Intelligence Really Does Predict Job Performance: A Long-Needed Reply to Richardson and Norgate

Peter Zimmer[1] and Emil Ole William Kirkegaard[2]

[1] Independent Researcher, USA
[2] Ulster Institute of Social Research, London, UK

Abstract

One commonly studied aspect of the importance of IQ is its validity in predicting job performance. Previous research on this subject has yielded impressive results, regularly finding operational validities for general mental ability exceeding 0.50. In 2015, Ken Richardson and Sarah Norgate criticized the research on the relationship between IQ and job performance, reducing it to virtually nothing. Their assessment of this topic has enjoyed little criticism since its publication despite the crux of their arguments being undermined by readily available empirical evidence and thirty years of replication of the contrary. This article replies to their main criticisms, including the construct validity of IQ tests and supervisory ratings, the validity of the Hunter-Schmidt meta-analytic methods, and possible psychological confounders.

***Keywords**:* industrial-organizational psychology, job performance, intelligence, IQ, g-factor, cognitive ability, general mental ability, meta-analysis, predictive validity

Intelligence Really Does Predict Job Performance: A Long-Needed Reply to Richardson and Norgate

Introduction

Richardson & Norgate (2015) presented a detailed critique of the dense literature on the relationship between job performance and IQ test scores. Their review primarily targeted the meta-analytic procedures introduced and popularized by John Hunter and Frank Schmidt, as well as the construct and predictive validity of IQ tests. Few papers replying to Richardson and Norgate have been published. To our knowledge, only two commentaries (Sternberg, 2015; Kaufman & Kaufman, 2015) have been published in reference to Richardson and Norgate (2015), both of which view the paper more positively than negatively. The latter commentary is a father-and-son response wherein each writer took a different position on the article. The more critical response, the father's, mainly attacked Richardson and Norgate's views on the construct validity of IQ tests, conceding to the claim that meta-analytic procedures drastically overestimate the true association between IQ and job performance. The generally positive outlook on the paper suggests that it is accurate throughout; we argue this is not the case.

Richardson and Norgate's main arguments can be summarized as follows: (1) IQ tests are indirect measures of poorly-defined concepts, causing them to lack construct validity, (2) the supposed predictive validity of IQ tests is a poor defense as the correlations argued to support predictive validity are built into the tests, (3) supervisory ratings, the primary measurement of job performance, are unlikely to be measures of actual job performance but are, rather, a product of biases in supervisor judgement, (4) meta-analytic results are riddled with uncertainty and the procedures meant to reduce error in meta-analyses are error-prone in themselves, (5) the large report produced by the National Academy of Sciences (NAS; Hartigan & Wigdor, 1989) showed the relationship between job complexity, job performance, and IQ to be much smaller than previously estimated by Hunter and colleagues, and (6) non-cognitive causes drive the relationship between general mental ability and job performance.

While Richardson and Norgate did not attack any specific paper, much of their article was in response to an article by Hunter & Hunter (1984). This article — often regarded as seminal in test validity research — led the way for decades of additional research on job performance and IQ and had 3423 citations on Google Scholar as of writing[3]. Most meta-analyses to date have found an average operational validity for general mental ability of over 0.50, with lower operational validity typically observed in the least complex jobs along with greater operational validity in more complex jobs. Schmidt (2002) summarized such findings:

> On the basis of meta-analysis of over 400 studies, Hunter and Hunter (1984) estimated the validity of GCA for supervisor ratings of overall job performance to be .57 for high-complexity jobs (about 17% of U.S. jobs), .51 for medium-complexity jobs (63% of jobs), and .38 for low-complexity jobs (20% of jobs). These findings are consistent with those from other sources (Hunter & Schmidt, 1996; validities are larger against objective job sample measures of job performance, Hunter, 1983a). For performance in job training programs, a number of large databases exist, many based on military training programs. Hunter (1986) reviewed military databases totaling over 82,000 trainees and found an average validity of .63 for GCA. This figure is similar to those for training performance reported in various studies by Ree and his associates (e.g., Ree and Earles, 1991), by Thorndike (1986), by Jensen (1986), and by Hunter and Hunter (1984) (p. 190).

[3] https://scholar.google.com/scholar?cites=13884901685679622391&as_sdt=2005&sciodt=0,5&hl=en

Job performance aside, IQ has also been found to predict changes in occupational status, (Schmitt, Gooding, Noe, &Kirsch, 1984), self-employment (de Wit & van Winden, 1989), and training success (Hülsheger, Maier, & Stumpp, 2007). The relationship between job performance and IQ has played a substantial role in public policy debate and psychology, and the critiques proposed by Richardson and Norgate seemed to turn what was previously well-established on its head. However, fallacious arguments, errant assumptions about meta-analysis, and pure speculation drove a significant portion of their paper. If strong conclusions are to be made from it, a number of relevant replies should be considered first.

Construct Validity of IQ Tests

As has been well established, IQ tests indicate a powerful and sociologically evident construct, general mental ability, *g*, which is often used to support their validity as a mental test (eg. Carroll, 1993). The authors may argue that the fact the original test creators were aiming to create similar tests forced a positive manifold and hence the tests are highly *g*-loaded. As it happens, they did not go to any length to make this argument within their article. Regardless, tests created without the intention of being *g*-loaded and even tests developed to discredit *g* theory have nonetheless ended up with a high *g*-loading (Dalliard, 2013).

Famously illustrating this phenomenon, Thurstone (1938) attempted to develop a test measuring seven independent facets of mental ability. Shortly after publishing his work on the test, Eysenck (1939) found these seven facets of mental ability all actually loaded onto *g*. Later, the British Ability Scales were developed in order to measure multiple independent mental abilities. However, when the data for the British Ability Scales was analyzed by Elliott (1986), the scales still gave rise to a higher order *g* factor. Finally, one could also point to the Cognitive Ability Scales (CAS) battery. This was based on the Planning, Attention-Arousal, Simultaneous and Successive (PASS) theory of intelligence, which was intentionally meant to combat *g* theory (Naglieri, 2001). Despite this, Keith, Kranzler, and Flanagan (2001) found that the CAS battery is not a valid measure of PASS, and it is actually a measure of *g*.

Recent studies have shown that performance on video games correlates very strongly with intelligence (latent correlations from .60 to .93), especially when prior practice on the games is relatively uniform and many games are used to extract a general gaming ability (Quiroga et al., 2015; Quiroga, Diaz, Román, Privado, & Colom, 2019). Similarly, a meta-analysis by Burgoyne et al. (2016) found that chess skill is correlated with various measures of *g*. Indeed, at the national level, smarter nations perform better across a wide range of mental games, even when adjusting for variation in internet prevalence, and adding regional dummies ($r = .79$, Kirkegaard, 2019). Overall, the evidence suggests the intercorrelation of IQ tests and their loading onto a higher-order *g* factor is not an artifact of test construction.

How does one determine the validity of a test? In the case of the *g* factor and IQ tests, the traditional method has been to use factor analysis. This method is prone to error (Cooper, 2019), but can be useful as a foundation for the validity of a given construct. Lubinski & Humphreys (1997) argued “a measure’s meaning (technically, its construct validity) is found in its network of causes and correlates, not in the unique aspects of its item content or label.” A similar definition was given by Nunnally (1978) who argued a measure can be considered construct valid if it either strongly correlates with other measures of said construct or if the predictive validity of the measure is similar to the predictive validity of other measures of the same construct. Campbell & Fiske (1959), in their landmark paper on construct validity, argued the validity of a construct is assessed through its correlations with construct-relevant variables. Even further back, the people who originally formulated the concept of construct validity, Cronbach & Meehl (1955), argued there is no single method to construct validation. They argued the correlations of two tests presumed to measure the same construct could be semi-sufficient evidence for construct validity. Thus, given the prior evidence that IQ tests are all correlated, there is certainly some construct validity to IQ tests, even if it is limited.

Richardson & Norgate note that there is no accepted theory of intelligence, and hence IQ tests are not built like many other forms of measurement, such as a breathalyzer (p. 154). They argue that IQ tests "rely on correlations of scores with those from other IQ or achievement tests as evidence of validity" and therefore cannot be construct valid. It should be noted that breathalyzers are measurements of internal, biological criteria, whereas virtually all psychological tests measure traits indirectly as of now. Furthermore, while the intercorrelation of tests is not perfect evidence of construct validity, it is surely useful as a foundation. There are more fitting methods to construct validation one could use, though Richardson & Norgate predictably disagree with the utility of these.

As a primary example, reaction times have a long-standing relationship with IQ tests (Der & Deary, 2017). But Richardson & Norgate dismiss this primarily on the ground that the correlations are small and may be confounded by psychological variables. These may be true, but the correlations between IQ tests and various measures of reaction times correlate in a way that would be predicted if IQ tests measured mental speed. For example, as Der & Deary (2017) show, the correlation between IQ tests and choice reaction times is stronger than that between IQ tests and simple reaction times. Furthermore, there is a Jensen effect on the relationship between IQ and reaction times, meaning the subtests which load the highest on *g* are more strongly correlated with reaction times (Jensen, 1998). Since this is true, we may predict the relationship is at least partially due to differences in mental ability (further discussion on the causes of this relationship are given by Jensen [1993] and Jensen [1998]).

They also argue there is no accepted internal theory of intelligence which allows us to properly interpret IQ tests. They cite Haier et al. (2009) to this respect, but it is worth noting much of Haier's (2016) book is dedicated to proving the PFIT theory of intelligence. Understanding the neurological basis of intelligence has proven difficult, though some theories have survived more than others. For example, while substantial criticism has been given to the idea that IQ tests measure mental efficiency, enough revision has been made to the theory to show it generally has some validity (see Haier, 2016). Furthermore, researchers have found Jensen effects in the relation of IQ tests to various biological variables (Jensen, 1998; Gignac, Vernon, Wickett, 2003), further providing evidence of construct validity. Finally, since Richardson & Norgate deny the validity of IQ tests to measure mental ability, they argue we must address the predictive validity of IQ tests.

Predictive Validity of IQ Tests

Richardson & Norgate's next arguments are against the predictive validity of IQ testing. The authors argue the cause of the correlation between educational achievement and IQ is an artifact of test construction, rather than a function of intelligence influencing educational outcomes. Richardson & Norgate say,

> Since the first test designers such as Binet, Terman, and others, test items have been devised, either with an eye on the kinds of knowledge and reasoning taught to, and required from, children in schools, or from an attempt to match an impression of the cognitive processes required in schools. This matching is an intuitively-, rather than a theoretically-guided, process, even with nonverbal items such as those in the Raven's Matrices. (p. 154)

If these processes are required in schools and such processes are truly mental ability, then Richardson & Norgate's argument is entirely circular. Their claim is essentially the same as anyone's who defends IQ testing: IQ and educational achievement are correlated because mental ability is required for school.

Richardson & Norgate argue the relationship is partially due to the fact that such mental processes required for intelligence tests are taught in modern curriculum. This is not actually true for many early childhood tests such as Piagetian tests. Yet factor analysis of such tests show that they measure the same thing as ordinary intelligence tests (Lasker 2022). In discussing age-related changes in the relationship between intelligence test scores and scholastic tests, Richardson & Norgate argue that the increase seen with age fits with their model. That may be true in the abstract, but if one checks the citation, it goes to Sternberg et al (2001) and then to McGrew and Knopik (1993) for the actual results. However, interpretation of this study is difficult because the authors adopted a correlated factors model, i.e., no general factor, so it is difficult to say whether the intelligence score increased its correlation with their achievement tests (two tests of writing) with age. The multiple R (square root of the more common r^2)

increased with age, but is this because g becomes more correlated with writing ability, or because non-g group factors increase their importance at later ages. Analysis of their reported regression coefficients (in their Table 3) suggests that it isn't g's correlation to the writing tests that's increasing: it is mainly the crystalized ability (gc) that increases its relation to the writing test over time. The other ability factors, such as long term memory (glr), fluid ability (gf) show no increases. One simple interpretation here is then that as children accumulate school knowledge with age, this leads to an increasing overlap between the gc factor and the writing tests. As such, there is no need to invoke the interpretation proposed by Richardson & Norgate. It would be informative to reanalyze this study using modern structural equation methods.

Going further, longitudinal studies by Watkins, Lei, & Canivez (2007) and Watkins & Styck (2017) both found that a model where *g* causes educational achievement was best fit compared to vice versa. Similarly, while education has been shown to raise IQ scores (Ritchie & Tucker-Drob, 2018), Ritchie, Bates, and Deary (2015) found that the effect of years of education on IQ was only on specific skills rather than on *g* (i.e., schooling improved some broad abilities, but not general intelligence). These studies imply that the relationship between IQ and educational achievement is driven by individual differences in general mental ability. Another important criticism of Richardson & Norgate's theory is that *g* and education have discriminant validity. Lubinski & Humphreys (1997) & Lubinski (2009), for example, showed that *g* is a better predictor of health outcomes than education (also see Gottfredson [2004] for a critical review of this topic). Gensowski, Heckman, & Savelyev (2011) found IQ predicted income beyond its relationship with education. If intelligence was just a proxy for social class or educational background, it is a mystery why employers would keep paying smarter people more money if they were not also more productive Indeed, numerous studies find correlations between income and intelligence, including controlling for parental social status (Marks 2022). Also in the economics literature, Altonji & Pierret (2001) found that education predicted income well at the beginning a person's career. But as people's careers progressed, intelligence became a better predictor. This finding is in line with a model where employers use education a proxy for intelligence and other traits, but over time, they learn an employee's characteristics, so they don't need the education proxy as much. Much more detail about signaling vs. human capital models of the value of education can be found in Bryan Caplan's book length treatment (Caplan, 2018).

The authors also support their belief by pointing to the fact that the relationship between IQ and educational achievement increases with age, but they give no reason why this should support the point that the correlation between IQ and educational achievement is built into the tests. On the contrary, there are two pro-*g* interpretations of this finding. The first would be an increasing role of cognitive ability at higher levels. As students continue through education, it becomes more difficult and cognitive ability plays a greater role in determining educational achievement. The second is the accumulative effect of *g* over time. Scholastic tests are based on the amount individuals learn in school over a period of years. This is essentially a measure of the average learning rate over increasing spans of time, therefore increasing the correlation between *g* and educational achievement. This interpretation would make sense given the fact that highly crystallized tests like those which measure verbal ability are often more g-loaded than fluid intelligence tests (Carroll, 1993).

Richardson & Norgate provide the argument that parental drive correlates with IQ. But, the effects of parental drive seem to disappear by adulthood whether looked at through a genetic lens (Bouchard, 2013) or through an environmental lens (Dickens & Flynn, 2001). This is because the effect of shared environment on phenotypic IQ almost entirely disappears by adulthood. Any correlation of parental drive on IQ by adulthood could be easily explained as a gene-environment correlation: higher IQ parents, who share half their genes with their children, likely motivate their kids more in addition to putting them in better schools. Putting all of this together, Richardson & Norgate make no convincing argument that the correlation between educational achievement and IQ is an artifact of test construction.

Richardson & Norgate's argument to deconstruct the relationship between IQ and occupational level and income are entirely contingent on their argument about IQ and educational achievement being accurate. Since, as we have shown, it is not, they have little ground to stand on, and their claims about predictive validity for occupational and income are not substantiated either.

Supervisory Ratings

The authors move on to challenge what supervisory ratings mean and what value they may provide. While they are correct that most studies on IQ and job performance are done using supervisory ratings, Richardson & Norgate seemingly ignore the related finding that IQ scores correlate with job performance as measured by objective work tests too. In fact, the latter correlation seems to be stronger, whereas by Richardson & Norgate's model, this is expected to be weaker or null. This has been shown in analyses by McHenry, Hough, Toquam, Hanson, &

Ashworth (1990) and Ree, Earles, & Teachout (1994). The former found validities of 0.63-0.65 for predicting on-the-job military performance with general mental ability. The latter found a validity of 0.45 in predicting on-the-job military performance. In a meta-analysis by Nathan & Alexander (1988), many forms of criteria were used, including ratings, rankings, work samples, and production quantities. General mental ability maintained high validity in predicting all of these. Hunter (1986) also used large military databases and found an operational validity of 0.63 for IQ in predicting job performance, using both supervisory ratings and objective job performance. Other reviews have used military service deaths rather than any sort of rating system and found even this is related to intelligence (Laurence & Ramsberger, 1991).

While supervisor ratings have relatively low correlations with results of work sample tests, both appear to be valid measures of job performance. The relatively small correlation between work sample tests and supervisory ratings is due to "notable criterion deficiency inherent in objective records and problems of unreliability [in objective job performance measures]" (Ones, Viswesvaran, & Schmidt, 2008). Still, other studies find statistically significant, positive correlations between the two variables (cf. Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995; Heneman, 1986; Viswesvaran, 2002). Assessments of job performance at the supervisory level are correlated with job performance assessments at the peer level (Harris and Schaubroeck, 2006; Viswesvaran, Schmidt. & Ones, 2002). However, supervisory ratings have greater reliability than peer ratings (Viswesvaran, Ones & Schmidt. 1996). Since these are true, Richardson & Norgate's argument that halo effects may substantially bias the correlation between IQ and job performance is unsound.

Addressing other biases, taller people do seem to be objectively better at their jobs and are good at advancing, likely partially due to a greater sense of self-esteem (see Rosenberg, 2009), and the evidence shows height correlates with IQ as well (Pearce, Deary, Young, and Parker, 2005). The effect of height on wages (which is correlative of supervisory rating) is non-linear and exponential, meaning the effect is primarily among the tallest people (Kim & Han, 2017). The most flawed bias argument relates to the effect of race in supervisory ratings. Dejung & Kaplan (1962) found black supervisors rated black employees higher than white employees, whereas white employees did not rate white employees better than black employees. A meta-analysis by McKay & McDaniel (2006) looked at studies on both objective and subjective ratings of job performance in blacks and whites. Objective measures closer favored that of the white supervisory ratings. Interestingly, a study by Roth, Huffcutt, Bobko (2003) found that objective measures of job performance actually predict a larger racial difference in job performance than do subjective measures. This is the exact opposite of what Richardson & Norgate would predict if there were meaningful bias in supervisory ratings. As a consequence, this would also mean that the effect of racial bias in supervisory ratings creates an *underestimation* of the job performance and IQ correlation. Bobko & Roth (2013) found differences in job performance between blacks and whites are mediated by job knowledge and are largest in the most complex jobs, largely contradicting the discrimination hypothesis. To cement things further, a study analyzed job predictors too, and found that predictors that are more correlated with intelligence are the ones with larger black-white gaps on them (Dahlke & Sackett, 2017).

Finally, it is worth mentioning a meta-analysis done by Viswesvaran, Schmidt, & Ones (2005). The authors analyzed research from over 90 years and found, after controlling for three different forms of measurement error and halo error, there remained a general factor of job performance ratings. Similar results are found in a military database by Vance, MacCallum, Coovert, & Hedge (1988). Overall, supervisory ratings remain a useful measure of job performance, and regardless, more objective measures of job performance have even greater correlations with IQ. The various other findings from related research support the important causal role of intelligence in explaining job performance.

Meta-analytic Procedures

Richardson & Norgate cast a large amount of doubt on the reliability of meta-analysis to create non-biased results. Their first argument is that meta-analysis lumps large amounts of low-quality studies in with high quality studies and this can cause the results produced by meta-analysis to be ultimately skewed. Similar arguments were made at the conception of meta-analysis (see Greco, Zangrillo, Biondi-Zoccai, & Landoni, 2013). However, meta-analyses usually weigh studies by quality and sample size/precision in order to give the best studies the most say in the final result (Hunter & Schmidt, 2004; also see Borenstein, Hedges, Higgins, & Rothstein, 2009). Additionally, many of the studies on job performance and IQ are high quality, large-scale studies rather than meta-analyses, typically done in the military with objective measures (e.g. McHenry et al., 1990; Ree et al., 1994). These studies are thus not subject to the criticism that Richardson & Norgate put forward yet still produce the same finding: *g* predicts job performance fairly well.

Richardson & Norgate also express concern with the tests used in studies on IQ and job performance, primarily that one meta-analysis (Salgado et al., 2003) classified tests as either 'g-tests' or 'batteries', thereby suggesting they don't measure the same thing. In fact, the authors of that study don't seem to make much of this distinction because none of their results are broken down by it, and it is not mentioned after the methods section. So we are curious as to why Richardson & Norgate make much of this somewhat odd phrasing. Each standalone test or battery of tests measures some mix of general intelligence, group factors, and more specific abilities as well as motivation indirectly (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011[4]; Gignac, Bartulovich, and Salleo, 2019). No single test is known to measure *g* and nothing else (aside from random error), but it is known that batteries of diverse tests measure the same g (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Johnson, Nijenhuis, & Bouchard, 2008). Generally, longer and more diverse tests provide better measures of *g* in the sense that they better capture the full construct and have higher reliabilities. For instance, in the Johnson et al. (2008) study, the *g* factor from the Cattell Culture Fair Test (CCFT) was less strongly correlated with the *g* factors from the other batteries even accounting for reliability. This is because the CCFT is a nonverbal battery with 4 types of matrix tests, and thus does not capture general intelligence variation related to e.g. verbal or 3D spatial abilities and thus is missing some of the construct variance (in other words, it lacks perfect construct validity). McHenry et al. (1990) found adding additional, cognitively demanding tests to the ASVAB battery only marginally increased the validity of the battery. So, while there is some variation in the construct measures by different tests and batteries of tests, they are relatively

[4] It should be noted Duckworth's meta-analysis was partially based on studies by Stephen Breuning who has been accused of fraud in his scientific literature (Witkowski, 2014).

minor and thus of little importance to researchers interested in the relationship between job performance and general intelligence.

Richardson & Norgate's more concrete replies come through their criticism of how much Hunter & Schmidt (and others) have corrected for restriction and error in their meta-analyses. While some (Kaufman & Kaufman, 2015) have agreed that the job performance corrections are likely too large, Richardson & Norgate certainly over-estimate the degree to which this is true. The primary analysis cited by Richardson & Norgate to defend their argument is the report done by Hartigan & Wigdor (1989). The analysis done by Hartigan & Wigdor (1989) was commissioned by the National Academy of Sciences to investigate the relationship between job performance and IQ. Hartigan & Wigdor primarily argue that inter-rater reliability should be estimated at about 0.80 rather than the 0.60 used by Hunter & Hunter (1984), and that correcting for range restriction causes a large upward bias because the meta-analyses on IQ and job performance are typically limited to specific job sectors.

It is interesting that Richardson & Norgate are willing to accept supervisory ratings when Hartigan & Wigdor used them for their analysis, but not in the case of Hunter & Hunter. Many studies have come out since Hartigan & Wigdor's analysis showing that their estimate of inter-rater reliability was too high. Most studies find inter-rater reliability of around 0.50 to 0.60, somewhat lower than what Hunter and Hunter (1984) used. Reviews of this nature include Schmidt, Hunter, & Caplan (1981), Shen, Cucina, Walmsley, & Seltzer (2014), Hirsh, Northrop, & Schmidt (1986), Rothstein (1990), Salgado and Anderson (2003), Salgado et al. (2003), Salgado, Anderson, Moscoso, Bertua, & De Fruyt (2003), Salgado and Moscoso (1996), and Viswesvaran et al. (1996). Presumably, Richardson & Norgate have read Viswesvaran et al.

(1996) as they cite it within their own article. All of the meta-analyses of inter-rater reliability find Hunter & Hunter's original estimate of 0.60 was not only correct, but probably an overestimate. The now-accepted validity coefficient for inter-rater reliability is 0.52 (Shen et al., 2014). As noted by Anderson, Ones, Kepir Sinangil, & Viswesvaran (2014), if the inter-rater reliability they found were applied to the Hartigan & Wigdor analysis, the mean operational validity would be 0.38, which is substantially closer to that estimated by Hunter & Hunter. In reviewing the evidence, Viswesvaran et al. (1996) noted that the probability of the 0.80 figure that Hartigan & Wigdor used being accurate is only 0.0026. Viswesvaran, Ones, Schmidt, Le, & Oh (2014), Shen et al. (2014), Brown (2014), and Sackett (2014) have also provided replies to common criticisms of correcting for measurement error. If Richardson & Norgate wish to seriously criticize the issue of inter-rater reliability, they will need a much stronger basis than the Hartigan & Wigdor analysis. The vast majority of research in this area is in strong disagreement with them.

In order to correct for range restriction, Hunter & Hunter had to estimate the standard deviations of job applicants' test results by assuming the careers which they could find this data for were generalizable to the entire United States population. Hartigan & Wigdor found this troubling and argued this assumption will strongly bias the results upward. Sackett & Ostgaard (1994) replied to Hartigan & Wigdor's analysis, which excluded a correction for range restriction, by empirically estimating the standard deviations for applicants of a wide range of jobs. Based on this analysis, Hunter & Hunter's correction for range restriction was justified. The authors argued that Hartigan & Wigdor wrongly excluded their correction for measurement error as it would lead to a much larger downward bias than the upward bias created by Hunter &

Hunter. Furthermore, until 2004, corrections were not made for indirect range restriction in meta-analyses on job performance and intelligence (Hunter & Schmidt, 2004). The method of correcting for indirect range restriction has been shown to provide more reliable estimates of validity. Schmidt, Shaffer, & Oh (2008) and Sjöberg, Sjöberg, Näswall, & Sverke (2012) found the traditional method of simply correcting for direct range restriction has resulted in underestimates in the validity of intelligence and personality measures in predicting job performance. Schmidt, Oh, & Le (2006) found that the operational validities for job performance measures were underestimated by 21 percent due to failure to correct for indirect range restriction.

Richardson & Norgate accuse Hunter & Hunter of over-correcting for sampling error. In their view, the true variability due to sampling error is three quarters or less of the size that Hunter & Hunter reported. However, their reasoning for this is ill-founded. First, they say that the studies used in IQ-job performance studies would be limited to a specific sort which are "willing to have employees tested and finding supervisors willing to rate them" (p. 158). Conversely, Hunter & Hunter (1984) reported on studies which showed that the validity of GMA in job performance holds across virtually all sorts of careers using both supervisory ratings and work sample tests. While it is true that IQ testing will be more liberally used in more complex careers, this is the result of the greater association between IQ and job performance in more complex occupations and a simple analysis of the cost of testing compared to the incremental increase in productivity.

Richardson & Norgate correctly note that data is often not available from older, poorer studies to correct them individually for sampling error, and that correcting after averaging the results could lead to some bias. However, Hunter & Schmidt (1994) found correcting the correlations individually may bias the estimation of sampling error further, hence the “average correlation” method is preferable. This is likely because correcting each study using its own estimates of reliability introduces another source of sampling error (in the reliability coefficient) into the estimate of the association between *g* and job performance, whereas using artifact distributions or averages avoids this source of variance, but at the cost of missing some true variation in reliability.

Since the report by Hartigan & Wigdor, even larger analyses have come out further supporting Hunter & Hunter’s validity estimates. Kuncel, Ones, & Sackett (2010) summarized these in a review concerning the role of intelligence in life outcomes. The most notable is a meta-analysis by Ones, Viswesvaran, & Dilchert (2005) which reviewed over 20,000 studies and a sample of over 5,000,000 people. They found the validity of cognitive ability in predicting job performance is around 0.50-0.60. Job complexity also correlated with the validity of IQ in predicting job performance, but even in low-complexity jobs, the validity coefficients ranged from 0.30-0.40. Overall, Richardson and Norgate’s criticism of standard meta-analytic procedures used in industrial and organizational psychology falls short.

Job Complexity

In order to criticize the position that job complexity correlates with the validity of GMA in predicting job performance, Richardson & Norgate refer, once again, to the Hartigan & Wigdor analysis. This is flawed for the same reasons discussed in the previous section. Hartigan & Wigdor did not correct for measurement error or range restriction. Once again, Ones et al.

(2005) replicated the Hunter & Hunter finding on job complexity with a much larger amount of studies and properly correcting for measurement error and range restriction. Any occupation which is more complex will demand more of the employee in a variety of ways, so there is no reason why the same shouldn't happen for intelligence. For example, job knowledge is correlated with job performance to a greater degree in more complex jobs as well (Dye, Reck, & McDaniel, 1993).

Richardson & Norgate also bring up how psychological variables may be confounding the correlation, such as self-esteem and the fact that people in jobs of lower complexity communicate less with their managers. However, the relationship between job complexity, job performance, and IQ has been shown on work sample tests as well (Salgado & Moscoso, 2019), i.e. objective tests not based on the opinions of supervisors or peers. Another reason to be wary of Richardson & Norgate's criticism is that people who are higher in intelligence do tend to have higher self-esteem, but this did not translate to greater confidence in job ability in a study by Lynch & Clark (1985).

Richardson & Norgate are concerned that there is more communication between supervisors and employees in more complex positions which may cause the correlation to be artifactual. This is improbable for multiple reasons. As we stated before, the role of job complexity is even shown on work sample tests as well (Salgado & Moscoso, 2019). However, a more telling reason why Richardson & Norgate are incorrect about this is that conscientiousness, the willingness to do tasks thoroughly, actually has greater validity in lower complexity jobs, as shown by Wilmot & Ones (2019). If Richardson & Norgate were correct that interaction with supervisors confounds this relationship, then there should always be greater validity at higher levels of job complexity.

Finally, there are a few more reasons why job complexity should partially mediate the relationship between job performance and IQ. Schmidt & Hunter (2004), for example, showed the standard deviations in IQ are smaller in more complex jobs. Ganzach, Gotlibobski, Greenberg, & Pazy (2013) found occupational complexity mediated the relationship between IQ and income. The correlation between IQ and educational achievement increases at higher levels of education (Arneson, Sackett, & Beatty 2011). An older study used the subjectively assessed degree of intelligence required for an occupation and found it is correlated to a remarkably strong degree ($r = 0.91$) with its subjectively rated level of prestige (Jensen, 1980: 340). Since the latter is true, it seems that IQ is important in predicting how well one can perform more complex occupations. When comparing simple reaction times to choice reaction times (the latter being more complex), choice reaction times have a greater correlation with IQ (Der & Deary, 2017). Intelligence becomes more predictive at higher ranges of complexity in a wide range of mental tasks, so there is no reason to assume the relationship wouldn't be the same for job performance.

Supposed Non-Cognitive Causes

Richardson & Norgate attribute any leftover relationship between IQ and job performance to be confounded by other psychological traits. However, Richardson & Norgate make some major errors in their analysis of this topic. They cite a study which showed the relationship between cognitive ability and job performance was entirely mediated by job knowledge (Palumbo, Miller, Shalin, & Steele-Johnson, 2005). However, the argument assumes IQ is not a cause of how easily and quickly individuals can attain job knowledge. As Schmidt & Hunter (2004) explained:

> As can be seen, in both data sets, the major effect of GMA is on the acquisition of job knowledge, and job knowledge in turn is the major determinant of job performance (measured using hands-on job sample tests). GMA does have a direct effect on job performance independent of job knowledge in both data sets but this effect is smaller than its indirect effect through job knowledge. . . These results also show that supervisory ratings of job performance are determined in both data sets by both job knowledge and job sample performance (p. 170).

The path analyses referred to by Schmidt & Hunter (2004) are reproduced in Figure 1. It is more likely that intelligence predicts job knowledge. James & Carretta (2002) noted that before people can perform occupational tasks, they need to learn what to do and how to do it. This requires the ability to retain and apply knowledge within the real world. Even job knowledge has its limits. For example, Joseph (1997) tested the success of “Right-to-know” training programs (additional information provided about toxic substances so as to reduce workplace injury). These programs had no significant effect on related workplace injuries for people with an IQ below 70. Periodic assessments over many years show the validity of *g* in predicting job knowledge, supervisory ratings, and performance on objective work sample tests does not decline (Schmidt, Hunter, & Outerbridge, 1986; Schmidt, Hunter, Outerbridge, & Goff, 1988). If job knowledge were more important or just as important in predicting occupational performance, the validity of *g* should decline over time as one becomes more familiar with their job, their duties, and their occupational network.

[INSERT FIGURE 1 HERE]

Much of their argument is based around a view that *g* is invalid or that IQ lacks construct validity. There is not much reason to further argue this case past what we detailed in Section 2, and what Kaufman & Kaufman (2015) wrote in their commentary of Richardson & Norgate's paper. Richardson & Norgate also point to evidence showing that IQ test performance can be improved through "presumably knowledge based - experience with compatible cognitive tasks" (p. 162). It is unlikely such IQ gains are on g, rather than specific skills though (see te Nijenhuis, Voskuijl, & Schijve, 2001; te Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis, Jongeneel-Grimen, & Kirkegaard, 2014; Ritchie et al., 2015;).

The authors also assert that "emotional intelligence" is a better predictor of job performance than IQ, citing a review by Goleman (2000) which found "emotional competence mattered twice as much" compared to IQ. The issue with this argument is that some have argued emotional intelligence is a very vague concept (Locke, 2005). Emotional intelligence could also be characterized as downstream of personality, which has been shown to predict job performance as well (Judge, Rodell, Klinger, Simon, & Crawford, 2013). Schulte, Ree, & Carretta (2004) found emotional intelligence is mostly just a measure of *g* and personality. As a consequence, the addition in incremental validity from using emotional intelligence measures is unimpressive (O'Boyle, Humphrey, Pollack, Hawver, & Story, 2011). As detailed in a review by Antonakis (2004), emotional intelligence does not appear to hold any validity beyond its relationship to IQ and personality in predicting leadership effectiveness as well.

In defense of the position that occupational structure and networks are more important in predicting job performance, Richardson & Norgate point to a study (Groysberg, 2010) which showed that high performers on Wall Street who switched firms suffered a decline in performance. However, pushing this as a major argument seems detached from reality – the pure existence of this decline does not mean that intelligence doesn't remain an important factor. Taking the high performers of Wall Street doesn't really tell us anything about the general American population either. The results could potentially be an example of regression to the mean, though the primary source does not seem to investigate this.

Finally, the authors present some arguments concerning anxiety, motivation and test scores. A study by Gignac et al. (2019) found that motivation had a modest correlation with IQ, but the effect was non-linear and entirely centered in the low-moderate levels of intelligence. Primarily less intelligent people are uninterested in taking IQ tests as they are not personally relevant to them (cf. Dang, Xiao, & Dewitte, 2015), so greater motivation would result in slightly better performance. Reeve & Lam (2007) found the effect of motivation on IQ was not on *g*. Furthermore, greater *g* predicted greater test motivation, as would be expected.

If motivation were truly a confounding variable, it would have to predict job performance as well as IQ. Many studies have found motivation is unrelated to educational achievement and job performance (Gagné & St Père, 2001; Bloom, 1976; McHenry et al., 1990; Schmidt & Hunter, 1998; Terborg, 1997). One study also finds proactive behavior explains less than one percent of the variance in objective sales performance (Pitt, Ewing, & Berthon, 2002). Since, its relation to IQ is minimal, it doesn't predict *g*, and it has no relation to job performance, motivation is not going to be a confounding psychological variable for the association between intelligence and job performance. Finally, since the IQ tests would be taken for employment purposes, the potential employees are likely to be more motivated than normal.

The effect of anxiety on IQ scores in general seems to be up for debate. In questioning the existing literature at the time, Jensen (1980) wrote,

> In brief, many studies have reported generally low but significant negative correlations between various measures of the subject's anxiety level, such as the Taylor Manifest Anxiety Scale and the Sarason Test Anxiety Scale, and performance on various mental ability tests. Many nonsignificant correlations are also reported, although they are in the minority, and are usually rationalized by the investigators in various ways, such as atypical samples, restriction of range on one or both variables, and the like (e.g., Spielberger, 1958). I suspect that this literature contains a considerably larger proportion of "findings" that are actually just Type I errors (i.e., rejection of the null hypothesis when it is in fact true) than of Type II errors (i.e., failure to reject the null hypothesis when it is in fact false). Statistically significant correlations are more often regarded as a "finding" than are nonsignificant results, and Type I errors are therefore more apt to be submitted for publication. Aside from that, sheer correlations are necessarily ambiguous with respect to the direction of causality. Persons who, because of low ability, have had the unpleasant experience of performing poorly on tests in the past may for that reason find future test situations anxiety provoking—hence a negative correlation between measures of test anxiety and ability test scores (p. 615).

There are other issues with the literature. For one, a bivariate correlation between IQ and anxiety does not determine causality. Second of all, there must be a distinction between trait anxiety and state anxiety. The former would be recognizable through a typical questionnaire but the latter is aroused in specific situations. Jensen & Figueroa (1975) noted that digit span scores are associated with state anxiety rather than trait anxiety. Additionally, research shows test anxiety and motivation are negatively correlated with one another (Rajiah, 2014). If those taking IQ tests for employment purposes are more motivated, which they may very well be, it is likely they are less anxious as well. Therefore, anxiety is unlikely an important confounder. Overall, there is no reason to take these non-cognitive causes as major detriments to the prior research on intelligence and job performance.

Summary

Ken Richardson & Sarah Norgate's case that IQ is an invalid predictor of job performance is uncompelling on account of its ignorance of the general literature that surrounds job performance, and a misinterpretation of intelligence assessments and their academic validity. Their first arguments relied on assuming face validity as a necessity for construct validity, which is an incredibly stringent requirement. As we demonstrate, the intercorrelation of IQ tests and their relation to a higher-order g factor, is sufficient enough to make a judgement as to the construct they measure. As a large body of research now shows, IQ is also predictive of life outcomes (Strenze, 2015; Herrnstein & Murray, 1994), and these correlations are not built into the tests. For the test to seriously be considered a measure of social class rather than intelligence, as Richardson might do, one has to deal with the facts that 1) the correlation between parental socioeconomic status and child IQ is relatively low (Hanscombe et al., 2012; $r = 0.08$-0.37, differing by age of child and when socioeconomic status was estimated), 2) multiple studies have found individual IQ is nearly as predictive of life outcomes within families as it is within cohorts (Murray, 1998; Murray, 2002; Frisell, Pawitan, & Långström, 2012; Hegelund, Flensborg-Madsen, Dammeyer, Mortensen, & Mortensen, 2019), and 3) IQ test scores correlate with neurological variables, both structural (e.g. whole brain volume) and activation patterns, implying processes are occurring within the brain to determine someone's score on the test (Haier, 2016).

Richardson & Norgate spent a short amount of time addressing certain biases which could occur in the workplace, influencing supervisory ratings. However, this appears to be a distraction. The authors take little to no time to investigate the correlation between IQ and work sample tests or other objective criteria for measuring job performance. In tackling biases, Richardson & Norgate fail to address contemporary literature which has dealt with these objections already. The end result is an insubstantial argument with little, empirically, to hold itself up.

The longest section of Richardson & Norgate's argument concerns meta-analytic procedures. Whereas some have applauded them for their discussion on this matter, it is largely flawed. This, even they should know, as their own sources seem to contradict the arguments they put forward. Most notably, they cite the argument from Hartigan & Wigdor (1989) that Hunter & Hunter (1984) largely underestimated the inter-rater reliability used for supervisor ratings. However, Viswesvaran et al. (1996), who Richardson & Norgate have cited in their paper, conducted a meta-analysis disproving this. Their usage of the Hartigan & Wigdor analysis is also strange for other reasons. First, it was viewed by the authors to be a positive replication of the work done by Hunter & Hunter. Second, Richardson & Norgate cite the General Aptitude Test Battery as a test used as a 'battery' rather than a 'g-test', questioning its psychometric validity, however Hartigan & Wigdor dedicated an entire chapter to showing its psychometric validity. These patterns indicate further omission of important details from the primary authors.

Richardson & Norgate end their paper by arguing there should be far greater skepticism regarding the relationship between job performance and IQ. However, as we have shown, Richardson & Norgate's position on the matter is on the fringes. Aside from smaller studies, Hunter & Hunter (1984) have enjoyed positive replication for over thirty years. Even before the paper by Hunter & Hunter, the results were the same. The most notable critic of these was McClelland (1973), whose criticisms are very similar to those of Richardson & Norgate. This was later responded to by Barrett & Depinet (1991) as well as Barrett, Kramen, & Leueke (2003), showing that none of the major claims about the predictive validity of IQ tests, biases in the workplace and in testing, and the underlying cause of the IQ and job performance relationship held up. This all goes to show Richardson & Norgate's arguments are not new, ignore relevant counter-evidence, and are not well-supported by data.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The study did not receive external funding.

References

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. The quarterly journal of economics, 116(1), 313-350.

Anderson, N., Ones, D. S., Kepir Sinangil, H., & Viswesvaran, C. (2014). *Handbook of Industrial, Work and Organizational Psychology* (Vol. 1).

Antonakis, J. (2004). On why "emotional intelligence" will not predict leadership effectiveness beyond IQ or the "big five": An extension and rejoinder. *Organizational Analysis*, 171–182.

Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the more-is-better and the good-enough hypotheses. *Psychological Science*, *22*(10), 1336–1342. https://doi.org/10.1177/0956797611417004

Barrett, G. V., Kramen, A. J., & Lueke, S. B. (2003). New Concepts of Intelligence. Their Practical and Legal Implications for Employee Selection. *The Scientific Study of General Intelligence: Tribute to Arthur Jensen*, 411–439. https://doi.org/10.1016/B978-008043793-4/50058-1

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 13.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill

Bobko, P., & Roth, P. L. (2013). Reviewing, Categorizing, and Analyzing the Literature on Black–White Mean Differences for Predictors of Job Performance: Verifying Some perceptions and Updating/Correcting Others. *Personnel Psychology*, *66*(1), 91–126. https://doi.org/10.1111/peps.12007

Bommer, W. H., Johnson, J., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587–605.

Bouchard, T. J. (2013). The Wilson Effect: The increase in heritability of IQ with age. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *16*(5), 923–930. https://doi.org/10.1017/thg.2013.54

Brown, R. D. (2014). In Defense of the Accuracy of the Criterion Reliability Adjustment of Bivariate Correlations. *Industrial and Organizational Psychology*, *7*(4), 524–526. https://doi.org/10.1111/iops.12188

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*, *59*, 72–83. https://doi.org/10.1016/j.intell.2016.08.002

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Caplan, B. (2018). The case against education. In The Case against Education. Princeton University Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (pp. ix, 819). Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52(4)*, 281–302. https://doi.org/10.1037/h0040957

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *The Journal of Applied Psychology*, *102*(10), 1403–1420. https://doi.org/10.1037/apl0000234

Dalliard. (2013, April 3). Is Psychometric g a Myth? *Human Varieties*. https://humanvarieties.org/2013/04/03/is-psychometric-g-a-myth/

Dang, J., Xiao, S., & Dewitte, S. (2015). Commentary: "Poverty impedes cognitive function" and "the poor's poor mental power." *Frontiers in Psychology*, *6*.

de Wit, G., & van Winden, F. A. A. M. (1989). An empirical analysis of self-employment in the Netherlands. *Small Business Economics*, *1*(4), 263–272. https://doi.org/10.1007/BF00393805

Dejung, J. E., & Kaplan, H. (1962). Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. *Journal of Applied Psychology*, *46*(5), 370–374. https://doi.org/10.1037/h0048376

Der, G., & Deary, I. J. (2017). The relationship between intelligence and reaction time varies with age: Results from three representative narrow-age age cohorts at 30, 50 and 69 years. *Intelligence*, *64*, 89–97. https://doi.org/10.1016/j.intell.2017.08.001

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*(2), 346–369. https://doi.org/10.1037/0033-295x.108.2.346

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, *108*(19), 7716–7720. https://doi.org/10.1073/pnas.1018601108

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The Validity of Job Knowledge Measures. *International Journal of Selection and Assessment*, *1*(3), 153–157. https://doi.org/10.1111/j.1468-2389.1993.tb00103.x

Elliott, C. D. (1986). The factorial structure and specificity of the British Ability Scales. *British Journal of Psychology*, *77*(2), 175–185. https://doi.org/10.1111/j.2044-8295.1986.tb01992.x

Eysenck, H. J. (1939). Primary Mental Abilities. *British Journal of Educational Psychology*, *9*(3), 270–275. https://doi.org/10.1111/j.2044-8279.1939.tb03214.x

Frisell, T., Pawitan, Y., & Långström, N. (2012). Is the Association between General Cognitive Ability and Violent Crime Caused by Family-Level Confounders? *PLOS ONE*, *7*(7), e41783. https://doi.org/10.1371/journal.pone.0041783

Gagné, F., & St Père, F. (2001). When IQ is controlled, does motivation still predict achievement? *Intelligence*, *30*(1), 71–100. https://doi.org/10.1016/S0160-2896(01)00068-X

Ganzach, Y., Gotlibobski, C., Greenberg, D., & Pazy, A. (2013). General Mental Ability and pay: Nonlinear effects. *Intelligence*, *41*(5), 631–637. https://doi.org/10.1016/j.intell.2013.07.015

Gensowski, M., Heckman, J., & Savelyev, P. (2011). *The Effects of Education, Personality, and IQ on Earnings of High-Ability Men*. 84.

Gignac, G. E., Bartulovich, A., & Salleo, E. (2019). Maximum effort may not be required for valid intelligence test score interpretations. *Intelligence*, *75*, 73–84. https://doi.org/10.1016/j.intell.2019.04.007

Gignac, G., Vernon, P. A., & Wickett, J. C. (2003). Factors Influencing the Relationship Between Brain Size and Intelligence. In *The Scientific Study of General Intelligence* (pp. 93–106). Elsevier. https://doi.org/10.1016/B978-008043793-4/50042-8

Gottfredson, L. S. (2004). Intelligence: Is It the Epidemiologists' Elusive "Fundamental Cause" of Social Class Inequalities in Health? *Journal of Personality and Social Psychology*, *86*(1), 174–199. https://doi.org/10.1037/0022-3514.86.1.174

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, Lung and Vessels*, *5*(4), 219–225.

Groysberg, B. (2012). *Chasing Stars: The Myth of Talent and the Portability of Performance* (Reprint edition). Princeton University Press.

Haier, R. J. (2016). *The Neuroscience of Intelligence*. Cambridge University Press.

Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E., & Head, K. (2009). Gray matter and intelligence factors: Is there a neuro-g? *Intelligence*, *37*(2), 136–144. https://doi.org/10.1016/j.intell.2008.10.011

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic Status (SES) and Children's Intelligence (IQ): In a UK-Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0030320

Harris, M., & Schaubroeck, J. (2006). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43–62. https://doi.org/10.1111/j.1744-6570.1988.tb00631.x

Hartigan, J., & Wigdor, A. (1 C.E.). *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. https://doi.org/10.17226/1338

Hegelund, E. R., Flensborg-Madsen, T., Dammeyer, J., Mortensen, L. H., & Mortensen, E. L. (2019). The influence of familial factors on the association between IQ and educational and occupational achievement: A sibling approach. *Personality and Individual Differences*, *149*, 100–107. https://doi.org/10.1016/j.paid.2019.05.045

Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology, 39*, 811–826.

Herrnstein, R. J., & Murray, C. (1996). *The Bell Curve: Intelligence and Class Structure in American Life* (1st Free Press pbk. ed edition). Free Press.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology*, *39*(2), 399–420. https://doi.org/10.1111/j.1744-6570.1986.tb00589.x

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of General Mental Ability for the Prediction of Job Performance and Training Success in Germany: A meta-analysis1. *International Journal of Selection and Assessment*, *15*(1), 3–18. https://doi.org/10.1111/j.1468-2389.2007.00363.x

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–362. https://doi.org/10.1016/0001-8791(86)90013-8

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. https://doi.org/10.1037/0033-2909.96.1.72

Hunter, J., & Schmidt, F. (1970). Intelligence and job performance: Economic and Social Implications. *Psychology, Public Policy, and Law*, *2*, 447–472. https://doi.org/10.1037/1076-8971.2.3-4.447

Hunter, J. E., & Schmidt, F. L. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. *Journal of Applied Psychology*, *79*(2), 171–177. https://doi.org/10.1037/0021-9010.79.2.171

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. SAGE.

James, M., & Carretta, T. R. (2002). G2K. *Human Performance*, *15*(1–2), 3–23. https://doi.org/10.1080/08959285.2002.9668081

Jensen, A. R. (1993). Why Is Reaction Time Correlated with Psychometric g? *Current Directions in Psychological Science*, *2*(2), 53–56. JSTOR.

Jensen, A. R., & Figueroa, R. A. (1975). Forward and backward digit span interaction with race and IQ: Predictions from Jensen's theory. *Journal of Educational Psychology*, *67*(6), 882–893. https://doi.org/10.1037/0022-0663.67.6.882

Jensen, A. R. (1980). *Bias in mental testing*. Free Press.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107. https://doi.org/10.1016/S0160-2896(03)00062-X

Johnson, W., Nijenhuis, J. te, & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95. https://doi.org/10.1016/j.intell.2007.06.001

Joseph, A. J. (1997). Right-to-know training of workers with IQ less than 70: A pilot study. *American Journal of Industrial Medicine*, *32*(4), 417–420. https://doi.org/10.1002/(SICI)1097-0274(199710)32:4<417::AID-AJIM14>3.0.CO;2-6

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875–925. https://doi.org/10.1037/a0033901

Kaufman, J. C., & Kaufman, A. S. (2015). It Can Be Very Tempting to Throw Out the Baby With the Bathwater: A Father-and-Son Commentary on “Does IQ Really Predict Job Performance?” *Applied Developmental Science*, *19*(3), 176–181. https://doi.org/10.1080/10888691.2015.1008922

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3rd edition). *School Psychology Review*, *30*(1), 89–119.

Kim, T. H., & Han, E. (2017). Height premium for job performance. *Economics & Human Biology*, *26*, 13–20. https://doi.org/10.1016/j.ehb.2017.01.002

Kirkegaard, E. O. W. (2019). Is National Mental Sport Ability a Sign of Intelligence? An Analysis of the Top Players of 12 Mental Sports. *The Mankind Quarterly*, *59*.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences*, *49*(4), 331–336. https://doi.org/10.1016/j.paid.2010.03.042

Lasker, J. (2022). Are Piagetian scales just intelligence tests? Intelligence, 95, 101702. https://doi.org/10.1016/j.intell.2022.101702

Laurence, J. H., & Ramsberger, P. F. (1991). *Low-Aptitude Men in the Military: Who Profits, Who Pays?* Praeger.

Locke, E. A. (2005). Why emotional intelligence is an invalid concept. *Journal of Organizational Behavior*, *26*(4), 425–431. https://doi.org/10.1002/job.318

Lubinski, D. (2009). Cognitive epidemiology: With emphasis on untangling cognitive ability and socioeconomic status. *Tool & Stankov Intelligence*, *37*, 625–633. https://doi.org/10.1016/j.intell.2009.09.001

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*, *24*(1), 159–201. https://doi.org/10.1016/S0160-2896(97)90016-7

Lynch, A. D., & Clark, P. (1985). Relationship of Self-Esteem, IQ, and Task Performance for a Sample of USA Undergraduates. *Psychological Reports*, *56*(3), 955–962. https://doi.org/10.2466/pr0.1985.56.3.955

Marks, G. N. (2022). Cognitive ability has powerful, widespread and robust effects on social stratification: Evidence from the 1979 and 1997 US National Longitudinal Surveys of Youth. Intelligence, 94, 101686.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, *28*(1), 1–14. https://doi.org/10.1037/h0034092

McGrew, K. S., & Knopik, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. School Psychology Review, 22(4), 687–695. https://doi.org/10.1080/02796015.1993.12085684

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, *43*(2), 335–354. https://doi.org/10.1111/j.1744-6570.1990.tb01562.x

McKay, P. F., & McDaniel, M. A. (2006). A reexamination of black-white mean differences in work performance: More data, more moderators. *The Journal of Applied Psychology*, *91*(3), 538–554. https://doi.org/10.1037/0021-9010.91.3.538

Murray, C. (1998). *Income Inequality and IQ*. AEI Press, c/o Publisher Resources Inc.

Murray, C. (2002). IQ and Income Inequality in a Sample of Sibling Pairs from Advantaged Family Backgrounds. *American Economic Review*, *92*, 339–343. https://doi.org/10.1257/000282802320191570

Naglieri, J. A. (2001). Cognitive Assessment System (CAS). In W. I. Dorfman & M. Hersen (Eds.), *Understanding Psychological Assessment* (pp. 235–257). Springer US. https://doi.org/10.1007/978-1-4615-1185-4_12

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. *Personnel Psychology*, *41*(3), 517–535. https://doi.org/10.1111/j.1744-6570.1988.tb00642.x

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Boyle, E. H., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, *32*(5), 788–818. https://doi.org/10.1002/job.714

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive Ability in Selection Decisions. In *Handbook of understanding and measuring intelligence* (pp. 431–468). Sage Publications, Inc. https://doi.org/10.4135/9781452233529.n24

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2008). No New Terrain: Reliability and Construct Validity of Job Performance Ratings. *Industrial and Organizational Psychology*, *1*(2), 174–179. https://doi.org/10.1111/j.1754-9434.2008.00033.x

Palumbo, M. V., Miller, C. E., Shalin, V. L., & Steele-Johnson, D. (2005). *The Impact of Job Knowledge in the Cognitive Ability-Performance Relationship*. 8.

Pearce, M. S., Deary, I. J., Young, A. H., & Parker, L. (2005). Growth in early life and childhood IQ at age 11 years: The Newcastle Thousand Families Study. *International Journal of Epidemiology*, *34*(3), 673–677. https://doi.org/10.1093/ije/dyi038

Pitt, L. F., Ewing, M. T., & Berthon, P. R. (2002). Proactive behavior and industrial salesforce performance. *Industrial Marketing Management*, *31*(8), 639–644. https://doi.org/10.1016/S0019-8501(01)00171-7

Quiroga, M. A., Diaz, A., Román, F. J., Privado, J., & Colom, R. (2019). Intelligence and video games: Beyond "brain-games." *Intelligence*, *75*(C), 85–94.

Quiroga, M. A., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., Hernández, M., Gallego, B., & Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence*, *53*, 1–7. https://doi.org/10.1016/j.intell.2015.08.004

Rajiah, K. (2014). Relationship of Test Anxiety, Psychological Distress and Academic Motivation among First Year Undergraduate Pharmacy Students. *International Journal of Applied Psychology*, *4*, 68–72.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, *79*(4), 518–524. https://doi.org/10.1037/0021-9010.79.4.518

Reeve, C. L., & Lam, H. (2007). Consideration of g as a common antecedent for cognitive ability test performance, test motivation, and perceived fairness. *Intelligence*, *35*(4), 347–358. https://doi.org/10.1016/j.intell.2006.08.006

Richardson, K., & Norgate, S. H. (2015). Does IQ Really Predict Job Performance? *Applied Developmental Science*, *19*(3), 153–169. https://doi.org/10.1080/10888691.2014.983635

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is Education Associated With Improvements in General Cognitive Ability, or in Specific Skills? *Developmental Psychology*, *51*(5), 573–582. https://doi.org/10.1037/a0038981

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, *29*(8), 1358–1369. https://doi.org/10.1177/0956797618774253

Rosenberg, I. B. (2009). Height Discrimination in Employment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1344817

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology*, *88*(4), 694–706. https://doi.org/10.1037/0021-9010.88.4.694

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology*, *75*(3), 322–327. https://doi.org/10.1037/0021-9010.75.3.322

Sackett, P. R., & Ostgaard, D. J. (1994). Job-specific applicant pools and national norms for cognitive ability tests: Implications for range restriction corrections in validation research. *Journal of Applied Psychology*, *79*(5), 680–684. https://doi.org/10.1037/0021-9010.79.5.680

Sackett, P. R. (2014). When and Why Correcting Validity Coefficients for Interrater Reliability Makes Sense. *Industrial and Organizational Psychology*, *7*(4), 501–506. https://doi.org/10.1111/iops.12185

Salgado, J. F., & Moscoso, S. (2019). Meta-Analysis of the Validity of General Mental Ability for Five Performance Criteria: Hunter and Hunter (1984) Revisited. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02227

Salgado, Jesús, & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology*, *12*(1), 1–17. https://doi.org/10.1080/13594320244000292

Salgado, Jesus, Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International Validity Generalization of GMA & Cognitive Abilities. *Personnel Psychology*, *56*, 573–605. https://doi.org/10.1111/j.1744-6570.2003.tb00751.x

Salgado, Jesus, Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, *88*, 1068–1081.

Salgado, Jesus, & Moscoso, S. (1996). Meta-analysis of Interrater Reliability of Job Performance Ratings in Validity Studies of Personnel selection. *Perceptual and Motor Skills*, *83*, 1195–1201. https://doi.org/10.2466/pms.1996.83.3f.1195

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262 – 274.

Schmidt, F., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86*, 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmidt, F. L. (2002). The Role of General Cognitive Ability and Job Performance: Why There Cannot Be a Debate. *Human Performance*, *15*(1–2), 187–210. https://doi.org/10.1080/08959285.2002.9668091

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, *66*(3), 261–273. https://doi.org/10.1037/0021-9010.66.3.261

Schmidt, F. L., Hunter, J. E. & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 432-439.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology*, *73*(1), 46–57. https://doi.org/10.1037/0021-9010.73.1.46

Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the Accuracy of Corrections for Range Restriction: Implications for Selection Procedure Validities and Other Research Results. *Personnel Psychology*, *59*(2), 281–305. https://doi.org/10.1111/j.1744-6570.2006.00065.x

Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, *61*(4), 827–868. https://doi.org/10.1111/j.1744-6570.2008.00132.x

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta Analyses of Validity Studies Published Between 1964 and 1982 and the Investigation of Study Characteristics. *Personnel Psychology*, *37*(3), 407–422. https://doi.org/10.1111/j.1744-6570.1984.tb00519.x

Schulte, M., Ree, M., & Carretta, T. (2004). Emotional intelligence: Not much more than G and personality. *Personality and Individual Differences*, *37*, 1059–1068. https://doi.org/10.1016/j.paid.2003.11.014

Sjöberg, S., Sjöberg, A., Näswall, K., & Sverke, M. (2012). Using individual differences to predict job performance: Correcting for direct and indirect restriction of range: Using individual differences. *Scandinavian Journal of Psychology*, *53*(4), 368–373. https://doi.org/10.1111/j.1467-9450.2012.00956.x

Shen, W., Cucina, J. M., Walmsley, P. T., & Seltzer, B. K. (2014). When Correcting for Unreliability of Job Performance Ratings, the Best Estimate Is Still .52. *Industrial and Organizational Psychology*, *7*(4), 519–524. https://doi.org/10.1111/iops.12187

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. Merrill-Palmer Quarterly, 47(1), 1–41. https://www.jstor.org/stable/23093686

Sternberg, R. J. (2015). Competence Versus Performance Models of People and Tests: A Commentary on Richardson and Norgate. *Applied Developmental Science*, *19*(3), 170–175. https://doi.org/10.1080/10888691.2015.1008920

Strenze, T. (2015). Intelligence and Success. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of Intelligence* (pp. 405–413). Springer New York. https://doi.org/10.1007/978-1-4939-1562-0_25

te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E. O. W. (2014). Are Headstart gains on the g factor? A meta-analysis. *Intelligence*, *46*, 209–215. https://doi.org/10.1016/j.intell.2014.07.001

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, *35*(3), 283–300. https://doi.org/10.1016/j.intell.2006.07.006

te Nijenhuis, J., Voskuijl, O. F., & Schijve, N. B. (2001). Practice and Coaching on IQ Tests: Quite a Lot of g. *International Journal of Selection and Assessment*, *9*(4), 302–308. https://doi.org/10.1111/1468-2389.00182

Terborg, J. R. (1977). Validation and extension of an individual differences model of work performance. *Organizational Behavior and Human Performance, 18*, 188 – 216.

Thurstone, L. L. (1969). *Primary Mental Abilities* (New impression edition). Univ. Chicago P.

Vance, R. J., MacCallum, R. C., Coovert, M. D., & Hedge, J. W. (1988). Construct validity of multiple job performance measures using confirmatory factor analysis. *Journal of Applied Psychology, 73(1)*, 74–80. https://doi.org/10.1037/0021-9010.73.1.74

Viswesvaran, C. (2002). Absenteeism and measures of job performance: A meta-analysis. *International Journal of Selection and Assessment, 10*, 12–17.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*(5), 557–574. https://doi.org/10.1037/0021-9010.81.5.557

Viswesvaran, C., Ones, D. S., Schmidt, F. L., Le, H., & Oh, I.-S. (2014). Measurement Error Obfuscates Scientific Knowledge: Path to Cumulative Knowledge Requires Corrections for Unreliability and Psychometric Meta-Analyses. *Industrial and Organizational Psychology*, *7*(4), 507–518. https://doi.org/10.1017/S1754942600006799

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, *87*(2), 345–354. https://doi.org/10.1037/0021-9010.87.2.345

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *The Journal of Applied Psychology*, *90*(1), 108–131. https://doi.org/10.1037/0021-9010.90.1.108

Watkins, M. W., Lei, P.-W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*(1), 59–68. https://doi.org/10.1016/j.intell.2006.04.005

Watkins, M. W., & Styck, K. M. (2017). A Cross-Lagged Panel Analysis of Psychometric Intelligence and Achievement in Reading and Math. *Journal of Intelligence*, *5*(3). https://doi.org/10.3390/jintelligence5030031

Witkowski, T. (2014, May 17). From the Archives of Scientific Fraud – Stephen Breuning. *Psychology Gone Wrong*. https://forbiddenpsychology.wordpress.com/2014/05/17/from-the-archives-of-scientific-fraud-stephen-breuning/

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences*, *116*(46), 23004-23010.