MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„You (don't) know best what you can do: A meta-analysis of self-assessed and psychometric intelligence"

verfasst von / submitted by

Sabine Patzl, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Jakob Pietschnig, Privatdoz.

Danksagung (Acknowledgments)

Wie ein bekanntes Sprichwort bereits sagt: „ It takes a village to write a master's thesis." So oder so ähnlich. Daher möchte ich an dieser Stelle einigen Personen danken, die mich beim Verfassen dieser Arbeit unterstützt haben:

Mein Dank gebührt zunächst meinem Masterarbeitsbetreuer Jakob Pietschnig für seine großartige Betreuung und seine Unterstützung bei der Verwirklichung meiner wissenschaftlichen Karriere.

Ich danke auch insbesondere meiner Freundin, Leidensgenossin, Studienkollegin und Nachbarin, Lisa Bucher. Danke für all die lustigen Abende, an denen wir Bier getrunken und über die Wissenschaft und Forschungsmethoden diskutiert haben. Danke für die gefühlt 1000 Seiten, die du Korrektur gelesen hast. Und einfach danke dafür, dass du genauso weird und nerdig bist, wie ich.

Meinen größten Dank muss ich natürlich meinem Freund Stefan Berndl aussprechen. Dafür, dass er es immer schafft mich zu beruhigen, wenn der Stress einmal doch zu viel wird und einfach immer die richtigen Worte findet. Dafür, dass er mein größter Supporter ist und natürlich dafür, dass er mich daran erinnert auch auf mich selbst zu schauen, wenn ich wieder einmal nur am Arbeiten bin und vergesse Pausen einzulegen.

Abschließend danke ich natürlich auch meiner Familie. Dafür, dass sie sich stundenlange Ausführungen und Erzählungen über meine Masterarbeit angehört hat und dafür, dass sie meine Freuden teilt, auch wenn die Wissenschaft gar nicht so ihr Ding ist. Insbesondere danke ich auch meiner Mutter, die in allen Lebensbereichen mein größtes Vorbild ist, und meinem Vater, der mich zu der willenstarken Person gemacht hat, die ich heute bin und der unglaublich stolz gewesen wäre. Und um mit seinen Worten zu schließen: „Lossts eich ned ärgern"; viel Spaß beim Lesen dieser Arbeit und danke vielmals für Alles,

Sabine

Abstract

Self-assessed intelligence (= SAI) plays an important role in high-stakes situations, for example, when making career decisions. Research shows that professional interests are primarily a function of self-assessed abilities rather than “true” abilities. However, previous accounts revealed that self-assessed intelligence and psychometric intelligence only correlate moderately. However, ever-increasing publication numbers of studies investigating this topic and the development of novel and more refined research synthesis methods have rendered the available meta-analytical evidence outdated. Moreover, as recent findings suggest potential bias due to the virtual ubiquitous declining effects in empirical research, an update seems necessary. Consequently, a total of 242 effect sizes from 98 studies (N = 54,566) were analyzed by applying a three-level meta-analytic model. Applying the Hedges and Olkin approach yielded a correlation of $r = .31$. A Hunter and Schmidt–typed approach was applied to investigate the extent the result was influenced by measurement unreliability, which yielded a correlation of $r = .39$. Associations with global intelligence self-assessments differed significantly from specific ones (i.e., correlations were highest for numerical ability followed by global, spatial, and other less-known cognitive abilities, e.g., naturalistic intelligence), whilst the type of self-assessment method did not significantly affect results. A total of eight methods were applied to test for potential dissemination bias. The observed summary effect strength must be considered to be somewhat inflated, as there was evidence for some bias. In all, we show evidence for a moderate correlation between SAI and psychometric intelligence, which generalizes over assessment methods, order of assessment, sex of participants, year of publication but is differentiated according to ability type.

Keywords: Intelligence, Meta-analysis, Effect inflation.

Contents

Introduction

Theoretical framework

Self-assessed intelligence (= SAI) in general, especially its accuracy, remains a topic of great interest (e.g., Gerstenberg et al., 2013; Jacobs & Roodenburg, 2014; Furnham, 2018; Zajenkowski et al., 2020). While one could argue that no one knows your abilities better than you do, previous research suggests that, in fact, this might not be the case: For example, women have been reported to significantly underestimate, while men overestimate their abilities (Storek & Furnham, 2012). Previous meta-analytical accounts which focused on the relationship between SAI and psychometric intelligence (Mabe & West, 1982; Freund & Kasten, 2012) revealed a moderate correlation.

The relevance of cognitive abilities (i.e., intelligence), in particular, stems, for example, from their importance in different high-stakes situations, e.g., when making career decisions as they predict performance outcomes in most jobs and situations (Schmitt, 2014). Moreover, a good "fit" between job demands and personal resources correlates with job satisfaction, organizational commitment, and (negatively) with the intention to quit, while a "misfit" is associated with indicators of strain (Kristof-Brown et al., 2005). Furthermore, recent findings indicate that professional interests are primarily a function of self-assessed rather than „true" abilities (Neubauer & Hofer, 2021). Therefore, the importance of self-assessed cognitive abilities becomes apparent.

Over the past century, many researchers attempted to define intelligence: Ebbinghaus (1908) argued that "intelligence means organization of ideas, manifold interconnection of all those ideas which ought to enter into a unitary group because of the naturals relations of the objective facts represented by them" (p. 150). Famously, Boring (1923) described intelligence as "what the [intelligence] tests test" (p. 35). More recently, 52 academic researchers in fields associated with intelligence issued a statement known as *Mainstream Science on Intelligence* (Gottfredson, 1997): "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. It is not merely book-learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings - “catching on”, “making sense” of things, or “figuring out’ what to do.” (p. 13)

One unifying theoretical framework that has been used widely for the interpretation, foundation, and organization of intelligence is the Cattell-Horn-Carroll (=CHC) theory of human cognitive abilities (Flanagan & Dixon, 2014): To date being the most essential and empirically supported model of intelligence, it integrates three previous models proposed by Raymond Cattell, John Horn, and John Carroll.

The CHC theory divides intelligence into three strata (Schneider & McGrew, 2018): At the bottom of the model are test specific abilities, which can be directly measured. Specific abilities are highly correlated and can therefore be clustered into narrow abilities (stratum I). Broad abilities (stratum II) are clusters of narrow abilities, which are considerably more correlated with each other than with abilities from different broad abilities (e.g., Gf for fluid reasoning). At the top of the hierarchy is the most comprehensive level, representing the general g factor of intelligence.

However, laypeople typically have their own definition and understanding of constructs like intelligence. Individuals either perceive intelligence as a fixed entity that cannot be changed (= entity theory) or as something malleable which can be increased (= incremental theory; Dweck & Legget, 1988). Additionally, people seem to distinguish between different cognitive abilities in terms of changeability: In a recent study, verbal, naturalistic, and intra-personal intelligence were seen as more malleable than creative and musical intelligence (Furnham, 2014).

Almost ten years ago, Freund and Kasten (2012) conducted a meta-analysis focused on the correlation between people’s SAI and IQ. They included 154 effect sizes which were distributed over 41 studies and yielded a moderate overall correlation of r = .33. However, previous meta-analytic findings have become outdated due to the ever-increasing number of publications addressing the relationship between SAI and IQ. Furthermore, a meta-analytic update might also be helpful for three reasons:

First, recent findings suggest potential bias in effect size estimation due to empirical research's virtual ubiquitous declining effects (Pietschnig et al., 2019). Therefore, initial study effect size estimates might have been inflated due to strategic research practices. Furthermore, this might have directly influenced *a priori* power estimations of replicators and, again, could have led to the non-dissemination of non-replications. However, eventually, replication sample sizes are expected to increase over time, leading to more accurate and therefore increasing estimates. This could, of course, also have influenced previous meta-analytical findings (see Pietschnig et al., 2019, for an overview).

Second, Hunter and Schmidt (2004) argue that observed values might deviate from "true" values due to measurement error variance (i.e., unreliability). This might lead to inflated estimates of heterogeneity and moderator effects, bias the mean effect size towards zero, and confound performed publication bias and sensitivity analyses (Wiernik & Dahlke, 2020). However, the so-called Hunter and Schmidt (2004) approach to meta-analysis has been criticized in the past as some authors argue the goal of meta-analyses "is to teach us better what *is* not what might some day be in the best of all possible worlds when all our independent and dependent variables are perfectly measured" (Rosenthal, 1991, p. 25). For this reason, some researchers prefer the (uncorrected) Hedges and Olkin–typed (1996) approach to meta-analysis.

Finally, more refined research methods, e.g., for the investigation of dissemination bias, have been developed since the last meta-analysis on SAI and IQ was conducted. Dissemination bias can be defined as a systematic error that occurs from the non-dissemination of results and studies, which can have various reasons like the selective inclusion of English studies (= language bias) studies that are readily available to the researcher (= availability bias; Rothstein et al., 2005). Moreover, under extreme publication pressure, some researchers might engage in some sort of questionable research practices (QRPs) like *p*-hacking (i.e., the use of different strategies to achieve statistical significance; Head et al., 2015). In all, this typically leads to inflated effect size estimations because non-significant analyses with smaller effect sizes are less likely to be published (Rothstein et al., 2005). A previous study reported no evidence for the presence of dissemination bias in their meta-analysis of SAI and IQ (Freund & Kasten, 2012). However, only a visual funnel-plot analysis was performed, which has been criticized for being heavily subjective in the past (Simmonds, 2015). Over the years, various methods have been developed to detect potential dissemination bias. Some methods investigate the so-called small study effect, including Sterne and

Egger's regression approach (Sterne & Egger, 2005), Begg & Mazumdar's rank correlation test (Begg & Mazumdar, 1994), Duval an Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000b, 2000a), and PET-PEESE (Stanley & Doucouliagos, 2014). Other methods are based on the reported *p*-values, including the *p*-curve analysis (Simonsohn et al., 2014a, 2014b), *p*-uniform (van Assen et al., 2015), and *p*-uniform* (van Aert & van Assen, 2021).

The present study

This study aims to investigate the relationship between self-assessed intelligence and psychometric intelligence, replicating and updating the previous meta-analysis by Freund and Kasten (2012).

Additionally, a total of nine standard and more modern dissemination bias detection methods will be applied. Potential bias due to measurement unreliability will be addressed by comparing results of a Hedges and Olkin–typed approach (Hedges & Olkin, 1996) to the results reported by the Hunter and Schmidt (2004) method. Finally, study preregistration plays an essential role in helping to improve the credibility of scientific research (Pietschnig et al., 2019). Therefore all moderator variables and their corresponding hypotheses, the study design, the sampling plan (including the exact data collection procedures and coding manual), all measured variables, and the exact analysis plan were previously preregistered and can be found under the following link: https://osf.io/xa2gp

Hypotheses

General effect

Two meta-analyses that previously focused on the relationship between SAI and IQ reported a positive overall relationship (Mabe & West, 1982; Freund & Kasten, 2012). Therefore, the following is hypothesized:

H1*: There will be a significant, positive relationship between self-assessed and psychometrically assessed intelligence.*

Moderator variables

Based on previous studies (e.g., Freund & Kasten, 2012; Zell & Krizan, 2014), the average true effect size is not expected to be the same for all included studies.

Instead, it is assumed that there will be a significant amount of between-study variance. The goal is to explain part of this variation by a total of nine moderator variables. Six out of nine potential moderator variables were included in a previous meta-analysis (Freund & Kasten, 2012). Namely, self-assessment methods, ability type, order of assessment, sex of participants, sample composition, and year of publication. Additionally, three new potential moderators will be included: neuroticism, self-efficacy, and selection process.

Self-assessment methods

There are different methods and scales, which can be used to obtain self-estimates. Some studies use absolute scales, which feature labels with absolute terms, for example, a scale ranging from *high ability* to *low ability* or absolute numerical scores (e.g., Dislich et al.,2012), while relative scales use relative terms as labels (like *below average* and *above average;* e.g., Jacobs & Roodenburg, 2014). Furthermore, some studies use relative scales with the explicit mention of a specific reference group, to which participants have to compare themselves (e.g., Furnham, 2018; participants had to estimate their abilities in comparison to a peer and while knowing their groups mean). Finally, mixed scales are a hybrid of absolute and relative scales and might therefore feature, for example, a relative middle category (e.g., *average intelligence*) but absolute anchor labels (like *low* and *high intelligence;* e.g., Gerstenberg et al., 2013).

A previous meta-analysis found that correlations differentiated according to the method of self-assessment (Freund & Kasten, 2012). Using relative scales led to more valid self-estimates compared to estimates obtained from an absolute scale. The explicit mention of a specific reference group increased the relationship between SAI and IQ even further. Estimates obtained from mixed scales did not affect the correlation.

Therefore, the following is hypothesized:

H2.1.: *The relationship between SAI and psychometrically assessed intelligence will be higher when SAI is assessed using relative scales compared to absolute scales.*

H2.2.: *The correlation will be even higher if an explicit reference group is named.*

H2.3.: *The correlation between self-estimated and psychometrically assessed intelligence will not increase significantly when using a mixed scale compared to an absolute one.*

(Hypothesis 2.1 – 2.3 were previously preregistered as one hypothesis. For clarity, H2 was split and now presented as three hypotheses.)

Ability type

As mentioned above, nowadays, intelligence is described less as a one-dimensional construct but instead divided into many different cognitive abilities (Schneider & McGrew, 2018). It was previously suggested that laypeople consider spatial, numerical, and, to some extent, verbal ability to be the essence of intelligence (Furnham, 2000, 2001). Therefore, it was previously hypothesized that estimating these "standard abilities" and specific abilities would lead to higher correlations than general cognitive ability, whereas estimating other non-standard abilities would lead to lower correlations compared to general cognitive ability (Freund & Kasten, 2012).

Over the past years, results concerning the usefulness of the specificity and prominence of abilities for self-estimates' validity remain inconclusive. Therefore, the following is hypothesized:

H3.1: *The correlation between self-assessed and psychometrically assessed numerical, verbal and spatial abilities will be higher than for self-assessed and psychometrically assessed general cognitive ability.*

H3.2: *Additionally, the correlation will be higher for well-known "standard skills" (numerical, verbal, and spatial abilities) compared to other less-known cognitive abilities.*

(Hypothesis 3.1 and 3.2 were previously preregistered as one hypothesis. For clarity, H3 was split and now presented as two hypotheses.)

Order of assessment

Another possible moderator variable is the order in which the psychometric test was taken, and people estimated their intelligence (i.e., whether people estimated their cognitive abilities before or after they took the psychometric test).

It was previously hypothesized that self-estimates would be more accurate when they were made after the psychometric assessment, as people would take the preceding test situation as a reference to estimate their ability, which might result in more accurate self-estimates (Freund & Kasten, 2012). However, previous results have not been as conclusive as expected. Therefore, the following is hypothesized:

H4: *There will be significant differences in the correlation between SAI and psychometrically assessed intelligence depending on the order of assessment.*

(For H4, the null hypothesis was preregistered.)

Sex of participants

Previous studies reported that men had higher self-assessed logical and spatial intelligence than women while reporting no sex differences in verbal intelligence (e.g., Stieger et al., 2010; Syzmanowicz & Furnham, 2011). This effect has been referred to as the Hubris-Humility Effect, based on female participants' underestimation (humility) and male participants' overestimation (hubris) of specific cognitive abilities (Storek & Furnham, 2012). In the past, this effect has often been associated with stereotype threat, which occurs when members of stigmatized groups (e.g., women) are at risk of confirming negative stereotypes about their group (Steele & Aronson, 1995).

A previous meta-analysis investigated whether relying on a women-only or men-only sample significantly influenced the correlation between SAI and IQ compared to a mixed sample (Freund & Kasten, 2012). However, there were no significant differences in validity, which was attributed partly to the lack of statistical power (most included effect sizes came from mixed-sex samples).

Moreover, recent meta-analytical findings further suggest that the effect of stereotype threat on women's performance in specific tasks might be smaller than initially expected (Picho-Kiroga et al., 2021). Therefore, as previous accounts are not as clear as expected, the present meta-analysis will compare studies relying on men-only samples to women-only and mixed samples. Furthermore, the following is hypothesized:

H5: *There will be significant sex differences in the correlation between SAI and psychometrically assessed intelligence.*

(For H5, the null hypothesis was preregistered.)

Sample composition

One problem when it comes to correlation coefficients is that they highly depend on the sample at hand. The squared correlation coefficient represents the percentage of shared variance between two variables. This variance reflects the extent to which all individuals differ from each other. Therefore, correlation is directly associated with and influenced by the variability in the data sample. Regarding psychology, studies typically heavily rely on student samples: From 2004 to 2007, almost 70% of studies published in *JPSP* (Journal of Social and Political Psychology) consisted of undergraduate students (Arnett, 2008). The percentage shrank from 2014 to 2018 to about 40% (Thalmayer et al., 2021), which is still a substantial amount. Hence, it was previously assumed that student/academic samples would show less variability concerning the variables of interest, as they should have more test-taking experience than general/non-academic samples (Freund & Kasten, 2012). Homogeneous samples generally lead to lower correlations. Therefore, the following is hypothesized:

H6: *For non-academic participants, the correlation between self-assessed and psychometrically assessed intelligence will be higher than for academic participants.*

Publication years

It was previously assumed that a better understanding of the concept of intelligence would result in more recent studies reporting higher correlations compared to older studies (Freund & Kasten, 2012). However, there have been reports of declining effects in psychological research (Pietschnig et al., 2019). Declining effects may be caused by low initial study power due to strategic researcher practices. Subsequently, inflated initial effects could directly influence *a priori* power analysis of replicators and, again, lead to the publication of inflated replications and the non-dissemination of non-replications. However, replication sample sizes are expected to increase over time, leading to more accurate and therefore increasing estimates (see Pietschnig et al., 2019, for an overview).

For these reasons, the influence of the year of publication is not as clear as expected. Therefore, the following is hypothesized:

H7: *The year of publication will influence the correlation between SAI and psychometrically assessed intelligence.*

(For H7, the null hypothesis was preregistered.)

Selection process

When it comes to self-ratings, researchers have expressed their concerns about the validity of such measures, as they might be affected by socially desirable responses, which is the tendency to portray oneself more favorably on self-report questionnaires (Tracey, 2016). It seems logical that individuals are especially tempted to answer in a somewhat desirable way when their answers influence their chances of getting a job. This reflects in the finding that participants tend to rate their personality more favorable when self-assessments are part of a selection process, but only when they show interest in the job they applied for (Feeney, 2018). Thus, one can assume that the exact mechanisms would influence self-assessed intelligence. Moreover, past research shows that measurements of self-enhancement bias predict SAI (Leising et al., 2016). Therefore, the following is hypothesized:

H8: *If intelligence was assessed as part of a selection/application process, the correlation between SAI and psychometrically assessed intelligence will be lower compared to "normal" test conditions.*

Personality

Another potential moderator, which might influence the association between SAI and IQ is the participants' personality, such as self-efficacy or neuroticism.

Self-efficacy. Self-efficacy refers to an individual's beliefs that they can successfully perform a task (e.g., Bandura, 1977). Therefore, it seems intuitive that an individual's self-efficacy beliefs would be associated with their self-assessed cognitive abilities. However, empirical evidence has so far been inconclusive. Some studies reported a significant correlation between people's self-estimated abilities and self-efficacy beliefs (e.g., Ackerman & Wolman, 2007; Brady-Amoon & Fuertes, 2011), while others reported no relationship between SAI and self-efficacy (Ng & Earl, 2008). Therefore further investigation on the influence of people's self-efficacy beliefs on the relationship between SAI and psychometric intelligence is necessary. Therefore, the following is hypothesized:

H9: *The higher a person's self-efficacy beliefs are, the higher the correlation between SAI and psychometric intelligence will be.*

Neuroticism. Neuroticism, also referred to as Emotional Stability, is a personality trait that stands for an individual's predisposition to experience negative emotions such as fear, anger, shame, and dejection (Costa & McCrae, 2011). Neuroticism has, together with Openness and Agreeableness, in the past been shown to account for a significant amount of variance when it comes to self-assessed fluid intelligence (Jacobs et al., 2012). Moreover, SAI has been reported to relate to neuroticism negatively and positively to extraversion and narcissism (Howard & Cogswell, 2018).

The aforementioned sex differences in SAI have also been hypothesized to stem from personality differences between the sexes as men's higher self-estimated spatial and logical abilities can partly be explained by women's higher neuroticism scores (Stieger et al., 2010). Therefore, the following is hypothesized:

H10: *The higher a person's neuroticism score, the lower the correlation between SAI and psychometric intelligence will be.*

Method

Literature search

First, for the years between 1915 and 2010, all studies/effect sizes previously reported by Freund and Kasten (2012) were included for further screening.

Second, the online databases PsychINFO, PSYNDEX, Scopus, and ISI Web of Science were searched systematically from 2011 to 2020, using the following search string: *(cognitive ability OR intelligenc*) AND (estimate* OR perceive* OR self-apprais* OR self-assess* OR self-estimate* OR self-evaluat* OR self-perceive* OR self-rate*).*

Third, the first 100 hits from Google Scholar and the Open Access Theses and Dissertation database (OATD) were searched for the keywords mentioned above to identify potentially relevant studies from grey literature between 2011 and 2020.

Fourth, a cited reference search was performed in the ISI Web of Science for all studies previously included by Freund and Kasten (2012) in the ISI Web of Science.

Fifth, from the 18,000 initial hits, studies from technical fields (i.e., computer science, artificial intelligence, engineering, electronics, robotics, physics, and mathematics) were excluded, it was considered rather unlikely that the correlation between SAI and IQ might be a research topic of interest in technical fields. This led to a total sample of 13,993 potentially relevant studies: 13,941 hits identified by database searching and 52 through other sources, all of which were screened in the next step. When studies were not published in English, German, or French, they were translated to English using an online tool (deepl.com).

Finally, after excluding all studies which did not meet the inclusion criteria (see next chapter), reference list screenings based on all eligible studies were used to identify potentially relevant studies that might have been missed up to this point. This resulted in the inclusion of two more studies.

Inclusion criteria and eligibility

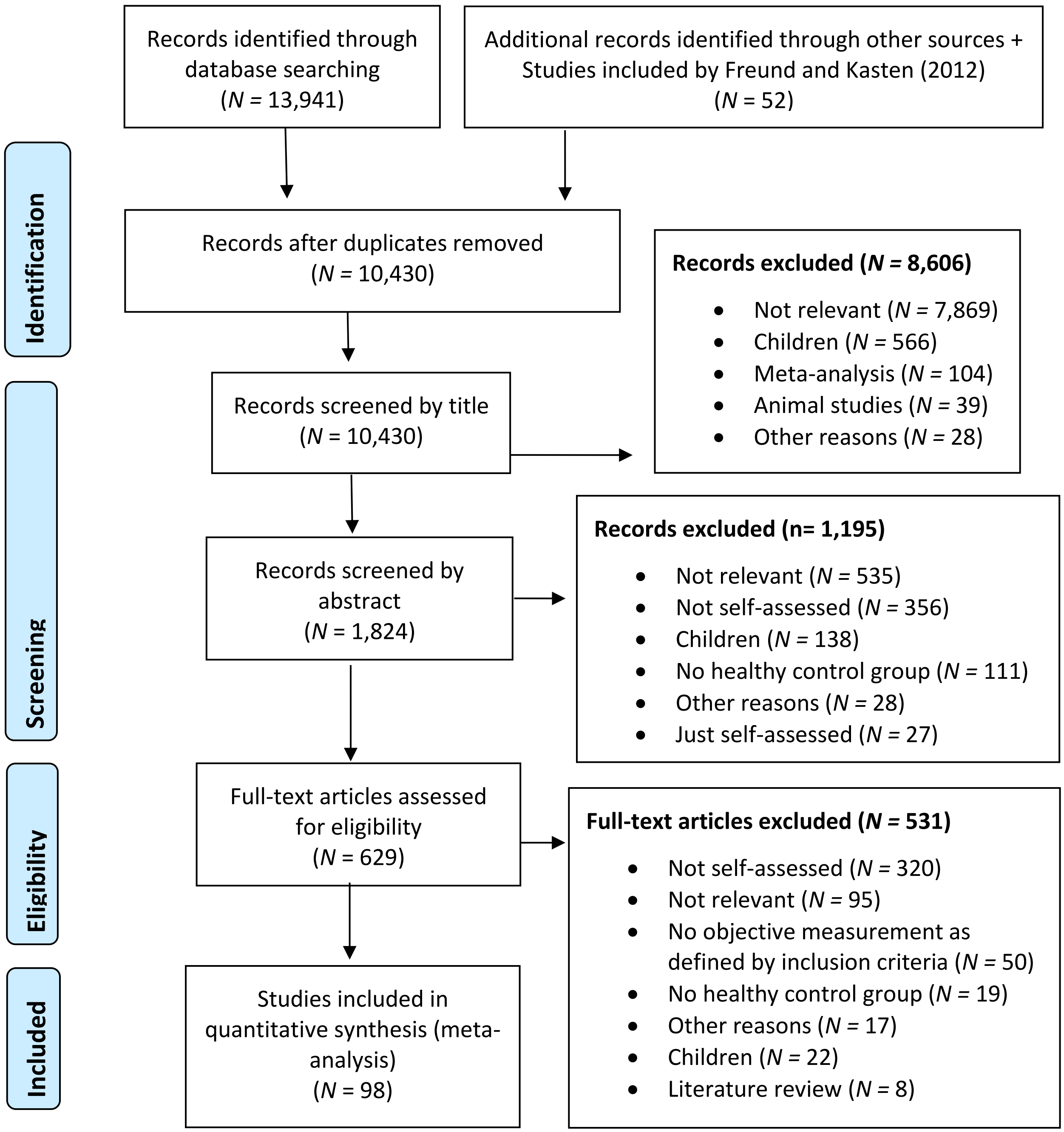
To be included, all studies had to meet five inclusion criteria, some of which were previously used in a previous meta-analysis (Freund and Kasten, 2012): They had to 1) associate self-assessed cognitive ability (intelligence) with psychometrically assessed cognitive ability. 2) Intelligence had to be collected using a standardized cognitive ability test (i.e., the test was administered in a standardized way, was objectively scored, and offered norms that allowed comparisons to a norm sample, which were either included in the test manual or derived from the data collected from the study sample). 3) A direct measure of self-assessed cognitive ability must have been used (studies with indirect measures, for example, when scores were drawn from other related constructs such as self-assessed cognitive impairment, were, therefore, omitted).

In addition to the abovementioned inclusion criteria, the following two criteria had to be met: 4) All included studies, except those included by the previous meta-analysis (Freund and Kasten, 2012), must have been conducted between 2011-2020. 5) Only studies that included healthy adult participants were considered (i.e., mean age > 18 years, decimals were rounded up).

The author performed the screening of title, abstract, and full texts as well as data extraction. The screening and literature search results and relevant details are displayed in the following PRISMA Flowchart (Figure 1). The main reason for the exclusion of a study was that it was considered irrelevant/off-topic: Over 60% of all rejected studies had to be dismissed for this reason. Other reasons were that no objective measurement was used to assess intelligence psychometrically or that participants did not self-estimate their intelligence. In addition, some studies had to be excluded from further analysis because participants were children.

Figure 1

Prisma 2009 Flow Diagram (Moher et al., 2009)



Coding

All studies were coded by the author using a standardized coding scheme. Correlation coefficients, total sample sizes, reliabilities of self-assessment as well as intelligence test measures, participants' sexes (1 = female-only sample, 2 = male-only sample, 3 = mixed sample), sample compositions (1 = mixed sample, 2 = student sample, 3 = general sample), publication years, methodology of self-assessments (1 = absolute scale, 2 = mixed scale, 3 = relative scale including a specific reference group, 4 = relative scale), ability types (1 = general cognitive ability, 2 = numerical ability, 3 = other cognitive ability, 4 = spatial ability, 5 = verbal ability), order of assessments (1 = estimate first, 2 = test first, 3 = order unknown), whether intelligence was assessed as part of a selection process (1 = no, 2 = yes), reported self-efficacy and neuroticism means, and current publication statuses (1= yes, 2 = no) were recorded.

Relevant variables from the studies already included in the previous meta-analysis (Freund & Kasten, 2012) were taken directly from the latter, if available. If an included study did not report a direct correlation coefficient between IQ and SAI, but a *t* value, *t* to *r* transformations (Borenstein et al., 2009) were applied. If participants estimated their cognitive abilities multiple times, an overall mean of all reported correlation coefficients was calculated.

Since different studies used different scaled and different item types to measure neuroticism and self-efficacy, the reported means needed to be made comparable. Therefore, a relative mean for neuroticism and self-efficacy was computed by multiplying the number of answer options by the number of questions. Subsequently, the average raw score was divided by that value. In order to do so, the respective number of answer options, maximum of achievable points, and the number of items were extracted.

10% of all new studies published between 2011 and 2020 were randomly selected and coded again independently by the author. A kappa statistic was calculated, which can be defined as the extent to which raters assign the same score to the same variable (McHugh, 2012). Kappa values were satisfying as the average Kappa was at Κ = .96 (median: Κ = 1.00), and values ranged from 0.47 to 1.00. All discrepancies were investigated and, if necessary, reviewed by a third round of coding.

Data Analyses

Hedges and Olkin (1996) vs. Hunter and Schmidt (2004)

In the past, it has been argued that the observed effect sizes deviate from the "true" effect sizes due to different "artifacts", one being measurement unreliability (Hunter & Schmidt, 2004). The included studies used several different measures to assess SAI and IQ. Because unreliability of measures can bias the overall effect size towards zero (amongst other things), unreliability correction seemed reasonable (Wiernik & Dahlke, 2020).

First, all calculations (except for dissemination bias) were calculated using the unreliability-corrected Hunter and Schmidt (2004)–typed approach.Therefore, a measurement correction of effect sizes (i.e., Pearson's *r*) and standard errors was performed by calculating the unreliability: $\sqrt{(reliability_{IQ} \times reliability_{SAI})}$. Next, the effect size (i.e., correlation) and standard error were divided by the total unreliability to get the adjusted effect size and standard error (Hunter & Schmidt, 2004). Finally, the R package psychmeta (Dahlke & Wiernik, 2019) was used to impute missing reliability values and correct IQ and SAI measures individually.

Second, the Hunter and Schmidt (2004) –typed approach has also been criticized in the past (e.g., Rosenthal, 1991) and in order to check to what extent the results were influenced by measurement unreliability, the Hedges and Olkin–typed (1996) approach to meta-analysis was applied. Therefore, all reported correlation coefficients were transformed into Fisher's *z*. Next, the standard error of *z* was calculated, which is approximately $1/\sqrt{N-3}$.

Finally, the results yielded by the Hedges and Olkin–typed (1996) and the Hunter and Schmidt (2004)–typed approach were compared.

Heterogeneity and sensitivity analyses

When synthesizing effects across studies into one single effect, heterogeneity has to be considered. Heterogeneity can be defined as the extent to which effect sizes vary within a meta-analysis (Harrer et al., 2021). Study heterogeneity was described by reporting the Q statistic, τ, τ^2, and I^2. Due to high observed study heterogeneity, a random-effects-model meta-analysis with a restricted (residual) maximum-likelihood estimator (REML) was conducted. In order to detect which effect sizes influence the overall estimate the most, leave-one-out sensitivity analyses were performed by excluding one study/effect size at a time. Furthermore, the contribution of each study/effect size to the overall heterogeneity was inspected using a baujat plot analysis (Baujat et al., 2002).

Summary effect and moderator analyses

To cope with dependencies between effect sizes (single studies reported multiple effect sizes) and in line with the previous meta-analysis (Freund & Kasten, 2011), the hierarchical linear modeling (HLM) approach to meta-analysis (Raudenbush & Bryk, 2002) was applied. This so-called “Multi-level” meta-analysis allows effect sizes to vary between participants (Level 1), outcomes (Level 2), and studies (Level 3) (Assink & Wibbelink, 2016). In addition, the three-level meta-analysis can also be used to explain within and between-study heterogeneity by including potential moderator variables if evidence for heterogeneity is present (Assink & Wibbelink, 2016). Therefore, subgroup analyses were performed by adding regression terms to the multilevel model, which led to a three level-mixed effects model (Harrer et al., 2021):

$$\hat{\theta}_{ij} = \beta_0 + \zeta_{(2)ij} + \zeta_{(3)ij} + \epsilon_{ij}$$

At Level 1 (base level or participants level), the model was a random-effects model. At this level, participants are nested within effect sizes. At Level 2, moderators at the relationship/outcome level (all moderators except for “publication year”) were included (within-study model). Level 3 included all moderators at the study level (i.e., “publication year”; between-study model). Before inclusion, all potential moderator variables were mean-centered (for continuous variables) or dummy-coded (for categorial variables). Before performing the multiple meta-regression (including all potential moderator variables), each moderator was individually tested for significance to address possible intercorrelations between moderators (Assink & Wibbelink, 2016; Hox, 2010). Therefore, the three-level mixed-effects model was used to perform subgroup analyses in the form of multiple single meta-regressions.

Summary effects, the amount of variance each level contributed to the total variance, and moderator analyses were calculated using the R package 'metafor' (Viechtbauer, 2010).

Since multiple studies found that there might be a relationship between self-efficacy and neuroticism (e.g., Ackerman & Wolman, 2007; Judge & Ilies, 2002), multicollinearity between these two moderators was expected. Therefore, two separate three-level mixed-effects models were formulated to perform multiple meta-regression: Model one included all other potentially relevant moderators in addition to neuroticism, and model 2 included all other potentially relevant moderators in addition to self-efficacy.

Model fit

For both, the models with potential moderator variables and the model without potential moderator variables, model fit criteria were compared to determine whether a three-level model explained more variance than a two-level model. Therefore, level 2 and level 3 were removed separately (i.e., variance on the respective level was set to zero), and each reduced model was compared to the full three-level model. (Assink & Wibbelink, 2016) The exact process was repeated for the full three-level moderator model.

Dissemination bias

It was planned to investigate potential dissemination bias using the following methods: *p*-curve (Simonsohn et al., 2014a, 2014b), *p*-uniform (van Assen et al., 2015), *p*-uniform* (van Aert & van Assen, 2021), PET-PEESE (Stanley & Doucouliagos, 2014), Begg and Mazumdar's rank-order correlation (1994), Sterne and Egger's regression approach (2005), moderator analysis of published vs. unpublished literature (Card, 2012), Duval & Tweedie's trim-and-fill procedure (2002b, 2002a), and test for excess significance following the approach of Ioannidis and Trikalinos (2007). Furthermore, studies were displayed in an enhanced funnel plot (Peters et al., 2008) to inspect the funnel plot asymmetry. All methods for dissemination bias detection were interpreted using current guidelines (see Siegel et al., 2021, for an overview).

Results

Sample

In total, 242 effect sizes from 98 studies were included, which ranged from r = -.25 to r = .85. About 95% of these reported effect sizes were positive. The accumulated sample size consisted of 54,566 participants, ranging from N = 13 to N = 13,690 participants. All included studies/effect sizes are displayed in Table A1 (Appendix).

Summary effect

First, a Hedges and Olkin–typed (1996) unconditional meta-analytic model without any moderators but with error terms at the relationship and study level was fitted. The estimated z was .320 with a standard error SE of .016, which yielded a correlation of r = .310. The summary effect was significant (t (241) = 20.440, p < .001; τ^2 = .016, σ^2 = .011), and the 95% confidence interval ranged from .290 to .351.

Next, all r values and their corresponding standard errors were corrected for unreliability (Hunter & Schmidt, 2004), and the model was fitted a second time. The relationship between self-estimated and psychometrically assessed intelligence was estimated at r = .385 with a corresponding standard error SE of .018. Again, the summary effect was significant (t (241) = 21.681, p < .001; τ^2 = .019 and σ^2 = .014) and the 95% confidence interval ranged from .350 to .420.

Taken together, this provides evidence for there being a moderate but significant positive relationship between self-assessed and psychometrically assessed intelligence.

Furthermore, the amount of observed heterogeneity was significant for both models (Hedges and Olkin–typed model (1996): Q (241) = 1319.704, p < .001; Hunter and Schmidt (2004) –typed model: Q (241) = 1257.423, p < .001), indicating that the inclusion of moderator variables might help explain parts of this variation among effect sizes.

Distribution of total variances

The next step was to determine the amount of variance each level contributed to the total variance. For the Hedges and Olkin–typed approach (1996), 15.36% of the variance was attributed to level 1 (participants level), 50.03% to level 2 (effect size level), and 34.61% to level 3 (study level). Thus, between-study heterogeneity I^2 was estimated at 84.64%, which can be interpreted as a substantial amount of heterogeneity (Higgins et al., 2003).

For the Hunter and Schmidt (2004) –typed model, 16.48% attributed to level 1 (participants level), 48.15% to level 2 (effect size level), and 36.38% to level 3 (study level). Again, there was substantial between-study heterogeneity (Higgins et al., 2003) as I^2 was estimated at 83.52%.

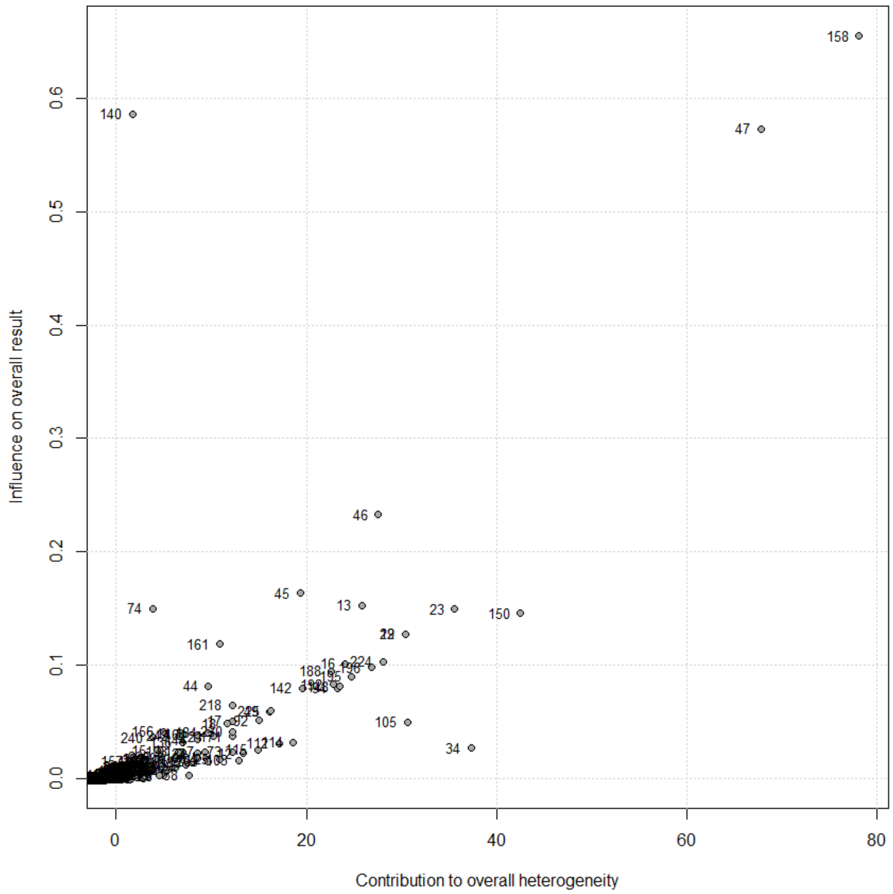
Influence analyses

To detect which effect sizes influence the overall estimate the most, leave-one-out sensitivity analyses were performed by excluding one effect size at a time (Viechtbauer & Cheung, 2010). “Meneghetti et al. (2014).3” (effect size ID = 158) was identified as influencing the effect size estimation. Leaving it out from analysis increased the overall effect size from z =.321 to z = .322. Heterogeneity shrank from I^2 = 84.64% to I^2 = 84.00%. Because omitting “Meneghetti et al. (2014).3” did not influence the interpretation heavily, it was decided not to be excluded from analyses.

Furthermore, a baujat plot (Baujat et al., 2002) was used to investigate each effect size’s contribution to total heterogeneity (Figure 2). An effect size is considered problematic if it contributes much to overall heterogeneity while not influencing the overall effect (Harrer et al., 2021). However, Figure 2 shows that this is not the case for any effect size in the data. On the contrary, Figure 2 demonstrates that heterogeneity seems to arise from the variety of effect sizes included in the data rather than from a specific effect size.

Figure 2

Baujat plot (Baujat et al., 2002).



Model fit

The removal of level 2 (within-study variance) from both the three-level Hedges and Olkin (1996) and Hunter and Schmidt (2004)–typed model showed better fit criteria (i.e., AIC and BIC) for the three-level model. Results, summarized in Table 1, show a significant amount of within-study variance ($p < .001$).

Results for the between-study variance are summarized in Table 2. Removing level three (between-study variance) again indicated that the full model has a better fit than the reduced model as both model-fit criteria (i.e., AIC and BIC) were lower for this model.

Table 1

Model fit: Likelihood-ratio-test for level two (within-study variance).

	df	LRT	AIC	BIC	*p*	Q
Hedges and Olkin (1996) – typed Model						
Full	3		-131.089	-120.635		1319.704
Reduced	2	206.569	73.480	80.450	< .001	1319.704
Hunter and Schmidt (2004) – typed Model						
Full	3		-77.826	-67.372		1257.423
Reduced	2	190.410	110.584	117.553	< .001	1257.423

Note. Full = full three-level meta-analytic model, Reduced = two-level model without within-study variance. LRT = log-likelihood test, Q = heterogeneity (Cochran's Q).

Table 2

Model fit: Likelihood-ratio-test for level three (between-study variance).

	df	LRT	AIC	BIC	*p*	QE
Hedges and Olkin (1996) – typed Model						
Full	3		-131.089	-120.635		1319.704
Reduced	2	32.151	-100.938	-93.969	< .001	1319.704
Hunter and Schmidt (2004) – typed Model						
Full	3		-77.826	-67.372		1257.423
Reduced	2	35.466	-44.361	-37.391	< .001	1257.423

Note. Full = full three-level meta-analytic model, Reduced = two-level model without between-study variance. LRT = log-likelihood test, *Q* = heterogeneity (Cochran's Q).

Moderator analyses

Individual tests of potential moderators

Before testing all potential moderator variables in a single model, each moderator was evaluated individually to address possible intercorrelations between moderators (Assink & Wibbelink, 2016; Hox, 2010). Therefore, the three-level mixed-effects model was used to perform subgroup analyses in the form of multiple single meta-regressions.

First, all categorical predictors were dummy-coded and continuous predictors were mean-centered. Next, each potential moderator was tested, using the Hedges and Olkin–typed (= H&O) approach (1996) and Hunter and Schmidt (2004)–typed (= H&S) model. Methodology of self-evaluation ($F_{H\&O}$ (3, 238) = 3.308, $p_{H\&O}$ =.021; $F_{H\&S}$ (3, 238) = 3.369, $p_{H\&S}$ = .019), ability type ($F_{H\&O}$ (4, 237) = 12.768, $p_{H\&O} < .001$; $F_{H\&S}$ (4, 237) = 13.572, $p_{H\&S} < .001$), order of assessment ($F_{H\&O}$ (2, 239) = 5.483, $p_{H\&O}$ =.005; $F_{H\&S}$ (2, 239) = 4.517, $p_{H\&S}$ = .012) and sample composition ($F_{H\&O}$ (1,240) =5.892, $p_{H\&O}$ = .016; $F_{H\&S}$ (1,240) = 6.664, $p_{H\&S}$ = .010) showed significant results, when tested separately.

Sex of participants ($F_{H\&O}$ (2,239) = 0.662, $p_{H\&O}$ = .517; $F_{H\&S}$ (2,239) = 0.892, $p_{H\&S}$ = .411) and year of publication ($F_{H\&O}$ (1,240) = 1.689, $p_{H\&O}$ =.195; $F_{H\&S}$ (1,240) = 1.632, $p_{H\&S}$ = .203) did not show significant influence on the correlation between SAI and IQ.

Self-efficacy and neuroticism were only reported for ten and 47 effect sizes, respectively. However, when it comes to self-efficacy, those ten reported effect sizes stem from only two studies. Therefore, it was decided to exclude self-efficacy from all further calculations. Neuroticism was investigated in the respective subset, but did not show significant results ($F_{H\&O}$ (1, 45) = 1.972, $p_{H\&O}$ =.167; $F_{H\&S}$ (1, 45) = 0.517, $p_{H\&S}$= .476). No cases in which intelligence was assessed as part of a selection process were found. Therefore, the moderator variable selection process had to be excluded from all analyses.

Some authors recommend that only potential moderators, which were previously identified as significant in the univariate analyses, should be included in the total moderator model (Assink & Wibbelink, 2016; Hox, 2010). However, in line with the previous meta-analysis (Freund & Kasten, 2012), all potential moderator variables were included in the three-level mixed-effects model for the multiple meta-regression analyses.

Multiple meta-regression

Table 3 and Table 4 summarize the moderator models' results for Hedges and Olkin–typed (1996) and Hunter and Schmidt (2004) –typed model. For the Hedges and Olkin–typed (1996) moderator model, variance at the effect size level was estimated at $\tau^2 = .012$ and $\sigma^2 = .008$ at the study level. The amount of observed heterogeneity was significant ($Q(228) = 969.113$, $p < .001$). The moderators now condition τ^2 at the effect size level (all moderator variables except for the year of publication), and σ^2 is dependent on the moderator variable on the study level (year of publication; Freund and Kasten, 2012).

For the Hunter and Schmidt (2004) –typed moderator model, variance at the effect size level was estimated at $\tau^2 = .013$ and $\sigma^2 = .010$ at the study level, and the amount of observed heterogeneity was significant ($Q(228) = 921.180$, $p < .001$).

Method of self-evaluation. There were no statistically significant differences among the four methods of self-evaluation (see Tables 3 and 4). Neither self-evaluations made on a relative scale nor on a scale including an explicit reference group showed an improvement over estimates made on an absolute scale. As expected, estimates obtained from a mixed scale did not show a significant improvement as well.

Table 3

Results of multiple meta-regression: Hedges and Olkin (1996)–typed meta-analytic model.

Variable	*m*	Coefficient	*SE*	95% CI	*p*
Fixed effect					
Intercept		0.368	0.065	[0.241, 0.495]	< .001
Level 2 moderator variables					
Methodology of self-assessment [a]					
Relative Scale	53	0.038	0.041	[-0.043, 0.118]	.178†
Relative Scale including reference group	64	0.058	0.045	[-0.030, 0.146]	.098†
Mixed Scale	62	0.002	0.038	[-0.072, 0.076]	.959
Ability Type [b]					
Numerical Ability	19	0.162	0.040	[0.084, 0.240]	< .001†
Spatial Ability	31	-0.065	0.037	[-0.139, 0.008]	.041†
Verbal Ability	27	0.030	0.036	[-0.040, 0.101]	.200†
Other Cognitive Abilities	52	-0.084	0.033	[-0.149, -0.020]	.005†
Order of Assessment [c]					
Standardized Test First	75	0.050	0.035	[-0.019, 0.120]	.156
Mixed/ Unknown	70	-0.034	0.034	[-0.101, 0.033]	.320
Sex of Participants [d]					
Females-only sample	17	-0.047	0.063	[-0.170, 0.076]	.451
Mixed sample	208	-0.055	0.061	[-0.176, 0.066]	.369
Sample Composition [e]					
General Sample	81	-0.044	0.030	[-0.104, 0.016]	.074†
Level 3 moderator variable					
Mean year of publication = 2006		-0.000	0.001	[-0.002, 0.002]	.852
	Variance com-ponent	*LRT*		*df*	*p*
Random effect					
Within study, τ^2	0.012	126.658		1	< .001
Between Studies, σ^2	0.008	24.187		1	< .001

Note. Effects are reported in Fishers z metric; sample size m is for the number of effect sizes.
[a] Reference category: Absolute scales (*m* = 63). [b] Reference category: General cognitive ability (*m* = 113). [c] Reference category: Self-estimation first (*m* = 97). [d] Reference category: Male-only sample (*m* = 17). [e] Reference category: Students (*m* = 158). † *p*-value halved due to one-tailed test.

Table 4

Results of multiple meta-regression: Hunter and Schmidt (2004) –typed meta-analytic model.

Variable	*m*	Coefficient	*SE*	95% CI	*p*
Fixed effect					
Intercept		0.442	0.073	[0.299, 0.585]	< .001
Level 2 moderator variables					
Methodology of self-assessment [a]					
Relative Scale	53	0.053	0.046	[-0.037, 0.143]	.124†
Relative Scale including Reference Group	64	0.069	0.050	[-0.030, 0.168]	.085†
Mixed Scale	62	0.010	0.042	[-0.073, 0.093]	.813
Ability Type [b]					
Numerical Ability	19	0.180	0.044	[0.094, 0.267]	< .001†
Spatial Ability	31	-0.078	0.042	[-0.160, 0.004]	.031†
Verbal Ability	27	0.023	0.040	[-0.056, 0.102]	.285†
Other Cognitive Abilities	52	-0.104	0.036	[-0.176, -0.033]	.003†
Order of Assessment [c]					
Standardized Test First	75	0.040	0.039	[-0.037, 0.118]	.303
Mixed/ Unknown	70	-0.039	0.038	[-0.114, 0.037]	.314
Sex of Participants [d]					
Females-only sample	17	-0.081	0.070	[-0.220, 0.058]	.252
Mixed sample	208	-0.064	0.069	[-0.200, 0.072]	.354
Sample Composition [e]					
General Sample	81	-0.050	0.034	[-0.117, 0.017]	.073†
Level 3 moderator variable					
Mean year of publication = 2006		-0.000	0.001	[-0.003, 0.002]	.850
	Variance comp-onent	*LRT*		*df*	*p*
Random effect					
Within study, τ^2	0.013	113.3962		1	< .001
Between Studies, σ^2	0.010	27.9896		1	< .001

Note. Effects are reported in Pearson's *r* metric; sample size m is for the number of effect sizes. [a] Reference category: Absolute scales (*m* = 63). [b] Reference category: General cognitive ability (*m* = 113). [c] Reference category: Self-estimation first (*m* = 97). [d] Reference category: Male-only sample (*m* = 17). [e] Reference category: Students (*m* = 158). † *p*-value halved due to one-tailed test.

Ability type. Associations with general self-assessed cognitive ability differed significantly from most specific cognitive abilities (see Tables 3 and 4). Consistent with the hypothesis, the self-estimation of numerical ability showed an improvement compared to general cognitive ability. On the other hand, the estimation of verbal ability did not affect the validity of self-estimates. Surprisingly, the estimation of spatial ability even led to a decline in validity compared to the estimation of general cognitive ability. As predicted, the estimation of other less-known cognitive abilities (e.g., naturalistic intelligence) showed the lowest correlation.

Order of assessment. The validity of self-assessed intelligence was not affected by order of assessment (see Tables 3 and 4): When the standardized test was taken first, the correlation was not significantly higher or lower.

Sex of participants. 208 out of 242 (86%) were estimated using a mixed-sex sample. Comparing the women-only sample and the mixed sample to the men-only sample did not show significant differences between those groups (see Tables 3 and 4).

Sample composition. 158 out of 242 effect sizes reported a student sample (65%). There was no significant difference between the student and the general sample (see Tables 3 and 4).

Year of publication. The average year of publication was 2006. There was no significant effect concerning the year of publication (see Tables 3 and 4).

Neuroticism. Forty-seven effect sizes assessed neuroticism. In this subsample, neuroticism did not influence the correlation between self-estimated and psychometrically assessed intelligence ($b_{\text{H\&O}}$ = -0.213, $p_{\text{H\&O}}$ = .746; $b_{\text{H\&S}}$ = -0.273; $p_{\text{H\&S}}$ = .669). However, in this subsample, no information was available on the subgroups “relative scale” from the moderator variable “methodology of self-evaluation” and the moderator variable “sex of participants”.

Selection process and self-efficacy. As previously mentioned, “self-efficacy”, could not be included in the overall moderator model because two reported ten effect sizes (from two studies). No cases in which intelligence was assessed as part of a selection process were found. Therefore, the moderator variable selection process had to be excluded from all analyses.

Model fit

Like the unconditional model without moderators, model fit criteria were compared to determine whether a three-level moderator model explained more variance than a two-level model. Again, removing level two (within-study variance, Table 5) and level three (between-study variance, Table 6) showed that the full model does indeed offer the best fit. moderator model.

Table 5

Model fit: Likelihood-ratio-test for level two (within-study variance) – moderator model.

	df	LRT	AIC	BIC	*p*	Q
Hedges and Olkin (1996) – typed Model						
Full	16		-147.913	-93.043		969.113
Reduced	15	126.658	-23.255	28.186	< .001	969.113
Hunter and Schmidt (2004) – typed Model						
Full	16		-99.039	-44.170		921.180
Reduced	15	113.396	12.357	63.797	< .001	921.180

Note. Full = full three-level meta-analytic moderator model, Reduced = two-level model without within-study variance. LRT = log-likelihood test, Q = heterogeneity (Cochran's Q).

Table 6

Model fit: Likelihood-ratio-test for level three (between-study variance) – moderator model.

	df	LRT	AIC	BIC	*p*	Q
Hedges and Olkin (1996) – typed Model						
Full	16		-147.913	-93.043		969.113
Reduced	15	24.187	-125.726	-74.286	< .001	969.113
Hunter and Schmidt (2004) – typed Model						
Full	16		-99.039	-44.170		921.180
Reduced	15	35.466	-73.050	-21.609	< .001	921.180

Note. Full = full three-level meta-analytic moderator model, Reduced = two-level model without between-study variance. LRT = log-likelihood test, Q = heterogeneity (Cochran's Q).

Dissemination bias

Common guidelines were used to interpret the different dissemination bias methods, with the typical inference criterion being $\alpha = .10$ (see Siegel et al., 2021, for an overview).

First, different methods investigating small study effects were applied. Both, Sterne and Egger's (2005) regression approach ($z = 2.560$, $p < .10$) and Begg & Mazumdar's (1994) rank correlation test, (Kendall's $\tau = .176$, $p < .10$) indicated funnelplot asymmetry. Next, potentially missing effect sizes were imputed using Duval and Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000b, 2000a). Results showed that the difference between the adjusted and the previously estimated summary effect did not exceed 20% and, therefore, did not indicate the presence of bias (Siegel et al., 2021). Next, the PET-PEESE procedure (Stanley & Doucouliagos, 2014) was applied, which is based on two random-effects models where either the standard error (PET) or the variance (PEESE) serve as predictors. PEESE was interpreted as the intercept's coefficient was significant ($p < .10$). However, the adjusted and the previously estimated summary effect difference did not exceed 20% (Siegel et al., 2021; Stanley, 2017).

Additionally, a contour-enhanced funnel-plot analysis (Peters et al., 2008) was performed by using the R package "metaviz" (Kossmeier et al., 2020). In Figure 3, the red dotted line represents Egger's regression line (Egger et al., 1997), which again shows a substantial amount of asymmetry. The dark dots represent those effect sizes imputed by Duval and Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000a, 2000b). Most imputed effect sizes were non-significant, which might indicate a potential file-drawer effect. The file-drawer effect refers to the fact that non-significant, small or neutral effects are less likely to be published and hidden in the symbolic "file-drawer". (Rosenthal, 1979)

Second, methods based on *p*-values were performed. Figure 4 shows the results of the *p*-curve analysis (Simonsohn et al., 2014a, 2014b). The *p*-curve is a histogram, showing all significant effect sizes in the meta-analysis at hand. In theory, the *p*-curve would be left-skewed when there was a lot of *p*-hacking present in the data (i.e., many *p*-values would be just below .05) or flat when there is no true effect present. The *p*-

curve at hand is significantly right-skewed ($p < .001$) and not flat ($p > .999$). These results indicate a true effect in the data (Harrer et al., 2021). The *p*-curve analysis also comes with a corrected overall summary effect, estimated at $z = .295$ ($r \sim .286$). However, some authors recommend that this estimator should not be interpreted when a significant amount of heterogeneity ($I^2 > 50\%$) is present in the data (van Aert et al., 2016). As this is the case, it was interpreted. For *p*-uniform (van Assen et al., 2015), the test yielded a *p*-value of .997, and for *p*-uniform* (van Aert & van Assen, 2021), *p* was at .273. Thus, both tests did not indicate the presence of dissemination bias.

Figure 3

Funnel plot of effect sizes and standard errors.

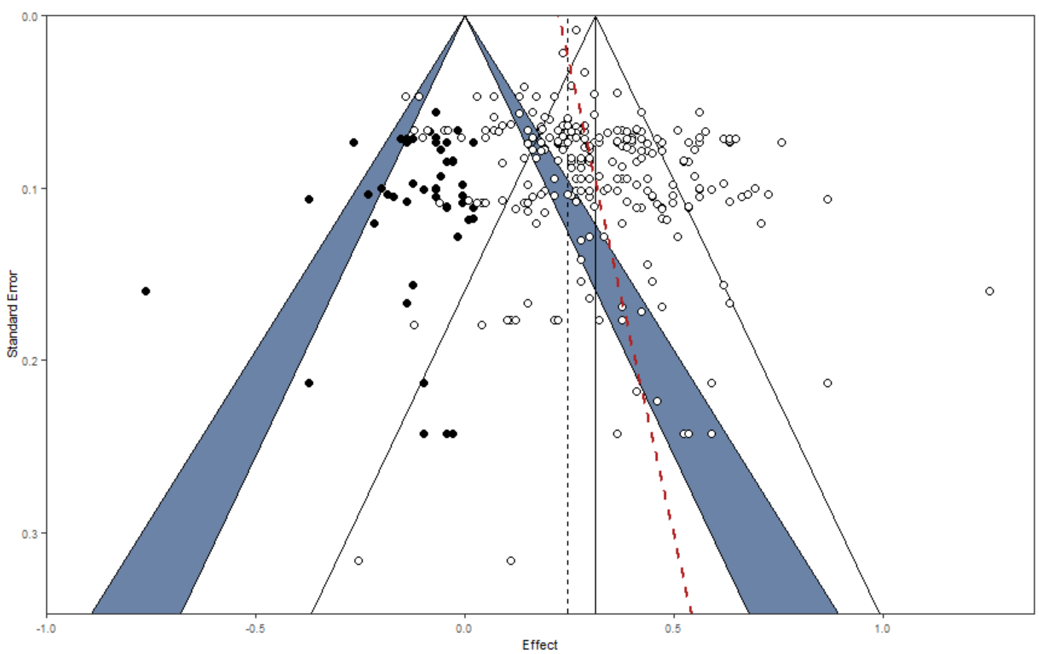
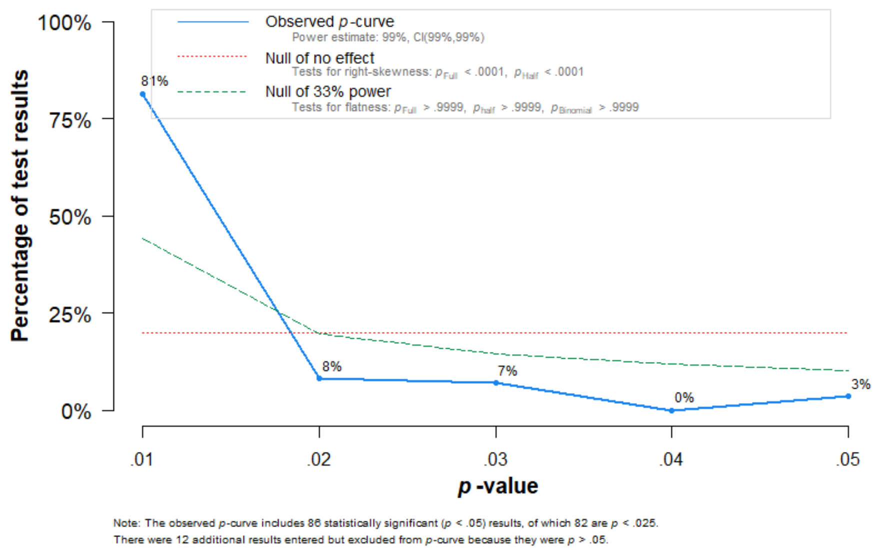


Figure 4

Results from the p-curve analysis.



Third, the expected significant effect sizes were compared to the observed number by applying the test of excess significance (Ioannidis & Trikalinos, 2007). However, no significant difference was observed (p = .207).

Finally, it was planned to investigate dissemination bias by performing a moderator analysis of published versus unpublished studies. However, no relevant unpublished studies were found during the literature search.

Overall, there was no indication of p-hacking in this data. Some results indicated the presence of a substantial amount of funnel plot asymmetry. However, the difference between the adjusted and previously estimated summary effect did not exceed 20%. Therefore, the missing effects do not seem to impact the overall effect size notably.

Discussion

The present meta-analysis focussed on the relationship between self-assessed and psychometrically assessed intelligence. However, with the growing number of publications addressing this very issue, previous meta-analytic findings (e.g., Freund & Kasten, 2012; Mabe & West, 1982) have become outdated. Moreover, a variety of more refined research methods (e.g., for the investigation of dissemination bias) have since been implemented. Furthermore, recent studies' findings report a potential bias due to virtual ubiquitous declining effects (Pietschnig et al., 2019). Therefore, this study aimed to investigate the relationship between self-assessed intelligence and psychometric intelligence, replicating and updating the previous meta-analytical findings while using a total of eight standard and more modern methods to address dissemination bias.

In total, 242 effect sizes from 98 studies were synthesized, which yielded a total correlation of $r = .31$ when the Hedges and Olkin (1975) approach was used. As expected, using a Hunter and Schmidt (2004)–typed approach resulted in a larger correlation of $r = .39$. The correlation between SAI and IQ, therefore, appears to be moderate but remarkably stable. However, when comparing these results to previous meta-analytical accounts ($r = .34$, Mabe & West, 1982; $r = .33$, Freund & Kasten, 2012), estimations seem to be declining slightly. Recent findings propose suboptimal initial study power due to questionable research practices as the main reason for misestimations in initial studies and the resulting declining effects (Pietschnig et al., 2019).

Additionally, the application of eight dissemination bias methods showed a substantial amount of funnel plot asymmetry (Begg & Mazumdar's rank correlation test (1994): Kendall's $\tau = .176$, $p < .10$; Sterne & Egger's approach (2005): $z = 2.56$, $p < .10$). Even though differences between the adjusted and estimated summary effect were not meaningful (Duval & Tweedie's trim and fill procedure, 2000a, 2000b, and Stanley & Doucouliagos' PET PEESE method, 2014), the results must be considered to be somewhat inflated. Therefore, one could argue that the summary effect yielded by the Hedges and Olkin–typed analysis (1996) might provide a more accurate estimator, as the Hunter and Schmidt (2004) –typed approach could inflate the effect even more.

Additionally, a significant amount of heterogeneity was observed in the data. Therefore, a total of seven potential moderators were included to explain parts of this variation. In line with previous meta-analytical findings (Freund & Kasten, 2012), moderator analyses showed that the correlation between SAI and IQ seems to be remarkably stable across moderators, as it generalized over the order of assessment, participant sex, and sample composition. However, contrary to previous findings by Freund and Kasten (2012), there were no differences between the four methods of self-assessment. They reported that relative scales with the explicit mention of a references group led to the most valid self-estimates, followed by “normal” relative scales (compared to absolute scales); using a mixed scale did not significantly improve over absolute scales. In this study, no significant differences between the four methods of self-assessment were observed.

Regarding the ability type, it was hypothesized that more common cognitive abilities like numerical, verbal, and spatial ability would lead to more valid self-estimates compared to general cognitive ability. In line with the hypothesis, estimates were most valid when the numerical ability was assessed while estimating other “less known” cognitive abilities decreased validity. However, contrary to previous assumptions, the assessment of verbal ability did not increase the validity compared to general cognitive ability. Surprisingly, the assessment of spatial ability even led to a slight decrease in validity. For reference: the single model (without all other moderator variables) following the Hedges and Olkin approach (1996) estimated the mean effect for numerical, general, spatial, and other less-known abilities (e.g., naturalistic intelligence) to be r = .452, r = .323, r = .240, and r = .224 respectively.

A possible explanation concerns laypeople’s definition of certain constructs. Laypeople typically tend to have their own definitions of constructs like intelligence and the associated cognitive abilities. For example, some abilities like verbal intelligence are seen as more malleable than creative and musical intelligence (Furnham, 2014). This assumption of malleability might further come with uncertainty concerning self-assessment of abilities. Therefore, people might have an easier time estimating abilities, which they perceive as constant, than abilities they perceive as malleable (e.g., verbal ability). It was previously argued that tests assessing numerical ability are often similar to math problems (e.g., numerical sequence tests) with which students are confronted early at school (Freund & Kasten, 2012). Therefore, most people might find it easier to estimate numerical ability than, for example, spatial ability.

Moreover, tests for spatial ability (e.g., mental rotation tasks) tend to be somewhat abstract and less salient. This might result in problems regarding self-estimates as laypeople might lack the ability to connect real-life situations, where they have to use their spatial abilities (e.g., reading maps) to the definition of spatial ability at hand. Therefore, the importance of detailed test instructions containing an exact definition of the construct and the necessity of practical, real-life examples becomes apparent.

It was further hypothesized that self-estimated would be more accurate when people estimated their abilities after the psychometric assessment. However, correlations were not significantly influenced by order of assessment. One possible explanation is that individuals with negative self-perceptions might perceive the given tasks to be more challenging (Zell & Krizan, 2014). Furthermore, this would result in biased self-perceptions of people with negative self-views, despite showing the same objective test performance as people with positive self-views. Therefore, this could cancel out the positive effects of the assessment order.

Another explanation is that people might have previously developed a fixed image of their abilities, which might not be influenced by a single testing experience (Freund & Kasten, 2012).

Regarding sample composition, it was assumed that for general samples (“non-academic”), the correlation between self-assessed and psychometrically assessed intelligence would be higher than for student samples (“academic”). However, there was no significant difference between the student sample and the general sample regarding the validity of self-estimates. One possible explanation could be that even when students have

Limitations of Existing Research

First, a fair amount of studies from technical fields were excluded during the literature search: The search strategy resulted in almost 18,000 potentially relevant studies published between 2011 and 2020. For comparison: The previous meta-analysis only reported to have found 238 potentially relevant studies (Freund & Kasten, 2012). As literature screening was performed solely by the author, it would have been nearly impossible to screen this amount of studies in a reasonable time frame. Therefore, it was decided to exclude all studies, which stem from technical fields (e.g., computer science or electrical engineering). However, this still led to the screening of almost 14,000 potentially relevant studies. Furthermore, it was considered rather unlikely that the correlation between SAI and IQ might be a research topic of interest in technical fields.

Second, it was planned to include grey literature in this meta-analysis as well. However, screening the first 100 hits of Google Scholar and the Open Access Theses and Dissertation database (OATD) did not result in any eligible studies.

Third, all studies were screened and coded solely by the author. Thus, even though Kappa values were satisfactory and screening was performed independently, this might still be a possible source of error.

Fourth, heterogeneity might have influenced some methods for dissemination bias detection. As previously mentioned, a significant amount of heterogeneity (~84%) was present in the data. In the past, some methods of dissemination bias detection have been shown to be less valid in the presence of between-study heterogeneity. For example, Duval and Tweedie's trim-and-fill procedure (2002b, 2002a) has been reported to not perform well in the presence of heterogeneity (van Assen et al., 2015), and it is not recommended to use the test for excess in such cases (Ioannidis & Trikalinos, 2007). Furthermore, both, *p*-curve (Simonsohn et al., 2014a, 2014b) and *p*-uniform (van Assen et al., 2015) have shown increasing false-positive rates (Carter et al., 2019) and overestimation of summary effects (Carter et al., 2019; van Aert et al., 2016) with increasing heterogeneity.

Finally, most reliability values used for the Hunter and Schmidt (2004) correction had to be estimated. For 173 out of 242 effect sizes, no reliability values were reported for the psychometric test, 135 of which were imputed using the R package psychmeta (Dahlke & Wiernik, 2019). If a sufficient number of studies reported reliability values for a specific psychometric test, those values were averaged and used in the case of missing data (this was the case for 38 effect sizes). Concerning the reliability of self-assessed intelligence, 147 out of 242 reliabilities had to be imputed (Dahlke & Wiernik, 2019). Although the imputation of missing reliability values generally works well and yields accurate effect sizes estimations and mean standard errors (Dahlke & Wiernik, 2019; Schmidt & Hunter, 2015), this might still be a possible error source.

Conclusion

This study presents evidence for a modestly sized but robust association of IQ and SAI. It is remarkably stable across moderators as it generalizes over assessment order, participants' sex, sample composition, and, contrary to previous findings, self-assessment methods but is differentiated according to ability type. When compared to the estimation of general cognitive ability, correlations were highest for numerical ability. Estimating spatial and other less frequently investigated cognitive abilities (such as naturalistic intelligence) led to a decrease in validity.

As expected, using the unreliability corrected Hunter and Schmidt (2004)–typed approach (2004) led to a larger summary effect than the Hedges and Olkin approach (1996). Moreover, most methods suggested a significant amount of funnel plot asymmetry. Therefore, results have to be considered somewhat inflated. Even though Duval and Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000b, 2000a) did not show substantial bias, in general, bias appears to be present. Therefore, one could argue that the summary effect provided by the Hedges and Olkin–typed analysis (1996) might deliver a more accurate estimator, as the Hunter and Schmidt (2004) –typed approach could inflate the effect even further.

Practical implications that emerge from this study relate, for example, to career counseling. It has been reported that professional interests are primarily a function of self-assessed rather than „true" abilities (Neubauer & Hofer, 2021). Furthermore, career counseling is seldom based on actual cognitive abilities but self-assessed abilities. For example, the Public Employment Service Austria offers a free online test designed to help adolescents make the right career decision. However, all recommendations are primarily based on self-estimates of interests and abilities. In the light of the results present, one could assume that this would (at least in parts) results in somewhat inaccurate estimations of an individual's abilities. Moreover, a "misfit" between a person's abilities and job demands has been associated with strain indicators (Kristof-

Brown et al., 2005) in the past. Therefore, career counselors and individuals should be made aware of potentially faulty/biased self-estimates.

In all, this study presents evidence for a modestly sized but reproducible association of psychometric and self-assessed intelligence, which is remarkably stable across moderators but differentiates according to ability type.

References

Marked (*) references were included in the meta-analysis.

*Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*(4), 587–605. https://doi.org/10.1016/S0191-8869(01)00174-X

*Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*(4), 797–825. https://doi.org/10.1037//0022-0663.93.4.797

*Ackerman, P. L., & Ellingsen, V. J. (2014). Vocabulary overclaiming-A complete approach: Ability, personality, self-concept correlates, and gender differences. *Intelligence*, *46*, 216–227. https://doi.org/http://dx.doi.org/10.1016/j.intell.2014.07.003

*Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, *1*(4), 270–304. https://doi.org/1076-898X/95/$3.00

*Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, *13*(2), 57–78. https://doi.org/10.1037/1076-898X.13.2.57

*Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there sex differences in confidence and metacognitive monitoring accuracy for everyday, academic, and psychometrically measured spatial ability? *Intelligence*, *70*, 42–51. https://doi.org/10.1016/j.intell.2018.08.001

Arnett, J. J. (2008). The Neglected 95%: Why american psychology needs to become less american. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Assink, M., & Wibbelink, C. (2016). Fitting three-level meta-analytic models in R : A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

*Bacon, E., Huet, N., & Danion, J.-M. (2011). Metamemory knowledge and beliefs in patients with schizophrenia and how these relate to objective cognitive abilities. *Consciousness and Cognition*, *20*(4), 1315–1326. https://doi.org/10.1016/j.concog.2011.02.017

*Bailey, K. G., & Lazar, J. (1976). Accuracy of self-ratings of intelligence as a function of sex and level of ability in college students. *The Journal of Genetic Psychology*, *129*(2), 279–290. https://doi.org/10.1080/00221325.1976.10534039

*Bailey, R. C., & Bailey, K. G. (1974). Self-perceptions of scholastic ability at four grade levels. *The Journal of Genetic Psychology*, *124*(2), 197–212. https://doi.org/10.1080/00221325.1974.10532281

*Bailey, R. C., & Mettetal, G. W. (1977). Perceived intelligence in married partners. *Social Behavior and Personality: An International Journal*, *5*(1), 137–141. https://doi.org/https://doi.org/10.2224/sbp.1977.5.1.137

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1007/978-3-319-75361-4

Baujat, B., Mahé, C., Pignon, J. P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, *21*(18), 2641–2652. https://doi.org/10.1002/sim.1221

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

*Bipp, T., & Kleingeld, A. (2012). Self-estimates of intelligence: Interaction effects of the comparison to a specific reference group and neuroticism. *Psychological Reports*, *110*(2), 403–415. https://doi.org/10.2466/04.07.09.PR0.110.2.403-415

*Bipp, T., Steinmayr, R., & Spinath, B. (2012). A functional look at goal orientations: Their role for self-estimates of intelligence and performance. *Learning and Individual Differences*, *22*(3), 280–289. https://doi.org/10.1016/j.lindif.2012.01.009

*Borella, E., Meneghetti, C., Ronconi, L., & De Beni, R. (2014). Spatial Abilities Across the Adult Life Span. *Developmental Psychology*, *50*(2), 384–392. https://doi.org/10.1037/a0033818

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). I*ntroduction to Meta-Analysis*. John Wiley & Sons. https://doi.org/https://doi.org/10.1002/9780470743386

Boring, E. G. (1923). Intelligence as the test tests it. *The New Republic*, *36*, 35–37.

*Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings and measured intelligence. *Journal of Personality and Social Psychology*, *65*(3), 546–553. https://doi.org/0022-3514/93/$3.00

Brady-Amoon, P., & Fuertes, J. N. (2011). Self-efficacy, self-rated abilities, adjustment, and academic performance. *Journal of Counseling & Development*, *89*(4), 431–438. https://doi.org/10.1002/j.1556-6676.2011.tb02840.x

*Bressan, P. (2018). Systemisers are better at maths. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-30013-8

*Brim, O. G. (1954). College grades and self-estimates of intelligence. *Journal of Educational Psychology*, *45*(8), 477–484. https://doi.org/10.1037/h0057492

Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guilford Press.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

*Chamorro-Premuzic, T., & Furnham, A. (2006). Self-assessed intelligence and academic performance. *Educational Psychology*, *26*(6), 769–779. https://doi.org/10.1080/01443410500390921

*Chamorro-Premuzic, T., Furnham, A., & Moutafi, J. (2004). The relationship between estimated and psychometric personality and intelligence scores. *Journal of Research in Personality*, *38*(5), 505–513. https://doi.org/10.1016/j.jrp.2003.10.002

*Chamorro-Premuzic, T., Moutafi, J., & Furnham, A. (2005). The relationship between personality traits, subjectively-assessed and fluid intelligence. *Personality and Individual Differences*, *38*(7), 1517–1528. https://doi.org/10.1016/j.paid.2004.09.018

*Cogan, L. C., Conklin, A. M., & Hollingworth, H. L. (1915). An experimental study of self-analysis, estimates of associates, and the results of tests. *School & Society*, *2*, 171–179. https://doi.org/10.1037/10729-007

Costa, P. T., & McCrae, R. R. (2011). The five-factor model, five-factor theory, and interpersonal psychology. In L. M. Horowitz, & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 91–104). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118001868.ch6

Dahlke, J. A., & Wiernik, B. M. (2019). psychmeta: An R package for psychometric meta-analysis. *Applied Psychological Measurement*, *43*(5), 415–416. https://doi.org/10.1177/0146621618795933

*de Keersmaecker, J., Onraet, E., Lepouttre, N., & Roets, A. (2017). The opposite effects of actual and self-perceived intelligence on racial prejudice. *Personality and Individual Differences*, *112*, 136–138. https://doi.org/10.1016/j.paid.2017.02.057

*DeNisi, A. S., & Shaw, J. B. (1977). Investigation of the uses of self-reports of abilities. *Journal of Applied Psychology*, *62*(5), 641–644. https://doi.org/10.1037//0021-9010.62.5.641

*Deuling, J. K., Denissen, J. J. A., van Zalk, M., Meeus, W., & Aken, M. V. (2011). Perceived influence in groups over time: How associations with personality and cognitive ability can change over time. *Journal of Research in Personality*, *45*(6), 576–585. https://doi.org/10.1016/j.jrp.2011.07.005

*Diener, C., Bornschlegl, M., Menke, B., & Petermann, F. (2016). Subjektive Einschätzung und objektive kognitive Leistung bei psychisch Gesunden. *Zeitschrift Fur Neuropsychologie*, *27*(3), 147–157. https://doi.org/10.1024/1016-264X/a000186

*Dislich, F. X. R., Imhoff, R., Banse, R., Altstötter-Gleich, C., Zinkernagel, A., & Schmitt, M. (2012). Discrepancies between implicit and explicit self-concepts of intelligence predict performance on tests of intelligence. *European Journal of Personality*, *26*(3), 212–220. https://doi.org/10.1002/per.827

*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*(3), 537–572. https://doi.org/10.1111/j.1467-6494.2011.00742.x

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Ebbinghaus, H. (1908). *Psychology: An elementary text-book.* D C Heath & Co Publishers. https://doi.org/10.1037/13638-000

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(629), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Feeney, J. (2018). Assessing and Preventing Applicant Faking on Personality Tests. *Electronic Thesis and Dissertation Repository*, *5424*. https://ir.lib.uwo.ca/etd/5424

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. In Reynolds, C. R., Vannest, K. J., & Fletcher-Janzen, E. (Eds.), *Encyclopedia of Special Education* (pp. 368–382). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ese0431

*Fragkiadaki, S., Kontaxopoulou, D., Beratis, I. N., Andronas, N., Economou, A., Yannis, G., Papanicolaou, A., & Papageorgiou, S. G. (2016). Self-awareness of cognitive efficiency: Differences between healthy elderly and patients with mild cognitive impairment (MCI). *Journal of Clinical and Experimental Neuropsychology*, *38*(10), 1144–1157. https://doi.org/10.1080/13803395.2016.1198469

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*(2), 296–321. https://doi.org/10.1037/a0026556

Furnham, A. (2000). Parents' estimates of their own and their children's multiple intelligences. *British Journal of Developmental Psychology*, *18*(4), 583–594. https://doi.org/10.1348/026151000165869

Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, *31*(8), 1381–1405. https://doi.org/10.1016/S0191-8869(00)00232-4

*Furnham, A. (2005). Self-estimated intelligence, psychometric intelligence and personality. *Psychologia*, *48*(3), 182–192. https://doi.org/10.2117/psysoc.2005.182

*Furnham, A. (2009). The validity of a new, self-report measure of multiple intelligence. *Current Psychology*, *28*(4), 225–239. https://doi.org/10.1007/s12144-009-9064-z

Furnham, A. (2014). Increasing your intelligence: Entity and incremental beliefs about the multiple “intelligences.” *Learning and Individual Differences*, *32*, 163–167. https://doi.org/10.1016/j.lindif.2014.03.001

*Furnham, A. (2018). Estimating one’s own and other’s psychological test scores. *Psychology*, *9*(8), 2231–2249. https://doi.org/10.4236/psych.2018.98127

*Furnham, A., & Chamorro-Premuzic, T. (2004). Estimating one’s own personality and intelligence scores. *British Journal of Psychology*, *95*(2), 149–160. https://doi.org/10.1348/000712604773952395

*Furnham, A., & Dissou, G. (2007). The relationship between self-estimated and test-derived scores of personality and intelligence. *Journal of Individual Differences*, *28*(1), 37–44. https://doi.org/10.1027/1614-0001.28.1.37

*Furnham, A., & Fong, G. (2000). Self-estimated and psychometrically measured intelligence: A cross-cultural and sex differences study of British and Singaporean students. *North American Journal of Psychology*, *2*(2), 191–200.

*Furnham, A., & Grover, S. (2020). Correlates of self-estimated intelligence. *Journal of Intelligence*, *8*(1). https://doi.org/10.3390/jintelligence8010006

*Furnham, A., Kidwai, A., & Thomas, C. (2001). Personality, psychometric intelligence, and self-estimated intelligence. *Journal of Social Behavior and Personality*, *16*(1), 97–114.

*Furnham, A., Moutafi, J., & Chamorro-Premuzic, T. (2005). Personality and intelligence: Gender, the big five, self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, *13*(1), 11–24. https://doi.org/10.1111/j.0965-075X.2005.00296.x

*Furnham, A., & Rawles, R. (1999). Correlations between self-estimated and psychometrically measured IQ. *The Journal of Social Psychology*, *139*(4), 405–410. https://doi.org/10.1080/00224549909598400

*Furnham, A., Zhang, J., & Chamorro-Premuzic, T. (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, Cognition and Personality*, *25*(2), 119–145. https://doi.org/10.2190/530v-3m9u-7uq8-fmbg

*Gabriel, M. T., Critelli, J. W., & Ee, J. S. (1994). Narcissistic illusions in self-evaluations of intelligence and attractiveness. *Journal of Personality*, *62*(1), 143–155. https://doi.org/10.1111/j.1467-6494.1994.tb00798.x

*Gerstenberg, F. X. R., Imhoff, R., Banse, R., Altstotter-Gleich, C., Zinkernagel, A., & Schmitt, M. (2013). How implicit-explicit consistency of the intelligence self-concept moderates reactions to performance feedback. *European Journal of Personality*, *27*(3), 238–255. https://doi.org/10.1002/per.1900

*Gerstenberg, F. X. R., Imhoff, R., Banse, R., & Schmitt, M. (2014). Discrepancies between implicit and explicit self-concepts of intelligence: Relations to modesty, narcissism, and achievement motivation. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00085

*Gignac, G. E. (2018). Socially desirable responding suppresses the association between self-assessed intelligence and task-based intelligence. *Intelligence*, *69*, 50–58. https://doi.org/10.1016/j.intell.2018.05.003

*Gignac, G. E., & Zajenkowski, M. (2019). People tend to overestimate their romantic partner's intelligence even more than their own. *Intelligence*, *73*, 41–51. https://doi.org/10.1016/j.intell.2019.01.004

*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. https://doi.org/10.1016/j.intell.2020.101449

Gottfredson, L. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis in R: A hands-on guide*. Chapmann & Hall/CRC Press.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). *The extent and consequences of p-hacking in science*. 1–15. https://doi.org/10.1371/journal.pbio.1002106

Hedges, L. V, & Olkin, I. (1996). *Statistical methods for meta-analysis* (6th ed.). Acad. Press. https://ubdata.univie.ac.at/AC02172465

*Herreen, D., & Zajac, I. T. (2017). The reliability and validity of a self-report measure of cognitive abilities in older adults: More personality than cognitive function. *Journal of Intelligence*, *6*(1), 1. https://doi.org/10.3390/jintelligence6010001

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7417), 557–560. https://doi.org/10.1136/bmj.327.7414.557

*Holling, H., & Preckel, F. (2005). Self-estimates of intelligence - Methodological approaches and gender differences. *Personality and Individual Differences*, *38*(3), 503–517. https://doi.org/10.1016/j.paid.2004.05.003

Howard, M. C., & Cogswell, J. E. (2018). The “other” relationships of self-assessed intelligence: A meta-analysis. *Journal of Research in Personality*, *77*, 31–46. https://doi.org/10.1016/j.jrp.2018.09.006

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.

*Huang, Z., & Maurer, J. (2019). Validity of self-rated memory among middle-aged and older Chinese adults: Results from the China Health and Retirement Longitudinal Study (CHARLS). *Assessment*, *26*(8), 1582–1593. https://doi.org/10.1177/1073191117741188

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage Publications, Inc.

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*(3), 245–253. https://doi.org/10.1177/1740774507079441

*Jacobs, K. E., & Roodenburg, J. (2014). The development and validation of the self-report measure of cognitive abilities: A multitrait-multimethod study. *Intelligence*, *42*(1), 5–21. https://doi.org/10.1016/j.intell.2013.09.004

*Jacobs, K. E., Szer, D., & Roodenburg, J. (2012). The moderating effect of personality on the accuracy of self-estimates of intelligence. *Personality and Individual Differences*, *52*(6), 744–749. https://doi.org/10.1016/j.paid.2011.12.040

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, *87*(4), 797–807. https://doi.org/10.1037/0021-9010.87.4.797

*Kajonius, P. J. (2014). Honesty-humility in contemporary students: Manipulations of self-image by inflated IQ estimations. *Psychological Reports*, *115*(1), 311–325. https://doi.org/10.2466/17.04.PR0.115c13z8

*Kornilova, T. V, Kornilov, S. A., & Chumakova, M. A. (2009). Subjective evaluations of intelligence and academic self-concept predict academic achievement: Evidence from a selective student population. *Learning and Individual Differences*, *19*(4), 596–608. https://doi.org/10.1016/j.lindif.2009.08.001

*Kornilova, T. V, & Novikova, M. A. (2011). САМООЦЕНКА В СТРУКТУРЕ ИНТЕЛЛЕКТУАЛЬНО- ЛИЧНОСТНОГО ПОТЕНЦИАЛА ЧЕЛОВЕКА [Self-estimate of men ’ s intellectual-personal potential strukture]. *ПСИХОЛОГИЧЕСКИЙ ЖУРНАЛ*, *2*(32), 25–35.

*Kornilova, T. V, & Novikova, M. A. (2012). Self-assessed intelligence, personality, and psychometric intelligence: Preliminary validation of a model with a selected student population. *Psychology in Russia: State of the Art*, *5*, 33–49. https://doi.org/10.11621/pir.2012.0002

*Kornilova, T. V, & Novikova, M. A. (2013). Self-assessed intelligence, psychometric intelligence, personality, and academic achievement: Two structural models. In M. Gowda, & A. Khanderia (Eds.), *Educational achievement: Teaching strategies, psychological factors and economic impact* (pp. 197–212).

Kossmeier, M., Tran, U. S., & Voracek, M. (2020). *Forest plots, funnel plots, and visual funnel plot inference for meta-analysis*. R package version 0.3.1.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

*Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the association of self-enhancement bias with self-ratings of personality and life satisfaction. *Assessment*, *23*(5), 588–602. https://doi.org/10.1177/1073191115590852

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. https://doi.org/10.1037//0021-9010.67.3.280

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemica Medica*, *22*(3), 276–282. https://hrcak.srce.hr/89395

*Meneghetti, C., Borella, E., Pastore, M., & De Beni, R. (2014). The role of spatial abilities and self-assessments in cardinal point orientation across the lifespan. *Learning and Individual Differences*, *35*, 113–121. https://doi.org/10.1016/j.lindif.2014.07.006

*Mitolo, M., Gardini, S., Caffarra, P., Ronconi, L., Venneri, A., & Pazzaglia, F. (2015). Relationship between spatial ability, visuospatial working memory and self-assessed spatial orientation ability: A study in older adults. *Cognitive Processing*, *16*(2), 165–176. https://doi.org/ 10.1007/s10339-015-0647-3

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Neubauer, A. C., & Hofer, G. (2021). Self-estimates of abilities are a better reflection of individuals' personality traits than of their abilities and are also strong predictors of professional interests. *Personality and Individual Differences*, *169*, Article 109850. https://doi.org/10.1016/j.paid.2020.109850

Ng, J. R., & Earl, J. K. (2008). Accuracy in self-assessment: the role of ability, feedback, self-efficacy and goal orientation. *Australien Journal of Career Development*, *17*(3), 39–50. https://doi.org/10.1177/103841620801700307

*Nyström, M. B. T., Sörman, D. E., Kormi-Nouri, R., & Rönnlund, M. (2019). To what extent is subjective well-being in late adulthood related to subjective and objective memory functioning? Five-year cross-lagged panel analyses. *Aging and Mental Health*, *23*(1), 92–99. https://doi.org/10.1080/13607863.2017.1394439

*Owczarek, K., Bazan, M., & Mirska, N. (2012). Norm-based subjective evaluation of cognitive processes in epileptic patients: A microgenetic approach. *Acta Neuropsychologica*, *10*(1), 81–95. https://doi.org/10.5604/17307503.995090

*Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality*, *66*(4), 525–554. https://doi.org/10.1111/1467-6494.00023

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Picho-Kiroga, K., Turnbull, A., & Rodriguez-Leahy, A. (2021). Stereotype threat and its problems: Theory misspecification in research, consequences, and remedies. *Journal of Advanced Academics*, *32*(2), 231–264. https://doi.org/10.1177/1932202X20986161

Pietschnig, J., Siegel, M., Eder, J. S. N., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in Psychology*, *10*, 1–13. https://doi.org/10.3389/fpsyg.2019.02874

*Proyer, R. T., & Ruch, W. (2009). Intelligence and gelotophobia: The relations of self-estimated and psychometrically measured intelligence to the fear of being laughed at. *Humor*, *22*(1/2), 165–181. https://doi.org/10.1515/HUMR.2009.008

*Rammstedt, B., & Rammsayer, T. H. (2002a). Die Erfassung von selbsteingeschätzter Intelligenz. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *23*(4), 435–446. https://doi.org/10.1024//0170-1789.23.4.435

*Rammstedt, B., & Rammsayer, T. H. (2002b). Self-estimated intelligence: Gender differences, relationship to psychometric intelligence and moderating effects of level of education. *European Psychologist*, *7*(4), 275–284. https://doi.org/10.1027//1016-9040.7.4.275

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Sage Publ. https://ubdata.univie.ac.at/AC03658798

*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. https://doi.org/10.1016/0191-8869(94)00148-L

Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage Publications, Inc. https://doi.org/10.4135/9781412984997

*Rothlind, J., Dukarm, P., & Kraybill, M. (2017). Assessment of self-awareness of cognitive function: Correlations of self-ratings with actual performance ranks for tests of processing speed, memory and executive function in non-clinical samples. *Archives of Clinical Neuropsychology*, *32*(3), 316–327. https://doi.org/10.1093/arclin/acw109

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons Ltd. https://doi.org/10.1002/0470870168

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of metaanalysis: Correcting error and bias in research findings* (3rd ed.). SAGE Publications, Ltd. https://doi.org/doi:10/b6mg

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 45–65. https://doi.org/10.1146/annurev-op-04-030217-100001

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan, & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.

*Schwert, C., Stohrer, M., Aschenbrenner, S., Weisbrod, M., & Schröder, A. (2018). Biased neurocognitive self-perception in depressive and in healthy persons. *Journal of Affective Disorders*, *232*, 96–102. https://doi.org/10.1016/j.jad.2018.02.031

Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2021). *Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices , prevalence rates , and predictors in industrial / organizational psychology*. PsyArXiv. https://doi.org/10.31234/osf.io/mtv2h

Simmonds, M. (2015). Quantifying the risk of error when interpreting funnel plots. *Systematic Reviews*, *4*(1), 1–7. https://doi.org/10.1186/s13643-015-0004-8

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve and effect size: correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666–681. https://doi.org/10.1177/1745691614553988

*Soh, L., & Jacobs, K. E. (2013). The biasing effect of personality on self-estimates of cognitive abilities in males and females. *Personality and Individual Differences*, *55*(2), 141–146. https://doi.org/10.1016/j.paid.2013.02.013

*Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition*, *182*, 193–212. https://doi.org/10.1016/j.cognition.2018.10.005

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, *8*(5), 581–591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H.R. Rothstein, A.J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* (pp. 99–110). Wiley. https://doi.org/10.1002/0470870168.ch6

Stieger, S., Kastner, C. K., Voracek, M., Von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2010). Independent effects of personality and sex on selfestimated intelligence: Evidence from Austria. *Psychological Reports*, *107*(2), 553–563. https://doi.org/10.2466/04.07.09.PR0.107.5.553-563

Storek, J., & Furnham, A. (2012). Gender and gender role differences in domain-masculine intelligence and beliefs about intelligence: A study with Mensa UK members. *Personality and Individual Differences*, *53*(7), 890–895. https://doi.org/10.1016/j.paid.2012.05.039

*Storek, J., & Furnham, A. (2013a). Gender, “g”, and fixed versus growth intelligence mindsets as predictors of self-estimated domain masculine intelligence (DMIQ). *Learning and Individual Differences*, *25*, 93–98. https://doi.org/10.1016/j.lindif.2013.03.007

*Storek, J., & Furnham, A. (2013b). Gender, g, gender identity concepts, and self-constructs as predictors of the self-estimated IQ. *Journal of Genetic Psychology*, *174*(6), 664–676. https://doi.org/10.1080/00221325.2013.772501

Syzmanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences*, *21*(5), 493–504. https://doi.org/10.1016/j.lindif.2011.07.001

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129. https://doi.org/10.1037/amp0000622

*Torres, I. J., Mackala, S. A., Kozicky, J.-M., & Yatham, L. N. (2016). Metacognitive knowledge and experience in recently diagnosed patients with bipolar disorder. *Journal of Clinical and Experimental Neuropsychology*, *38*(7), 730–744. https://doi.org/10.1080/13803395.2016.1161733

Tracey, T. J. G. (2016). A note on socially desirable responding. *American Psychological Association*, *63*(2), 224–232. https://doi.org/http://dx.doi.org/10.1037/cou0000135

van Aert, R. C. M., & van Assen, M. A. L. M. (2021). *Correcting for publication bias in a meta-analysis with the p-uniform* method* [Manuscript submitted for publication]. Department of Methodology and Statistics, Tilburg University. https://osf.io/preprints/metaarxiv/zqjr9/download

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, *11*(5), 713–729. https://doi.org/10.1177/1745691616650874

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. https://doi.org/10.1037/met0000025

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

*Villado, A. J., Randall, J. G., & Zimmer, C. U. (2016). The effect of method characteristics on retest score gains and criterion-related validity. *Journal of Business and Psychology*, *31*(2), 233–248. https://doi.org/10.1007/s10869-015-9408-7

*Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What makes you think you're so smart? Measured abilities, personality, and sex differences in relation to self-estimates of multiple intelligences. *Journal of Individual Differences*, *29*(1), 35–44. https://doi.org/10.1027/1614-0001.29.1.35

* von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology*, *105*(2), 243–253. https://doi.org/10.1111/bjop.12031

*Webb, W. B. (1955). Self-evaluations, group evaluations, and objective measures. *Journal of Consulting Psychology*, *19*(3), 210–212. https://doi.org/10.1037/h0047190

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, *3*(1), 94–123. https://doi.org/10.1177/2515245919885611

*Wolff, R., & Wasden, R. (1969). Measured intelligence and estimates by nursing instructors and nursing students. *Psychological Reports*, *25*(1), 77–78. https://doi.org/10.2466/pr0.1969.25.1.77

*Zajenkowski, M., & Czarna, A. Z. (2015). What makes narcissists unhappy? Subjectively assessed intelligence moderates the relationship between narcissism and psychological well-being. *Personality and Individual Differences*, *77*, 50–54. https://doi.org/10.1016/j.paid.2014.12.045

*Zajenkowski, M., Czarna, A. Z., Szymaniak, K., & Dufner, M. (2019). What do highly narcissistic people think and feel about (their) intelligence? *Journal of Personality*, *88*(4), 703–718. https://doi.org/10.1111/jopy.12520

*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. https://doi.org/10.1016/j.intell.2018.07.003

*Zajenkowski, M., Jankowski, K. S., & Stolarski, M. (2019). Why do evening people consider themselves more intelligent than morning individuals? The role of big five, narcissism, and objective cognitive ability. *Chronobiology International*, *36*(12), 1741–1751. https://doi.org/10.1080/07420528.2019.1680559

*Zajenkowski, M., Leniarska, M., & Jonason, P. K. (2020). Look how smart I am! Only narcissistic admiration is associated with inflated reports of intelligence. *Personality and Individual Differences*, *165*(June), Article 110158. https://doi.org/10.1016/j.paid.2020.110158

*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. https://doi.org/10.1016/j.intell.2016.06.002

Zell, E., & Krizan, Z. (2014). Do People Have Insight Into Their Abilities? A Metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

List of Figures

List of Tables

Abbreviations

Abbreviation	Meaning
AIC	Akaike-Information-Criterion
BIC	Bayesian-Information-Criterion
CHC	Cattell-Horn-Carroll
H&S	Hunter and Schmidt
H&O	Hedges and Olkin
LRT	Likelihood-Ratio-Test
IQ	Intelligence Quotient
PE fit	Person-Environment fit
REML	Restricted Maximum Likelihood
SAI	Self-Assessed Intelligence

Appendices

Appendix A: Overview of all coded effect sizes

Table A1

Overview of all coded effect sizes

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Ackerman & Ellingsen (2014)	.38	.40	193	Student sample	Mixed	Mixed scale	Verbal	Estimate	Test battery	Estimated	Estimated	N/A	N/A
Ackerman & Wolman (2007)	.25	.26	142	Student sample	Mixed	Mixed scale	Verbal	Estimate	Test battery	0.92	Estimated	0.51	0.72
	.48	.52	142	Student sample	Mixed	Mixed scale	Numerical	Estimate	Test battery	0.93	Estimated	0.51	0.72
	.34	.35	142	Student sample	Mixed	Mixed scale	Spatial	Estimate	Test battery	0.89	Estimated	0.51	0.72
	.27	.28	142	Student sample	Mixed	Mixed scale	General	Estimate	Test battery	Estimated	Estimated	0.51	0.72
	.25	.26	142	Student sample	Mixed	Mixed scale	Verbal	Test	Test battery	0.91	Estimated	0.51	0.72
	.49	.54	142	Student sample	Mixed	Mixed scale	Numerical	Test	Test battery	0.95	Estimated	0.51	0.72
	.39	.41	142	Student sample	Mixed	Mixed scale	Spatial	Test	Test battery	0.91	Estimated	0.51	0.72
	.29	.30	142	Student sample	Mixed	Mixed scale	General	Test	Test battery	Estimated	Estimated	0.51	0.72
Ackerman et al. (1995)	.29	.30	93	Student sample	Mixed	Mixed scale	Spatial	Test	Test battery	Estimated	Estimated	0.55	N/A
	.42	.45	93	Student sample	Mixed	Mixed scale	Verbal	Test	Test battery	Estimated	Estimated	0.55	N/A
	.58	.66	93	Student sample	Mixed	Mixed scale	Numerical	Test	Test battery	Estimated	Estimated	0.55	N/A
Ackerman et al. (2001)	.51	.56	320	Student sample	Mixed	Mixed scale	Verbal	Test	Test battery	0.87	Estimated	N/A	N/A
	.40	.42	320	Student sample	Mixed	Mixed scale	Numerical	Test	Test battery	0.84	Estimated	N/A	N/A
	.16	.16	320	Student sample	Mixed	Mixed scale	Other	Test	Test battery	0.82	Estimated	N/A	N/A
Ackerman et al. (2002)	-.05	-.05	228	General sample	Mixed	Mixed scale	Verbal	Test	Test battery	0.72	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	.47	.51	228	General sample	Mixed	Mixed scale	Numerical	Test	Test battery	0.81	Estimated	N/A	N/A
	.05	.05	228	General sample	Mixed	Mixed scale	Other	Test	Test battery	0.80	Estimated	N/A	N/A
	-.09	-.09	228	General sample	Mixed	Mixed scale	Other	Test	Test battery	0.88	Estimated	N/A	N/A
	.36	.38	228	General sample	Mixed	Mixed scale	Verbal	Test	Test battery	0.72	Estimated	N/A	N/A
	.18	.18	228	General sample	Mixed	Mixed scale	Numerical	Test	Test battery	0.81	Estimated	N/A	N/A
	-.09	-.09	228	General sample	Mixed	Mixed scale	Other	Test	Test battery	0.80	Estimated	N/A	N/A
	-.12	-.12	228	General sample	Mixed	Mixed scale	Other	Test	Test battery	0.88	Estimated	N/A	N/A
Ariel et al. (2018)	.48	.52	144	Student sample	Male	Absolute	Spatial	Estimate	Revised Purdue Spatial Visualization Test	0.78	0.89	N/A	N/A
	.56	.63	89	Student sample	Female	Absolute	Spatial	Estimate	Revised Purdue Spatial Visualization Test	0.78	0.89	N/A	N/A
	.22	.22	144	Student sample	Male	Absolute	Spatial	Estimate	Spatial Relations test	0.78	0.87	N/A	N/A
	.37	.39	89	Student sample	Female	Absolute	Spatial	Estimate	Spatial Relations test	0.78	0.87	N/A	N/A
Bacon et al. (2011)	-.12	-.12	34	General sample	Mixed	Absolute	Other	Unknown	WMS-R	0.78	N/A	N/A	N/A
Bailey & Bailey (1974)	.40	.42	37	General sample	Male	Reference group	General	Estimate	Otis Quick-Scoring Test	0.32	Estimated	N/A	N/A
	.39	.41	24	General sample	Female	Reference group	General	Estimate	Otis Quick-Scoring Test	0.32	Estimated	N/A	N/A
	.29	.30	40	General sample	Male	Reference group	General	Estimate	Otis Quick-Scoring Test	0.76	Estimated	N/A	N/A
	.04	.04	34	General sample	Female	Reference group	General	Estimate	Otis Quick-Scoring Test	0.76	Estimated	N/A	N/A
	.55	.62	44	General sample	Male	Reference group	General	Estimate	Otis Quick-Scoring Test	0.97	Estimated	N/A	N/A
	.85	1.26	42	General sample	Female	Reference group	General	Estimate	Otis Quick-Scoring Test	0.97	Estimated	N/A	N/A
	.27	.28	45	Student sample	Male	Reference group	General	Estimate	Otis Quick-Scoring Test	0.78	Estimated	N/A	N/A
	.44	.47	75	Student sample	Female	Reference group	General	Estimate	Otis Quick-Scoring Test	0.78	Estimated	N/A	N/A
Bailey & Lazar (1976)	.48	.52	20	Student sample	Female	Reference group	General	Estimate	Concept Mastery Test	Estimated	Estimated	N/A	N/A
	.53	.59	20	Student sample	Male	Reference group	General	Estimate	Concept Mastery Test	Estimated	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Bailey & Mettetal (1977)	.49	.54	20	General sample	Female	Relative	General	Estimate	Otis Quick-Scoring Test	Estimated	Estimated	N/A	N/A
	.35	.37	20	General sample	Male	Relative	General	Estimate	Otis Quick-Scoring Test	Estimated	Estimated	N/A	N/A
Bipp & Kleingeld (2012)	.27	.28	115	Student sample	Mixed	Reference group	General	Estimate	Cattell's Culture Fair Test 3	Estimated	0.63	0.52	N/A
Bipp et al. (2012): Study 1	.26	.27	89	Student sample	Mixed	Reference group	General	Test	Cattell's Culture Fair Test 3	Estimated	0.71	N/A	N/A
Bipp et al. (2012): Study 2	.18	.18	165	Student sample	Mixed	Absolute	General	Estimate	Cattell's Culture Fair Test 3	Estimated	0.62	N/A	N/A
Borella et al. (2014)	.13	.13	454	General sample	Mixed	Absolute	Spatial	Unknown	Mental Rotations Test	0.82	0.86	N/A	N/A
	.07	.07	454	General sample	Mixed	Absolute	Spatial	Unknown	Minnesota Paper Forms Board Test	0.82	0.83	N/A	N/A
	.03	.03	454	General sample	Mixed	Absolute	Spatial	Unknown	Short Embedded Figure Test	0.82	0.82	N/A	N/A
	-.11	-.11	454	General sample	Mixed	Absolute	Spatial	Unknown	Short Object Perspective-Taking Task	0.82	0.87	N/A	N/A
Borkenau & Liebler (1993)	.32	.33	100	Mixed sample	Mixed	Absolute	General	Estimate	Leistungsprüfsystem	0.76	Estimated	N/A	N/A
Bressan (2018)	.51	.56	201	Student sample	Mixed	Absolute	Numerical	Estimate	WAIS	Estimated	Estimated	N/A	N/A
Brim (1954)	.43	.46	86	Student sample	Mixed	Mixed scale	General	Test	American Council Psychological Examination	Estimated	Estimated	N/A	N/A
Chamorro-Premuzic & Furnham (2006)	.41	.44	184	Student sample	Mixed	Relative	General	Test	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A
	.41	.44	184	Student sample	Mixed	Relative	General	Estimate	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A
Chamorro-Premuzic et al. (2004)	.44	.47	83	General sample	Mixed	Relative	Other	Test	Alice Heim Test	Estimated	Estimated	N/A	N/A
	.49	.54	83	General sample	Mixed	Relative	General	Test	Baddeley Reasoning Test	Estimated	0.79*	N/A	N/A
	.40	.42	83	General sample	Mixed	Relative	Spatial	Test	S&M Spatial Ability Test	Estimated	Estimated	N/A	N/A
	.39	.41	83	General sample	Mixed	Relative	General	Test	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Chamorro-Premuzic et al. (2005)	.22	.22	182	Student sample	Mixed	Relative	General	Estimate	Standard Progressive Matrices	Estimated	Estimated	N/A	N/A
Cogan et al. (1915)	.70	.87	25	Student sample	Male	Reference group	General	Estimate	Test battery	Estimated	Estimated	N/A	N/A
	.53	.59	25	Student sample	Male	Reference group	General	Estimate	Test battery	Estimated	Estimated	N/A	N/A
de Keeresmaecker et al. (2017)	.15	.15	183	General sample	Mixed	Reference group	General	Unknown	Wilde Intelligenz Test	Estimated	Estimated	N/A	N/A
DeNisi & Shaw (1977)	.36	.38	114	Student sample	Mixed	Absolute	Other	Unknown	Bennett Mechanical Comprehension Test	Estimated	Estimated	N/A	N/A
	.21	.21	114	Student sample	Mixed	Absolute	Spatial	Unknown	Minnesota Paper Forms Board Test	Estimated	0.78*	N/A	N/A
	.26	.27	114	Student sample	Mixed	Absolute	General	Unknown	Otis	Estimated	Estimated	N/A	N/A
	.29	.30	114	Student sample	Mixed	Absolute	Verbal	Unknown	Personnel Test for Industry	Estimated	Estimated	N/A	N/A
	.41	.44	114	Student sample	Mixed	Absolute	Numerical	Unknown	Personnel Test for Industry	Estimated	Estimated	N/A	N/A
	.36	.38	114	Student sample	Mixed	Absolute	Verbal	Unknown	SAT	Estimated	Estimated	N/A	N/A
	.37	.39	114	Student sample	Mixed	Absolute	Numerical	Unknown	SAT	Estimated	Estimated	N/A	N/A
	.35	.37	114	Student sample	Mixed	Absolute	General	Unknown	SAT	Estimated	Estimated	N/A	N/A
Deuling et al. (2011)	.22	.22	187	Student sample	Mixed	Mixed scale	General	Unknown	Advanced Progressive Matrices	0.75	0.71	N/A	N/A
Diener et al. (2016)	.24	.24	286	General sample	Mixed	Absolute	Other	Unknown	Neuropsychological Assessment Battery	0.82	Estimated	N/A	N/A
Dislich et al. (2012): Study 1	.45	.48	74	Student sample	Mixed	Absolute	General	Estimate	Mehrfachwahl-Wortschatztest	0.86	0.83	N/A	N/A
Dislich et al. (2012): Study 2	.41	.44	51	Student sample	Mixed	Absolute	General	Estimate	Advanced Progressive Matrices	0.84	0.79	N/A	N/A
Dislich et al. (2012): Study 3	.55	.62	108	Student sample	Mixed	Absolute	General	Estimate	IST 2000-R	0.89	0.88	N/A	N/A
Dufner et al. (2012): Study 1	.23	.23	2,048	General sample	Mixed	Absolute	Verbal	Estimate	Mehrfachwahl-Wortschatztest	0.93	0.68	N/A	N/A
Dufner et al. (2012): Study 2	.24	.24	188	Student sample	Mixed	Absolute	General	Estimate	Advanced Progressive Matrices	0.62	0.71	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Fragkiadaki et al. (2016)	.31	.32	35	General sample	Mixed	Reference group	Spatial	Test	Brief Visuospatial Memory Test-Revised	Estimated	Estimated	N/A	N/A
	.21	.21	35	General sample	Mixed	Reference group	Other	Test	Driving Scenes Test-Neuropsychological Assessment Battery	Estimated	Estimated	N/A	N/A
	.36	.38	35	General sample	Mixed	Reference group	Verbal	Test	Hopkins Verbal Learning Test-Revised	Estimated	Estimated	N/A	N/A
	.11	.11	35	General sample	Mixed	Reference group	Spatial	Test	Judgment of Line Orientation Test	Estimated	Estimated	N/A	N/A
	.22	.22	35	General sample	Mixed	Reference group	Other	Test	Psychomotor Vigilance Test	Estimated	Estimated	N/A	N/A
	.12	.12	35	General sample	Mixed	Reference group	Other	Test	Trail Making Test-A	Estimated	Estimated	N/A	N/A
	.10	.10	35	General sample	Mixed	Reference group	Other	Test	Trail Making Test-B	Estimated	Estimated	N/A	N/A
Furnham & Chamorro-Premuzic (2004)	.30	.31	187	Student sample	Mixed	Relative	General	Estimate	Wonderlic Personnel Test	Estimated	0.88*	0.42	N/A
Furnham & Dissou (2007)	.53	.59	101	Student sample	Mixed	Relative	General	Test	Baddeley Reasoning Test	Estimated	0.79*	N/A	N/A
	.51	.56	101	Student sample	Mixed	Relative	General	Test	Baddeley Reasoning Test	Estimated	0.79*	N/A	N/A
Furnham & Fong (2000)	.19	.19	172	Student sample	Mixed	Relative	General	Test	Standard Progressive Matrices	Estimated	N/A	N/A	N/A
Furnham & Gover (2020) †	.30	.31	475	General sample	Mixed	Absolute	General	Unknown	10 item scale (Gover, 2019) †	Estimated	N/A	N/A	N/A
Furnham & Rawles (1999)	.27	.28	53	Student sample	Male	Relative	General	Estimate	S&M Spatial Ability Test	Estimated	N/A	N/A	N/A
	.09	.09	140	Student sample	Female	Relative	General	Estimate	S&M Spatial Ability Test	Estimated	N/A	N/A	N/A
Furnham (2005)	.41	.44	100	Student sample	Mixed	Relative	General	Test	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A
Furnham (2009)	.44	.47	187	Student sample	Mixed	Relative	Verbal	Unknown	Multiple Intelligences Test	0.83	Estimated	0.39	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	.51	.56	187	Student sample	Mixed	Relative	Numerical	Unknown	Multiple Intelligences Test	0.60	Estimated	0.39	N/A
	.37	.39	187	Student sample	Mixed	Relative	Spatial	Unknown	Multiple Intelligences Test	0.51	Estimated	0.39	N/A
	.56	.63	187	Student sample	Mixed	Relative	Other	Unknown	Multiple Intelligences Test	0.80	Estimated	0.39	N/A
	.38	.40	187	Student sample	Mixed	Relative	Other	Unknown	Multiple Intelligences Test	0.61	Estimated	0.39	N/A
	.35	.37	187	Student sample	Mixed	Relative	Other	Unknown	Multiple Intelligences Test	0.75	Estimated	0.39	N/A
	.31	.32	187	Student sample	Mixed	Relative	Other	Unknown	Multiple Intelligences Test	0.51	Estimated	0.39	N/A
	.18	.18	187	Student sample	Mixed	Relative	Other	Unknown	Multiple Intelligences Test	0.64	Estimated	0.39	N/A
Furnham (2018): Study 1	.25	.26	95	Student sample	Mixed	Reference group	General	Test	Advanced Progressive Matrices	Estimated	0.75*	0.61	N/A
	.46	.50	95	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.61	N/A
	.42	.45	95	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.61	N/A
Furnham (2018): Study 2	.36	.38	72	Student sample	Mixed	Reference group	General	Test	Advanced Progressive Matrices	Estimated	0.75*	0.64	N/A
	.61	.71	72	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.64	N/A
	.17	.17	72	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.64	N/A
Furnham (2018): Study 3	.70	.87	91	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.59	N/A
Furnham (2018): Study 4	.50	.55	118	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.58	N/A
	.31	.32	118	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.58	N/A
Furnham (2018): Study 5	.39	.41	106	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.61	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	.46	.50	106	Student sample	Mixed	Reference group	Numerical	Test	Mental Arithmetic	Estimated	Estimated	0.61	N/A
	.15	.15	106	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.61	N/A
Furnham (2018): Study 6	.51	.56	102	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.64	N/A
	.60	.69	102	Student sample	Mixed	Reference group	Numerical	Test	Mental Arithmetic	Estimated	Estimated	0.64	N/A
	.33	.34	102	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.64	N/A
Furnham (2018): Study 7	.62	.73	96	Student sample	Mixed	Reference group	General	Test	Baddeley Reasoning Test	Estimated	0.79*	0.55	N/A
	.59	.68	96	Student sample	Mixed	Reference group	Numerical	Test	Mental Arithmetic	Estimated	Estimated	0.55	N/A
	.18	.18	96	Student sample	Mixed	Reference group	General	Test	Wonderlic Personnel Test	Estimated	0.88*	0.55	N/A
Furnham et al. (2001)	.35	.37	100	General sample	Mixed	Relative	Numerical	Estimate	Numerical Aptitude Test	Estimated	Estimated	N/A	N/A
	.29	.30	100	General sample	Mixed	Relative	Spatial	Estimate	Spatial Aptitude Test	Estimated	Estimated	N/A	N/A
	.26	.27	100	General sample	Mixed	Relative	Verbal	Estimate	Verbal Aptitude Test	Estimated	Estimated	N/A	N/A
Furnham et al. (2005): Study 1	.19	.19	100	Student sample	Mixed	Relative	General	Estimate	Baddeley Reasoning Test	Estimated	0.79*	0.43	N/A
	.27	.28	100	Student sample	Mixed	Relative	General	Estimate	Wonderlic Personnel Test	Estimated	0.88*	0.43	N/A
Furnham et al. (2005): Study 2	.27	.28	130	Student sample	Mixed	Relative	General	Estimate	Baddeley Reasoning Test	Estimated	0.79*	0.39	N/A
	.25	.26	130	Student sample	Mixed	Relative	General	Estimate	Wonderlic Personnel Test	Estimated	0.88*	0.39	N/A
Furnham et al. (2006)	.32	.33	64	Student sample	Mixed	Relative	General	Test	Baddeley Reasoning Test	Estimated	0.79*	N/A	N/A
	.29	.30	64	Student sample	Mixed	Relative	General	Test	Standard Progressive Matrices	Estimated	Estimated	N/A	N/A
	.47	.51	64	Student sample	Mixed	Relative	General	Test	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A
Gabriel et al. (1994)	.27	.28	62	Student sample	Male	Reference group	General	Estimate	Shipley Institute of Living Scale	Estimated	Estimated	N/A	N/A
	.30	.31	84	Student sample	Female	Reference group	General	Estimate	Shipley Institute of Living Scale	Estimated	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Gerstenberg et al. (2013): Study 1	.36	.38	177	Student sample	Mixed	Mixed scale	General	Estimate	Mehrfachwahl-Wortschatztest	0.83	0.89	N/A	N/A
Gerstenberg et al. (2013): Study 2	.41	.44	165	Student sample	Mixed	Mixed scale	General	Estimate	Mehrfachwahl-Wortschatztest	0.87	0.87	N/A	N/A
Gerstenberg et al. (2013): Study 3	.36	.38	132	Student sample	Mixed	Mixed scale	General	Estimate	Mehrfachwahl-Wortschatztest	0.82	0.91	N/A	N/A
Gerstenberg et al. (2014)	.49	.54	84	Student sample	Mixed	Mixed scale	General	Estimate	Mehrfachwahl-Wortschatztest	0.89	0.81	N/A	N/A
Gignac & Zajenkowski (2019)	.26	.27	218	General sample	Female	Mixed scale	General	Estimate	Advanced Progressive Matrices	Estimated	0.86	N/A	N/A
	.33	.34	218	General sample	Male	Mixed scale	General	Estimate	Advanced Progressive Matrices	Estimated	0.88	N/A	N/A
Gignac & Zajenkowski (2020)	.28	.29	929	General sample	Mixed	Mixed scale	General	Estimate	Advanced Progressive Matrices	Estimated	0.75*	N/A	N/A
Gignac (2018)	.11	.11	253	Student sample	Mixed	Absolute	General	Estimate	WAIS	0.70	0.66	N/A	N/A
Herreen & Zajac (2017)	.16	.16	93	General sample	Mixed	Mixed scale	General	Test	Test battery	0.89	Estimated	N/A	N/A
	.51	.56	93	General sample	Mixed	Mixed scale	General	Test	Test battery	0.90	Estimated	N/A	N/A
Holling & Preckel (2005)	.46	.50	88	General sample	Mixed	Mixed scale	General	Unknown	IST 70	Estimated	Estimated	N/A	N/A
Huang & Maurer (2019)	.26	.27	13,690	General sample	Mixed	Absolute	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
Jacobs & Roodenburg (2014): Study 2	.28	.29	222	Student sample	Mixed	Relative	Other	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	Estimated	0.95	N/A	N/A
	.52	.58	222	Student sample	Mixed	Relative	Verbal	Estimate	Woodcock-Johnson Tests of cognitive abilities. third edition	Estimated	0.95	N/A	N/A
	.24	.24	222	Student sample	Mixed	Relative	Spatial	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	Estimated	0.81	N/A	N/A
	.08	.08	222	Student sample	Mixed	Mixed scale	Other	Estimate	Woodcock-Johnson Tests of cognitive	0.85	0.95	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
									abilities, third edition				
	.37	.39	222	Student sample	Mixed	Mixed scale	Verbal	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	0.88	0.95	N/A	N/A
	.25	.26	222	Student sample	Mixed	Mixed scale	Spatial	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	0.88	0.81	N/A	N/A
Jacobs et al. (2012)	.35	.37	189	Student sample	Mixed	Relative	Other	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	Estimated	0.95	N/A	N/A
	.56	.63	189	Student sample	Mixed	Relative	Verbal	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	Estimated	0.95	N/A	N/A
	.22	.22	189	Student sample	Mixed	Relative	Spatial	Estimate	Woodcock-Johnson Tests of cognitive abilities, third edition	Estimated	0.81	N/A	N/A
Kajonius (2014)	.64	.76	187	Student sample	Mixed	Absolute	General	Unknown	Illustrerad Vetenskap IQ test	Estimated	0.66	N/A	N/A
Kornilova & Novikova (2011)	.24	.24	96	Student sample	Mixed	Relative	General	Unknown	ROADS battery	Estimated	Estimated	N/A	N/A
Kornilova & Novikova (2012)	.24	.24	96	Student sample	Mixed	Reference group	General	Unknown	ROADS battery	Estimated	Estimated	N/A	N/A
Kornilova & Novikova (2013)	.25	.26	600	Student sample	Mixed	Relative	General	Unknown	Wonderlic Personnel Test	Estimated	0.88*	N/A	N/A
Kornilova et al. (2009)	.23	.23	184	Student sample	Mixed	Reference group	General	Estimate	IST 70	Estimated	0.67	N/A	N/A
Leising et al. (2016)	.26	.27	201	General sample	Mixed	Absolute	General	Estimate	Test battery	0.79	Estimated	N/A	N/A
Meneghetti et al. (2014)	.17	.17	450	General sample	Mixed	Absolute	Spatial	Unknown	Short Embedded Figure Test	0.80	0.90	N/A	N/A
	.21	.21	450	General sample	Mixed	Absolute	Spatial	Unknown	Short Mental Rotations Test	0.80	0.81	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	-.14	-.14	450	General sample	Mixed	Absolute	Spatial	Unknown	Short Object Perspective Taking Test	0.80	0.80	N/A	N/A
Mitolo et al. (2015)	.09	.09	90	Student sample	Mixed	Absolute	Spatial	Test	Embedded Figure Test	0.83	0.85	N/A	0.53
	.01	.01	90	Student sample	Mixed	Absolute	Spatial	Test	Minnesota Paper Forms Board Test	0.83	0.73	N/A	0.53
Nyström et al. (2019)	.14	.14	586	General sample	Mixed	Mixed scale	Other	Estimate	Test battery	0.58	0.66	N/A	N/A
Owczarek et al. (2012)	.28	.29	246	Student sample	Mixed	Absolute	Other	Estimate	Bourdon test	Estimated	Estimated	N/A	N/A
	.23	.23	246	Student sample	Mixed	Absolute	Verbal	Estimate	Rey 15-Item Memory Test	Estimated	Estimated	N/A	N/A
	.09	.09	246	Student sample	Mixed	Absolute	Other	Estimate	WAIS	Estimated	Estimated	N/A	N/A
	.24	.24	246	Student sample	Mixed	Absolute	Other	Estimate	WAIS	Estimated	Estimated	N/A	N/A
Paulhus et al. (1998)	.20	.20	274	Student sample	Mixed	Absolute	General	Estimate	Wonderlic Personnel Test	0.46	0.88	N/A	N/A
	.24	.24	274	Student sample	Mixed	Absolute	General	Estimate	Wonderlic Personnel Test	0.68	0.88	N/A	N/A
	.23	.23	241	Student sample	Mixed	Absolute	General	Estimate	Wonderlic Personnel Test	0.39	0.88	N/A	N/A
	.26	.27	241	Student sample	Mixed	Absolute	General	Estimate	Wonderlic Personnel Test	0.65	0.88	N/A	N/A
Proyer & Ruch (2009)	.19	.19	168	General sample	Mixed	Absolute	Verbal	Estimate	IST 2000-R	Estimated	0.84*	N/A	N/A
	.50	.55	168	General sample	Mixed	Absolute	Numerical	Estimate	IST 2000-R	Estimated	0.84*	N/A	N/A
	.16	.16	168	General sample	Mixed	Absolute	Spatial	Estimate	IST 2000-R	Estimated	0.84*	N/A	N/A
	.23	.23	168	General sample	Mixed	Absolute	Other	Estimate	IST 2000-R	Estimated	0.84*	N/A	N/A
	.38	.40	168	General sample	Mixed	Absolute	Other	Estimate	Standard Progressive Matrices	Estimated	Estimated	N/A	N/A
	.29	.30	168	General sample	Mixed	Absolute	Verbal	Estimate	Wortschatztest	Estimated	Estimated	N/A	N/A
Rammstedt & Rammsayer (2002a)	.17	.17	150	Student sample	Mixed	Mixed scale	Other	Unknown	Berliner Intelligenzstrukturtest	0.82	Estimated	N/A	N/A
	.40	.42	150	Student sample	Mixed	Mixed scale	Verbal	Unknown	Leistungsprüfsystem	0.88	Estimated	N/A	N/A
	.14	.14	150	Student sample	Mixed	Mixed scale	Verbal	Unknown	Leistungsprüfsystem	0.81	Estimated	N/A	N/A
	.27	.28	150	Student sample	Mixed	Mixed scale	Spatial	Unknown	Leistungsprüfsystem	0.80	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	.14	.14	150	Student sample	Mixed	Mixed scale	Other	Unknown	Leistungsprüfsystem	0.78	Estimated	N/A	N/A
	.25	.26	150	Student sample	Mixed	Mixed scale	Other	Unknown	Leistungsprüfsystem	0.85	Estimated	N/A	N/A
	.29	.30	150	Student sample	Mixed	Mixed scale	Numerical	Unknown	Wilde Intelligenz Test	0.81	Estimated	N/A	N/A
Rammstedt & Rammsayer (2002b)	.39	.41	228	Student sample	Mixed	Mixed scale	Verbal	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	.07	.07	228	Student sample	Mixed	Mixed scale	Verbal	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	.35	.37	228	Student sample	Mixed	Mixed scale	Numerical	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	.27	.28	228	Student sample	Mixed	Mixed scale	Spatial	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	.15	.15	228	Student sample	Mixed	Mixed scale	Other	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	-.04	-.04	228	Student sample	Mixed	Mixed scale	Other	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
	.22	.22	228	Student sample	Mixed	Mixed scale	Other	Estimate	Leistungsprüfsystem	Estimated	Estimated	N/A	N/A
Reilly & Mulhern (1995)	.42	.45	45	Student sample	Male	Relative	General	Test	WAIS	Estimated	Estimated	N/A	N/A
	.15	.15	80	Student sample	Female	Relative	General	Test	WAIS	Estimated	Estimated	N/A	N/A
Rothlind et al. (2017)**	.55	.62	199	General sample	Mixed	Reference group	Spatial	Test	Brief Visuospatial Memory Test-Revised	Estimated	Estimated	N/A	N/A
	.43	.46	199	General sample	Mixed	Reference group	Verbal	Test	California Verbal Learning Test	Estimated	Estimated	N/A	N/A
	.40	.42	199	General sample	Mixed	Reference group	Verbal	Test	Controlled Oral Word Association Test	Estimated	Estimated	N/A	N/A
	.56	.63	199	General sample	Mixed	Reference group	Other	Test	Rey-Osterrieth Complex Figure	Estimated	Estimated	N/A	N/A
	.57	.65	199	General sample	Mixed	Reference group	Other	Test	Short Category Test	Estimated	Estimated	N/A	N/A
	.40	.42	199	General sample	Mixed	Reference group	Other	Test	Symbol Digit Modalities Test Oral Version	Estimated	Estimated	N/A	N/A
	.28	.29	199	General sample	Mixed	Reference group	Other	Test	Symbol Digit Modalities Test Written Version	Estimated	Estimated	N/A	N/A
	.37	.39	199	General sample	Mixed	Reference group	Other	Test	Trail Making Test-B	Estimated	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Schwert et al. (2018)	.04	.04	88	General sample	Mixed	Absolute	Other	Unknown	Figural Memory	0.87	0.71	N/A	N/A
	.05	.05	88	General sample	Mixed	Absolute	Other	Unknown	Go/no-go	0.87	0.71	N/A	N/A
	.14	.14	88	General sample	Mixed	Absolute	Other	Unknown	Tower of London†	0.87	0.57	N/A	N/A
	-.06	-.06	88	General sample	Mixed	Absolute	Other	Unknown	Trail Making Test-B	0.87	0.81	N/A	N/A
	.03	.03	88	General sample	Mixed	Absolute	Other	Unknown	WAFA	0.87	0.86	N/A	N/A
	.15	.15	88	General sample	Mixed	Absolute	Other	Unknown	WAFG	0.87	0.89	N/A	N/A
Soh & Jacobs (2013)	.44	.47	165	General sample	Mixed	Relative	General	Estimate	Test battery	Estimated	0.70	N/A	N/A
	.40	.42	165	General sample	Mixed	Relative	Verbal	Estimate	Test battery	Estimated	0.70	N/A	N/A
Sokolowski et al. (2019)	.37	.39	175	Student sample	Mixed	Absolute	Numerical	Unknown	Kit of Factor-Referenced Cognitive Tests	Estimated	Estimated	N/A	N/A
	.28	.29	175	Student sample	Mixed	Absolute	Spatial	Unknown	Mental Rotation Task	0.83	Estimated	N/A	N/A
Storek & Furnham (2013a)	.12	.12	82	Student sample	Female	Relative	General	Estimate	Baddeley Reasoning Test	0.76	0.79	N/A	N/A
	.15	.15	39	Student sample	Male	Relative	General	Estimate	Baddeley Reasoning Test	0.76	0.79	N/A	N/A
	.44	.47	82	Student sample	Female	Relative	General	Estimate	Wonderlic Personnel Test	0.76	0.86	N/A	N/A
	.56	.63	39	Student sample	Male	Relative	General	Estimate	Wonderlic Personnel Test	0.76	0.86	N/A	N/A
Storek & Furnham (2013b)	.19	.19	79	Student sample	Female	Relative	General	Unknown	Wonderlic Personnel Test	0.71	0.88*	N/A	N/A
	.43	.46	23	Student sample	Male	Relative	General	Unknown	Wonderlic Personnel Test	0.71	0.88*	N/A	N/A
Torres et al. (2016)	.44	.47	38	Mixed sample	Mixed	Reference group	Other	Test	California Verbal Learning Test 2nd Edition	0.55	0.62	N/A	N/A
	.36	.38	38	Mixed sample	Mixed	Reference group	General	Estimate	Test battery	0.79	0.74	N/A	N/A
Villado et al. (2016)***	.07	.07	288	Student sample	Mixed	Absolute	General	Unknown	Advanced Progressive Matrices	0.75	0.75	N/A	N/A
	.19	.19	288	Student sample	Mixed	Absolute	General	Unknown	Wonderlic Personnel Test	0.75	0.89	N/A	N/A
Visser et al. (2008)	.31	.32	200	Student sample	Mixed	Reference group	Verbal	Unknown	Test battery	Estimated	Estimated	N/A	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
	.05	.05	200	Student sample	Mixed	Reference group	Spatial	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	.38	.40	200	Student sample	Mixed	Reference group	Numerical	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	.16	.16	200	Student sample	Mixed	Reference group	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	-.10	-.10	200	Student sample	Mixed	Reference group	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	-.01	-.01	200	Student sample	Mixed	Reference group	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	.20	.20	200	Student sample	Mixed	Reference group	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	.25	.26	200	Student sample	Mixed	Reference group	Other	Unknown	Test battery	Estimated	Estimated	N/A	N/A
	.20	.20	200	Student sample	Mixed	Reference group	General	Unknown	Test battery	Estimated	Estimated	N/A	N/A
von Stumm (2014)	.29	.30	176	General sample	Mixed	Relative	General	Test	Advanced Progressive Matrices	Estimated	0.66	N/A	N/A
	.49	.54	186	General sample	Mixed	Absolute	Other	Test	Advanced Progressive Matrices	Estimated	0.66	N/A	N/A
Webb (1955)	.21	.21	95	General sample	Mixed	Reference group	General	Estimate	Otis Quick-Scoring Test	Estimated	Estimated	N/A	N/A
Wolff & Wasden (1969)	-.25	-.26	13	General sample	Female	Absolute	General	Estimate	WAIS	Estimated	Estimated	N/A	N/A
	.11	.11	13	General sample	Female	Absolute	General	Test	WAIS	Estimated	Estimated	N/A	N/A
Zajenkowski & Czarna (2015): Study 1	.23	.23	205	Student sample	Mixed	Mixed scale	General	Unknown	Advanced Progressive Matrices	Estimated	0.84	N/A	N/A
Zajenkowski & Gignac (2018): Study 1	.30	.31	303	Student sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	0.78	0.61	N/A
Zajenkowski & Gignac (2018): Study 2	.40	.42	225	Student sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	0.79	0.42	N/A
Zajenkowski et al. (2016): Study 1	.42	.45	160	Student sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	Estimated	N/A	N/A
Zajenkowski et al. (2019a): Study 1	.39	.41	232	Student sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	0.9	0.57	N/A
Zajenkowski et al. (2019a): Study 2	.18	.18	241	Student sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	0.71	N/A	N/A
Zajenkowski et al. (2019b): Study 1	.35	.37	504	General sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	Estimated	0.57	N/A

Authors	*r*	z	N	Sample composition	Sex	Method of self-assessment	Ability type	Order of assess-ment	Test	Self-assess-ment Reliability	Test reliability	*M* Neuro-ticism	*M* Self-efficacy
Zajenkowski et al. (2019b): Study 2	.18	.18	232	General sample	Mixed	Mixed scale	General	Unknown	Test battery	Estimated	Estimated	N/A	N/A
Zajenkowski et al. (2020)	.13	.13	311	General sample	Mixed	Mixed scale	General	Estimate	Advanced Progressive Matrices	Estimated	0.60	N/A	N/A

Note. Self-assessment methods: Absolute = absolute scale, Relative = relative scale, Reference group = relative scale including explicit mention of a reference group. Ability type: General = general cognitive ability, Numerical = numerical ability, Spatial = spatial ability, Verbal = verbal ability, Other = other ability. Order of assessment: Estimate = self-estimate first, Test = psychometric test first. Test: † = it could not be determined with absolute certainty that test norms were available. Self-assessment Reliability: Estimated = was estimated using psychmeta (Dahlke & Wiernik, 2019). Test Reliability: Estimated = was estimated using psychmeta (Dahlke & Wiernik, 2019), * = estimated by averaging the given test reliabilities in the dataset. *M* Neuroticism = calculated as described in chapter Method / Coding. *M* Self-Efficacy: calculated as described in chapter Method / Coding.
** Correlations do not stem from healthy participants. However, the healthy and the reported sample were comparable.
*** Reported correlations were z-transformed and averaged, as multiple correlations were reported.
N/A = not applicable.

Appendix B: Abstract (English)

Self-assessed intelligence (= SAI) plays an important role in high-stakes situations, for example, when making career decisions. Research shows that professional interests are primarily a function of self-assessed abilities rather than “true” abilities. However, previous accounts revealed that self-assessed intelligence and psychometric intelligence only correlate moderately. However, ever-increasing publication numbers of studies investigating this topic and the development of novel and more refined research synthesis methods have rendered the available meta-analytical evidence outdated. Moreover, as recent findings suggest potential bias due to the virtual ubiquitous declining effects in empirical research, an update seems necessary. Consequently, a total of 242 effect sizes from 98 studies (N = 54,566) were analyzed by applying a three-level meta-analytic model. Applying the Hedges and Olkin approach yielded a correlation of $r = .31$. A Hunter and Schmidt–typed approach was applied to investigate the extent the result was influenced by measurement unreliability, which yielded a correlation of $r = .39$. Associations with global intelligence self-assessments differed significantly from specific ones (i.e., correlations were highest for numerical ability followed by global, spatial, and other less-known cognitive abilities, e.g., naturalistic intelligence), whilst the type of self-estimation method did not significantly affect results. A total of eight methods were applied to test for potential dissemination bias. The observed summary effect strength must be considered to be somewhat inflated, as there was evidence for some bias. In all, we show evidence for a moderate correlation between SAI and psychometric intelligence, which generalizes over assessment methods, order of assessment, sex of participants, year of publication but is differentiated according to ability type.

Keywords: Intelligence, Meta-analysis, Effect inflation.

Appendix C: Abstract (German)

Selbsteingeschätzte Intelligenz (= SAI) spielt eine wichtige Rolle in vielen Situationen wie z. B. bei der Berufswahl: Berufliche Interessen scheinen in erster Linie von selbsteingeschätzten Fähigkeiten abzuhängen und nicht von den "wahren" Fähigkeiten einer Person. Vorangegangene Studien zeigten jedoch, dass SAI und psychometrische Intelligenz nur mäßig korrelieren. Durch die ständig steigende Anzahl an Veröffentlichungen von Studien, die dieses Thema untersuchen, und die Entwicklung neuer und verfeinerter Methoden zur Forschungssynthese sind die verfügbaren meta-analytischen Daten jedoch inzwischen veraltet. Da neuere Erkenntnisse zudem auf potenzielle Verzerrungen aufgrund des „Decline Effects“ in der empirischen Forschung hindeuten, erscheint eine Aktualisierung vorangegangener meta-analytischer Untersuchungen notwendig. In der vorliegenden Studie wurden insgesamt 242 Effektgrößen aus 98 Studien (N = 54,566) unter Anwendung eines dreistufigen metaanalytischen Modells analysiert. Unter der Anwendung des Hedges und Olkin Ansatzes korrelierten die beiden Werte zu r = .31. Zusätzlich wurde überprüft, inwieweit das Ergebnis durch die Reliabilität der Messinstrumente beeinflusst wird. Dafür wurde die Korrelation anhand des Ansatzes nach Hunter und Schmidt erneut geschätzt: r = .39. Die Zusammenhänge unterschieden sich weiters abhängig davon, ob die allgemeine Intelligenz oder spezifische kognitive Fähigkeiten eingeschätzt wurden: Die Korrelation war am höchsten für numerische Fähigkeiten, gefolgt von allgemeinen, räumlichen und allgemein weniger bekannten kognitiven Fähigkeiten (z.B. naturalistische Intelligenz). Weiters beeinflusste die Methode der Selbsteinschätzung das Ergebnis nicht signifikant. Insgesamt wurden acht Methoden verwendet, um auf einen möglichen „Dissemination bias“ zu testen. Es gab Hinweise auf eine gewisse Verzerrung, was darauf hindeutet, dass die beobachteten Gesamteffekte als etwas überschätzt angesehen werden müssen. Die Arbeit gibt Hinweise auf einen moderaten Zusammenhang zwischen SAI und IQ, der hinsichtlich der Methoden der Selbsteinschätzung, der Reihenfolge der Testung, des Geschlechts der Teilnehmer*innen, des Publikationsjahrs und generalisiert jedoch nach Fähigkeitstyp differenziert werden kann.

Schlagwörter: Intelligenz, Meta-Analyse, Effektüberschätzung.