motivation and intelligence test performance; Importance = self-reported test-taking importance; Effort = self-reported test-taking effort; IQg = general intelligence test performance.
Appendix A

Student Option Scale – Trait Version

Please respond to the items below on the response scale provided:

<table>
<thead>
<tr>
<th>Item #</th>
<th>Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Doing well on tests is important to me.</td>
</tr>
<tr>
<td>2.</td>
<td>I engage in a good effort when tested.</td>
</tr>
<tr>
<td>3.</td>
<td>I am not curious about how I do on tests relative to others.*</td>
</tr>
<tr>
<td>4.</td>
<td>I am not concerned about the scores I receive on tests.*</td>
</tr>
<tr>
<td>5.</td>
<td>Tests are important to me.</td>
</tr>
<tr>
<td>6.</td>
<td>I give my best effort on tests.</td>
</tr>
<tr>
<td>7.</td>
<td>While taking examinations, I could work harder on them.*</td>
</tr>
<tr>
<td>8.</td>
<td>I like to know how well I do on tests.</td>
</tr>
<tr>
<td>9.</td>
<td>I usually do not give tests my full attention while completing them.*</td>
</tr>
<tr>
<td>10.</td>
<td>While taking tests, I am able to persist to completion of the tasks.</td>
</tr>
</tbody>
</table>

Note. The Student Opinion Scale (Trait Version) was adapted from the Sundre (1999) Student Opinion Scale (State Version); * negatively keyed item; response scale: Strongly Disagree = 1; Disagree = 2; Neutral = 3; Agree = 4; Strongly Agree = 5; test-taking importance subscale = item 1, item 3, item 4, item 5, item 8; test-taking effort subscale = item 2, item 6, item 7, item 9, item 10.