

Research Articles

Is Trying Harder Enough? Causal Analysis of the Effort-IQ Relationship Suggests Not

Timothy Bates, Ph.D.¹ 

¹ Psychology, University of Edinburgh

Keywords: intelligence, IQ, motivation, effort, incentive

<https://doi.org/10.65550/001c.142071>

Intelligence & Cognitive Abilities

Vol. 1, Issue 1, 2025

Claims that effort increases cognitive scores are now under great doubt. What is needed is randomized controlled trials optimized for testing causal influence and avoiding confounding of self-evaluation of performance with feelings of good effort. Here we report three large studies using unconfounded measures of effort and instrumental analysis to isolate any causal effect of effort on cognitive score. An initial study ($N = 393$) validated an appropriate effort measure, demonstrating excellent external and convergent validity ($\beta = .61$). Study 2 ($N = 500$, preregistered) randomly allocated subjects to a performance incentive, using an instrumental variable analysis to detect causal effects of effort. The incentive successfully manipulated effort ($\beta = .18$, $p = .001$). However, the causal effect of effort on scores was near-zero and non-significant ($\beta = .04$, $p = .886$). Study 3 ($N = 1,237$) replicated this null result with preregistered analysis and an externally developed measure of effort: incentive again raised reported effort ($\beta = .17$, $p < .001$), but effort had no significant causal effect on cognitive score ($\beta = .27 [-0.07, 0.62]$), $p = .15$). Alongside evidence of research fraud and confounding in earlier studies, the present evidence for the absence of any causal effects of effort on cognitive scores, effort research should shift its focus to goal setting – where effort is useful – rather than raising basic ability, which it appears unable to do.

Introduction

Strong claims that effort is a major cause of differences in cognitive ability scores (Duckworth et al., 2011) have recently been cast in doubt (Duckworth et al., 2025), and, now, retracted (“Retraction for Duckworth et al., Role of Test Motivation in Intelligence Testing,” 2025). The authors note in this retraction notice that “*Adjusting for publication bias using trim-and-fill reduces the estimated effect size to a more plausible effect size ($g = 0.16$) but now, with this reduced sample size, the confidence interval for the publication-bias-adjusted estimate includes zero (i.e., the publication-bias-corrected estimate no longer achieves statistical significance)*”. The claim that effort raises ability scores has substantial implications for theories of ability and motivation, and, importantly, for educational policy and the validity of cognitive testing (Bates & Gignac, 2022; Inzlicht et al., 2021; Warne, 2022). The purpose of the present paper was to resolve the question of whether effort exerts any causal influence on cognitive ability scores. We do so by using an instrumental variable approach, allowing randomized controlled causal test of effects of effort on ability, with effects of effort assessed directly in terms of changes in effort, and replicating this causal design in two large-scale tests of effort manipulated via incentives. We first background the literature.

Test-taking effort is typically operationalised in scales such as the Student Opinion Scale (SOS: Sundre & Thelk, 2007), e.g., “*I engaged in good effort throughout these tests*”. In observational studies, such post-test measures of effort correlate around .28 with test scores (Silin et al., 2020). This sets an upper bound on the causal effect of effort (Bates & Gignac, 2022), but is confounded by feelings of efficacy developed during task performance. This confound is important, because self-appraisal very strongly reflects objective performance (Hart & Staveland, 1988), i.e., “*good effort*” essentially means “*expecting a good score*”.

A meta-analysis of studies manipulating small monetary incentives concluded effort lifts test scores by a very large 0.64 SDs (Duckworth et al., 2011) – an exceptional effect in social research and over double the limit set by the cross-sectional association.

However, the studies included in this meta-analysis suffered from significant limitations, including low statistical power and compromised designs. Critically, some of the largest effect sizes were from fraudulent research that was later retracted (Warne, 2022). Recognizing these issues, the meta-analysis authors themselves issued a note of concern, acknowledging that withdrawing this dubious data greatly shrinks the overall effect (Duckworth et al., 2025).

Recent large within-subjects repeat-measures randomised control tests of monetary incentive effects have yielded much smaller effects ($d = .17$) and no evidence of

dose dependency (Bates & Gignac, 2022). However most or even all this modest association of effort with cognitive performance may reflect reverse causation. Note, reverse causation occurs when participants infer their level of effort from their perceived success, leading higher cognitive performance to cause reports of greater effort, rather than the other way around. Contrarily, it is possible in existing designs to suggest that small apparent effects of effort in fact mask strong effects of effort due to weak coupling of the incentive to effort. This can be remedied by an instrumental modelling approach, which distinguishes between the effect of the randomized instrument on the proposed causal variable and the effect of this causal variable on the outcome, allowing the effect to be measured in units of the variable of interest. The instrumental variable method is explained in more detail in Study 2 (see also Fig 2). Very briefly, this method enables causal analysis of the effects of variables such as effort which the researcher cannot directly vary, but instead are varied indirectly by an instrument (in this using incentive as an instrument to manipulate effort (Bates & Gignac, 2022; Duckworth et al., 2011).

Current data, then, are compatible with one or more of a) effort causally increases cognitive ability; b) cognitive ability increases effort (i.e., a rational choice to invest); and c) complete reverse causation. Note, reverse causation occurs when higher cognitive performance leads participants to report greater effort, as they infer effort from their perceived success, rather than effort causing better performance.

What is needed to test the prediction that effort does not causally raise cognitive ability is a method capable of isolating causal effects of effort on cognitive performance. Doing this requires 1) A statistically significant experimental manipulation of effort via incentive; 2) Randomized allocation of subjects to this incentive; 3) A prospective measure of effort on the cognitive task; and 4) Modelling effects of effort in a way capable of estimating the causal effect of effort on scores (Sanderson et al., 2022). We do this in Study 2 with a pre-test measure of effort, analysing the “effort score” causal path within an instrumental variable design not susceptible to confounding. We begin by validating a prospective measure of effort (study 1). Study 2 tests the causal effect of effort in a large study, and Study 3 replicates study 2 using larger N, pre-registered analysis, and an independent measure of effort.

Study 1

The conventional framework for effort research is task motivation theory. Task motivation is understood as task value $\times$ expectancy of success (Eccles et al., 1983; Eccles & Wigfield, 2023). In this formulation, increased task value can increase effort, contingent on prospects of success. Goal setting theory (Locke & Latham, 2002) further predicts that after completion of a motivated task, subjects evaluate their performance, using this to update self-efficacy. Therefore incentive can increase effort, but self-efficacy may increase effort (investment), and, moreover, if effort is measured after performance, reports of effort are

contaminated by subjective awareness of performance, often to very substantial degree (Dong et al., 2013).

Awareness of problems with post-task reports of effort has lead motivation researchers to suggest that it is critical to use pre-task measures of effort in order to avoid confounding (Weiner, 2000). Thus both theory and practice imply that to maximise validity and avoid confounding, effort should be measured prospectively (Myers & Finney, 2019). Despite this most research has used measured effort post-test. This is a critical distinction, as the gap between confounded correlational findings and causally-identified effects can be stark. For instance, while observational studies of financial literacy report a correlation of around $\beta = .15$ with financial behavior, large-scale causal interventions often find zero effect (Fernandes et al., 2014). Prospective measures of effort, then, are recommended, and can avoid confounding of subjective attributions with information about performance and allows us to randomly vary incentives and resolve the resulting unambiguous measures using instrumental variable models to estimate the causal effects of effort.

A large body of work suggests that assessing effort prior to performance can remove the confound of post-hoc performance evaluation while retaining validity. The validity of effort measures both prior to and during testing has been compared. In a large representative sample of German ninth-graders (Penk & Schipolowski, 2015) found that a pre-test and post-test assessment correlated highly. Moreover, disclosure of intended effort induces commitment (Carver & Scheier, 1998), and increases test performance (Myers & Finney, 2019), increasing the efficacy of effort. Pre-test effort measures, then, avoid confounds while retaining validity. Jointly, this information implies that, to avoid potential confounding, effort can and should be measured prospectively (Penk & Richter, 2017) to protect against mistaken models of causation (Myers & Finney, 2019).

We set out, therefore, to validate a prospective effort scale which correlated strongly with the post-test Student Opinion Scale (SOS: Sundre & Thelk, 2007) – the most widely validated and used instrument (Silm et al., 2020). The SOS is completed after cognitive testing and has separate scales the importance with which the subject views their score and the effort they exerted. As noted by Myers and Finney (2019), the SOS lends itself straightforwardly to translation into a prospective measure. For study 1, therefore, we built a pre-test version of the SOS re-wording items to measure prospective effort.

We had four predictions. First, we predicted that pre-test SOS would correlate very strongly with post-test SOS, showing face validity of the new instrument. Second, we predicted that the pre-test SOS would show external validity, correlating with ratings of satisfactory performance as a subject, and number of projects completed on prolific (our subject recruitment service). Third we assumed that pre-test SOS would correlate significantly with cognitive ability scores, and at around the level shown for the SOS. Fourth, in-line with the confounding hypothesis, we predicted that, controlling for post-test score, pre-test effort would fail to

predict cognitive score as post-test score could control for effort. Study 1 hypotheses, methods, and participant numbers were pre-registered ([aspredicted.org](https://aspredicted.org/#103900) #103900).

Method

Subjects

Subjects were recruited from Prolific Academic, a high quality online service for recruiting participants and payment handling for participation (Douglas et al., 2023). We recruited 393 adult subjects (age $M = 44.45$, $SD = 13.83$, 198 male and 195 female. Ethics approval was granted by the University of Edinburgh Psychology ethics committee (#4-2122/2).

Materials

Post-test effort was assessed using the SOS. An example item is “I gave my best effort on these tests.”. Subjects respond on a 5-point Likert scale with anchors from ‘Strongly Disagree’ to ‘Strongly Agree’. ω -t in our sample was 0.72. Note, only the effort scale was scored and used here, but to preserve the normal format for this test, all 10 items of the SOS were administered, i.e., including items assessing how important the tests were to the participant.

Pre-test effort was measured by modifying the item tense of the SOS items. Effort items were:

1. I will engage in good effort throughout this test.
2. I will give my best effort on this test.
3. If I had to, I likely could work harder this test than I will. (R)
4. I will not give this test my full attention while completing it. (R)
5. I will persist to completion of this test.

This scale proved reliable in the present sample (Omega- ω -t = 0.78).

The response range of the pre-test effort items was extended from 1-5 to a 1-7 to increase precision. Anchors were 1: “Disagree Strongly” to 7: “Agree Strongly”. ω -t reliability was 0.78. As for the SOS, modified importance items administered but are not analysed.

Cognitive ability was assessed using the 32-item Form B of the Baddeley (1968) sentence verification test. Items require grammatical transformation to evaluate the truth of a simple sentence, e.g. “A does not follow B: **AB**” (TRUE). Subjects had 90 seconds to complete as many items as possible. Coefficient ω in our sample was 0.91.

Procedure

Testing took place online using the Prolific academic and Qualtrics platforms. Testing took around 5 minutes. After providing informed consent, subjects completed the pre-test effort scale, followed by the first ability test followed by the SOS measure of effort and the rating of test-difficulty. Finally, subjects estimated the difficulty of the cognitive test items on a 5-option scale from “very easy” to “very difficult”. As part of pilot collection for research on alter-

Table 1. Descriptive statistics and correlations (and SE) of the key variables in Study 1.

	Mean (SD)	Pre-effort	Baddeley B
Pre-effort	21.99 (2.31)		
Baddeley B	16.82 (6.63)	0.03 (0.05)	
SOS effort	21.6 (2.62)	0.61 (0.03)	0.24 (0.05)

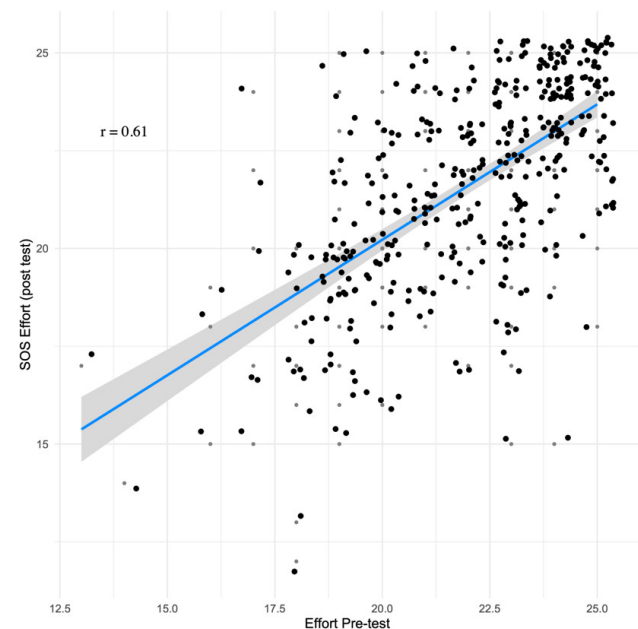


Figure 1. Validation of pre-test measure of effort against SOS.

native measurement formats this was followed by 3 slider-response items asking subjects to specify their motivation, intended effort and attention (not analysed here).

Results

All analyses were conducted in R (R Development Core Team, 2024) and *umx* (Bates et al., 2019). Descriptive statistics are given in Table 1.

The prediction that pre-test SOS would correlate very strongly with post-test SOS effort was tested by regressing pre-test effort on post-test SOS effort. This revealed a strong relationship between the two measures ($\beta = .61$ [0.531, 0.688], $t = 15.25$, $p < 0.001$), showing excellent convergent validity of the new instrument (see Fig 1). We bolstered this by using the pre-test effort measure to predict two independent quality measures available from Prolific academic: Subject quality score (number of strikes against test-taker in earlier participations) and number of approvals of participation. Both were significantly associated with our new effort measure ($\beta = .16$ [0.06, 0.26], $t = 3.24$, $p = .001$ and $\beta = .14$ [.05, .24], $t = 2.87$, $p = .004$ respectively). This real-world effort supports the external validity of the new instrument.

We had assumed that pre-test SOS would correlate significantly with cognitive ability scores. Strikingly, and in contrast to the post-test measure, a regression predicting

cognitive score from pre-test effort was not statistically significant ($\beta = .03$ [-0.07, 0.13], $t = .64$, $p = .524$). Controlling for age and sex made no material difference. By contrast, in the same data, and despite the strong correlation of the pre and post effort measures, the post-test SOS measure was significantly related to test score ($\beta = .24$ [0.14, 0.33], $t = 4.86$, $p < 0.001$). These last two results are, perhaps, a harbinger of a lack of causal effect of effort on cognitive ability test performance.

Discussion

Four findings were of interest. The pre-completion effort measure was reliable and correlated strongly with SOS post-test effort ($r = .61$). The pre-test measure of effort was also validated by independent objective indicators of effort in the form of data quality and an independent measure of workload ($r = .14$ & $.16$). Finally, the pre-test measure of effort correlated with cognitive scores only weakly and non-significantly, compared to the confounded measure of effort recorded after the subjects had completed the task.

This non-significant correlation suggests that the prospective effort measure, unlike post-test scales, may be free from performance-driven confounds, hinting at a potential lack of causal influence of effort on scores. We next moved to study 2, where SOS_{pre} could be used to test the causal effect of effort on cognitive score in an instrumental variable framework targeting the question of causation for effort increasing cognitive scores.

Study Two

Having developed the prospective effort measure in study 1, in study 2 we use it to test the causal effect of effort on cognitive score. We begin by briefly introducing the instrumental variables framework, within which RCTs exist as a conceptual subset (Sanderson et al., 2022).

Because effort cannot be directly varied, researchers rely on manipulating effort by varying incentive (Bates & Gignac, 2022; Duckworth et al., 2011). However, the key independent variable of interest is effort, not incentive, as it is effort which is proposed to exert a causal effect on cognitive ability. This problem of manipulating an inaccessible parameter via a variable under control of the researcher is a common one, and the best-developed approach is the Instrumental Variable (IV) framework – a class of statistical model estimating the causal effect of a putative causal variable (in our case effort) on an outcome (in our case ability score), using a third “instrumental” variable (in our case incentive) to perturb the causal variable at random with respect to confounders (Wright, 1928). This approach is drawn as a path model testing causality in Fig 2.

As can be seen (Fig 2), the instrumental variable affects the putative causal variable but 1) must have no direct effect on the outcome, and 2) is not under control of any of the confounders (which in this case is assured because the incentive is set at random by the researchers). By using the instrumental variable, researchers can estimate the effect of the independent variable on the dependent variable, in units of the causal variable, and while still modelling un-

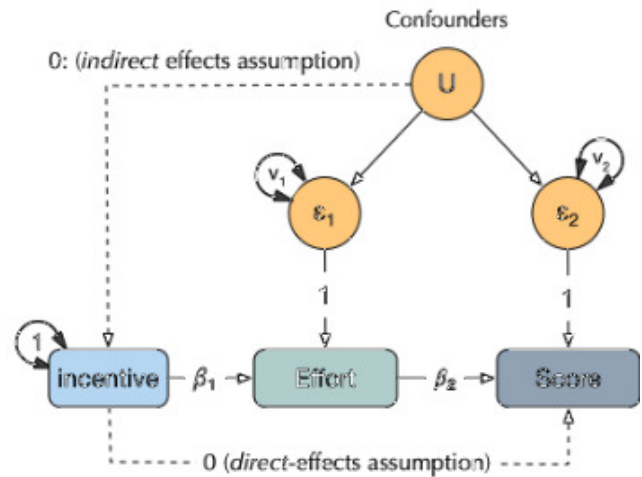


Figure 2. Randomized controlled trial within an instrumental variable framework (modified from Bates et al., 2019).

measured confounders (U). These properties make it an attractive and widely deployed design.

A key strength of the IV design is its ability to detect a large underlying causal effect even if the instrument (incentive) has only a modest impact on the causal variable (effort). Therefore, if effort does causally improve scores as claimed by Duckworth et al. (2011, p. 7717) the estimated effect of effort on performance (β_2 in Fig 2) can be large.

Here, we manipulate effort with a £2 incentive to improve scores by 1 item or more correct compared to an initial baseline. If the design is valid, then subjects in the incentive condition will show higher effort scores (i.e., significant β_1 , validating the incentive). The effort theory of cognitive ability predicts that effort will have a significant and substantial effect on cognitive score (i.e., significant β_2). Based on funding, we ran 500 subjects.

Method

Sample

Five hundred adult subjects (age $M = 39.5$, $SD = 13.25$) were recruited on Prolific Academic (248 male and 252 female). The sample was predominantly white (Asian 23; Black 17; Other 8; Southeast Asian 2; White 450). Subjects were paid £1.35 for their participation (excluding £2 bonus).

Materials

Intelligence was measured using parallel forms of the Visual Paper Folding test (Ekstrom et al., 1976). Each form included 10-items completed with a 3-minute time limit and the forms are designed to be equally difficult. For further details, see Study 1. For the total sample, reliability (ω -t) was estimated at .73 and .67 for Forms A and B respectively. Effort was measured using the pre-test measure of effort developed in Study 1. Internal consistency ω -t reliability was .75 and .78 for the two effort measures. After each

Table 2. Mean and SD and intercorrelations among key variables in Study 2.

	Mean (SD)	1	2	3	4	5
1 Paper Fold A	4.79 (2.27)					
2 Paper Fold B	5.21 (2.06)	0.62				
3 SOS _{Pre} 1	21.57 (2.31)	-0.05	0.01			
4 SOS _{Pre} 2	21.47 (2.54)	-0.01	0.09	0.70		
5 SOS 1	21.13 (2.96)	0.11	0.14	0.60	0.73	
6 SOS 2	21.45 (2.96)	0.06	0.19	0.58	0.74	0.73

test, subjects also completed the 10-item Sundre and Thelk (2007) Student Opinion Scale. Reliability was $\omega = .80$ and $.81$ for the two SOS tests.

Procedure

Subjects completed testing online. After consenting, subjects completed Form A of the paper folding test, followed by the completion of the effort questionnaire. Next, half the subjects at random were selected for the motivation condition, receiving the message, “*In this second part of the study, you will be given the opportunity to earn an additional £2 if you can improve your test performance by one or more items. The test is very similar to the first test you completed. You will also have the same amount of time to complete the test. We will score both tests and if your score is higher on the second test, you will be awarded a bonus £2 pounds.*” An attention check (identifying if the incentive was nothing, £2, or 10 pence) had to be passed before proceeding to complete the prospective effort measure, followed by the second cognitive test, followed by the SOS. All participants (100%) passed the attention check, confirming awareness of the incentive condition.

Results

For descriptive statistics, see Table 1. Before continuing to the causal modelling, we began by simply examining in linear models the association of effort with cognitive scores, with effort assessed post-test-performance using the SOS.

For both the first and second block of cognitive assessment, SOS effort was associated with cognitive score ($\beta = .11$ [0.03, 0.2], $t = 2.55$, $p = .011$ for block 1, and 0.19 [0.1, 0.28], $t = 4.35$, $p < 0.001$ for block 2). However, for the pre-test measures of effort intention, this association was absent, with a non-significant association in the wrong direction for block 1 ($\beta = -0.05$ [-0.14, 0.04], $t = -1.12$, $p = .262$) for block 1 and non-significant for block 2 ($\beta = .09$ [0, 0.17], $t = 1.92$, $p = .055$).

We evaluated the core hypothesis that effort causally increases cognitive ability using the model shown in Fig 2. The two relevant parameters are β_1 and β_2 – the paths “*incentive effort*” and “*effort score*” respectively (see Table 2).

Validating the instrument (incentive) as having a strong and statistically significant increasing effect on effort, β_1 was estimated at 0.18 and this was highly significant ($\chi^2(1)$

$= 10.80$, $p = .001$). The effect of effort on scores, however, was estimated near-zero ($\beta_2 = .04$) and could be fixed at zero no statistically significant loss of fit ($\chi^2(1) = .02$, $p < .886$). This simpler model also showed the lowest AIC (see Table 2). A model with no effect of effort on cognitive scores was therefore the preferred model (see Figure 3 for a graphical version of this final model).

As we had administered the SOS, we could also apply this causal model to effort assessed using the SOS. Interestingly, this model supported the same conclusions: Incentive increased reported SOS effort ($\chi^2(1) = 16.12$, $p < 0.001$), but the path from SOS effort to cognitive score could be dropped without significant loss of fit ($\chi^2(1) = .02$, $p = .89$). The bivariate confounder path (ϵ in effort with ϵ in cognitive scores) was estimated at .19. This causal model for the SOS data indicates two things. First, there is zero causal effect of effort on scores. Second, the confounder path captures the reverse-causal association of post-test effort with performance as a strong cognitive performance influences a participant’s subsequent report of their effort.

Discussion

Study 2 showed a significant effect of the monetary incentive increasing effort. However, no statistically significant effect of effort on cognitive score was seen (see Figure 3). The results, then, disconfirm the claim that effort causally raises cognitive scores. This was also supported by the analysis of the confounded post-test SOS measure, which showed no evidence of causal impact for this measure of effort. It is worth noting also, that, in line with Bates and Gignac (2022), the *differences between scores on Forms A and B of the paper folding test suggests a small practice effect*.

The analysis of the SOS is consistent both with a lack of causal effects of effort on this post-test measure, and with evidence that this measure is confounded by awareness of performance. For instance Myers and Finney (2019) found that retrospective effort scores exhibited increased variance, suggesting that the retrospective measure blended clean prospective effort information with confounding from performance evaluation. They found that the unconfounded measure in their tasks showed a 55% decline in effect compared to the retrospective association (standardized beta of .16 versus .29). Together with other work, then, the present result indicates that the SOS is contaminated by subjects’ evaluation of their performance and that non-

Table 3. Table of Model Comparisons Study 2

Model	EP	Δ Fit	Δ df	p	AIC	Δ AIC	Baseline
Full model	10				3534.517	0	
Drop Incentive Effort	9	10.80	1	0.001	3543.317	8.801	Full model
Drop Effort Score	9	0.02	1	0.886	3532.537	-1.98	Full model

Note: EP = Estimated (i.e., free) parameters; Δ -2LL = change in $-2 \times$ Log-Likelihood of the model; Δ df = Change in degrees of freedom with respect to the comparison model; Δ AIC = Change in Akaike Information Criterion; 'Baseline' = The baseline model for each comparison.

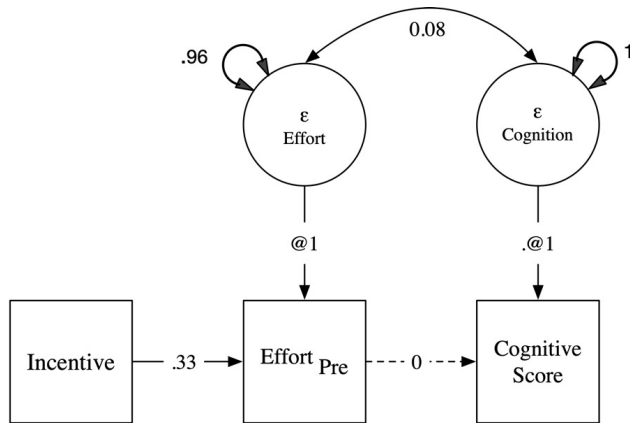


Figure 3. Instrumental variable preferred model, showing the significant effect of incentive on effort ($p = .001$) and the lack of effect of effort on cognitive score ($p = .89$) in Study 2.

confounded measures of effort support zero causal effects on performance.

Jointly, the results suggest that links of effort with cognitive scores in designs with post-task measures of effort is illusory, reflecting reverse causation from performance impacting feelings of goodness of effort. We conclude that interpretations of effort-performance correlations as causal were mistaken. The present result instead supports the conclusion that effort does not increase test scores. However, we wished to confirm this finding in independent and larger dataset, and to evaluate its robustness. Therefore, we replicated the finding in study 3 below.

Study 3

The goal of study 3 was to replicate study 2 in an independent larger sample, testing robustness to the specific measure used to assess effort. We increased the sample size from 500 for study 2 to over 1,200 for study 3. We note that we had pre-registered that we would use to SOS to show once more that it correlates $r \sim .28$ with cognitive score. On reflection, as this was by now already well-demonstrated, we made a decision after registration but before data collection began to omit the SOS. Doing this reduced testing time by approximately 2 minutes per participant, enabling an increase in sample size from 1,000 to 1,237. Given prior evidence of SOS's correlation with cognitive scores ($r \sim .28$, Silm et al., 2020), its inclusion was deemed unnecessary. Finally, to show robustness of the null finding to measures

and to verify the outcome was not restricted to using the scale developed in study 1, we varied the measure of effort to one developed independently. In study 3, effort was measured using a 6-item effort measure based on a scale developed for the large Swedish TIMSS study (Eklöf, 2010) and modified by Penk and Richter (2017) who also validated this as a pre-test measure of effort.

The pre-registered hypotheses of Study 3 ([aspredicted.org #124312](https://aspredicted.org/#124312)) were as follows:

1. Incentive (randomised reward for increasing score over that achieved on the matched baseline test) will increase effort* (parameter β_1).
2. Causal effect of effort on cognitive score (parameter β_2) will be non-significant (strong version of hypothesis).

We also registered a fallback from claim 2, namely that if any causal effect of effort were to be found on cognitive score, that it would be small (0 to 2 IQ points).

Methods

Sample

We recruited adult participants on prolific academic ($N = 1,237$, age $M = 42$, $SD = 14.1$, 613 male and 613 female). The sample was Asian 74; Black 36; Other 41; Southeast Asian 9; White 1,077. Subjects received £1.20 for their participation (excluding bonus). We excluded participants from earlier studies using the Prolific Academic feature to filter IDs to ensure no overlap.

Materials and Procedure

Effort was measured using the 6-item scale developed by Penk and Richter (2017). An example item is "I feel motivated to do my best on this test". Responses on a scale from 1: strongly disagree to 4: strongly agree. Reliability of the effort measure was $\omega = .80$. The same cognitive tests as used in study 2 were used. In this sample, ω -t reliabilities were .76 and .67 for Form A and B respectively.

The procedure was identical to that used in study 2 with the exception that the SOS was not administered. After testing, subjects complete a very brief effort questionnaire (NASA TL-X: Hart & Staveland, 1988) piloting use of this measure for unrelated studies on conscientiousness.

Table 4. Means and SD for each of the key variables, and variable inter-correlation matrix (SE in brackets).

	Mean (SD)	Paper- Folding A	Paper- Folding B
Paper-Folding A	4.49 (2.44)	--	--
Paper-Folding B	4.93 (2.21)	0.69 (0.02)	--
TIMSS effort for folding B	21.68 (4.42)	0.27 (0.03)	0.36 (0.02)

Results

Descriptive statistics are presented in Table 3.

Analyses were pre-registered and conducted in the same manner as in study two – using the instrumental variable structural model to evaluate parameters β_1 (the effect of incentive on effort) and β_2 (the effect of effort on scores) parameters with χ^2 comparisons with 1 degree of freedom to the baseline model.

The pre-planned tests were 1) Test if β_1 is significant. If not, the analyses would end, and we would seek a more effective incentive to increase effort. 2) Given a significant effect of the incentive on effort, we would test if β_2 was non-significant (could be dropped without significant loss of fit). We predicted β_2 would be small and non-significant. If so, our strongest prediction would be supported: that effort has no causal effect on cognitive score. A fallback was that, if β_2 could not be dropped, then its size would be assessed; If small ($\leq .2$), then a less restrictive prediction would be supported, namely that effort has only a small causal effect on cognitive score.

The results for each of these pre-registered hypotheses are shown in Table 4. A model setting β_1 to zero showed a highly statistically significant effect of the incentive on effort ($\chi^2(1) = 28.05$, $p < .001$). This validated the incentive manipulation and permitted moving to testing hypothesis two. This was assessed by setting β_2 to zero and evaluating the effect on model fit. As seen in Table 4, setting the effect of effort on cognitive ability to zero had no significant effect on model fit ($\chi^2(1) = 2.08$, $p = .149$). The final preferred model, then, was one in which incentive raised effort, but effort did not increase cognitive ability (see). The point estimate for b_2 was larger, but the wide confidence interval included zero ($\beta_2 = .27 [-0.07, 0.62]$). These results closely mirrored those obtained in study 2, showing replicability and robustness to the measure of effort, and rejecting for a second time any effect of effort on cognitive scores – combining the results for the effect of effort on cognitive score across study 2 and study 3 using Fisher's method (Fisher, 1925), yields an overall p-value of 0.40.

Discussion

Supporting the reliability and robustness of the results to the specific measure of effort, study 3 closely replicated study 2. The pre-planned test that incentive effectively in-

creased effort was supported, and no effect of effort on cognitive test score was found. While the point estimate for β_2 was 0.27, the wide confidence interval includes zero. The wide confidence interval in study three suggests to us that the alternative effort measure was not as precise or efficient at estimating effort as that developed by us in Study 1. This measure may introduce more noise, leading to less precise estimates. In future work, researchers should directly compare the psychometric properties of these pre-test measures in an independent replication.

Jointly, these findings support the hypothesis that effort does not causally raise cognitive score. Both studies, then, showed that, while incentives reliably and substantially manipulated effort, increased effort did not manifest in any statistically or theoretically significant causal effect on cognitive scores. We next discuss the overall findings and their implications for research and policy around effort and cognitive ability.

General Discussion

The results from these two large studies, yielding a combined p-value of .40 for the causal effect of effort on cognitive score, compel a major reinterpretation of the relationship between effort and cognitive performance. Alongside retractions for fraud (Warne, 2022) of studies driving earlier claims of large associations of effort with cognitive scores, modern, large, studies already indicated that the maximum association is around .16 (Bates & Gignac, 2022). These studies suggested that the causal fraction of this association may be smaller than even this, and perhaps zero. The present result, adopting a framework capable of testing causality and avoiding confounds present in earlier designs. The results suggest that any causal impact of effort on cognitive scores is negligible or absent, consistent with a true effect size near zero.

At least three conclusions follow from this work. First, the validity of cognitive scores as predictors of performance, while already supported by overwhelming external validity data (Ones & Viswesvaran, 2023), is given additional support. Moreover, the null role of effort indicates that the IQ-performance association is due to ability rather than confounds such as, say, conscientiousness. Users of ability tests should be assured that ability tests effectively recruit reasoning and accurately assess intelligence when used as designed.

Second, influential meta-analytic claims of causal effects should be abandoned as being both discredited (Warne, 2022), and now, refuted. Finally, the results are important for what they imply about the nature of cognition and of effort. Just as Francis Crick's (1958) '*Central Dogma of Molecular Biology*' importantly conjectured that information could not flow back from protein to DNA, the present results suggest a potential 'central dogma of cognition': that volitional effort can direct cognitive resources but cannot fundamentally alter or bypass the efficacy of the underlying cognitive systems themselves.

These results, then, move beyond prior empirical tests to imply that effort may have no mechanistic means of raising cognitive function. In the language of cognitive neuropsych-

chology, cognitive ability is ‘encapsulated’ from the influence of effort. By analogy with perceptual systems, while motivation can direct where one looks, exhorting an individual to ‘try harder’ cannot alter the fundamental sensitivity of their visual system if a stimulus is below their perceptual threshold.

Limitations & Future Research

The present work should not be interpreted as implying that effortful goal-directed action and conscientiousness do not cause valuable results – conscientiousness has universally positive outcomes (Moffitt et al., 2011), and the role of effort on tasks where ability is not a limiting factor but which instead rely on ambitious goals and persistence in pursuing these over time is robust (Bates et al., 2023). Recent work studying physical (Bates, 2025) and cognitive effort (Bates, 2024) suggests that conscientiousness decouples effort from the frustration that otherwise acts to terminate effort. The present research does not contradict this body of work. Instead, it addresses the remarkable-if-true question of whether the desirable side-effects of effort extend to improving intelligence. This would be remarkable, but here, we rule this out.

A strength of the design was that subjects in both arms of the study are asked about intended effort, balancing any effects of the measure on performance. Importantly, asking people to verbalise intended higher effort has been shown to *increase* effort via behavioral commitment effects (Manstead & Eekelen, 1998). Thus the measure in the experimental group should have induced larger not smaller effects of the incentive on performance (Carver & Scheier, 1998).

A key implication of these findings is that educators seeking to improve cognitive outcomes should focus on methods with proven efficacy, such as spaced recall and systematic time on task (Hattie, 2023), rather than expecting that simply motivating students to try harder will raise their underlying ability. A second implication is that effort research should be emphasised for its beneficial effects on “non-cognitive” pathways to success – for instance improved goal setting (Bates et al., 2023). It is also worth noting that the present lack of effect of effort on ability may inform research on the inverse question: whether doing cognitive work reduces the ability to do future effort, the so-called resource model of willpower. Lin, Saunders, Friese, Evans, and Inzlicht (2020), for instance, found that, prior effort does not reduce the accuracy of a subsequent information processing task. This null result is compatible

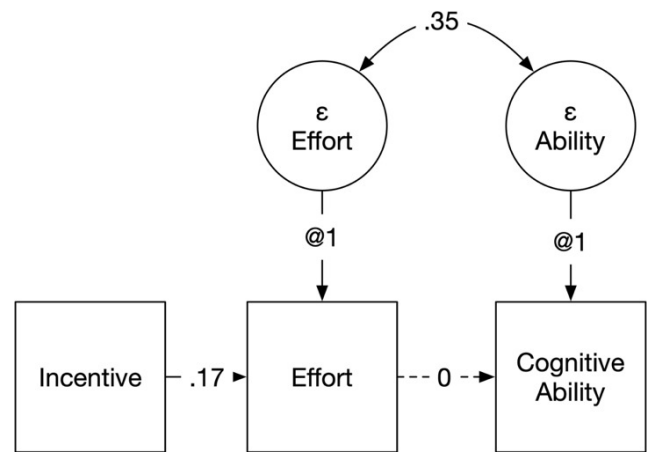


Figure 4. Significant effect of incentive on effort, but no effect of effort on cognitive score, Study 3.

with the present data in supporting the impermeability of cognition to effort, whether prior or contemporary.

Conclusion

The present studies, based on instrumental variable analysis of an appropriate RCT design in two large samples, supported the absence of effects of effort on cognitive performance. The results suggest earlier results reflected confounded effects of ability on effort and post-test re-evaluation of effort. While effort has important motivational functions, enhancing intelligence or intellectual performance is not one of them.

Funding

Support for this project was provided in part by The John Templeton Foundation (US GF13666.154364).

Data Availability

Pre-registrations are available at <https://research-box.org/790>

Submitted: May 20, 2025 CST. Accepted: July 09, 2025 CST.

Published: August 02, 2025 CST.

Table 5. Table of Causal Model Comparisons, Study 3.

Model	EP	Δ Fit	Δ df	p	AIC	Δ AIC	Compared to
Full model	10	--	--	--	8558.139	--	--
Drop Incentive Effort	9	28.052	1	< 0.001	8584.191	26.052	Full model
Drop Effort Score	9	2.085	1	0.149	8558.224	0.085	Full model

Note: EP = Estimated (i.e., free) parameters; Δ-2LL = change in $-2 \times \text{Log-Likelihood}$ of the model; Δ df = Change in degrees of freedom with respect to the comparison model; Δ AIC = Change in Akaike Information Criterion; ‘Baseline’ = The baseline model for each comparison



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341–342. <https://doi.org/10.3758/BF03331551>
- Bates, T. C. (2024). Conscientiousness makes effort less frustrating. *Personality and Individual Differences*, 226. <https://doi.org/10.1016/j.paid.2024.112690>
- Bates, T. C. (2025). Conscientiousness makes physical effort less frustrating. *Personality and Individual Differences*.
- Bates, T. C., Enkhbat, T., Gray, E., Lee, J., & Zakharin, M. (2023). How to get things done: Tight linkage of conscientiousness with twelve mechanisms of Goal Setting Theory. *Personality and Individual Differences*, 214. <https://doi.org/10.1016/j.paid.2023.112331>
- Bates, T. C., & Gignac, G. E. (2022). Effort impacts IQ test scores in a minor way: A multi-study investigation with healthy adult volunteers. *Intelligence*, 92. <https://doi.org/10.1016/j.intell.2022.101652>
- Bates, T. C., Maes, H., & Neale, M. C. (2019). umx: Twin and Path-Based Structural Equation Modeling in R. *Twin Research and Human Genetics*, 22(1), 27–41. <https://doi.org/10.1017/thg.2019.2>
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139174794>
- Crick, F. (1958). On Protein Synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.
- Dong, Y., Stupnisky, R. H., & Berry, J. C. (2013). Multiple causal attributions: An investigation of college students learning a foreign language. *European Journal of Psychology of Education*, 28(4), 1587–1602. <https://doi.org/10.1007/s10212-013-0183-4>
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One*, 18(3), e0279720. <https://doi.org/10.1371/journal.pone.0279720>
- Duckworth, A. L., Quinn, P. D., ... Lynam, D. R. (2025). Editorial Expression of Concern for Duckworth et al., Role of test motivation in intelligence testing. *Proc Natl Acad Sci U S A*, 122(8), e2502439122. <https://doi.org/10.1073/pnas.2502439122>
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), 7716–7720. <https://doi.org/10.1073/pnas.1018601108>
- Eccles, J. S., Adler, T. E., Fatterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). Freeman.
- Eccles, J. S., & Wigfield, A. (2023). Expectancy-value theory to situated expectancy-value theory: Reflections on the legacy of 40+ years of working together. *Motivation Science*, 9(1), 1–12. <https://doi.org/10.1037/mot0000275>
- Eklöf, H. (2010). *Student motivation and effort in the Swedish TIMSS Advanced field study* [Paper]. 4th meeting of the IEA International Research Conference, Gothenburg, Sweden.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests*. ETS Research and Development.
- Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial Literacy, Financial Education, and Downstream Financial Behaviors. *Management Science*, 60(8), 1861–1883. <https://doi.org/10.1287/mnsc.2013.1849>
- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland.
- Hattie, J. A. (2023). *Visible Learning: The Sequel A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement*. Routledge. <https://doi.org/10.4324/9781003380542>
- Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology*, 72, 319–345. <https://doi.org/10.1146/annurev-psych-061020-105721>
- Lin, H., Saunders, B., Friese, M., Evans, N. J., & Inzlicht, M. (2020). Strong Effort Manipulations Reduce Response Caution: A Preregistered Reinvention of the Ego-Depletion Paradigm. *Psychol Sci*, 31(5), 531–547. <https://doi.org/10.1177/0956797620904990>
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>
- Manstead, A. S. R., & Eekelen, S. A. M. (1998). Distinguishing Between Perceived Behavioral Control and Self-Efficacy in the Domain of Academic Achievement Intentions and Behaviors. *Journal of Applied Social Psychology*, 28(15), 1375–1392. <https://doi.org/10.1111/j.1559-1816.1998.tb01682.x>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2693–2698. <https://doi.org/10.1073/pnas.1010076108>

- Myers, A. J., & Finney, S. J. (2019). Does It Matter if Examinee Motivation Is Measured before or after A Low-Stakes Test? A Moderated Mediation Analysis. *Educational Assessment*, 26(1), 1–19. <https://doi.org/10.1080/10627197.2019.1645591>
- Ones, D. S., & Viswesvaran, C. (2023). A response to speculations about concurrent validities in selection: Implications for cognitive ability. *Industrial and Organizational Psychology*, 16(3), 358–365. <https://doi.org/10.1017/iop.2023.43>
- Penk, C., & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation and Accountability*, 29(1), 55–79. <https://doi.org/10.1007/s11092-016-9248-7>
- Penk, C., & Schipolowski, S. (2015). Is it all about value? Bringing back the expectancy component to the assessment of test-taking motivation. *Learning and Individual Differences*, 42, 27–35. <https://doi.org/10.1016/j.lindif.2015.08.002>
- R Development Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.1). R Foundation for Statistical Computing. <http://www.R-project.org>
- Retraction for Duckworth et al., Role of test motivation in intelligence testing. (2025). *Proc Natl Acad Sci U S A*, 122(18), e2508040122. <https://doi.org/10.1073/pnas.2508040122>
- Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., ... Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, 2(1). <https://doi.org/10.1038/s43586-021-00092-5>
- Silm, G., Pedaste, M., & Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review*, 31. <https://doi.org/10.1016/j.edurev.2020.100335>
- Sundre, D. L., & Thelk, A. D. (2007). *The Student Opinion Scale (SOS): A Measure of Examinee Motivation: Test Manual*.
- Warne, R. T. (2022). *More Articles by Stephen Breuning That Need Retraction*. <https://russellwarne.com/2022/02/22/more-articles-by-stephen-breuning-that-need-retraction/>
- Weiner, B. (2000). Attributional Thoughts about Consumer Behavior. *Journal of Consumer Research*, 27(3), 382–387. <https://doi.org/10.1086/317592>
- Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils*. (Vol. Appendix B: Effects of a Duty on Price and Output with Special Reference to Butter and Flaxseed). MacMillan.