Intelligence 96 (2023) 101717

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

# Reevaluating the Dunning-Kruger effect: A response to and replication of Gignac and Zajenkowski (2020)

Curtis S. Dunkel [a,*], Joseph Nedelec [b], Dimitri van der Linden [c]

[a] Western Illinois University, Department of Psychology, Macomb, IL 61455, United States of America
[b] School of Criminal Justice, University of Cincinnati, United States of America
[c] Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, the Netherlands

ARTICLE INFO

*Keywords:*
Dunning-Kruger effect
General intelligence
Heteroscedasticity

ABSTRACT

As applied to general intelligence, the Dunning-Kruger effect (DK) is the phenomenon in which individuals at the lower end of the intellectual ability distribution are more likely to overestimate their intelligence. In a recent article in *Intelligence* it was suggested that the DK is primarily a statistical artifact and, indeed, the application of more appropriate analyses led to a failure to replicate a significant effect. When some of the limitations (namely sample representativeness) were addressed and the more appropriate statistical methods were used in the current study, our analyses illustrated a statistically significant DK effect. However, the magnitude of the effect was minimal; bringing its meaningfulness into question. In conclusion, it is recommended that the conditions that result in a significant DK be further explored.

## 1. Introduction

Rarely does a research finding in Cognitive Psychology become part of the common parlance. The Dunning-Kruger effect (DK) is an exception (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999). Named after the psychological scientists who discovered the phenomenon, the DK refers to the inverse relationship between one's actual aptitude and one's ability to accurately estimate said aptitude. In other words, while people generally exhibit some positive bias in assessing their own ability, this bias is heightened in those at the lower end of the distribution. It is thought that the second component of this "double curse" (Dunning et al., 2003) of inaccurate self-assessment, occurs due to a deficit in meta-cognition. This deficit in meta-cognition results in the failure to grasp what one knows and does not know.

However, as several critics have indicated the original series of studies were conducted using a small and cognitively elite student sample (i.e., participants were enrolled at an Ivy League university). Such a skewed sample leads to questions about the generalizability of the results (Krajc & Ortmann, 2008; also see Schlösser, Dunning, Johnson, & Kruger, 2013). Although, the DK effect has been replicated with non-cognitively elite and representative samples (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Jansen, Rafferty, & Griffiths, 2021; Lyons, Montgomery, Guess, Nyhan, & Reifler, 2021; Schlösser et al., 2013).Additional criticisms surround alternative explanations for the findings. Krueger and Mueller (2002) suggested that the DK could be caused by regression toward the mean and/or the better than average effect (see Kruger & Dunning, 2002 for a reply). For example, if, for the majority of individuals, the default response to the question of one's ability is that they are slightly better than average, then it automatically stands that those at the lower end of the distribution of ability will be the least accurate (i.e., exhibit the greatest upward bias). Counter to this criticism the DK effect has also been found when individuals were tasked with judging their absolute (not just relative) performance (Ehrlinger et al., 2008; Jansen et al., 2021), although inconsistently (Schlösser et al., 2013). In sum, distinguishing between alternative possibilities for the DK and the original explanation of an indiscriminate deficit in meta-cognition has proven remarkably difficult.

To address this point, recently Gignac and Zajenkowski (2020) reevaluated the DK with regards to self-assessed intelligence. They contended that the DK effect may actually be a manifestation of a combination of alternative factors, including the better-than-average effect, regression toward to the mean, and they also argued that the result of the specific statistical analyses used may have contributed to finding the effect. Regarding the latter, they criticized the customary

* Corresponding author.
*E-mail address:* c-dunkel@wiu.edu (C.S. Dunkel).

https://doi.org/10.1016/j.intell.2022.101717
Received 10 August 2022; Received in revised form 17 November 2022; Accepted 17 November 2022
Available online 23 November 2022

approach for assessing the DK which entails the categorization of participants into four discrete groups based on objective IQ test scores (high IQ, mid-high IQ, mid-low IQ, low IQ) and the subsequent testing of group differences in difference scores between objective and self-assessed intelligence (SAI). The DK is said to occur when significant differences between groups in the accuracy of the SAI emerge, in which the lower IQ groups are less accurate because they overestimate their objective IQ, and more so than those at higher objective IQ levels. Gignac and Zajenkowski (2020) state that this statistical procedure is strongly confounded with the better than average effect and regression toward the mean described in the previous paragraph. The two tests they advocate that do not suffer from these issues, are the Glejser (1969) test of heteroscedasticity and using linear regression techniques to test quadratic effects. Thus, Gignac and Zajenkowski expected that the DK effect would appear when using the traditional statistical approach, but that the effect would fail to materialize when using the improved alternative methods. Indeed, their results were as anticipated. The customary approach in which categories of participants are formed based on IQ scores yielded a significant DK effect, yet the tests of heteroscedasticity and quadratic effects did not.

Nevertheless, despite these findings, we believe there may be reasons to suspect a small, but significant, DK after some adjustments to the methods and analyses as used by Gignac and Zajenkowski (2020). While their sample size was substantial, especially in comparison to the original DK studies (Kruger & Dunning, 1999), it was still primarily composed of university students, reducing the full range of possible IQ scores.[1] As mentioned, the positive skew in intelligence exhibited by a university sample causes several methodological concerns with regards to the relationship between SAI and objective intelligence (Krajc & Ortmann, 2008) that may impact the DK. Recently, Gignac (2022) addressed these limitations when testing the DK by using a more representative sample. However, the focus of the Gignac (2022) was on financial literacy and not cognitive ability. We wish to specifically focus on cognitive ability and, therefore, to address these concerns in our replication/extension of the Gignac and Zajenkowski (2020) findings, we utilized a large nationally representative data set.

Additionally, we maintain that the analyses used by Gignac and Zajenkowski (2020) confound two aspects of the DK. While the DK refers to the overestimation of ability for those on the far left end of the distribution, it is also found that those on the far right end of the distribution exhibit a slight tendency to *underestimate* their relative performance (Dunning et al., 2003; Kruger & Dunning, 1999). Thus, a test of heteroscedasticity may be attenuated because the absolute residuals would not only be higher on the low end of the distribution, but also slightly higher on the high end of the IQ distribution, albeit in the opposite direction. Likewise, Gignac and Zajenkowski (2020) used hierarchical multiple regression to test for the DK by regressing objective IQ on SAI and testing for quadratic effects. However, given that deviation from linearity should not only occur at the low end of the IQ distribution, but also slightly at the high end of the IQ distribution, testing for cubic effects may be more appropriate. For these reasons we believe it is valuable to reexamine the possibility of a DK in the relationship between IQ and SAI.

## 2. Method

### 2.1. Sample

To test the hypotheses of the current study, data from the restricted version of the National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris et al., 2009) were employed. The Add Health was initiated with the administration of an in-school questionnaire when participants were in grades 7 to 12 in 1994. To date, there have been five subsequent waves of data collection. For more information about the Add Health study see Harris et al. (2009). For the current study, data from the in-home interview that occurred in the third wave (conducted in 2001–2002) were used.

The analytical sample for the current study was based on cases with valid data for both measures of intelligence and with an objective IQ score of 64 or greater (see below). Consequently, the analytical sample ($n = 13{,}977$) includes participants who were between the ages of 18 and 28 ($\overline{X} = 22.32$, $SD = 1.82$) at the time of data collection. Additionally, the analytical sample includes 46.88% ($n = 6553$) males, 53.12% ($n = 7424$) females, and is comprised of participants who self-identified their race as White (68.95%; $n = 9504$) or Person of Color (31.05%; $n = 4280$ [193 cases were missing on race]).

### 2.2. Measures

#### 2.2.1. Self-assessed intelligence (SAI)

The *SAI* measure was comprised of two items. First, participants responded to the question "Compared to other people your age, how intelligent are you?". Responses were coded using a six-point Likert scale (1 = moderately below average; 2 = slightly below average; 3 = about average; 4 = slightly above average; 5 = moderately above average; 6 = extremely above average). Second, participants were later asked "How intelligent are you?" with responses indicated on a four-point Likert scale (1 = very intelligent; 2 = moderately intelligent; 3 = slightly intelligent; 4 = not at all intelligent)[2]

A multi-step process was followed to create the *SAI* measure. First, item 2 ("How intelligent are you?") was reverse coded to match the coding direction of item 1. Second, both items were standardized (i.e., $z$-transformed) and then summed. Third, to produce estimated IQ scores, the total scores were then standardized once again, with a mean of 100 and a standard deviation of 15.

#### 2.2.2. Objective IQ

Intelligence was measured using an abridged version of the Peabody Picture Vocabulary Test (PPVT; Dunn, 1981) administered in wave 3 (for details about the PVT in the Add Health, aka., AHPVT, see Halpern, Joyner, Udry, & Suchindran, 2000). We used the standardized scores calculated by the Add Health study team that mirror traditional IQ test scores ($\overline{X} = 100.37$, $SD = 16.75$; min., max.: 9, 123). However, the distribution was not normal (skew = −2.25, kurtosis = 12.23). To ameliorate the non-normality and accord with the method employed by Gignac and Zajenkowski (2020), we removed any outlier cases identified using a 1.5 multiplier (using Stata's *extreme* command with the *iqr* option). This process identified cases with an IQ score below 64 as outliers; these cases ($n = 260$) were removed from the sample. After removing these cases, the intelligence measure approached normality (skew = −0.46, kurtosis = 2.38).[3]

## 3. Results

All analyses were completed using Stata SE 17 (StataCorp., 2021). The summary statistics of the SAI ($\overline{X} = 100.13$, $SD = 14.86$; min., max.:

[1] It is important to note that the samples used by Gignac and Zajenkowski (2020) also included couples recruited from the broader Warsaw population and this clearly offers some remediation of the criticism of an unrepresentative sample. However, the recruitment strategy (inclusion criteria) for the community couples sample presents its own potential problems.

[2] See Table S1 for summary information for both items; the correlation between the items was $r = 0.42$, $p < .001$.

[3] The AHPVT was also conducted in Wave 1. To assess reliability of the IQ measure, we first removed outliers on the Wave 1 AHPVT measure following the same process indicated for the Wave 3 AHPVT measure and then we estimated the correlation between the two measures ($r = 0.68$, $p < .0001$, $n = 13{,}216$).

45.65, 127.32; skew = −0.07, kurtosis = 2.57) and objective IQ ($\overline{X}$ = 102.01, $SD$ = 12.62; min., max.: 64, 123) illustrated expected variability given the nature of the sample (large *n*-size and derived from nationally representative sample). In contrast to an expected DK effect, the mean SAI was less than the mean objective IQ, $t_{(13976)} = -13.00$, $p < .001$, Cohen's $d = -0.14$. However, as expected with a DK effect the association between SAI and objective IQ was positive and statistically significant, $r = 0.24$, $p < .001$ [95%CI: 0.22, 0.25]. The magnitude of the observed association is lower than the average reported in meta-analyses (i.e., $r = 0.326$ [95%CI: 0.284, 0.368] in Freund & Kasten, 2012). However, the 95% confidence intervals of the association in the present study are somewhat close to the overall average in Freund and Kasten (2011) and the magnitude is within the distribution of estimates observed in their meta-analyses.[4]

Following Gignac and Zajenkowski (2020) we first used what they identified as the standard method for testing the DK effect. The analytical sample was divided into four groups based on their objective IQ score: low (≤ 85), low average (86–100), high average (101–115), and high (≥ 116). Difference scores were then computed by subtracting objective IQ from SAI (i.e., SAI – objective IQ). Thus, positive scores reflect an overestimation, while negative scores reflect an underestimation of one's objective intelligence. Summary statistics for the intelligence measures and the difference scores across the four IQ groups are displayed in Table 1.

The first indication of an ostensible DK effect derived from the standard method of testing can be observed in Fig. 1. As displayed in Panel A of Fig. 1, a decrease in overestimation occurs across the four objective IQ groups and the classic cross-over of the objective and subjective intelligence assessments occurs in the higher IQ groups. Consequently, we see that the low and low average IQ groups *over*estimated their intelligence, and the high average and high IQ groups *under*-estimated their intelligence. This pattern was reinforced by the next assessment: a one-way analysis of variance (ANOVA) with objective IQ group membership acting as the independent variable and the difference scores as the dependent variable was performed. Overall, the mean difference scores varied by IQ group ($F_{(3,\ 13{,}973)} = 1757.04$, $p < .0001$) and the effect size was strong ($\eta^2_p = 0.27$). Furthermore, all post-ANOVA pairwise comparisons (Tukey's HSD correction) indicated statistically significant contrasts across all groups ($ps < 0.0001$). Such differences can be observed in Panel B of Fig. 1 which illustrates the mean difference scores across the objective IQ groups. Thus, consistent with the DK effect lower IQ was associated with an overestimation of intelligence and higher IQ was associated with an underestimation of intelligence.

Thus far the results align with the DK effect and what Gignac and Zajenkowski (2020) initially observed. Following their next step, we conducted the Glejser test of heteroscedasticity to assess the presence of a DK effect. First, SAI was regressed on the objective IQ scores in a linear bivariate regression. The residuals were saved and then transformed into absolute values.[5] Finally, the absolute values of the residuals were then correlated with the objective IQ scores. A negative correlation would indicate that as cognitive ability decreased, so too does the accuracy of the SAI. The correlation between the absolute residuals and objective IQ was $r = -0.07$, $p < .0001$ [95%CI: −0.087, −0.054]. Consequently, using Gignac and Zajenkowski's first recommended unconfounded assessment, our analyses indicated a slight, yet statistically significant, DK effect.[6]

Gignac and Zajenkowski's second recommended assessment, curvilinear regression, was also conducted. A hierarchical multiple regression analysis was estimated using Stata's *hireg* command wherein SAI (dependent variable) was regressed on objective IQ in Step 1 and a quadratic term (objective IQ × objective IQ) was created and entered in Step 2. When the interaction term was entered ($b = 0.009$, $\beta = 1.52$, [95%CI of $b$: 0.007, 0.010]; model $R^2 = 0.066$), there was a modest, yet significant $\Delta R^2 = 0.011$ ($F_{(1,\ 13{,}974)} = 162.10$, $p < .001$). The positive beta for the interaction term indicates that the slope becomes more positive as SAI increases.

Finally, to account for the possibility that the pattern between objective IQ and SAI could be better represented by a cubed effect a third step in the hierarchical regression model was estimated wherein a cubed term (objective IQ × objective IQ × objective IQ) was entered. Inclusion of the cubed term ($b = 0.00013$, $\beta = 3.30$, [95%CI of $b$: 0.00004, 0.00022]; model $R^2 = 0.067$) resulted in a very slight yet statistically significant increase in explained variance, $\Delta R^2 = 0.001$ ($F_{(1,\ 13{,}973)} = 7.889$, $p = .005$), suggesting that the cubed effect may fit the data slightly better than the quadratic effect.[7] The nonlinear nature of the data is evinced in Panel C of Fig. 1, which displays both a linear line of best fit (using Stata's *lfit* command) a line of best fit derived from polynomial regression (using Stata's *lpoly* command: Epanechnikov kernel function, a rule-of-thumb bandwidth estimator, and a polynomial degree of three).

## 4. Discussion

Gignac and Zajenkowski (2020) identified several concerns with the methods and analyses underlying the DK; thus bringing the validity of the effect into question. In correcting for these deficiencies, Gignac and Zajenkowski failed to find support for the DK. However, we believe that Gignac and Zajenkowski's methods and analyses themselves had some limitations. Namely, while the sample used by Gignac and Zajenkowski included community members, over 53% of the sample still included undergraduate students.[8] Thus, their sample likely had a positively skewed intelligence distribution and the analyses they conducted failed to account for the tendency of individuals at the right tail of the distribution to underestimate their ability. An additional limitation was that their analytical sample was comprised of sub-samples generated via convenience sampling techniques (social media and personal connections) and stringent inclusion criteria that may have biased their sample in some way (e.g., couples in long-term [6+ months] romantic relationships; Gignac & Zajenkowski, 2019; Zajenkowski & Gignac, 2018). Consequently, the extent to which their analytical sample allows for generalization is arguably limited and while Gignac (2022) recently addressed the sampling issue the focus of the study was on financial literacy and not cognitive ability.

To address these methodological concerns, our analyses included two main improvements on the prior literature. First, we addressed the

[4] Freund and Kasten's (2011) analyses illustrated that social comparison measures of SAI exhibit higher correlations with objective measures. In the current study, the association between the first SAI item (which is a social comparison measure) and objective IQ is $r = 0.26$, $p < .001$ [95%CI: 0.25, 0.28). Thus, the inclusion of the second SAI item (which is not a social comparison measure) attenuated the association between the aggregated SAI measure and the objective IQ measure. However, when the analyses in the current study were repeated with only the social comparison SAI measure the results were virtually identical to those using the two-item measure of SAI. Thus, to maximize variability we chose to maintain the two-item measure of SAI reported herein.

[5] The absolute residuals approached normality (skew = 0.83, kurtosis = 3.49).

[6] More recently, Gignac (2022) has recommended against using the Glejser test in assessing potential DK effects.

[7] All analyses were replicated correcting for sample weights and the survey design of the Add Health. The results were virtually identical to those presented herein. Given that inclusion of the sample weights reduced the sample size by 798 cases (from $n$ = 13,977 to $n$ = 13,179) and the results did not differ, we opted to present the non-weighted results. See supplemental materials for more information.

[8] Based on the data provided by the authors to the Open Science Framework (osf.io/dg547).

**Table 1**
Summary statistics for intelligence measures by intelligence group and the full analytical sample.

| | Objective IQ | | | SAI | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|
| Intelligence Group | Mean | SD | *n* | Mean | SD | *n* | Mean | SD | *n* |
| Low (IQ ≤ 85) | 78.87 | 4.42 | 1535 | 96.59 | 16.21 | 1535 | 17.73 | 16.49 | 1535 |
| Low average (IQ 86–100) | 92.32 | 3.97 | 3982 | 96.54 | 14.90 | 3982 | 4.22 | 15.29 | 3982 |
| High average (IQ 101–115) | 108.13 | 3.82 | 6395 | 100.86 | 13.98 | 6395 | −7.26 | 13.92 | 6395 |
| High (IQ ≥ 116) | 119.00 | 1.98 | 2065 | 107.45 | 13.34 | 2065 | −11.55 | 13.46 | 2065 |
| Analytical sample | 102.01 | 12.62 | 13,977 | 100.13 | 14.86 | 13,977 | −1.88 | 17.08 | 13,977 |

Notes: Objective IQ min., max.: 64, 123; SAI min., max: 45.65, 127.32; Difference min., max: −65.35, 63.32; $F_{(3,\ 13,973)} = 1757.04$, $p < .0001$, $\eta_p^2 = 0.27$; post-ANOVA pairwise comparison (Tukey's HSD correction) indicated differences across all group comparisons ($p < .0001$).

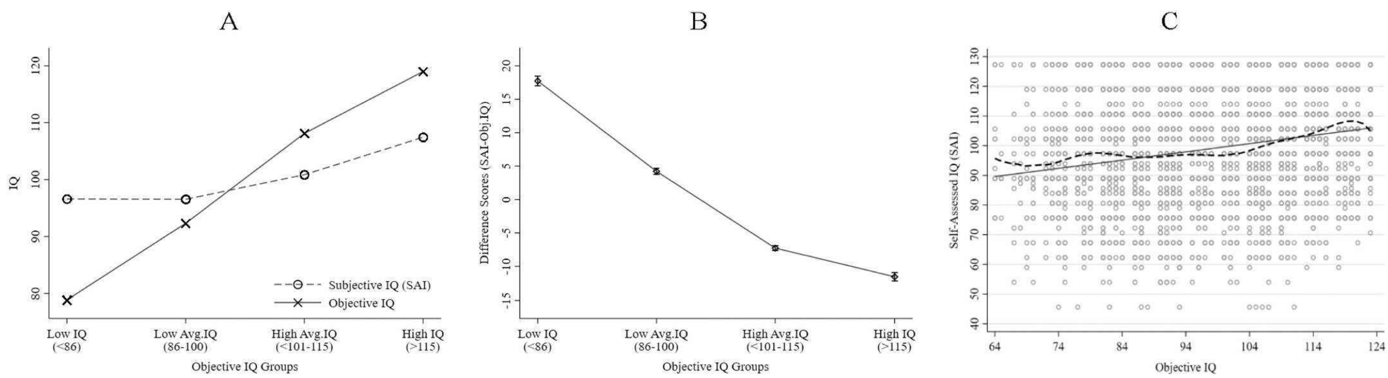


**Fig. 1.** Panel A: Subjective (SAI) and objective IQ means across objective IQ groups; panel B: mean difference scores (SAI – Objective IQ) across objective IQ groups; panel C: scatterplot with linear fit (solid) and line of best fit using local polynomial regression (dashed) for objective IQ and SAI (*lpoly* command in Stata; Epanechnikov kernel function, ROT bandwidth estimator, and a polynomial degree of 3).

problem of representativeness across the ability spectrum by using data from a large nationally representative sample (i.e., Add Health). Second, in addition to repeating each analysis conducted by Gignac and Zajenkowski (2020), a cubed term was added to the hierarchical regression allowing for the underestimation of those at the right tail of ability distribution to be modeled.

Similar to Gignac and Zajenkowski (2020) the traditional categorization of participants into ability groups yielded a significant DK. In contrast, however, the analyses designed to remediate the issues with the categorization method (i.e., Glejser test of heteroscedasticity and hierarchical regression) also resulted in a significant DK, albeit with a small effect size. While the addition of the cubed term to the regression model also explained a statistically significant amount of additional variance, the effect size was so small as to bring its meaningfulness into question. Gignac and Zajenkowski (2020) recommended that changes in $R^2$ of 2% to 4% of additional variance explained is required to be indicative of substantial significance. In our analyses, the quadratic term increased the explained variance in self-assessed intelligence by 1.1% while the cubed term increased the model $R^2$ by <1%. Thus, while our analyses illustrated a statistically significant DK effect, the magnitude of the effect appeared to be minimal.

The conclusions of the current study should be tempered by at least two limitations. First, while the objective measure of intelligence employed in the Add Health is a well-validated measure it is focused on verbal cognitive abilities and exhibits a restricted range of scores. To the extent that other measures of intelligence with greater breadth and range would result in varied findings from the current study awaits future assessment. Second, the measure of self-assessed intelligence was based on two rather vague categorical items which limited the variation relative to continuous measures. The low reliability of the SAI measure could have any number of effects on the results. For example, if the reliability itself varies across the ability spectrum, then the results may mimic those of a significant DK. Thus, future research is encouraged to replicate the current findings using continuous measures of self-assessed intelligence within a nationally representative sample. In sum, the use of a more representative sample resulted in a significant DK, even with the analyses prescribed by Gignac and Zajenkowski (2020). Thus, it may be premature to conclude that the DK is mostly a statistical artifact and additional research should focus on the variables and conditions that impact the reliability and strength of the effect.

## Declaration of Competing Interest

None.

## Data availability

The authors do not have permission to share data.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2022.101717.

## References

Dunn, L. M. (1981). *Peabody picture vocabulary test-revised: Manual for forms L and M.* Circle Pines, NM: American Guidance Service.
Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.
Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware? Further explorations of (lack of) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98–121.
Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*, 296–321.
Gignac, G. E. (2022). The association between objective and subjective financial literacy: Failure to observe the Dunning-Kruger effect. *Personality and Individual Differences, 184*, Article 111224.
Gignac, G. E., & Zajenkowski, M. (2019). People tend to overestimate their romantic partner's intelligence even more than their own. *Intelligence, 73*, 41–51.
Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence, 80*, Article 101449.

Glejser, H. (1969). A new test for heteroskedasticity. *Journal of the American Statistical Association, 64*, 316–323.

Halpern, C. T., Joyner, K., Udry, J. R., & Suchindran, C. (2000). Smart teens don't have sex (or kiss much either). *Journal of Adolescent Health, 26*, 213–225.

Harris, K. M., Florey, F., Tabor, J., Bearman, P. S., Jones, J., & Udry, J. R. (2009). *The national longitudinal study of adolescent health: Research design*. Chapel Hill, NC: Carolina Population Center, University of North Carolina [WWWDocument] http://www.cpc.unc.edu/projects/addhealth/design.

Jansen, R. A., Rafferty, A. N., & Griffiths, T. L. (2021). A rational model of the Dunning-Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behavior, 5*, 756–763.

Krajc, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology, 29*, 724–738.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*, 180–188.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Kruger, J., & Dunning, D. (2002). Unskilled and unaware—But why? A reply to Krueger and Mueller. *Journal of Personality and Social Psychology, 82*, 189–192.

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academic of Science, 118*, Article e2019527118.

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. *Journal of Economic Psychology, 39*, 85–100.

StataCorp.. (2021). *Stata statistical software: Release 17*. College Station, TX: StataCorp LLC.

Zajenkowski, M., & Gignac, G. E. (2018). Why do angry people overestimate their intelligence? Neuroticism as a suppressor of the association between trait-anger and subjectively assessed intelligence. *Intelligence, 70*, 12–21.