Intelligence 80 (2020) 101449

Contents lists available at ScienceDirect

Intelligence

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

# The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data

Gilles E. Gignac[a,*], Marcin Zajenkowski[b]

[a] University of Western Australia, Australia
[b] Faculty of Psychology, University of Warsaw, Poland

ARTICLE INFO

*Keywords:*
Dunning-Kruger effect
Intelligence
Self-assessed intelligence

ABSTRACT

The Dunning-Kruger hypothesis states that the degree to which people can estimate their ability accurately depends, in part, upon possessing the ability in question. Consequently, people with lower levels of the ability tend to self-assess their ability less well than people who have relatively higher levels of the ability. The most common method used to test the Dunning-Kruger hypothesis involves plotting the self-assessed and objectively assessed means across four categories (quartiles) of objective ability. However, this method has been argued to be confounded by the better-than-average effect and regression toward the mean. In this investigation, it is argued that the Dunning-Kruger hypothesis can be tested validly with two inferential statistical techniques: the Glejser test of heteroscedasticity and nonlinear (quadratic) regression. On the basis of a sample of 929 general community participants who completed a self-assessment of intelligence and the Advanced Raven's Progressive Matrices, we failed to identify statistically significant heteroscedasticity, contrary to the Dunning-Kruger hypothesis. Additionally, the association between objectively measured intelligence and self-assessed intelligence was found to be essentially entirely linear, again, contrary to the Dunning-Kruger hypothesis. It is concluded that, although the phenomenon described by the Dunning-Kruger hypothesis may be to some degree plausible for some skills, the magnitude of the effect may be much smaller than reported previously.

## 1. Introduction

The Dunning-Kruger hypothesis states that the misestimation of ability is larger at the lower end of the spectrum of objectively measured ability than at the higher end of the spectrum of objectively measured ability (Kruger & Dunning, 1999). Several empirical investigations across a variety of abilities and skills have ostensibly supported the Dunning-Kruger hypothesis, on the basis of plotting the difference between the self-assessed and objectively measured ability means across four levels (quartiles) of the objectively measured ability (Dunning, 2011).

Some work critical of the Dunning-Kruger effect has suggested that the apparent phenomenon is likely a statistical artefact, i.e., the better-than-average effect and regression toward the mean (Krajc & Ortmann, 2008; Krueger & Mueller, 2002). However, the work critical of the Dunning-Kruger effect has arguably not gained broad awareness, as research ostensibly supportive of the Dunning-Kruger hypothesis continues to be published, without any reference to the critical work (e.g., Mahmood, 2016; Sullivan, Ragogna, & Dithurbide, 2018; West & Eaton, 2019). Perhaps not coincidently, easy to execute statistical approaches to testing the Dunning-Kruger hypothesis in a valid manner with individual differences data have not yet been described in the literature.

Consequently, the purpose of this investigation was to demonstrate with a basic simulation that the commonly reported Dunning-Kruger effect can be observed on the basis of data simulated to represent only the better-than-average effect and the regression toward the mean effect. Furthermore, we propose that the Dunning-Kruger hypothesis can be tested validly on individual differences data with two statistical approaches: an informative test of heteroscedasticity and/or nonlinear regression. Finally, we used these methods to test the Dunning-Kruger hypothesis with self-assessed intelligence scores and objectively measured intelligence scores.

### 1.1. The Dunning-Kruger hypothesis: Background

A substantial amount of empirical research has estimated the association between self-assessed ability and objectively measured ability across a variety of abilities. On the basis of a quantitative review of the meta-analyses in the area (e.g., academic achievement, intelligence, sports ability), Zell and Krizan (2014) reported a mean correlation of

* Corresponding author at: School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia.
*E-mail address:* gilles.gignac@uwa.edu.au (G.E. Gignac).

https://doi.org/10.1016/j.intell.2020.101449
Received 3 December 2019; Received in revised form 27 February 2020; Accepted 21 March 2020

0.29 between self-assessed ability and objectively measured ability. Such a value is essentially consistent with the meta-analytically estimated correlation of 0.33 between self-assessed intelligence and objectively measured intelligence (Freund & Kasten, 2012).

Although the $r \approx 0.30$ correlation between self-assessed intelligence may be considered relatively large for individual differences research (Gignac & Szodorai, 2016), the correlation is not sufficiently large to consider self-assessed intelligence as a respectable proxy of objectively measured intelligence (Paulhus, Lysy, & Yik, 1998). Nonetheless, research on self-assessed cognitive ability scores is considered important, as self-estimates of cognitive ability are often used in vocational counselling settings (Holling & Preckel, 2005). Additionally, people who underestimate their ability tend not to pursue careers in which they may reasonably be expected to succeed (Chipman, Krantz, & Silver, 1992; Ehrlinger & Dunning, 2003). By contrast, people who overestimate their abilities tend to cause more accidents and tend to make a greater number of poor decisions in life (Moore & Healy, 2008; Plumert, 1995; van de Venter & Michayluk, 2008). Consequently, work relevant to the self-assessment of abilities may be regarded as important.

To date, several moderators of the magnitude of the association between self-assessed ability and objectively measured ability has been identified (Freund & Kasten, 2012; Mabe & West, 1982). Arguably, the most famous proposed moderator of the self-assessed and objectively measured ability association is a person's objective ability on the dimension of interest. The moderation is known as the Dunning-Kruger effect (Dunning, 2011; Kruger & Dunning, 1999). The Dunning-Kruger effect is a type of cognitive bias, whereby less skilled/able people tend to overestimate the level to which they possess the skill/ability in question to a degree greater than people with more substantial levels of the skill/ability (Kruger & Dunning, 1999). This effect has been studied with regard to various skills, including logical reasoning, grammar, and personal sense of humor, for example (Dunning, 2011). It has been suggested that the reason the effect occurs is because the relative absence of the ability diminishes the capacity to evaluate the degree to which one possesses the ability (Dunning, 2011; Kruger & Dunning, 1999). Thus, the Dunning-Kruger effect is essentially considered to be a problem of systematic individual differences in meta-cognition (Schlösser, Dunning, Johnson, & Kruger, 2013).

### 1.2. Statistical approaches to the testing the Dunning-Kruger effect

A small number of approaches have been employed to test the Dunning-Kruger hypothesis statistically. Perhaps the most commonly used approach is the method used in the original paper by Kruger and Dunning (Kruger & Dunning, 1999). First, they measured the ability of interest via self-assessment (i.e., subjective ability) followed by the behavioural task (objective ability). Next, on the basis of quartile values, Kruger and Dunning (1999) created four categories of objective ability: low, low-average, high-average, and high. Then, they plotted the self-assessed ability and objectively measured ability means across the four objective ability categories. Finally, Kruger and Dunning (1999) showed that the magnitude of the difference between the self-assessed ability means and the objective ability means was larger at the lower end of the objective ability spectrum, in comparison to the higher objective ability spectrum. Stated alternatively, there was negative correlation between the four objective ability categories and the magnitude of the difference between self-assessed ability means and the objective ability means. Several researchers have employed the same approach to ostensibly support the Dunning-Kruger hypothesis across various skills and abilities (e.g., Pennycook, Ross, Koehler, & Fugelsang, 2017; Sullivan et al., 2018; West & Eaton, 2019).

An essentially identical approach involves calculating self-assessed ability and objective ability difference scores. Then, an oneway between-subjects ANOVA is conducted on the difference scores with the four objective ability categories (quartiles) as the independent variable (e.g., von Stumm, 2014). The observation of a significant difference in the means, with a downward slopping trend in the means, has been suggested to be supportive of the Dunning-Kruger hypothesis (Schlösser et al., 2013). A similar approach involves estimating the correlation between the self-assessed ability minus objective ability difference scores and objective ability (e.g., Adam & Vogel, 2017). The observation of a statistically significant negative correlation, in this context, implies ostensible support for the Dunning-Kruger hypothesis.

### 1.3. Criticisms and simulation

Krueger and Mueller (2002)[1] questioned whether the interpretation of the Dunning-Kruger effect, as conventionally tested, was due to a disproportionate lack of insight on the part of those who possessed less of the ability measured objectively. Specifically, they contended that the apparent statistical evidence may be due to a combination of two other phenomena: the better-than-average effect and regression toward the mean. The better-than-average effect represents the fact that the majority of the general population considers themselves above average across a number of skills and abilities (Mabe & West, 1982). In fact, only about 5% of the general population rates themselves below average in intelligence, for example (Gignac & Zajenkowski, 2019). Furthermore, on average, people tend to self-report their IQ to be approximately 115 (Gold & Kuhn, 2017) or possibly higher (Gignac & Zajenkowski, 2019). Thus, within the context of cognitive skills and ability, the better-than-average effect tends to be substantial. Furthermore, with respect to the Dunning-Kruger effect, specifically, if people, on average, self-report their IQ at 115, then it necessarily implies that the portion of the sample below an objectively determined IQ of 100 will have, on average, larger self-rated IQ versus objectively measured IQ discrepancy scores (i.e., greater overestimation), in comparison to the people who have an objective IQ above 100, assuming the correlation between the self-reported IQ scores and the objectively measured IQ scores is less than 1.0 (Krueger & Mueller, 2002). An imperfect correlation between the two variables opens up the possibility for regression toward the mean.

Regression toward the mean is said to occur when relatively distant values from the mean on $X$ are observed to be closer to the mean on $Y$ (Nesselroade, Stigler, & Baltes, 1980). Regression toward the mean is expected to occur when two variables are found to correlated imperfectly (i.e., $<r = 1.0|$; Campbell & Kenny, 1999). Krueger and Mueller (2002) pointed out that self-assessed and objectively measured abilities tend to correlate far from perfectly. Consequently, they argued that substantial regression toward the mean effects would be expected to occur within the data typically analysed within the Dunning-Kruger effect literature. In their original paper, Kruger and Dunning (1999) did acknowledge that regression toward the mean may have impacted their results, to some degree. However, they did not believe that the regression toward the mean effect could be so substantial as to account entirely for the meta-cognitive bias effect they proposed. Although Krueger and Mueller (2002) provided references for the importance of regression toward the mean, they did not provide a clear demonstration of the combination of the better-than-average effect and regression toward the mean as an alternative explanation for the Dunning-Kruger effect. It may be for this reason that researchers continue to report results ostensibly supportive of the Dunning-Kruger effect on the basis of the statistical approach originally employed by Kruger and Dunning (1999).

Fortunately, it is simple to demonstrate the better-than-average effect and regression toward the mean with simulated data. Specifically, we simulated data ($N = 1000$) for two variables: $X$ (say, objectively measured IQ) with a mean of 100 and a standard deviation of 15 and $Y$

[1] Not to be confused with Kruger, the co-originator of the Dunning-Kruger hypothesis.

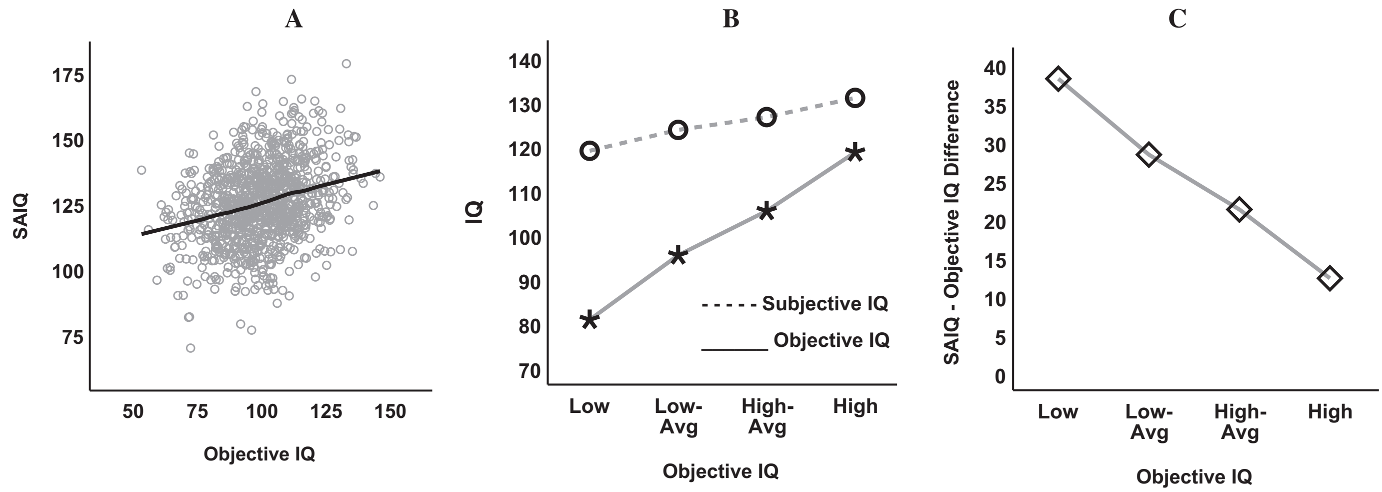


**Fig. 1.** Data ($N = 1000$) simulated to reflect a correlation of 0.30 and a better-than-average effect ($g = 1.67$); panel A = scatter plot depicting a linear association between objective IQ and self-assessed intelligence (SAIQ); the line of best fit was estimated via LOESS estimation (Epanechnikov; span = 50%); panel B = plot of subjective IQ and objective IQ means across the spectrum of objective IQ; panel C = plot of the SAIQ and objective IQ difference score means across the spectrum of objective IQ.

(say, self-assessed IQ) with a mean of 125 and a standard deviation of 15. Thus, the difference between the IQ means of 25 IQ points reflected the better-than-average effect, a magnitude of effect approximately equal to previously reported research (e.g., Gignac & Zajenkowski, 2019). Furthermore, on the basis of empirical literature that has shown a correlation of approximately 0.30 between self-assessed IQ and objectively measured IQ (Freund & Kasten, 2012), we specified the correlation between $X$ and $Y$ at 0.30. As can be seen in Fig. 1 (panel A), the nature of the simulated association between $X$ (objective IQ) and $Y$ (self-assessed IQ) was linear.

Next, consistent with the analytical approach used by Kruger and Dunning (1999), we created four groups of data on the basis of the quartiles of the simulated objective IQ scores: low, low-average, high-average, and high. Furthermore, we plotted the objectively measured means and the self-assessed IQ means in a chart, consistent with Kruger and Dunning (1999). As can be seen in Fig. 1 (panel B), the Dunning-Kruger effect may be suggested to be present, as the magnitude of the difference between the simulated objectively measured IQ scores and the simulated self-assessed IQ (SAIQ) scores was larger at the lower end of the objective IQ simulated scores. Correspondingly, the plot of the SAIQ and objective IQ simulated difference score means suggested a negative correlation between objective IQ and ability to self-estimate, again, consistent with the Dunning-Kruger effect (see Fig. 1, panel C). Of course, it is impossible for the Dunning-Kruger effect to reside within these data, as they were simulated to reflect purely the better-than-average effect and regression toward the mean. However, it is remarkable that the patterns of effects depicted in Fig. 1 (panels B and C) are similar to that reported across several investigations that reported evidence ostensibly supportive of the Dunning-Kruger hypothesis (e.g., Pennycook et al., 2017; Sullivan et al., 2018; West & Eaton, 2019).

Although potentially valid approaches to testing the Dunning-Kruger hypothesis could possibly be devised, based on comparing simulated data means (reflecting the better-than-average effect and regression toward the mean) against the corresponding field data means, such approaches would not be efficient. First, simulated data would need to be generated. Secondly, continuously scored data would need to be artificially categorised, an approach to data analysis that has been criticised (McClelland, Lynch Jr, Irwin, Spiller, & Fitzsimons, 2015). A more ideal approach to testing the Dunning-Kruger hypothesis with individual differences data would not require the simulation of data, nor would it require the coarse categorisation of a continuously scored variable's scores. Finally, the analysis would need to be not confounded by the better-than-average effect and regression toward the mean. As we describe next, two related data analytic approaches satisfy such criteria: (1) testing for heteroscedasticity of residuals; and (2) testing for nonlinearity.

### 1.4. Heteroscedasticity and the Dunning-Kruger hypothesis

Recall that the Dunning-Kruger hypothesis states that people at the lower end of the spectrum of an objectively measured ability have less meta-cognitive insight into that ability, which impacts negatively their capacity to evaluate the degree to which they possess that ability. Within the context of regressing self-assessed ability scores onto objectively measured scores, such a phenomenon implies that the regression residuals (i.e., degree of misprediction) would be larger at the lower-end of the objectively measured spectrum of the ability, in comparison to the higher-end of the objectively measured spectrum of the ability. Stated alternatively, evidence supportive of the Dunning-Kruger hypothesis for a sample of field data would imply that the linear regression model solution residual variance would be observed to be statistically significantly heteroscedastic.

Although several inferential statistics have been developed to test heteroscedasticity in residuals (Kaufman, 2013), an especially insightful test of heteroscedasticity is the Glejser test (Glejser, 1969), as it provides information on the form of heteroscedasticity (Koutsoyiannis, 1973). The Glejser test consists of, first, estimating the residuals associated with a regression solution. For example, regressing self-assessed ability (dependent variable) onto objectively measured ability. Next, the residuals are converted into absolute values. Then, for relatively normally distributed residuals, a Pearson correlation is estimated between the predicted values and the absolute residuals. The observation of a statistically significant, negatively directed Glejser test correlation would be considered supportive of the Dunning-Kruger hypothesis, as it would imply that objective intelligence's predictive capacity of self-assessed intelligence increases (i.e., smaller residuals), as objective intelligence increases (use bootstrapping if asymmetry in the distribution of residuals is suspected; Godfrey, Orme, & Santos Silva, 2006). Thus, the direction of the Glejser test correlation provides information about the form of heteroscedasticity, unlike many other tests of heteroscedasticity (e.g., Breusch-Pagan test; see Koutsoyiannis, 1973). By contrast, a non-significant Glejser test correlation, or a significant, positive Glejser test correlation, would be non-supportive of the Dunning-Kruger hypothesis.

To our knowledge, the Dunning-Kruger hypothesis has never been tested with a test for heteroscedasticity. Furthermore, although a substantial number of studies have estimated the correlation between self-assessed intelligence and objectively measured intelligence, none of these studies have reported any results relevant to heteroscedasticity.

### 1.5. Nonlinear regression and the Dunning-Kruger hypothesis

In addition to testing the Dunning-Kruger hypothesis with an informative test of heteroscedasticity, perhaps the most insightful and straightforward test of the Dunning-Kruger hypothesis would involve determining whether the association between the objectively measured ability scores and the self-assessed scores is nonlinear. It is well-established that, in the context of regression, heteroscedasticity and nonlinearity are considered related statistical phenomena (Carroll & Spiegelman, 1992). That is, when a bivariate linear model is estimated from the data, and a nonlinear association exists between the *X* and *Y* variables, the model solution will yield unequal residual variances across the spectrum of the *X* and *Y* variables (Kaufman, 2013). However, it is also possible to observe heteroscedasticity for an entirely linear association between two variables (Wilcox & Muska, 2001). Thus, both statistical analyses can offer important, unique information, when a full evaluation of the nature of the association between self-assessed ability and objectively measured ability is sought.

Recall, the Dunning-Kruger hypothesis represents the notion that people at the lower end of the ability spectrum have lesser capacity at discerning the degree to which they possess the ability (i.e., meta-cognition). Such an effect implies that the magnitude of the correlation between self-assessed ability and objectively measured ability increases across the spectrum of objectively measured ability. For example, the correlation between self-assessed intelligence and objectively measured intelligence has been reported at $r \approx 0.30$ (Freund & Kasten, 2012). However, for people with less objective ability, and correspondingly less meta-cognition for that ability, the correlation may be expected to be closer to between 0.00 and 0.10 at the lower end of the objective ability spectrum, if the Dunning-Kruger effect is a plausible phenomenon. By comparison, at the higher end of the intellectual ability spectrum, the correlation between self-assessed and objectively measured intelligence may be expected to be closer to 0.35 to 0.45, if the Dunning-Kruger effect resides within the data. In statistical terms that are consistent with nonlinear regression analyses (Pedhazur, 1997), the Dunning-Kruger effect would be considered a plausible account of the data, if the nature of the association between self-assessed and objective measured ability were to be observed to be consistent with a statistically significant, positive, monotonic, quadratic effect, as depicted with the simulated data, in Fig. 2, for example.

Nonlinear effects, such as a quadratic effect (i.e., one bend in the line of best fit), can be tested via hierarchical multiple regression, where the linear term is entered at step 1 and the nonlinear (quadratic) term is entered at step 2 (Pedhazur, 1997). Typically, a quadratic term in nonlinear hierarchical regression analysis is represented by squared values of *X* (Pedhazur, 1997). A statistically significant change in $R^2$ would be considered supportive of a nonlinear (quadratic) effect between *X* and *Y*. Furthermore, a positively directed quadratic effect (i.e., positive beta-weight or positive semi-partial correlation) would imply that the magnitude of the positive association increases across the spectrum of the *X* and *Y* variables.

When we estimated a quadratic effect on the basis of the simulated data depicted in Fig. 2, the quadratic effect beta-weight, controlling for the linear effect, was found to be significant statistically, $R^2_{\text{change}} = 0.008$, $F(1, 997) = 9.35$, $p = .002$; $b = 0.004$, $\beta = 0.87$, semi-partial $r = 0.09$, $p < .001$. Correspondingly, the Glejser test yielded a statistically significant, negative correlation $r = -0.59$, $p < .001$. Thus, had the data depicted in Fig. 2 been collected from the field, both statistical results would have been supportive of the Dunning-Kruger hypothesis.

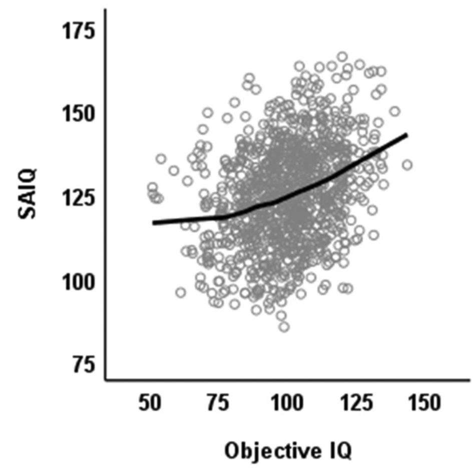


**Fig. 2.** Simulated data ($N = 1000$) to reflect a positive, monotonic, quadratic effect (i.e., supportive of the Dunning-Kruger hypothesis); SAIQ = self-assessed intelligence.

To our knowledge, no work within the Dunning-Kruger literature has tested the Dunning-Kruger hypothesis with a test of heteroscedasticity or nonlinear (curvilinear) regression on field data. Furthermore, within the self-assessed intelligence and objectively measured intelligence literature more broadly, little work has explored the precise nature of the association. In a rare exception, Holling and Preckel (2005) included a scatter plot for their self-assessed and objectively measured IQ scores. However, although the scatter plot bore some resemblance to a linear effect, they did not specifically test for the possibility of a nonlinear effect, statistically. Additionally, the Holling and Preckel's (2005) investigation was based on a sample of only 88 high school students, a sample size insufficiently powerful to detect typically reported moderator effects in the behavioural sciences (Shieh, 2009).

### 1.6. Summary and purpose

The Dunning-Kruger effect has been ostensibly replicated across a number of abilities and skills (Dunning, 2011). However, much, if not all, of the differential psychology results in the area are confounded by the better-than-average effect and regression toward the mean. We have contended above that the Dunning-Kruger hypothesis can be tested more validly on individual differences data with statistical techniques such as an informative test of heteroscedasticity (i.e., Glejser test) and nonlinear (quadratic) regression.

Consequently, the purpose of the following empirical investigation was to evaluate the nature of the association between self-assessed intelligence (SAIQ) and objectively measured IQ with a large, general community sample. First, we hypothesized that the SAIQ mean would be larger than the objectively measured IQ mean, consistent with the better-than-average effect. We also hypothesized that the correlation between the SAIQ scores and the objectively measured IQ scores would be positive and approximately 0.30 in magnitude. Finally, we investigated whether the nature of the association was supportive of the Dunning-Kruger hypothesis, on the basis of the Glejser test and nonlinear (quadratic) regression.

## 2. Method

### 2.1. Sample

To maximize power, we combined data across three samples, two of which have been analysed previously for different purposes (Gignac & Zajenkowski, 2019; Zajenkowski & Gignac, 2018). The overall sample

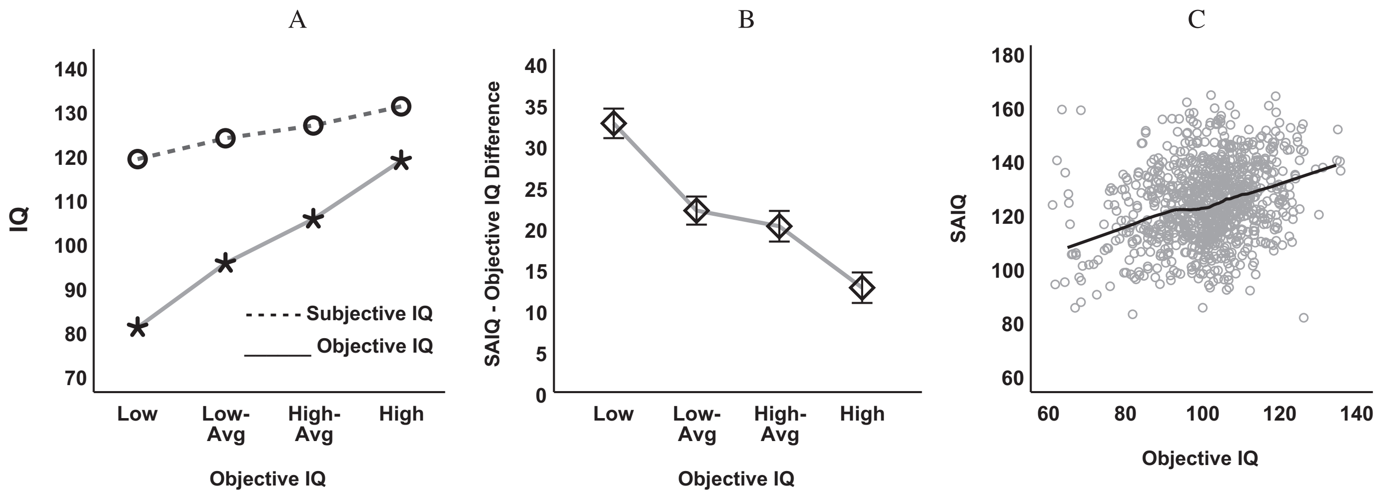


**Fig. 3.** Tests of the Dunning-Kruger hypothesis with field data ($N$ = 929); panel A = plot of subjective IQ and objective IQ means across the spectrum of objective IQ; panel B = plot of the self-assessed intelligence (SAIQ) and objective IQ difference score means across the spectrum of objective IQ (with 95% confidence intervals); panel C = scatter plot depicting a linear association between objective IQ and SAIQ ($N$ = 929); the line of best fit was estimated via LOESS (Epanechnikov; span = 50%).

consisted of 1071 unique cases. However, 139 cases were missing SAIQ scores. Additionally, three cases were missing age. Thus, the total working sample consisted of 929 cases. We identified four suspected outlying SAIQ values (3, 3, 5, and 6) on the basis of the outlier labelling rule with a 1.5 multiplier (Hoaglin & Iglewicz, 1987) and a combined inspection of the histogram. Consequently, we winsorized all four values to the lowest SAIQ value not suspected to be an outlier (i.e., a value of 10). An evaluation of the Ravens total scores did not identify any outlying observations. The total working sample consisted of 50.4% females (age $M$ = 26.92; $SD$ = 9.15; skew = 2.09). Complete data were uploaded to Open Science Framework and are available under the following address: osf.io/dg547.

### 2.2. Measures

#### 2.2.1. Subjectively assessed intelligence

Participants assessed their own intelligence on a scale ranging from 1 to 25 (see Zajenkowski, Stolarski, Maciantowicz, Malesza, & Witowska, 2016). Five groups of five columns were labelled as very low, low, average, high or very high, respectively (see Fig. S1). Participants' SAIQ was indexed with the marked column counting from the first to the left; thus, the scores ranged from 1 to 25. Prior to providing a response to the scale, the following instruction was presented:

"People differ with respect to their intelligence and can have a low, average or high level. Using the following scale, please indicate where you can be placed compared to other people. Please mark an X in the appropriate box corresponding to your level of intelligence."

In order to place the 25-point scale SAIQ scores onto a scale more comparable to a conventional IQ score (i.e., $M$ = 100; SD = 15), we transformed the scores such that values of 1, 2, 3, 4, 5… 21, 22, 23, 24, 25 were recoded to 40, 45, 50, 55, 60… 140, 145, 150, 155, 160. As the transformation was entirely linear, the results derived from the raw scale SAI scores and the recoded scale SAI scores were the same.

#### 2.2.2. Objectively assessed intelligence

Participants completed the Advanced Progressive Matrices (APM; Raven, Court, & Raven, 1994). The APM is a non-verbal intelligence test which consists of items that include a matrix of figural patterns with a missing piece. The goal is to discover the rules that govern the matrix and to apply them to the response options. The APM is considered to be less affected by culture and/or education (Raven et al., 1994). It is known as good, but not perfect, indicator of general intellectual functioning (Carroll, 1993; Gignac, 2015). We used the age-based norms published in Raven et al. (1994, p. 55) to convert the raw APM scores into percentile scores. We then converted the percentile scores into $z$-scores with the IDF.NORMAL function in SPSS. Then, we converted the $z$-scores into IQ scores by multiplying them by 15 and adding 100. Although the norms were relatively old, we considered them essentially valid, given evidence that the Flynn effect had slowed down considerably by 1980 to 1990 and may have even reversed to a small degree since the early 1990s (Woodley of Menie et al., 2018).

### 2.3. Procedure

Across all three samples of data, each participant was tested individually in a quiet laboratory at the University of Warsaw. Participants first completed a demographic survey and a set of self-report measures including SAIQ. Afterwards, participants were administered the Raven's test.

## 3. Results

All analyses were conducted with IBM SPSS (Version 25). The SAIQ (skew = −0.54; kurtosis = 0.89) and objective IQ (skew = 0.08; kurtosis = −0.26) scores were associated with essentially normally distributed scores. Consequently, parametric statistical analyses were considered appropriate. The SAIQ scores (range: 85/160; inter-quartile range: 115/135) and the objective IQ scores (range: 65/135; inter-quartile-range: 96/109) were also representative of a wide spectrum of ability, suggesting the sample was not disproportionately sampled from one end of the distribution in the population. The SAIQ mean ($M$ = 123.76; $SD$ = 14.19) was statistically significantly larger than the objective IQ mean ($M$ = 101.70; $SD$ = 11.63), $t(928)$ = 43.02, $p$ < .001, Cohen's $d$ = 1.71). Thus, on average, people estimated their IQ to be higher than that verified by their IQ measured objectively, as hypothesized. Furthermore, the Pearson correlation between SAIQ and objective IQ was positive and significant, $r(927)$ = 0.28, 95%CI: 0.22/0.34, $p$ < .001, as hypothesized.

Next, consistent with the procedure commonly used to test the Dunning-Kruger effect (i.e., that confounds the better-than-average effect and regression toward the mean), we separated the sample into four objective IQ categories: low ($M$ = 86.77), low-average ($M$ = 99.75), high-average ($M$ = 106.03), and high ($M$ = 115.68). As can be seen in Fig. 3 (panel A), we found ostensible evidence supportive of the Dunning-Kruger hypothesis, as the difference between the SAIQ and objective IQ means was larger at the lower end of the objective IQ spectrum (mean difference: low = 32.63; low-average = 22.05; high-average = 20.15; high = 12.68). Next, SAIQ and objective IQ difference scores were calculated (skew = 0.32; kurtosis = 0.88). A between-subjects oneway ANOVA test of the difference between the mean difference scores across the four levels of objective IQ was significant statistically, $F(3, 925) = 79.00$, $p < .001$, $\eta^2 = 0.20$. Furthermore, a linear contrast analysis was found to be significant statistically, $F(1, 925) = 222.16$, $p < .001$. As can be seen in Fig. 3 (panel B), the trend in the difference score means was downward slopping, suggesting ostensible support for the Dunning-Kruger hypothesis.

Next, in order to test the Dunning-Kruger hypothesis in a more valid (less confounded) manner, a linear, bivariate regression analysis was conducted, whereby the SAIQ scores were regressed onto the objective IQ scores, and the residuals were saved and converted into absolute values. The absolute residuals were essentially normally distributed (skew = 0.08; kurtosis = −0.26). The correlation between the objective IQ scores and the absolute residuals (i.e., the Glejser test correlation) was not found to be significant statistically, $r(927) = -0.05$, 95%CI: −0.11/0.02, $p = .132$, suggesting the data were homoscedastic, which did not support the Dunning-Kruger hypothesis.

Finally, we evaluated the nature of the association between objective IQ and SAIQ further with a curvilinear regression analysis. The hierarchical multiple regression failed to identify a statistically significant quadratic effect, $R^2_{\text{change}} \leq 0.01$, $F(1, 926) = 0.37$, $p = .545$; $b < 0.01$, $\beta = 0.20$, semi-partial $r = 0.02$, $p = .545$, suggesting, again, a failure to support the Dunning-Kruger hypothesis. Furthermore, as can be seen in Fig. 3 (panel C), the LOESS regression line[2] reflected an essentially linear effect across the entire spectrum of the objective IQ scores.

## 4. Discussion

We found evidence for the better-than-average effect for fluid intelligence. The magnitude of the effect ($d = 1.71$) would be considered large, based on Cohen's (1992) guidelines. Although a meta-analysis of the better-than-average effect for intelligence has not been conducted, the magnitude of the better-than-average effect observed in this investigation may be on the larger side to what may be expected, on the basis of previous empirical investigations in the area of cognitive abilities ($d \approx 0.50$ to 1.0; Gold & Kuhn, 2017; Reilly & Mulhern, 1995). Nonetheless, based on the results of this investigation, as well as the broader better-than-average effect literature (Alicke & Govorun, 2005), it is plausible to suggest that researchers of the Dunning-Kruger hypothesis should expect the better-than-average effect to influence their data in a non-negligible manner.

We also found that self-assessed intelligence and objectively measured intelligence correlated positively and statistically significantly at $r = 0.28$, which is comparable to the meta-analytically estimated correlations reported in the literature ($r = 0.33$; Freund & Kasten, 2012). Thus, it may be argued that people do self-assess their cognitive intelligence with some level of accuracy, even when self-assessed intelligence is measured with a single item. Of course, the correlation is not so large as to suggest that self-assessed intelligence may be used as a proxy of objectively measured intelligence (Paulhus et al., 1998). However, it is important to note that the magnitude of the correlation (i.e., substantially less than 1.0) implies that substantial regression toward the mean effects need to be considered, when testing the Dunning-Kruger hypothesis (Krueger & Mueller, 2002).

In fact, based on the typical approach to testing the Dunning-Kruger hypothesis, we observed an ostensibly clear Dunning-Kruger effect with the field intelligence data (see Fig. 3, panels A and B). That is, we observed more substantial mean differences between self-assessed IQ and objectively measured IQ at the lower end of the objectively measured IQ spectrum. Furthermore, the magnitude of the difference in the mean differences across the spectrum of objectively measured IQ was roughly comparable to that reported in other investigations putatively supportive of the Dunning-Kruger hypothesis (e.g., Pennycook et al., 2017; von Stumm, 2014; West & Eaton, 2019).

However, when we tested the Dunning-Kruger hypothesis less ambiguously with a test of homoscedasticity, we failed to find evidence supportive of the Dunning-Kruger hypothesis. Specifically, the Glejser test correlation was near zero ($r = -0.05$), suggesting that the degree to which people mispredicted their objectively measured intelligence was equal across the whole spectrum of objectively measured intelligence. Correspondingly, when we examined the field intelligence data with nonlinear regression, we failed to find any evidence to suggest an association consistent with a quadratic effect between objectively measured intelligence and self-assessed intelligence. In summary, the results associated with our recommended statistical analyses implied that the accuracy with which people self-assess their intelligence is essentially equal across the whole spectrum of objectively measured intelligence.

As few, if any, investigations in the area have tested the Dunning-Kruger hypothesis with a test of heteroscedasticity or nonlinear regression, we cannot compare our results directly with the results reported in other published field investigations. However, in light of our simulated data and field data results, we suspect a large number of empirical studies that have published ostensible support for the Dunning-Kruger effect on the basis of the conventional approach to testing the hypothesis would have likely failed to observe the Dunning-Kruger effect, had an analysis like the Glejser test or a test of nonlinear regression been performed. In our view, tests of heteroscedasticity and nonlinear regression, statistics with a well-established statistical background (Cohen, 1978; Cook & Weisberg, 1983), are much less ambiguous tests of the Dunning-Kruger hypothesis, as there has been no suggestion in the literature that these analyses are confounded by the better-than-average effect and/or regression toward the mean.

### *4.1. Alternative explanations for SAIQ and IQ discrepancies*

If, as we suggest below, the Dunning-Kruger effect is likely, at most, only a weak moderator of the association between objectively measured ability and self-assessed ability, the question of what factors might explain the divergence between how able people think they are and how able they actually are remains. There is some evidence to suggest that the psychological processes may be to some degree consciously motivated and/or relevant to personality type trait variance.

For example, Gold and Kuhn (2017) found that people self-assessed their intelligence, on average, five IQ points lower, after completing an intelligence test, in comparison to those who self-assessed their IQ before completing the intelligence test (Gold & Kuhn, 2017). Such a result suggests that people do recalibrate in a rational manner their self-perceived cognitive ability on the basis of experience (i.e., less better-than-average effect). However, Gold and Kuhn (2017) found that their participants' increase in insight into their cognitive ability disappeared, when the participants self-assessed their IQ again one week later (i.e., the five IQ point difference was no longer identified upon re-self-assessment). On the basis of the Gold and Kuhn (2017) results, the long-term benefits of the training (e.g., feedback; skills) provided in some studies designed to reduce the magnitude of the Dunning-Kruger effect may be questioned (e.g., Callender, Franco-Watkins, & Roberts, 2016; study 4 of Kruger & Dunning, 1999). In another relevant study, Shepperd (1993) found that when rewarded for self-perception accuracy, poor performers on the SAT reduced substantially the degree to which they overestimated their SAT scores.

[2] Locally estimated scatter plot smoothing (LOESS) is a non-parametric approach to estimating a line of best fit that is more sensitive to possible idiosyncrasies within a set of data (Cleveland & Devlin, 1988).

The results of the two studies reviewed above suggest that the discrepancy between self-assessed ability and objectively measured ability likely occurs, at least to some degree, through motivated processes, rather than ignorance of one's ignorance (i.e., unskilled and unaware of it). It is noteworthy that trait narcissism has also been found to be a substantial predictor of the discrepancy between self-assessed intelligence and objectively measured intelligence (e.g., Dufner et al., 2012; Zajenkowski, Czarna, Szymaniak, & Dufner, 2019). Additionally, self-deceptive enhancement and impression management have also been found to be explanatory factors of the misestimation of ability (Balcetis, 2008; Gignac, 2018). Importantly, narcissism, self-deceptive enhancement, and impression management tend not to be associated with objectively measured general intelligence (O'Boyle, Forsyth, Banks, & Story, 2013). Thus, from this perspective, the misestimation of one's cognitive ability appears to arise, at least in part, through processes entirely independent of one's objective ability, in distinct contrast to the Dunning-Kruger hypothesis. Of course, more factors need to be uncovered to account for all of the variability in cognitive ability misestimation.

### 4.2. Practical considerations

We note that all inferential statistical tests are sensitive to sample size (Cohen, 1988). In this investigation, the sample size was over 900, yielding substantial power to reject the homoscedasticity null hypothesis. However, sample sizes less than 200 to 300 might not be expected to achieve respectable power (say, $\geq$ 0.80), with respect to testing the Dunning-Kruger hypothesis, when tested with the Glejser test and/or nonlinear regression (Aguinis, 1995; Harvey & Phillips, 1974). Additionally, the objective IQ scores and the absolute residuals were relatively normally distributed in this investigation, which allowed for asymptotic normal theory hypothesis testing. Had some question about the symmetry of the scores been raised, Monte Carlo simulation research suggests that Glejser test of homoscedasticity and nonlinear regression terms can tested validly via bootstrapping (Godfrey & Orme, 1999; Hall & Pittelkow, 1990).

In addition to evaluating the weight of evidence (within either a frequentist or Bayesian paradigm), researchers should take an effect size approach to evaluating the Dunning-Kruger hypothesis, when tested via the Glejser test and the nonlinear regression test. In the absence of any other specifically published guidelines for the Glejser test, statistically significant correlations of 0.10, 0.20, and 0.30 may be considered relatively small, typical, and relatively large, in individual differences research (Gignac & Szodorai, 2016). Additionally, with respect to nonlinear regression, a quadratic term that accounts for 2 to 4% additional variance in self-assessed intelligence may be regarded as minimally substantively significant, on the basis of the unique predictive validity recommendations suggested by Hunsley and Meyer (2003).

We note the possibility that the Glejser test and the nonlinear regression (quadratic) test will not necessarily yield results consistent with each other, in all cases. For example, one analysis may be more powerful statistically, under certain conditions. Monte Carlo simulation research could help address this issue. Currently, we recommend the statistical significance of both statistical analyses, in order to support the Dunning-Kruger hypothesis. Finally, it may be useful to consider novel methods to test for curvilinearity in regression (e.g., Simonsohn, 2018), as opposed to the well-known polynomial (quadratic) regression approach adopted in this investigation.

### 4.3. Limitations

When analysing composite scores, internal consistency reliability is a psychometric characteristic known to impact the magnitude of an observed correlation (Nunnally & Bernstein, 1994). In the context of the methods proposed in this investigation to test the Dunning-Kruger hypothesis validly, it was assumed that the reliability of the test scores was approximately equal across the whole spectrum of ability. That is, if the APM test score reliability were relatively lower at the higher-end of the ability spectrum, then the observed correlation would have been proportionately smaller at the higher-end of the ability spectrum, in comparison to the corresponding true score correlation. Ultimately, if the APM test score reliability were substantially variant across the spectrum of the APM scores in our sample, then the observed score linear association depicted in Fig. 3 (panel C) would be misleading.

Methods to evaluate the test score reliability of composite scores across the continuum of ability are not yet well established (local structural equation modeling may offer some opportunities; Hildebrandt, Lüdtke, Robitzsch, Sommer, & Wilhelm, 2016). Nonetheless, in order to evaluate, in an approximate manner, the possibility that the APM scores were associated with differential levels of internal consistency reliability in our sample, we estimated the correlation between quasi-parallel forms of the Advanced Raven's Progressive Matrices (i.e., even-numbered and odd-numbered item halves). We found the correlation between the two halves to be essentially linear (see supplementary materials, Fig. S2). Consequently, we do not believe the possibility of varying reliability in the test scores played a meaningful role in this investigation.

Additionally, our investigation was based on individual differences data, consistent with the majority of the research that has tested the Dunning-Kruger hypothesis. Admittedly, the tests of heteroscedasticity and nonlinear regression we employed in this investigation may be regarded as only appropriate for such data. It is important to note, however, that some research interpreted to be supportive of the Dunning-Kruger hypothesis is experimental in nature (e.g., Nguyen, 2018). Consequently, the statistical artefact account of the Dunning-Kruger effect described in this investigation may not apply to those investigations. Additionally, we suspect that the magnitude of the Dunning-Kruger effect reported in some samples (e.g., study 1, Kruger & Dunning, 1999) appear to be so large that the better-than-average effect and regression toward the mean may not account fully for the pattern of results. Thus, at this stage, our contention is that the Dunning-Kruger effects reported in the literature are mostly the result of statistical artefacts, rather than entirely so. As studies that test the Dunning-Kruger hypothesis with the tests recommended in this investigation accumulate in the literature, a more precise partitioning of effects (genuine versus statistical artefact) will be possible.

We also acknowledge that we tested the Dunning-Kruger hypothesis with only a single measure of ability. Although progressive matrices tests may share as much as 50% of their variance with general intelligence, it should be acknowledged that Raven's is not identical with general intellectual functioning (Gignac, 2015). Furthermore, although Dunning (2011) suggested that the Dunning-Kruger effect may be a general process, Dunning (2011) also suggested that there may be some relevant specific variance associated with narrower competencies/skills (Dunning, 2011). Consequently, it is possible that the statistical techniques we employed in this investigation may confirm the Dunning-Kruger effect, when tested on one or more specific abilities/skills. Finally, although analysed with the conventional (less valid) approach, some research suggests that the Dunning-Kruger effect may be moderated by task difficulty (Burson, Larrick, & Klayman, 2006). Thus, it is possible that the Dunning-Kruger effect may be identified for some cognitive abilities not measured in this investigation, when tested with valid statistical approaches.

## 5. Conclusion

The Dunning-Kruger hypothesis states that incompetent individuals tend to overestimate their ability to a larger degree than more competent individuals. To date, individual differences studies ostensibly supportive of the Dunning-Kruger hypothesis have failed to take into consideration statistical artefacts, such as the better-than-average effect and regression toward the mean. Perhaps a key reason for this lack of consideration was based on the absence of any demonstrated and easily implemented statistical procedures to do so. With tests such as the Glejser test and nonlinear regression, appropriate statistical tests are available. When such valid statistical analyses are applied to individual differences data, we believe that evidence ostensibly supportive of the Dunning-Kruger hypothesis derived from the mean difference approach employed by Kruger and Dunning (1999) will be found to be substantially overestimated.

## Acknowledgements

Research funding: Grant no 2016/23/B/HS6/00312. from the National Science Centre in Poland.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2020.101449.

## References

Adam, K. C., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, & Psychophysics, 79*(5), 1506–1523.
Aguinis, H. (1995). Statistical power with moderated multiple regression in management research. *Journal of Management, 21*(6), 1141–1158.
Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. Dunning, & J. Krueger (Eds.). *The self in social judgment* (pp. 85–106). New York: Psychology Press.
Balcetis, E. (2008). Where the motivation resides and self-deception hides: How motivated cognition accomplishes self-deception. *Social and Personality Psychology Compass, 2*(1), 361–381.
Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.
Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning, 11*(2), 215–235.
Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: The Guilford Press.
Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.
Carroll, R. J., & Spiegelman, C. H. (1992). Diagnostics for nonlinearity and heteroscedasticity in errors-in-variables regression. *Technometrics, 34*(2), 186–196.
Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *Psychological Science, 3*, 292–295. https://doi.org/10.1111/j.1467-9280.1992.tb00675.x.
Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596–610.
Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin, 85*(4), 858–866.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. 2nd.* Hillsdale, NJ: Erlbaum.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika, 70*(1), 1–10.
Dufner, M., Denissen, J. J. A., van Zalk, M., Matthes, B., Meeus, W. H. J., van Aken, M. A. G., & Sedikides, C. (2012). Positive intelligence illusions: On the relation between intellectual self-enhancement and psychological adjustment. *Journal of Personality, 80*, 537–571.
Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. Olson, & M. P. Zanna (Vol. Eds.), *Advances in experimental social psychology. 44. Advances in experimental social psychology* (pp. 247–296). New York: Elsevier.
Ehrlinger, J., & Dunning, D. A. (2003). How chronic self views influence (and mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5–17. https://doi.org/10.1037/0022-3514.84.1.5.
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. https://doi.org/10.1037/a0026556.
Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence, 52*, 71–79.
Gignac, G. E. (2018). Socially desirable responding suppresses the association between self-assessed intelligence and task-based intelligence. *Intelligence, 69*, 50–58.
Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.
Gignac, G. E., & Zajenkowski, M. (2019). People tend to overestimate their romantic partner's intelligence even more than their own. *Intelligence, 73*, 41–51.
Glejser, H. (1969). A new test for heteroskedasticity. *Journal of the American Statistical Association, 64*(325), 316–323.
Godfrey, L. G., & Orme, C. D. (1999). The robustness, reliability and power of heteroskedasticity tests. *Econometric Reviews, 18*(2), 169–194.
Godfrey, L. G., Orme, C. D., & Santos Silva, J. M. C. (2006). Simulation-based tests for heteroskedasticity in linear regression models: Some further results. *The Econometrics Journal, 9*(1), 76–97.
Gold, B., & Kuhn, J. T. (2017). A longitudinal study on the stability of self-estimated intelligence and its relationship to personality traits. *Personality and Individual Differences, 106*, 292–297.
Hall, P., & Pittelkow, Y. E. (1990). Simultaneous bootstrap confidence bands in regression. *Journal of Statistical Computation and Simulation, 37*(1–2), 99–113.
Harvey, A. C., & Phillips, G. D. A. (1974). A comparison of the power of some tests for heteroskedasticity in the general linear model. *Journal of Econometrics, 2*(4), 307–316.
Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research, 51*(2–3), 257–258.
Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association, 82*(400), 1147–1149.
Holling, H., & Preckel, F. (2005). Self-estimates of intelligence—Methodological approaches and gender differences. *Personality and Individual Differences, 38*(3), 503–517.
Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*, 446–455.
Kaufman, R. L. (2013). *Heteroskedasticity in regression: Detection and correction.* Thousand Oaks, CA: Sage.
Koutsoyiannis, A. (1973). *Theory of econometrics: An introductory exposition of econometric methods.* London: MacMillan.
Krajc, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology, 29*(5), 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*(2), 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. https://doi.org/10.1037/0022-3514.77.6.1121.
Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280–296.
Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy, 10*(2), 3.
McClelland, G. H., Lynch, J. G., Jr., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, type II errors, and false–positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology, 25*(4), 679–689.
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517.
Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88*, 622–631.
Nguyen, T. B. (2018). *The Dunning-Kruger effect: Are the unskilled always unaware? (Doctoral dissertation).* Claremont, USA: The Claremont Graduate University.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
O'Boyle, E. H., Forsyth, D., Banks, G. C., & Story, P. A. (2013). A meta-analytic review of the dark triad–intelligence connection. *Journal of Research in Personality, 47*(6), 789–794.
Paulhus, D. L., Lysy, D. C., & Yik, M. S. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*(4), 525–554.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Melbourne: Wadsworth/ Thompson.
Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review, 24*(6), 1774–1784.
Plumert, J. M. (1995). Relations between children's overestimation of their physical abilities and accident proneness. *Developmental Psychology, 31*, 866–876.
Raven, J. C., Court, J. H., & Raven, J. (1994). *Manual for Raven's progressive matrices and Mill Hill vocabulary scales. Advanced progressive matrices.* Oxford: Oxford Psychologists Press.
Reilly, J., & Mulhern, G. (1995). Gender differences in self-estimated IQ: The need for care in interpreting group data. *Personality and Individual Differences, 18*(2), 189–192.
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.
Shepperd, J. A. (1993). Student derogation of the scholastic aptitude test: Biases in perceptions and presentations of college board scores. *Basic and Applied Social Psychology, 14*(4), 455–473.
Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. *Organizational Research Methods, 12*(3), 510–528.
Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science, 1*(4), 538–555.
von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology, 105*(2), 243–253.

Sullivan, P. J., Ragogna, M., & Dithurbide, L. (2018). An investigation into the Dunning–Kruger effect in sport coaching. *International Journal of Sport and Exercise Psychology,* 1–9.
van de Venter, G., & Michayluk, D. (2008). An insight into overconfidence in the forecasting abilities of financial advisors. *Australian Journal of Management, 32*(3), 545–557.
West, K., & Eaton, A. A. (2019). Prejudiced and unaware of it: Evidence for the Dunning-Kruger model in the domains of racism and sexism. *Personality and Individual Differences, 146*, 111–119.
Wilcox, R. R., & Muska, J. (2001). Inferences about correlations when there is heteroscedasticity. *British Journal of Mathematical and Statistical Psychology, 54*(1), 39–47.
Woodley of Menie, M.A, Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Figueredo, A. J. (2018). What causes the anti-Flynn effect? A data synthesis and analysis of predictors. *Evolutionary Behavioral Sciences, 12*(4), 276–295.
Zajenkowski, M., Czarna, A., Szymaniak, K., & Dufner, M. (2019). What do highly narcissistic people think and feel about (their) intelligence? *Journal of Personality* in press.
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.
Zajenkowski, M., Stolarski, M., Maciantowicz, O., Malesza, M., & Witowska, J. (2016). Time to be smart: Uncovering a complex interplay between intelligence and time perspectives. *Intelligence, 58*, 1–9.
Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111–125.