AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Applied Psychology

ISSN: 0021-9010

https://doi.org/10.1037/apl0001159

A Contemporary Look at the Relationship Between General Cognitive Ability and Job Performance

Paul R. Sackett, Saron Demeke, Isaac M. Bazian, Anne Marie Griebie, Reed Priest, and Nathan R. Kuncel
Department of Psychology, University of Minnesota, Minneapolis

The relationship between general cognitive ability (GCA) and overall job performance has been a long-accepted fact in industrial and organizational psychology. However, the most prominent data on this relationship date back more than 50 years. This meta-analysis examines the relationship between GCA and overall job performance using studies from the current century. Results across 153 samples and a total sample size of 40,740 show a mean observed validity of .16, with a residual *SD* of .09. Correcting for unreliability in the criterion and correcting predictive studies for range restriction produces a mean corrected validity of .22 and a residual *SD* of .11. While this is a much smaller estimate than the .51 value offered by Schmidt and Hunter (1998), that value has been critiqued by Sackett et al. (2022), who offered a mean corrected validity of .31 based on integrating findings from prior meta-analyses of 20th century data. We obtain a lower value (.22) for 21st century data. We conclude that GCA is related to job performance, but our estimate of the magnitude of the relationship is lower than prior estimates.

Keywords: cognitive ability, job performance, validity, meta-analysis

Supplemental materials: https://doi.org/10.1037/apl0001159.supp

The general cognitive ability (GCA)–overall job performance relationship is viewed as well-documented. Schmidt and Hunter (1998) have positioned GCA as the single best predictor of job performance. Their meta-analytic review of the validity of widely used predictors offered, for each predictor, its incremental validity over GCA. This frames GCA as the starting point; the value of other predictors is in terms of the increment provided. They report a mean operational validity (i.e., corrected for error of measurement in the criterion and for range restriction) of .51; the uncorrected value is .28.

However, recent developments are prompting a revisiting of this estimate of the validity of GCA. It is normative in the personnel selection field for a meta-analytic estimate of the mean predictor–criterion relationship to make use of three things: (a) an estimate of the mean observed validity, (b) an estimate of the mean criterion reliability, used to correct the observed validity estimate, and (c) an estimate of the mean amount of range restriction, used to further correct the validity estimate (Sackett et al., 2022).

Recent research has revisited two of these three. Regarding the second point, Zhou et al. (2022) noted that the most widely used mean interrater reliability estimate for supervisor ratings of job performance is the .52 value provided by Viswesvaran et al. (1996). That value is based on a meta-analysis of 42 studies, and the study did not consider job complexity as a moderator. Zhou et al.'s (2022) updated estimate is based on 136 studies and notes a marked difference in findings between supervisory/managerial jobs (mean reliability = .46) and nonsupervisor/managerial jobs (mean reliability = .61). Note that a lower reliability value produces a larger correction (e.g., an observed validity of .30 corrects to .44 if .46 is used as the reliability estimate and to .38 if .61 is used).

With respect to the third point, Sackett et al. (2022) reexamined the range restriction correction used to estimate operational validity for GCA in predicting overall job performance. They documented that (a) range restriction will generally be much smaller in concurrent validity studies (i.e., when the predictor is not used directly in screening) than in predictive studies (i.e., predictor used directly in screening), (b) the validity studies used in Schmidt and Hunter (1998) analysis were from settings where the predictor was not used in screening, and (c) the range restriction factor used in Schmidt and Hunter (1998) is not plausible in concurrent studies. Sackett et al.'s (2022) position has been challenged by Oh et al. (in press), who argue that conditions under which concurrent studies produce large range restriction are more common than Sackett et al. posit. Sackett et al. (2023) agree that it is possible to obtain sizable

Paul R. Sackett https://orcid.org/0000-0001-7633-4160

This research was supported by a grant from the Institute for Mental Chronometry to the University of Minnesota (Paul R. Sackett and Nathan R. Kuncel: principal investigators).

Paul R. Sackett played a lead role in conceptualization, writing–original draft, and writing–review and editing, a supporting role in formal analysis and software, and an equal role in data curation, funding acquisition, and project administration. Saron Demeke played a lead role in formal analysis and software, a supporting role in writing–review and editing, and an equal role in data curation. Isaac M. Bazian played a supporting role in formal analysis and writing–review and editing and an equal role in data curation. Anne Marie Griebie played a supporting role in formal analysis and writing–review and editing and an equal role in data curation. Reed Priest played a supporting role in formal analysis and writing–review and editing and an equal role in data curation. Nathan R. Kuncel played a supporting role in conceptualization, formal analysis, methodology, and writing–review and editing and an equal role in funding acquisition and project administration.

Correspondence concerning this article should be addressed to Paul R. Sackett, Department of Psychology, University of Minnesota, Minneapolis, Elliott Hall, 75 East River Parkway, Minneapolis, MN 55455, United States. Email: psackett@umn.edu

range restriction in concurrent studies but argue that is not plausible that these conditions make up the bulk of the cumulative meta-analytic data on validity. Combining Zhou et al.'s (2022) revised estimates of interrater reliability and Sackett et al.'s reassessment of range restriction results in a mean operational validity estimate for GCA of .31, considerably smaller than Schmidt and Hunter's (1998) mean value of .51.

The present article reexamines the first component of a meta-analytic validity estimate, namely the mean observed validity using GCA to predict overall performance, via a meta-analysis of studies reported between 2000 and 2021. We observe that the data used to estimate GCA–performance relationships are quite old. The Schmidt and Hunter (1998) summary, which is the go-to citation for GCA validity, relies solely on Hunter's (1983) evaluation of validity studies of a single measure (the GCA composite from the General Aptitude Test Battery [GATB]) conducted prior to 1972. Note that Schmidt and Hunter (1998) presented validity only for jobs of moderate complexity (Level 3 of a five-level job complexity scale); the mean observed validity across all jobs was .25 (Sackett et al., 2022). As mean validity for other predictors examined in both the Schmidt and Hunter (1998) and the Sackett et al. (2022) summaries of meta-analytic validity estimates across predictors are not limited by job type, we use an all-jobs estimate for GCA as well.

Beyond Schmidt and Hunter (1998), four additional meta-analyses of the GCA–job performance relationship have been reported. Hartigan and Wigdor (1989) analyzed 264 GATB validity studies more recently than those reported in Schmidt and Hunter (1998), namely from the 1970s and 1980s, reporting a mean observed validity of .21. Schmitt et al. (1984) analyzed 25 studies relating ability to supervisory ratings in the time period of 1964–1982, reporting a mean observed validity of .22. Bertua et al. (2005) reported a meta-analysis of 12 studies done in the United Kingdom, reporting a mean observed validity of .22. Despite the relatively recent publication date, all studies are pre-2000; in fact, 50% of the meta-analyzed studies are pre-1950. Salgado et al. (2003) reported a meta-analysis of European studies, reporting a mean observed validity of .29. Of 93 studies, only one study was post-2000, 19 were from the 1990s, and half were prior to 1960. So, our cumulative literature is quite old, and mean observed validity estimates range from .21 to .29.

Nye et al. (2022) presented a related meta-analysis. Their question was about the incremental validity of specific abilities over GCA in predicting task performance, organizational citizenship behavior, and counterproductive work behavior. However, unlike the work presented here, they did not examine overall job performance. In the spirit of the present work, they also focused on more recent research, examining studies between 1990 and the present.

Is there a conceptual reason to question whether prior findings still hold today? We believe it would be hard to offer any reason to expect that GCA is no longer related to performance. Prior work reports that GCA validity increases with job complexity, though useful relationships are found even for the least complex jobs (Hartigan & Wigdor, 1989). With the shift from a manufacturing economy to an information economy, there are conflicting predictions about the effect on the cognitive complexity of work in general. Advancements in technology have prompted three major hypotheses surrounding impacts on work demands: Increased automation will result in less skilled work by offloading major tasks to automated machinery (*deskilling*), advancements in technology will create new demands on workers and increase the overall complexity of work (*upgrading*), and last, technology will affect work in both directions and some occupations will be deskilled while others are upgraded such that occupations that fall in the midrange of skill complexity will decline (*polarization*). Due to the varying effects of technology across different dimensions of work, it is difficult to determine which hypothesis is best supported (National Academy of Sciences, 1999).

We do note two issues that could result in a difference between the earlier Schmidt and Hunter (1998) findings and more present studies. The first is that while modern conceptions of job performance view it as multifaceted, including task performance, organizational citizenship, and the avoidance of counterproductive work behaviors (Rotundo & Sackett, 2002), Sackett et al. (2017) documented that the GATB studies, which were used as the basis for the Schmidt and Hunter (1998) estimate, were exclusively measures of task performance. Conceptually, task performance would be expected to be more cognitively loaded than citizenship or counterproductive work behavior. A meta-analysis by Gonzalez-Mulé et al. (2014) supports this logic, finding much stronger relationships between GCA and task performance than between GCA and citizenship or counterproductive behavior. Thus, if newer research focuses on this broader conception of job performance, we would expect to see smaller relationships than with the prior narrower conceptualization of job performance as limited to task performance.

The second issue is that the nature of work has changed. There has been a large reduction in manufacturing jobs and an increase in jobs with public-facing and teamwork components, thus changing the nature of task performance. To the degree that core job tasks increasingly involve one-on-one (e.g., customer service) and team interactions, task performance takes on a larger interpersonal component. The argument here is not that cognitive ability is no longer important, but rather that the criterion space is broader, incorporating interpersonal skills to a larger degree, and lowering the portion of the criterion space that is driven by cognitive skills.

To illustrate change over time in the interpersonal demands of jobs, we compared an interpersonal skill score based on O*NET ratings collected between 2004 and 2021 of 15 generalized work activities in the interpersonal domain (Burrus & Way, 2017) with the single-item "People" rating from the *Dictionary of Occupational Titles* (*DOT*), the predecessor to O*NET, collected between 1977 and 1990. Across all occupations in the *DOT*, the mean People score was 1.61; across all occupations in O*NET, the mean interpersonal skills score was 3.16.[1] We interpret this as evidence of broad increases in the interpersonal skills requirements of work. Thus, to the degree that interpersonal skill increases as a variance component in task performance, and consequently, overall performance, the ability–performance correlation would be expected to decrease.

We suggest that revisiting the GCA–performance relationship is useful. After all, it is uncomfortable to rest one's conclusions on very old data. We propose a meta-analysis of studies reported in the

[1] O*NET contains a crosswalk to the *DOT* Data-People-Things ratings, and we found a correlation of .62 between the O*NET scale labeled Leading, Motivating, and Coordinating by Burrus and Way (2017) and the People score. Given that People is a single-item rating, we view this as a strong correlation between the two. As the O*NET score is on a 5-point low-to-high scale and People is on a 9-point high-to-low scale, we reverse-scored the People rating and transformed it to a comparable 5-point, 1–5 scale for purposes of comparison.

current century. With the large body of literature on the changing nature of work, we concluded that we wanted a contemporary estimate of validity. The move from the 20th to the 21st century seemed a natural break point, hence our decision to focus on 21st century studies. We believe that such an examination yields a validity estimate that is current and credible.

Our a priori expectation was that few researchers set out to study the relationship between GCA and performance, viewing that relationship as so well-established that new work is not a meaningful contribution. However, our expectation was that researchers examining other predictors (particularly novel predictors) may with considerable frequency include a GCA measure in their study to compare validity across predictors and to assess the incremental validity of a new predictor over GCA. Thus, while we searched broadly for studies, one aspect of our search strategy was to seek studies that focused on incremental validity of other predictors over GCA in predicting a job performance criterion, from which we extracted information about the validity of the GCA measure. Of the publicly available studies that we located, 69% were studies in which the incremental validity of another measure over GCA was examined.

Potential Moderator Variables

We examine eleven potential moderator variables, which we group into two categories. We first discuss five features that may serve as confounds, distorting our ability to gain insight into validity (e.g., studies of a single job vs. studies pooling data across multiple jobs). We then turn to six features that may identify sources of substantive differences in validity (e.g., use of overall performance vs. task performance as the criterion).

The Role of Study Designs Examining Incremental Validity

Readers of an earlier version of this work suggested that the extensive use of studies examining incremental validity may produce biased results. Their argument has several components. The first is that studies successful in finding incremental validity above and beyond GCA for an alternate predictor were more likely to be accepted for publication, while studies, where the new predictor of interest did not show incremental validity over GCA, would be seen as less of a contribution and thus be rejected for publication. The second component of the argument is that incremental validity for an alternate predictor would be more likely to be found in settings in which, due to sampling error alone, GCA validity was in the lower tail of the GCA validity distribution. Lower GCA validity makes it easier for a new predictor to show an increment in validity. If both aspects of this argument were true, then the set of studies available for meta-analysis would draw from the low end of the GCA validity distribution and thus underestimate GCA validity. In response to this concern, we examined two moderator variables. First, was incremental validity examined in the study or not? Second, among studies that examined incremental validity of an alternate predictor over GCA, was incremental validity found or not? Comparing GCA validity in these differing conditions will permit a response to the possibility that studies examining incremental validity produce biased results.

Overall Performance Versus Task Performance

We compare studies using a broader overall performance criterion with those focusing on a narrower conception of performance limited to task performance. One argument is that higher validity should be expected for studies using the narrower task performance criterion, as such a criterion excludes less cognitively loaded aspects of performance (e.g., citizenship, counterproductive work behavior). A contrary position is that with changes in work that add interpersonal and teamwork components to a great many jobs, task performance itself will include a variety of less cognitively loaded aspects of performance, resulting in no difference in validity for task performance criteria versus overall job performance criteria. Given these competing positions, we do not offer a specific hypothesis but pose a comparison between task performance and overall performance criteria as a research question.

Single Job Versus Data Pooled Across Jobs

We compare studies focusing on a single job with studies pooling data across multiple jobs. Pooling data are problematic if the jobs combined differ in their mean level on the predictor. Our sense is that supervisors evaluate performance relative to others doing the same job: Someone performing well on a less complex job is likely to receive the same performance rating as someone performing well on a more complex job. For example, the level of GCA needed to be an above-average computer programmer is higher than that needed to be an above-average data entry clerk. So, although high GCA is needed for high performance in some jobs, low ability can still yield high performance in other jobs. In short, if jobs differ in mean GCA but do not differ in mean performance rating, one will get a lower validity estimate if one pools the data and computes a single validity estimate than if separate estimates are computed within job and averaged. Ostroff (1993) offered a useful treatment of the effects of between-group and within-group variance on correlation coefficients.

We note that the database used for the classic Schmidt and Hunter (1998) estimate is exclusively single job. The data we compile here include a mixture of single- and multijob studies. We assert that a single-job estimate is really what the field wants. The question we are asking is "If multiple applicants present themselves for a given job, does GCA predict which of them will perform well and which will perform poorly?" Thus, we suggest it will be useful to differentiate single- and multijob validity studies, which is a distinction that we do not typically see drawn in meta-analyses of selection measures.

We do not offer a specific hypothesis about a difference in validity between single- and multijob studies, as an expectation of lower validity in multijob studies is contingent on predictor mean differences among the pooled jobs. We do not know whether the jobs pooled in our data set differ in predictor means, as only pooled means are reported. Thus, we view this as a research question. Should lower validity be observed in multijob studies, we would advocate reliance on the single-job studies.

Predictive and Concurrent Designs, With Applicants or Incumbents

We compare studies using predictive and concurrent designs, using applicants or incumbents. Sackett et al. (2022) showed that large restriction of range is less likely to occur in samples of current employees, where the predictor in question was not used for selection, but is potentially sizable in predictive studies with applicants, with the predictor used in selection. We note that the term "predictive" can refer either to applicant samples or to current employee samples where the criterion is obtained at a later point than that at which the predictor is administered to employees. This latter form of predictive design is also not likely to exhibit substantial range restriction, as the predictor was not used in selection. Thus, we code studies separately as predictive versus concurrent and applicant versus incumbent and hypothesize lower mean observed validity in predictive studies using applicants (Hypothesis 1).

Publicly Available Versus Consultant-Provided Samples

We compare publicly available studies (e.g., studies available via online search) with unpublished studies obtained by contacting members of the Society for Industrial and Organizational Psychology (SIOP) and consulting firms in the personnel selection domain. An initial version of our work was presented at an SIOP conference, focusing solely on publicly available studies (Griebie et al., 2022). The present version adds both additional public studies and the consultant-provided unpublished studies. We have no specific expectation as to differences in findings for public studies versus consultant-provided studies and report the comparison as an exploratory analysis.

Rating Versus Objective Criteria

We compare studies using supervisory ratings as criteria with those using objective measures. Prior meta-analytic work has reported higher validity for objective measures (Schmitt et al., 1984), and we hypothesize the same here (Hypothesis 2).

Managerial Versus Nonmanagerial Jobs

We compare studies using managerial samples with those using nonmanagerial samples. Prior meta-analytic work has reported higher validity for managerial jobs (Bertua et al., 2005: Salgado et al., 2003). This is consistent with work reporting higher validity for more cognitively complex jobs than for less complex jobs (e.g., Gutenberg et al., 1983). Thus, we hypothesize higher validity for managerial jobs than for nonmanagerial jobs (Hypothesis 3).

Cognitive, Interpersonal, and Physical Demands

Finally, we examine the relationship between job characteristics and validity. Prior research has focused on the cognitive demands of jobs and found higher validity for jobs higher in cognitive complexity (Gutenberg et al., 1983; Hunter, 1983). We hypothesize the same here (Hypothesis 4). While our a priori focus based on prior research was on cognitive demands, we also conducted exploratory analyses of interpersonal and physical demands as well, reflecting the "Data-People-Things" structure used in the *DOT*.

Method

Sample

Our goal was to locate studies from 2000 to 2021 that correlate a measure of GCA with a broad overall job performance criterion. Given that past meta-analyses treated measures of task performance as reflecting overall performance, we searched for studies using overall performance or task performance as a criterion to compare the two. We used four search strategies. First, we used Google Scholar to search for studies including the terms "cognitive ability," "job performance," and "incremental validity." This resulted in a set of 5,050 studies. Without the term "incremental validity," the search returned an unwieldy 1.6 million studies. Second, we searched the metaBUS database (Bosco et al., 2017) for the years 2000–2021 for studies including a measure of cognitive ability and a measure of job performance. This produced 86 samples. Third, we searched the ProQuest dissertation database for dissertations for the years 2000–2021 including a measure of cognitive ability and a measure of job performance. This produced 83 samples. Last, we solicited unpublished studies. We contacted the 306 individuals in the SIOP membership directory who listed personnel selection as an interest. We also examined the 2020 SIOP conference program and identified 28 consulting firms that advertised in the program. We examined the website of each firm and contacted all firms that reported engaging in selection system development and validation work. This outreach produced 34 samples.

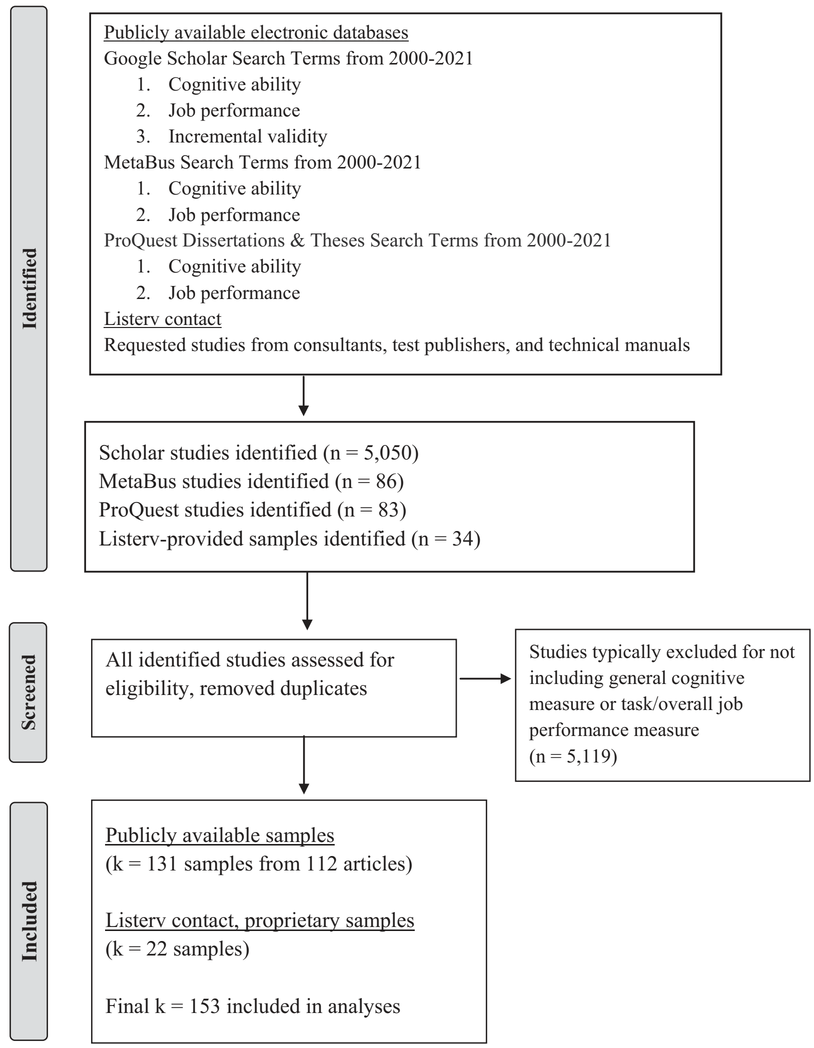
Many practitioners were willing to make study results and technical reports available to us only under the condition that the study not be identifiable, citing client's unwillingness to make findings public. We agreed to confidentiality and adopted a strategy of assigning each study a code (e.g., Consultant Study 1) rather than providing a complete citation for each study. This creates tension between two desirable attributes. On the one hand, anonymity made it possible for us to include these unpublished studies, thus increasing the scope of the meta-analysis. On the other hand, the technical reports are not available to readers of the meta-analysis, thus limiting their opportunity to review judgments we made about study features. To address this, we provide two separate sets of meta-analytic results: including and excluding these nonaccessible studies. Thus, readers believing that only publicly available studies should be included can focus only on findings from such studies.

These searches resulted in a total of 153 samples meeting the inclusion criteria below, with a total sample size of 40,740, a mean sample size of 266, and a median sample size of 153. The 153 samples were coded on the measures described below. Each sample was coded separately by two members of the project team, a consensus was made between coders, and coding was reviewed by the senior member of the team. The mean intraclass correlation coefficient, ICC(1) prior to consensus discussion was .90. The Appendix reports details on each study. Figure 1 contains a Preferred Reporting Items for Systematic Reviews and Meta Analyses diagram outlining the literature search procedures.

Inclusion Criteria

We included predictors that could be classified as measures of general, rather than specific, cognitive ability. Many studies used measures explicitly designed as measures of GCA (e.g., Wonderlic, Watson Glaser), which tap a range of specific abilities. Others used a composite of specific ability measures (e.g., the Armed Forces Qualification Test, made up of four subtests from the Armed Services Vocational Aptitude Battery). Yet others reported validity separately for several specific ability measures, in which case we composited the measures. We excluded studies using a measure of a single specific ability (e.g., numeric ability). We

Figure 1
Stages of Literature Search



excluded spatial, perceptual speed, specific domain knowledge, and highly speeded tests.

On the criterion side, we included studies with overall job performance or task performance as the criterion, either as rated by a supervisor or using an objective measure of work output. We excluded self, peer, or subordinate ratings, and studies targeting a single narrow facet of performance (e.g., citizenship, counterproductive behavior). We excluded studies targeting work outcomes other than performance (e.g., turnover, satisfaction, performance in training). We excluded studies where performance reflects a short time period (e.g., performing a job simulation for a day). We also excluded studies of performance in nonwork settings (e.g., academic performance) and on laboratory tasks.

We obtained a single GCA–performance correlation value from each sample. If multiple ability measures were used, we computed the correlation between a composite of the measures and the criterion when the needed information was available (i.e., the correlation between ability measures). Absent correlation among measures, we averaged the individual ability–criterion correlations. We used this same strategy in the case of multiple performance measures (e.g., ratings at multiple time periods). The use of a composite formula is clearly preferred, as averaging without taking the intercorrelation among predictors or criteria into account produces an underestimate of the composite validity. If a study reported validities for both task and overall job performance, we obtained both values and used a composite validity of overall performance in our analyses but do report the separate values for the particular analysis of the difference in validity between overall and task performance.

After applying these inclusion criteria, we had 153 samples available for analysis. Of these, 131 were publicly available, and 22 were obtained via our reachout to practicing psychologists and consulting firms. Twelve of the psychologists and consulting firms provided one or more studies, totaling four validity studies; 22 studies from eight of these respondents met our inclusion criteria. Four provided multiple useable studies (8, 5, 3, and 2); four provided a single study.

Test Reliability

To aid in comparisons with prior meta-analyses of GCA–performance relationships, it is useful to examine the reliability of the ability measures used. Fifty-five studies reported local internal consistency reliabilities, with a mean of .84 and *SD* of .06. Appendix tables the studies providing the reliability estimates. Of the prior meta-analyses, Schmitt et al. (1984) do not report reliability, Salgado et al. (2003) report a mean of .89 ($SD = 0.09$), and Bertua et al. (2005) report a mean of .85 ($SD = 0.05$). Thus, our set of studies appears roughly comparable to the studies in prior meta-analyses that contain data from multiple ability measures. Two prior analyses (Hartigan & Wigdor, 1989; Schmidt & Hunter, 1998) focus on a single test, namely the GATB. Using the Wang and Stanley (1970) formula for the reliability of a composite, we estimate GATB reliability as .93. Thus, modestly higher reliability is found for the GATB than for the typical measures used in other meta-analyses.

Measures

We coded the following variables for each sample:

N

We coded the sample size for each sample. If multiple correlations were combined with a single sample (e.g., correlations between ability and Year 1 and Year 2 performance), with differing sample sizes, we averaged the sample sizes.

Study Design

We coded each sample as predictive (test given at one point in time, with criteria collected at a later point in time) or concurrent (test and criteria obtained at the same point in time).

Study Participants

We coded each sample as administering tests to applicants or incumbents. Note partial overlap with the predictive/concurrent distinction above. All concurrent studies by definition involve incumbents, while predictive studies could be done with applicants or incumbents (e.g., test current incumbents and obtain criterion measures a year later).

Sample Type

We coded each sample as single- or multijob (e.g., a study reporting that it included "employees in a manufacturing organization" would be coded as multijob).

Criterion Type

We coded each study as relying on supervisory rating criteria or objective criteria.

Job Type

We coded the sample as managerial or nonmanagerial. Job titles including the terms "manager," "supervisor," or "executive" were coded as managerial.

Overall Versus Task Performance

For studies using rating criteria, we coded them as rating overall performance or task performance. We relied primarily on the label assigned by the study authors (e.g., "supervisor rating of overall performance," "objective measure of task performance"). In instances where authors reported criteria without labels, we examined the description and made a judgment. Settings where multiple dimensions of performance were rated were labeled overall performance if the set of dimensions was broad and wide-ranging and labeled task performance if focused on narrow facets of core job activities. Consensus among two coders was used, with a final review by a third coder.

Publicly Available Versus Technical Report From Consultant

Studies were coded as publicly available (i.e., obtainable via the internet) or as technical reports obtained through our reachout to practicing psychologists and to consulting firms.

Job Characteristics

As noted earlier, while our a priori hypothesis involved the relationship between the cognitive demands of a job and the validity of GCA measures, we also wanted to examine interpersonal and physical demands. We made use of scales developed by Burrus and Way (2017) that made use of 41 O*NET Generalized Work Activities. They conducted a principal components analysis of 41 O*NET Generalized Work Activities using as data the 117 most frequently held occupations in the United States. We replicated their analysis on the full set of occupations represented in O*NET. Four components were found, three of which we used here. The first is Working With Information, with 16 items loading (e.g., "analyzing data or information," "making decisions and solving problems," "organizing, planning, and prioritizing work"); coefficient α is .98. We used this scale as our measure of the cognitive complexity of a job. The second is Leading, Motivating, and Coordinating, with 15 items (e.g., "establishing and maintaining interpersonal relationships," "developing and building teams"); coefficient α is .97. The third is Manual and Physical Activities, with eight items (e.g., "controlling machines and processes," "handling and moving objects"); coefficient α is 93. These have conceptual correspondence to the Data, People, and Things scales used in the *DOT*. The fourth component, labeled Helping Others by Burrus and Way, has only two items and a coefficient α of .53; it is not examined further in our analysis.

Representativeness of Samples

Of interest is the question of the representativeness of the set of studies contributing to our meta-analytic database. We sought a measure that could be used to compare the samples in our data to the U.S. economy. We made use of the Burrus and Way (2017) Working With Information scale described above. We computed the mean and *SD* of this Working With Information scale for the 91 samples in our data reflecting a single job that could be matched to the O*NET database, as well as for the full set of 873 O*NET occupations. In our studies, we found a mean of 3.68 (on a

1–6 scale), with an *SD* of .72. In the full set of O*NET occupations, the mean was 3.85, with an *SD* of .80. As a result, we believe that our sample is reasonably representative of the world of work as a whole. In particular, the comparable *SD*s mean that tests of the relationship between information processing demands and validity are not affected by range restriction on the Working With Information scale.

Analysis

We conducted a psychometric meta-analysis, using the "psychmeta" package in R (Dahlke & Wiernik, 2019). The code used for analysis is available from the senior author. No studies reported interrater reliability for the performance criterion and thus we relied on artifact distributions compiled from prior studies of interrater reliability. Zhou et al. (2022) reported a mean interrater reliability of .46 and *SD* of .07 for managerial jobs and a mean interrater reliability of .61 and *SD* of .13 for nonmanagerial jobs; we apply these separate corrections to the subsets of managerial and nonmanagerial jobs, respectively. Ten studies reported the data needed for a range restriction correction (i.e., test standard deviations for applicants and incumbents, or corrected and uncorrected validity), permitting the computation of a *u* ratio (incumbent *SD*/applicant *SD*). We used nine of these to create an artifact distribution with a mean *u* ratio of .81 and an *SD* of .19. The tenth study was an extreme outlier, producing a *u* ratio of 1.78, which reflects the atypical finding of much more test score variance among those selected than in the applicant pool. That ratio is 5.1 *SD*s above the mean of the distribution obtained from other studies. We note that an earlier meta-analysis of cognitive ability–performance relationships by Salgado et al. (2003), reported *u* ratios for 30 studies, with none of these exhibiting range expansion (e.g., a *u* ratio greater than 1.0), further indicating the atypicality of this outlier. We applied this range restriction artifact distribution to studies using applicant samples; following Sackett et al. (2022), we did not make range restriction corrections for current employee samples for our main analysis. To document the effects of this methodological choice, we also repeated our overall meta-analysis applying our range restriction artifact distribution to all samples, both applicant and incumbent. Our main analysis uses the Case 2 range restriction correction, which has been the most widely used approach in meta-analyses in the selection field. This approach is a direct restriction correction, despite the expectation virtually all restriction is indirect, rather than direct. A Case 3 indirect correction requires knowledge of the correlation between the actual basis for selection and the predictor of interest, which is generally not known. Case 4 permits a correction for indirect range restriction without knowledge of this correlation, but it requires the restrictive assumption that the variable actually used for selection is only related to the criterion via its relationship with the predictor under investigation. Thus, we do not use Case 4 for our main analysis. However, to document the effects of this methodological choice, we repeated our overall meta-analysis applying Case 4 first only to applicant samples and then to all samples.

As different criterion reliability corrections were used for managerial and nonmanagerial samples, and different range restriction corrections were used for the applicant and current employee samples, we conducted four separate analyses and pooled the results to obtain our meta-analytic estimates (i.e., managerial jobs–applicant samples, managerial jobs–incumbent samples, nonmanagerial jobs–applicant samples, and nonmanagerial jobs–incumbent samples).

We also made use of metaregression. Analyses were conducted in R using the "metafor" and "mice" packages to account for missingness in moderators (Balduzzi et al., 2019; Van Buuren & Groothuis-Oudshoorn, 2011). We note that concerns about the use of metaregression have been raised (Schmidt, 2017), including low statistical power. Note that we use metaregression as a complement to, rather than a replacement for, the individual examination of moderators. As the majority of studies could be coded on all categorical moderators (131 out of 153), and as the number of studies was relatively large, we judged it useful to use metaregression to examine the effects of each potential moderator net of the others. Due to limited information in the 22 consultant samples, we could not determine whether these studies searched for or found incremental validity. Further, the three job characteristic variables (Working With Information; Leading, Motivating, and Coordinating; and Manual and Physical Activities) were available only for a subset of 91 single-job samples. Multiple imputation was used to deal with the missingness in these moderators and perform a metaregression on the full set of 153 studies. Using appropriate distributions for the moderator variables with missing information, 20 imputed data sets were generated. For each imputed data set, a mixed-effects metaregression model was fit with a restricted maximum likelihood approach; sampling variances were computed with the large-sample approximation using a weighted average of sample correlation coefficients. Estimates from these models were pooled into one final model. Values for three-level categorical incremental validity were imputed by polytomous regression. Missing O*NET information was imputed with predictive mean matching. Mixed-effects metaregression models were fit on the set of 20 imputed data sets; results were pooled across models following Rubin (2004).

We examined the potential effects of publication bias on the meta-analytic findings. Given the evidence from previous research that there is no single clear method to assess publication bias, we opted to evaluate bias in a series of analyses with multiple approaches and assessed their convergence (Carter et al., 2019). The following tests were used to evaluate publication bias: funnel plot asymmetry tests, test of excess significance, cumulative meta-analysis, and weighted average of the adequately powered studies (Ioannidis & Trikalinos, 2007; McDaniel, 2009; Stanley & Doucouliagos, 2014, 2017). All tests were conducted using the "metafor" and "meta" packages in R (Balduzzi et al., 2019; Viechtbauer, 2010).

Transparency and Openness

Below we describe our sampling plan, all data exclusions, and all measures in the study. We adhered to the *Journal of Applied Psychology* methodological checklist. The study's full data set is presented in the Appendix. Predictor and criterion means and *SD*s for all studies are included in the Supplemental Materials. Preregistration was not used.

Results

Overall results across 40,740 individuals and 153 independent samples yielded an observed mean correlation of .16 with performance, with a residual *SD* of .09 (see Table 1). Correcting for unreliability in the criterion and correcting predictive samples

Table 1
Meta-Analytic Correlations Between General Cognitive Ability and Job Performance

Variable	k	N	r_{obs}	SD_r	SD_{res}	95% CI_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	95% CI_ρ	80% CV_ρ
Overall	153	40,740	.16	0.11	0.09	[.14, .17]	.22	0.15	0.11	[.20, .23]	[.08, .35]
Incremental searched	91	22,568	.16	0.13	0.11	[.13, .18]	.21	0.18	0.16	[.18, .25]	[.01, .42]
Not found	24	4,526	.12	0.11	0.09	[.07, .16]	.16	0.17	0.14	[.10, .21]	[−.02, .33]
Found	67	18,042	.17	0.13	0.11	[.14, .20]	.23	0.17	0.14	[.19, .26]	[.05, .40]
Incremental not searched	40	13,758	.17	0.07	0.05	[.15, .19]	.24	0.13	0.08	[.21, .26]	[.14, .34]
Overall performance	110	23,555	.16	0.12	0.11	[.13, .18]	.22	0.19	0.14	[.19, .24]	[.03, .40]
Task performance	58	22,286	.17	0.09	0.08	[.14, .19]	.23	0.14	0.12	[.20, .26]	[.07, .38]
Multijob	55	17,514	.16	0.10	0.08	[.13, .18]	.22	0.16	0.12	[.19, .25]	[.06, .37]
Single job	98	23,226	.16	0.11	0.09	[.14, .18]	.22	0.17	0.12	[.19, .24]	[.06, .37]
Predictive	42	12,730	.13	0.09	0.08	[.10, .16]	.20	0.17	0.13	[.17, .24]	[.04, .37]
Applicant	26	10,220	.12	0.08	0.06	[.09, .15]	.20	0.16	0.12	[.15, .24]	[.04, .35]
Nonapplicant	16	2,510	.18	0.14	0.12	[.10, .25]	.23	0.18	0.13	[.17, .30]	[.07, .40]
Concurrent	111	28,010	.17	0.11	0.09	[.15, .19]	.22	0.15	0.10	[.20, .24]	[.09, .35]
Public	131	36,326	.16	0.11	0.09	[.14, .18]	.22	0.16	0.11	[.20, .24]	[.08, .37]
Consultant	22	4,414	.13	0.09	0.06	[.09, .17]	.17	0.14	0.05	[.15, .19]	[.11, .23]
Objective	10	889	.29	0.18	0.15	[.16, .41]	.38	0.29	0.25	[.22, .54]	[.05, .71]
Rating	143	39,851	.16	0.10	0.09	[.14, .17]	.21	0.15	0.10	[.20, .23]	[.08, .35]
Nonmanager	120	32,251	.16	0.11	0.09	[.14, .18]	.22	0.14	0.10	[.20, .24]	[.09, .35]
Manager	33	8,489	.14	0.11	0.09	[.10, .17]	.20	0.17	0.12	[.16, .24]	[.06, .35]

Note. k = number of independent samples; N = sample size; r_{obs} = observed sample size weighted mean correlations; SD_r = observed standard deviation of r; SD_{res} = residual standard deviation of r; 95% CI_r = 95% confidence interval around r; $\bar{\rho}$ = mean true-score correlation; SD_{r_c} = observed standard deviation of corrected correlations (r_c); SD_ρ = residual standard deviation of ρ; 95% CI_ρ = confidence interval around ρ; 80% CV_ρ = credibility interval around ρ. All correlations corrected using artifact distributions for unreliability in criterion. Applicant samples corrected for range restriction. For incremental searched versus not searched moderator, consultant studies are excluded. For overall/task performance, 15 samples provide both outcomes and are reported separately in this moderator analysis only. To maintain independence, the validity from these samples is collapsed as one overall performance outcome across other moderator rows and other analyses.

(but not concurrent samples) for range restriction produced a mean corrected correlation of .22, with a residual *SD* of .11. While we advocated the correction of only predictive studies for range restriction in these data, we repeated our analyses applying the range restriction artifact distribution to all samples, both predictive and concurrent. We present these analyses in Table 2. This analysis produced a mean corrected correlation of .25, with a residual *SD* of .13. We also repeated our analyses applying a Case 4 range restriction, rather than the Case 2 approach used in our primary analysis. Using Case 4 and correcting predictive samples (but not concurrent samples) for range restriction produced a mean corrected correlation of .22, with a residual *SD* of .11. Using Case 4 for all samples produced a mean corrected correlation of .26, with a residual *SD* of .13.

The Supplemental Material contains information comparable to the meta-analysis reported in Table 1, excluding the consultant studies that are not publicly available. In these 131 samples, we found an observed mean correlation of .16 with performance, with a residual *SD* of .09. Correcting for criterion unreliability and range restriction for the applicant sample yielded a mean corrected correlation of .22 with a residual *SD* of .11.

We first examined categorical moderators individually. Categorical moderators were generally in the hypothesized direction predicted by prior research, though differences were generally small. For all but one moderator (publicly available studies vs. consultant-provided studies), confidence intervals overlapped. Samples for which incremental validity was searched for, and found, produced ρ = .23, similar to a value of .24 produced by samples for which incremental validity was not examined. Samples for which incremental validity was searched for, but not found, produced lower mean correlations (ρ = .16), though confidence intervals overlapped. The use of overall performance ratings (ρ = .22) resulted in correlations comparable to those obtained using task performance ratings (ρ = .23). Samples that contained data across multiple jobs produced identical mean correlations as samples using data from single jobs (ρ = .22). On average, studies using concurrent designs produced larger average observed correlations (r = .17) than those with predictive designs (r = .13), consistent with Hypothesis 1, but correlations were comparable after correcting the predictive studies for range restriction (ρ = .21 for predictive studies and ρ = .22 for concurrent studies). Among studies using predictive designs, means were lower for applicant samples (ρ = .20) than for nonapplicant samples (ρ = .24). Publicly available samples produced higher correlations (ρ = .22) than consultant-provided samples (ρ = .17); this is the only categorical moderator for which confidence intervals did not overlap. Samples using objective criteria produced higher validity (ρ = .38) than samples using supervisory ratings (ρ = .21). While directionally consistent with Hypothesis 2, the confidence intervals did overlap. Managerial samples produced a lower mean correlation (ρ = .20) than did nonmanagerial samples (ρ = .22). This is in the opposite direction as posited by Hypothesis 2, but the confidence intervals did overlap.

A metaregression was conducted to examine the simultaneous effects of all of the categorical moderators and the three continuous job characteristics variables (see Table 3 for correlations among moderators Table 4 for means and *SD*s for the continuous job characteristics variables, and Table 5 for the metaregression results). We had hypothesized a relationship between the Working With Information scale derived from O*NET and validity but found a correlation of .00 with GCA validity, inconsistent with Hypothesis 4. Although our focus was on these Working With Information scales, based on prior work relating job characteristics to Graduate

Table 2
Overall Meta-Analysis Using Alternative Approaches to Range Restriction

Type of range restriction correction used	$\bar{\rho}$	SD_{r_c}	SD_{ρ}	95% CI_{ρ}	80% CV_{ρ}
Case 2 correction applied only to applicant studies	.22	0.15	0.11	[.20, .23]	[.08, .35]
Case 2 correction applied to all studies	.25	0.18	0.13	[.23, .27]	[.08, .42]
Case 4 correction applied only to applicant studies	.22	0.16	0.11	[.21, .24]	[.08, .36]
Case 4 correction applied to all studies	.26	0.18	0.13	[.24, .28]	[.09, .43]

Note. $k = 153$, $N = 40{,}740$. $\bar{\rho}$ = mean true-score correlation; SD_{r_c} = observed standard deviation of corrected correlations (r_c); SD_{ρ} = residual standard deviation of ρ; 95% CI_{ρ} = confidence interval around $\bar{\rho}$; 80% CV_{ρ} = credibility interval around $\bar{\rho}$. All correlations corrected using artifact distributions for unreliability in criterion and range restriction in predictor using either Case 2 or Case 4 corrections.

Management Admission Test validity, we also examined the O*NET-based factor scores based on Burrus and Way (2017), reflecting Leading, Motivating, and Coordinating and Manual and Physical Activities. A first finding of interest is a correlation of .82 between the O*NET Working With Information and Leading, Motivating, and Coordinating factors, indicating the strong tendency for these features to covary. In contrast, the Manual and Physical Activities factor had small correlations with both Working With Information (−.07) and Leading, Motivating, and Coordinating (−.14). The Leading, Motivating, and Coordinating scale did not correlate with GCA validity (−.08). The Manual and Physical Activities factor did show a statistically significant correlation with GCA validity (r = .22).

Net of other moderators, incumbent samples produce significantly higher observed validity than applicant samples, consistent with Hypothesis 1. Significantly higher validity was obtained using objective criteria rather than rating criteria, consistent with Hypothesis 2. Publicly available samples produce significantly higher validity than consultant-provided samples. Finally, higher validity was obtained for jobs higher in Manual and Physical Activities.

The results from the methods used to evaluate publication bias in the current data set do not suggest a strong threat of bias due to small-study effects. First, the funnel plot for all studies does not show large departures from symmetry, with studies in the sample roughly equally likely to have r values above and below the observed mean estimate of 0.16 (Figure 2). The studies are clustered near the top of the funnel plot and effect sizes appear heterogeneous. In this data set, most primary studies included large sample sizes (mean N = 266, Mdn = 153), resulting in more robust standard errors. Further, two versions of Egger's regression tests are nonsignificant (α = .05), suggesting no substantial asymmetry (precision-effect test: predicted using standard error, $z = 1.51$, $p = .13$; precision-effect estimate with standard error: predicted using variance, $z = 1.11$, $p = .27$). The estimated effect sizes from these methods, which adjust for large standard error and variance, were not significantly different from the unadjusted mean correlation (from precision-effect test, estimated $r = 0.14$, 95% CI [.08, .19]).

The contoured funnel plot does show that most studies resulted in significant effect sizes. We thus conducted a test of excess significance to assess publication bias. In the current data set, there were minimal differences between the number of expected (86.9) and observed (95 out of 153) significant findings and the difference between observed and expected is not statistically significant ($p = .09$).

In a cumulative meta-analysis using sample size as an indicator of precision, the estimated mean correlation is recomputed by the addition of each individual study, beginning with the largest study, and continuing in descending order of sample size (McDaniel, 2009). Our results showed that the estimated mean correlation only marginally drifts in the positive direction as smaller studies are added. By the time one includes the 60 largest studies, each with $N >$ 170 (cumulative N = 30,804), the estimated mean stabilizes at r = 0.16. As a follow-up analysis, we conducted a meta-analysis using the weighted average of the adequately powered studies. This analysis computes the meta-analytic estimates using only those studies with adequate statistical power to detect the initially estimated mean effect size. Simulation studies show this method can outperform standard random-effects model estimation in the presence of publication bias (Stanley & Doucouliagos, 2017). In the current data set, this analysis selected 41 samples with enough power to detect the observed mean effect size. After retaining only these studies, the recomputed effect size estimate is 0.16, which suggests that the initial analysis with all studies did not overestimate the mean estimate (0.16).

Overall, the publication bias methods applied to the current data set suggest minimal levels of bias due to small-study effects. One possible reason for this is that our literature search included multiple sources of unpublished findings, including dissertation, consultant-provided, and test publisher studies, which may decrease the likelihood of publication bias in our sample.

Discussion

The primary goal of this study was to examine the criterion-related validity of GCA measures using data from this century. Overall, GCA measures produce useful correlations with measures of job performance in the modern era, with a mean observed correlation of .16, and a mean corrected correlation of .22, with residual SD = 0.11. This is markedly lower than the .51 estimate produced by Schmidt and Hunter (1998), but more similar to the estimate of .31 produced by Sackett et al. (2022) based on revisiting prior meta-analyses. Following Sackett et al. (2022), we made range restriction corrections only for applicant samples. However, we reanalyzed our data in the traditional manner, applying the range restriction artifact distribution to all studies, both predictive and concurrent. This produced a mean corrected validity of .25, with residual SD = 0.13. While higher than our estimate of .22, it remains markedly lower than prior estimates based on older data. The above findings are based on using a Case 2 range restriction correction;

Table 3
Correlations Among Moderator Variables

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Incremental searched, not found (1) versus searched, found (2)	—										
2. Overall (0) versus task performance (1)	.09 [−.12, .29]	—									
3. Multiple (0) versus single job (1)	.07 [−.14, .27]	−.13 [−.28, .03]	—								
4. Concurrent (0) versus predictive (1)	−.03 [−.23, .18]	−.03 [−.19, .12]	.03 [−.13, .19]	—							
5. Applicant (0) versus incumbent (1)	.01 [−.21, .21]	.10 [−.06, .25]	−.05 [−.21, .11]	−.74 [−.80, −.65]	—						
6. Consultant (0) versus public (1)	—	.18 [.02, .33]	−.15 [−.30, .01]	.00 [−.16, .16]	−.04 [−.19, .12]	—					
7. Objective (0) versus rating (1)	−.08 [−.28, .13]	−.36 [−.49, −.21]	−.09 [−.24, .07]	−.13 [−.29, .03]	−.05 [−.21, .11]	−.11 [−.26, .05]	—				
8. Manager (0) versus nonmanager (1)	.06 [−.15, .26]	.12 [−.04, .28]	.10 [−.06, .26]	.00 [−.16, .16]	−.03 [−.18, .13]	.06 [−.10, .21]	.05 [−.11, .21]	—			
9. O*NET Working With Information	.04 [−.30, .25]	.00 [−.20, .21]	—	.11 [−.10, .31]	−.23 [−.42, −.03]	−.10 [−.30, .11]	.07 [−.14, .27]	−.28 [−.46, −.08]	—		
10. O*NET Leading, Motivating, and Coordinating	.04 [−.24, .31]	−.01 [−.21, .20]	—	.08 [−.13, .28]	−.21 [−.40, −.00]	−.10 [−.30, .11]	.04 [−.17, .24]	−.63 [−.74, −.49]	.82 [.74, .88]	—	
11. O*NET Manual and Physical Activities	.07 [−.21, .34]	−.13 [−.33, .07]	—	.14 [−.07, .34]	−.16 [−.36, .04]	−.20 [−.39, .01]	.20 [−.00, .39]	.04 [−.17, .24]	−.07 [−.27, .14]	−.14 [−.34, .06]	—

Note. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). Correlation between consultant/public and incremental searched/not searched is null. Due to limited information in consultant-provided studies, it was not possible to determine whether these studies examined incremental validity. Correlations between O*NET characteristics and single versus multiple jobs are null. Only single-job articles have O*NET information.

using Case 4 instead produced the same findings as Case 2 when corrections were made only to applicant samples (i.e., mean corrected correlation of .22, with residual $SD = 0.11$). Applying Case 4 to all samples raised the mean corrected correlation to .26, with a residual *SD* of .13. Thus, while methodological choices had an effect on the results, meta-analytic mean correlations remain lower than those obtained in prior research regardless of the choices made.

Earlier, we noted a meta-analysis by Nye et al. (2022) that focused on the incremental validity of specific abilities over GCA. They focused on task performance, rather than overall performance. However, as we found no meaningful difference in GCA validity using overall performance versus task performance as a criterion, we revisited Nye et al. (2022). They report a mean corrected validity of .23 for GCA against a task performance criterion, comparable to the value of .22 produced in our study.

Our study and Nye et al. (2022) both focus on newer studies. It is possible that temporal changes that we observed here for GCA validity may be found for other predictors as well. This raises a broader issue about comparing meta-analytic results that differ in the time period examined. Updating meta-analytic information in various domains is warranted.

We believe these findings have important practical implications. They suggest a more modest role for GCA ability in many settings. We do emphasize that our work focuses on overall job performance as the criterion. Settings with extensive training programs prior to moving onto the job may be examples of scenarios in which different criteria (e.g., training performance) are of great interest, and our work does not speak to GCA as a predictor of training performance. However, broad statements that GCA is our best predictor of job performance are not consistent with what we find here. We are not arguing against the use of measures of GCA, but rather for consideration of a range of possible predictors for a given setting.

We note that the mean validity is accompanied by a sizeable residual *SD* of .11 Thus, validity in a given setting can be considerably larger—or smaller—than this mean value. We view this as a finding broadly consistent with findings for other predictors of job performance. Comparable or larger residual *SD*s are found for predictors such as structured interviews, situational judgment tests, knowledge tests, integrity tests, and personality measures (Sackett et al., 2022). As Sackett et al. (2022) note, a focus on mean validity has led to an underemphasis on the variation in validity across settings. Additional work is needed to shed light on features predictive of higher versus lower validity values for a given setting.

Moderator Variables

We examined potential variables both individually (e.g., comparing mean validity by moderator category) and net of other variables in metaregression. The result is a nuanced set of findings, with significant moderator effects found primarily in the metaregression. Of interest was the role of incremental validity in each study. Some studies were designed to examine predictors other than GCA and examined whether the predictor of interest showed incremental validity over GCA. As outlined in the introduction, some readers had suggested that studies that did find incremental validity over GCA were more likely to be published, and finding incremental validity over GCA would be more likely in studies where, due to sampling error, the GCA validity estimate was in the

Table 4
*Means and Standard Deviations of O*NET Characteristics*

Variable	*M*	*SD*
O*NET Working With Information	3.68	0.72
O*NET Leading, Motivating, and Coordinating	3.08	0.89
O*NET Manual and Physical Activities	2.01	1.08

Note. k = 91, subset of single-job samples with O*NET coded job information.

lower tail of the GCA validity distribution. The result would be an underestimate of GCA validity. In response to these concerns, we sorted studies into three categories: incremental validity examined and found, incremental validity examined and not found, and incremental validity not examined. Critically, no meaningful differences in validity were found between the three categories, with all confidence intervals overlapping. If anything, there are suggestions of an opposite pattern than proposed: Studies that examined and found incremental validity had a higher mean validity than studies that did not find incremental validity.

We examined validity for samples using overall performance as a criterion and for samples using solely task performance. Here, we found comparable mean validities for overall and task performance. We note that we do not have detailed information about the domain covered by overall performance ratings and task performance ratings in most studies and thus cannot fully explain this finding. On conceptual grounds, we retain our expectation that a broad overall performance measure, incorporating task performance, citizenship, and counterproductive work behavior will produce a lower correlation with cognitive ability than a measure focused solely on task performance. Prior meta-analytic findings comparing the correlation between ability and the task, citizenship and counterproductive behavior domains show markedly lower correlations for citizenship and counterproductive behavior. Subsequently, a composite of task performance, citizenship, and counterproductive behavior would likely result in a lower correlation with ability than a measure of task performance alone (Gonzalez-Mulé et al., 2014). However, it is not clear how overall performance was operationalized in the studies we examine here. Without explicit instruction as to which features to consider when making an overall performance rating, raters may construe performance narrowly as the task performance domain.

Table 5
Metaregression Results for Moderators

Variable	Estimate	*SE*	*t*	*p*
Intercept	−0.04	0.13	−0.34	.73
Incremental not searched (0) versus searched, did not find (1)	−0.06	0.03	−1.70	.09
Incremental not searched (0) versus searched, found (1)	−0.00	0.03	−0.05	.96
Overall performance (0) versus task (1)	−0.01	0.03	−0.44	.66
Concurrent (0) versus predictive (1)	0.03	0.04	0.82	.42
Applicant (0) versus incumbent (1)	0.11	0.05	2.44	.02
Consultant (0) versus public (1)	0.08	0.03	2.48	.01
Objective (0) versus rating (1)	−0.12	0.05	−2.23	.03
Manager (0) versus nonmanager (1)	0.02	0.05	0.32	.75
O*NET Working With Information	0.03	0.04	0.79	.44
O*NET Leading, Motivating, and Coordinating	−0.01	0.04	−0.22	.82
O*NET Manual and Physical Activities	0.04	0.01	3.13	.00

Note. k = 153, all samples. Estimate = pooled estimated coefficients; *SE* = pooled standard errors of the coefficient estimates; t = test statistics for each model computed using Knapp and Hartung (2003) method adjusting for amount of residual heterogeneity then pooled; p = corresponding pooled p values. Single- /multijob moderator excluded.

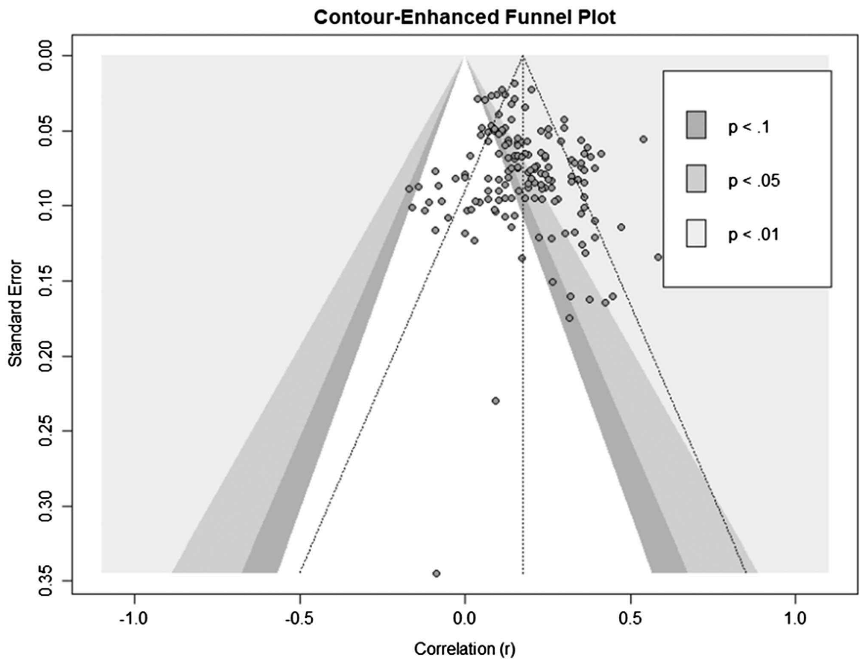
We also found no meaningful difference in validity for studies using a single job versus pooling data across jobs. Note that our conceptual argument was that pooling data across jobs with meaningful differences in mean test performance would result in lower validity if criterion means were comparable across the jobs. However, we do not know whether the multiple jobs pooled in our database differed in mean test scores. We do continue to caution against pooling data when predictor means do differ.

We found no meaningful differences in validity between managerial and nonmanagerial samples or between jobs using objective performance measures versus ratings as criteria. Relatively small numbers of studies using managerial samples and objective performance measures may have hindered attempts to evaluate these moderators.

Applicant status was the most clearly expected moderator, as it differentiates samples for which substantial range restriction is possible from samples where it generally is not. We expect less restriction in incumbent samples. We note that the term "predictive" is most commonly applied to applicant studies but is also used for current employee samples in which there is a time lag between testing employees and obtaining criteria. We view this as a source of confusion in interpreting prior research. For example, the GATB studies examined by Hartigan and Wigdor (1989) differentiate between predictive and concurrent samples, finding no difference in validity. However, all of the GATB studies made use of current employee samples, and the GATB was not used in selection. Thus, limited range restriction was expected in both predictive and concurrent samples. Schmitt et al. (1984) highlighted that the key distinction is between applicant samples, where the predictor of interest can be used for selection, and current employee samples, where it cannot. Thus, we examine applicant versus incumbent and predictive versus concurrent validation strategy as separate variables. As hypothesized, significantly lower mean observed validity was observed for applicant samples in our metaregression. See Barrett et al. (1981) and Guion and Cranny (1982) for useful treatments of comparing predictive and concurrent studies.

We note that predictive and concurrent designs also differ in other features. One is the concern that test-taking motivation may be lower in concurrent studies, as the incentive for performing well is much lower in a research setting than when seeking a job. But this would lead to a prediction of lower validity in concurrent studies, rather than the higher validity that is observed. Another is the passage of time: Conceptually, one might argue for higher validity when test and performance are measured concurrently. This concern is more pressing for predictors where a change in one's standing on the construct of interest is more readily changed via study and experience (e.g., measures of job knowledge). Given the relatively high stability of GCA in adult samples, at least in the short term, we do not view a gap of, say, a year between administering a GCA measure and obtaining a performance measure as a likely

Figure 2
Contour-Enhanced Funnel Plot



Note. The vertical dashed line is the estimated mean correlation (.16). Comparison with contours (p value bands) does not suggest a systematic selection of studies at $p < .05$ or $p < .01$.

explanation for differences in findings between predictive and concurrent studies in the GCA domain.

Although we had no strong a priori expectations, we found lower mean validity for consultant-provided studies than for publicly available studies, both in our categorical moderator examination and metaregression. We do note that comments on a prior version of this article suggested that higher validity would be found in consultant-provided studies, as many consultants conduct sizable numbers of validity studies for clients and thus are more skilled in doing so than, for example, a student conducting a study as part of a dissertation.

Prior research has found higher validity for jobs higher in cognitive complexity. Contrary to our hypothesis, we found no evidence for this in our data. The Working With Information scale derived from O*NET correlated .00 with validity. Our first suspicion was that perhaps the sample of jobs in our meta-analytic database was range restricted in terms of cognitive complexity. However, as noted earlier, we found comparable means and *SD*s for the Working With Information Scale for our data and for the full set of occupations in O*NET. This is a provocative finding that merits further exploration. Prior work, relying heavily on the GATB database, was based on samples with a very high representation of manufacturing jobs, yet only 9% of the samples in our database are manufacturing jobs. This is consistent with the move away from a manufacturing economy; perhaps the broader range of job types also entails a change in the relationship between cognitive complexity and validity.

Finally, contrary to our expectations, we found no differences in validity for managerial and nonmanagerial samples. We did find higher validity for samples using objective criteria, with the effect significant in the metaregression including job characteristics.

Possible Factors Contributing to the Lower Validity Estimate

We now discuss a number of possible contributing factors to the lower validity of these new studies. Potential factors include differences in the jobs studied, the criterion used, the tests used, and the candidate pools examined. We discuss each of these in turn.

First, the criteria used in prior work were limited to a narrow conceptualization of task performance, with a focus on quantity and quality of work accomplished. Much has been written about the changing nature of work, with working in team settings and coordinating efforts with others now widespread (e.g., Ilgen & Pulakso, 1999). This suggests a broadening of the definition of performance to include effectiveness in these domains. Note that this moves a variety of interpersonal factors, such as effectiveness in working as a member of a team, into the task performance domain. Thus, we suggest that an explanation for the lower validity of cognitive ability measures is the broadening of the performance space to include interpersonal skills, thus diminishing the portion of the performance domain predicted by ability. Also, as noted above, prior meta-analytic evidence for cognitive ability tests relies heavily on data from manufacturing jobs, which make up only a small portion of our current database. Only 9% of studies in our database are from manufacturing jobs. To the extent that manufacturing jobs have, on average, lower demands for interpersonal skills, this change in the representation of manufacturing jobs would contribute to the difference in validity.

Second, the studies we obtained used a mix of cognitive measures which were, on average, less reliable (mean internal consistency reliability = .84, *SD* = 0.06) than the highly reliable GATB (coefficient α = .93). Using the correction for attenuation, we estimate that our corrected mean validity would rise from .22 to .23 if the mean reliability in our set of studies had been comparable to that of the GATB.

We examined the possibility that the nature of the tests used has changed, perhaps not capturing GCA as well as well-established tests. We examined the cognitive ability measures used in all samples and sorted the studies into two categories: those using established or legacy measures like the Wonderlic and Watson Glaser ($k = 97$, mean observed $r = 0.16$) and those using newer measures ($k = 56$, mean observed $r = 0.15$). The confidence intervals overlapped, indicating no meaningful differences between the kinds of tests used in the set of studies available to us.

Third, there is the possibility that applicant pools have changed, such that the range of ability among individuals presenting themselves for consideration for a specific job is narrower now than in the past. There is some limited evidence in support of this. Lee and Steel (2019) report that, relative to broad national norms, job-specific applicant pools using the Wonderlic are more restricted now than was seen in earlier work by Sackett and Ostgaard (1994). Such narrower range applicant pools would result in lower validity estimates, all else equal.

Last, there are publication bias and file drawer issues with the initial set of 515 GATB validity studies analyzed by Hunter (1983). The U.S. Employment Service set out to develop job-specific test aptitude batteries for as many occupations as possible. The manual for the GATB (U.S. Employment Service, 1970) details the process for doing so. Four criteria contributed to identifying relevant aptitudes for an occupation: (a) a higher mean than other aptitudes, (b) a lower *SD* for the occupation than for other occupations, (c) significant criterion-related validity, and (d) high job analyst ratings of the importance of the aptitude. If this process succeeded in finding relevant aptitudes, the specific battery was published as a technical report. Importantly, Bemis (1968) reported that in about 10% of cases, the process did not succeed, and the battery was not published. We note that the GATB research was done in the era in which the role of sampling error was not widely recognized, and relatively small sample sizes were common in validation research. In fact, the GATB manual notes a target N of at least 50 when initiating a project to develop an occupation-specific aptitude battery. Thus, due to sampling error alone, it would not be uncommon to obtain both lower and nonsignificant validity estimates. To the extent that low validity values contributed to a study not being published, the mean validity based on published studies is an overestimate.

We note that this is not an issue with the newer set of 264 validity studies analyzed by Hartigan and Wigdor (1989), as that set of studies did not rely on the U.S. Employment Service publication process. Rather, raw data from all of these studies were made available to Hartigan and Wigdor (1989). They report being unable to find an explanation for the lower mean validity of .21 in the newer studies, relative to the mean of .25 in the older studies, but they appear unaware of the file drawer issue we identify here.

Conclusion

In conclusion, the goal of this study was to determine whether GCA measures continue to correlate with job performance using data from this century. Results across 153 studies and a total sample size of 40,740 show a mean observed validity of .16, with a residual *SD* of .09. Correcting for unreliability in the criterion and for range restriction produces a mean corrected validity of .22 and a residual *SD* of .11. We conclude that GCA is related to job performance, but our estimate of the magnitude of the relationship is lower than prior estimates. We note that GCA is also commonly used as a predictor of criteria other than overall job performance, such as performance in a training setting, and our findings are limited to the role of GCA in predicting overall performance.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Al Ali, O. E., Garner, I., & Magadley, W. (2012). An exploration of the relationship between emotional intelligence and job performance in police organizations. *Journal of Police and Criminal Psychology*, *27*(1), 1–8. https://doi.org/10.1007/s11896-011-9088-9

*Alexander, S. G. (2007). *Predicting long term job performance using a cognitive ability test* (Order No. 3288237) [Doctoral dissertation, University of North Texas]. ProQuest Dissertations & Theses Global.

Allen, J. (2016). *Conceptualizing learning agility and investigating its nomological network* [Doctoral dissertation]. Florida International University.

*Allworth, E., & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment*, *8*(3), 137–147. https://doi.org/10.1111/1468-2389.00142

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, *66*(1), 1–6. https://doi.org/10.1037/0021-9010.66.1.1

*Bedford, C. L. (2011). *The role of learning agility in workplace performance and career advancement* (Order No. 3465093) [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations & Theses Global.

Bemis, S. E. (1968). Occupational validity of the general aptitude test battery. *Journal of Applied Psychology*, *52*(3), 240–244. https://doi.org/10.1037/h0025733

*Benavidez, J. E. (2005). *Expanding the predictor and criterion space to reduce adverse impact in a public sector environment* (Order No. 3163310) [Doctoral dissertation, University of Oklahoma]. ProQuest Dissertations & Theses Global.

*Benjamin, K. (2006). *A criterion validation of the New Zealand Army Officer Selection Board: A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Psychology at Massey University* [Doctoral dissertation]. Massey University. https://mro.massey.ac.nz/handle/10179/10536

*Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, *14*(3), 223–235. https://doi.org/10.1111/j.1468-2389.2006.00345.x

*Bergner, S. (2020). Being smart is not enough: Personality traits and vocational interests incrementally predict intention, status and success of leaders and entrepreneurs beyond cognitive ability. *Frontiers in Psychology*, *11*, Article 204. https://doi.org/10.3389/fpsyg.2020.00204

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*(3), 387–409. https://doi.org/10.1348/096317905X26994

*Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. R. (2009). Construct and criterion-related validation of a measure of emotional reasoning skills: A two-study investigation. *International Journal of Selection and Assessment*, *17*(1), 101–118. https://doi.org/10.1111/j.1468-2389.2009.00455.x

*Bosco, F., Allen, D. G., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology*, *68*(4), 859–898. https://doi.org/10.1111/peps.12099

Bosco, F. A., Uggerslev, K. L., & Steel, P. (2017). MetaBUS as a vehicle for facilitating meta-analysis. *Human Resource Management Review*, *27*(1), 237–254. https://doi.org/10.1016/j.hrmr.2016.09.013

*Brantley, L. B. (2000). *Predicting job performance: An integrated model* (Order No. 9968804) [Doctoral dissertation, University of South Florida]. ProQuest Dissertations & Theses Global.

Burrus, J., & Way, J. (2017). Using O*NET to develop a framework of job characteristics to potentially improve the predictive validity of personality measures. *Personnel Assessment and Decisions*, *3*(1). Article 3. https://doi.org/10.25035/pad.2017.003

*Callans, M. C., Nguyen, D., & Wells, B. M. (2016). *Incremental validity of the wonderlic Motivation Potential Assessment (MPA)*. https://businessdocbox.com/Human_Resources/73010291-Case-study-incremental-validity-of-the-wonderlic-motivation-potential-assessment-mpa.html

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

*Cavazotte, F., Moreno, V., & Hickmann, M. (2012). Effects of leader intelligence, personality and emotional intelligence on transformational leadership and managerial performance. *The Leadership Quarterly*, *23*(3), 443–455. https://doi.org/10.1016/j.leaqua.2011.10.003

*Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*(3), 233–254. https://doi.org/10.1207/S15327043HUP1503_01

*Childers, O. K. (2016). *Biodata: A thing of the past? Examining the predictive validity and user reactions of rationally-selected, empirically keyed biodata* [Doctoral Dissertation]. University of Houston. https://uh-ir.tdl.org/handle/10657/3251

*Christiansen, N. D., Janovics, J. E., & Siers, B. P. (2010). Emotional intelligence in selection contexts: Measurement method, criterion-related validity, and vulnerability to response distortion. *International Journal of Selection and Assessment*, *18*(1), 87–101. https://doi.org/10.1111/j.1468-2389.2010.00491.x

*Cober, R. T., Cober, A. T., Lawrence, A. D., & O'Connell, M. S. (2003). *Predictors of multi-tasking ability for selection: Attitudes versus ability* [Conference session]. Society for Industrial and Organizational Psychology, Orlando, FL, United States.

*Connolly, J. J. (2001). *Assessing the construct validity of a measure of learning agility* (Order No. 3013189) [Doctoral dissertation, Florida International University]. ProQuest Dissertations & Theses Global.

*Côté, S., & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, *51*(1), 1–28. https://doi.org/10.2189/asqu.51.1.1

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

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

*David, E. M., & Witt, L. A. (2009). The effects of organizational citizenship behavior and general mental ability on task performance. *Revista de Psihologie Aplicată*, *11*(1), 7–13.

*Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, *55*(4), 827–849. https://doi.org/10.1111/j.1744-6570.2002.tb00131.x

*Dhliwayo, P., & Coetzee, M. (2020). Cognitive intelligence, emotional intelligence and personality types as predictors of job performance: Exploring a model for personnel selection. *SA Journal of Human Resource Management*, *18*, Article 1348. https://doi.org/10.4102/sajhrm.v18i0.1348

*Doucet, L., Shao, B., Wang, L., & Oldham, G. R. (2016). I know how you feel, but it does not always help: Integrating emotion recognition, agreeableness, and cognitive ability in a compensatory model of service performance. *Journal of Service Management*, *27*(3), 320–338. https://doi.org/10.1108/JOSM-11-2014-0307

Ejiogu, K. C., Rose, M., Yang, Z., & Trent, J. (2007). *Incremental validity of numerical reasoning over critical thinking* [Data set]. American Psychological Association. https://doi.org/10.1037/e680172007-001

*Ejiogu, K. C., Yang, Z., Trent, J., & Rose, M. (2006). *Understanding the relationship between critical thinking and job performance* [Poster presentation]. Society for Industrial and Organizational Psychology, Dallas, TX, United States.

*Faura, L. (2016). *Getting talent that fits: (Wm+g+h^2=performance)* (Order No. 10190159) [Doctoral Dissertation, Oklahoma State University]. ProQuest Dissertations and Theses. https://www.proquest.com/docview/1841275336/abstract/C18A976D45EA4A80PQ/1

*Ferris, G. R., Witt, L. A., & Hochwarter, W. A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology*, *86*(6), 1075–1082. https://doi.org/10.1037/0021-9010.86.6.1075

*Fertig, S. (2009). *The incremental validity of a Situational Judgement Test (SJT) relative to personality and cognitive ability to predict managerial performance* [Master's Thesis]. University of Stellenbosch. https://scholar.sun.ac.za:443/handle/10019.1/1769

*Fetzer, M. S. (2004). *An examination of the construct validity, criterion - related validity, and adverse impact of the Cognitive Behavior Inventory (CBI)* (Order No. 3165226) [Doctoral dissertation, The University of Southern Mississippi]. ProQuest Dissertations & Theses Global.

*Flannery, N. M. (2020). *Investigating the convergent, discriminant, and predictive validity of the mental toughness situational judgment test* [Doctoral dissertation]. Virginia Tech. https://vtechworks.lib.vt.edu/handle/10919/99062

*Fort, T. R. (2010). *Personality predictors of job performance: Implicit association tests versus self-report measures* (Order No. 3427327) [Doctoral dissertation, Walden University]. ProQuest Dissertations & Theses Global.

*Goldstein, H. W., Yusko, K. P., & Nicolopoulos, V. (2001). Exploring black–white subgroup differences of managerial competencies. *Personnel Psychology*, *54*(4), 783–807. https://doi.org/10.1111/j.1744-6570.2001.tb00232.x

Gonzalez-Mulé, E., Mount, M. K., & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology*, *99*(6), 1222–1243. https://doi.org/10.1037/a0037547

*Green, T., & Macqueen, P. (2008). *Cognitive ability: How important*. Compass Consultants. http://www.compassconsulting.com.au/icms_docs/31265_Cognitive_Ability_How_Important.pdf

Griebie, A., Bazian, I. M., Demeke, S., Priest, R., Sackett, P. R., & Kuncel, N. R. (2022). *A contemporary look at the relationship between general cognitive ability and job performance* [Paper presentation]. Society for Industrial and Organizational Psychology Conference, Seattle, WA, United States.

*Grim, A. M. (2010). *Use of situational judgment test to measure individual adaptability in applied settings* [Master's Thesis]. George Mason University. http://mars.gmu.edu/handle/1920/5793

Guion, R. M., & Cranny, C. J. (1982). A note on concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, *67*(2), 239–244. https://doi.org/10.1037/0021-9010.67.2.239

*Gunter, J. (2010). *How do situational judgments [sic] tests and situational interviews compare? An examination of construct and criterion-related validity* [Doctoral dissertation, University of Central Florida]. Electronic Theses and Dissertations. https://stars.library.ucf.edu/etd/1615

Gutenberg, R. L., Arvey, R. D., Osburn, H. G., & Jeanneret, P. R. (1983). Moderating effects of decision-making/information-processing job dimensions on test validities. *Journal of Applied Psychology*, *68*(4), 602–608. https://doi.org/10.1037/0021-9010.68.4.602

*Harlaar, M. (2013). *The validation of a video-based situational judgment test for the selection of call center employees* [Master's thesis]. Vrije Universiteit Amsterdam.

*Harris, A. M. (2017). *The interactive effects of personality and general mental ability on performance: revisiting theory and methodology* [Master's thesis]. University of Georgia.

Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. National Research Council.

*Harzer, C., Bezuglova, N., & Weber, M. (2021). Incremental validity of character strengths as predictors of job performance beyond general mental ability and the big five. *Frontiers in Psychology*, *12*, Article 518369. https://doi.org/10.3389/fpsyg.2021.518369

*Hattrup, K., O'Connell, M. S., & Labrador, J. R. (2005). Incremental validity of locus of control after controlling for cognitive ability and conscientiousness. *Journal of Business and Psychology*, *19*(4), 461–481. https://doi.org/10.1007/s10869-005-4519-1

*Hausdorf, P. A., & Risavy, S. D. (2015). Predicting training and job performance for transit operators. *International Journal of Selection and Assessment*, *23*(2), 191–195. https://doi.org/10.1111/ijsa.12107

*Hawes, S. R. (2000). *A comparison of biodata, ability, and a conditional reasoning test as predictors of reliable behavior in the workplace* (Order No. 9996356) [Doctoral dissertation, The University of Tennessee]. ProQuest Dissertations & Theses Global.

*Henderson, N. D. (2010). Predicting long-term firefighter performance from cognitive and physical ability measures. *Personnel Psychology*, *63*(4), 999–1039. https://doi.org/10.1111/j.1744-6570.2010.01196.x

*Hochwarter, W. A., Witt, L. A., & Kacmar, K. M. (2000). Perceptions of organizational politics as a moderator of the relationship between conscientiousness and job performance. *Journal of Applied Psychology*, *85*(3), 472–478. https://doi.org/10.1037/0021-9010.85.3.472

*Hodge, R. (2010). *Issues in validity generalization the criterion problem* [Doctoral dissertation, University of Central Florida]. Electronic Theses and Dissertations. https://stars.library.ucf.edu/etd/1528

*Huffcutt, A. I., Culbertson, S. S., Goebl, A. P., & Toidze, I. (2017). The influence of cognitive ability on interviewee performance in traditional versus relaxed behavior description interview formats. *European Management Journal*, *35*(3), 383–387. https://doi.org/10.1016/j.emj.2016.07.007

Hunter, J. E. (1983) *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery* (U.S. Employment Service Technical Report No. 45). Department of Labor. https://eric.ed.gov/?id=ED241577

*Hunthausen, J. M. (2000). *Predictors of task and contextual performance: Frame -of -reference effects and applicant reaction effects on selection system validity* (Order No. 9961790) [Doctoral dissertation, Portland State University]. ProQuest Dissertations & Theses Global.

*Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, *88*(3), 545–551. https://doi.org/10.1037/0021-9010.88.3.545

Ilgen, D. R., & Pulakso, E. D. (Eds.). (1999). *The changing nature of work*. Jossey-Bass.

*Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2012). Emotional Intelligence in personnel selection: Applicant reactions, criterion, and incremental validity. *International Journal of Selection and Assessment*, *20*(3), 347–358. https://doi.org/10.1111/j.1468-2389.2012.00605.x

Ioannidis, J. P., & 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

*Ion, A., Iliescu, D., Ilie, A., & Ispas, D. (2016). The emic–etic approach to personality measurement in personnel selection. *Personality and Individual Differences*, *97*, 55–60. https://doi.org/10.1016/j.paid.2016.02.082

*Ispas, D., Iliescu, D., Ilie, A., Sulea, C., Askew, K., Rohlfs, J. T., & Whalen, K. (2014). Revisiting the relationship between impression management and job performance. *Journal of Research in Personality*, *51*, 47–53. https://doi.org/10.1016/j.jrp.2014.04.010

*Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, *98*(2), 326–341. https://doi.org/10.1037/a0031257

*Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology*, *86*(5), 984–996. https://doi.org/10.1037/0021-9010.86.5.984

*Johnson, R. E., Silverman, S. B., Shyamsunder, A., Swee, H.-Y., Rodopman, O. B., Cho, E., & Bauer, J. (2010). Acting superior but actually inferior? Correlates and consequences of workplace arrogance. *Human Performance*, *23*(5), 403–427. https://doi.org/10.1080/08959285.2010.515279

*Klingner, Y., & Schuler, H. (2004). Improving participants' evaluations while maintaining validity by a work sample- intelligence test hybrid. *International Journal of Selection and Assessment*, *12*(1–2), 120–134. https://doi.org/10.1111/j.0965-075X.2004.00268.x

*Kluemper, D. H. (2006). *An examination of ability-based emotional intelligence in the structured employment interview* (Order No. 3211350) [Doctoral Dissertation, Oklahoma State University]. ProQuest Dissertations & Theses Global.

*Kluemper, D. H., DeGroot, T., & Choi, S. (2013). Emotion management ability: Predicting task performance, citizenship, and deviance. *Journal of Management*, *39*(4), 878–905. https://doi.org/10.1177/0149206311407326

Knapp, D. J., Heffner, T. S., & Campbell, R. C. (2003). *Recommendations for an Army NCO Semi-Centralized Promotion System for the 21st Century*. Human Resource Research Organization. https://doi.org/10.1037/e500162012-001

Knapp, D. J., & Tremble, T. R. (2007). *Concurrent validation of experimental army enlisted personnel selection and classification measures*. Human Resource Research Organization. https://apps.dtic.mil/sti/pdfs/ADA471963.pdf

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

*Konkin, J. (2013). *The moderating effect of emotional labor on the relationships of emotional intelligence and adjustment with managerial job performance* (Order No. 3562091) [Doctoral Dissertation, Alliant International University]. ProQuest Dissertations & Theses Global.

*Kostman, J. T. (2004). *Multi-dimensional performance requires multi-dimensional predictors: Predicting complex job performance using cognitive ability, personality and emotional intelligence assessment instruments as combinatorial predictors* (Order No. 3115266) [Doctoral dissertation, City University of New York]. ProQuest Dissertations & Theses Global.

*Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational and Organizational Psychology*, *79*(3), 411–432. https://doi.org/10.1348/096317905X68790

*Krishnakumar, S. (2008). *The role of emotional intelligence and job emotional requirements in job attitudes and behavior* [Doctoral dissertation]. Virginia Tech. https://vtechworks.lib.vt.edu/handle/10919/27849

*Kuncel, N. R., Rose, M., Ejiogu, K., & Yang, Z. (2014). Cognitive ability and socio-economic status relations with job performance. *Intelligence*, *46*, 203–208. https://doi.org/10.1016/j.intell.2014.06.003

*Law, K. S., Wong, C.-S., Huang, G.-H., & Li, X. (2008). The effects of emotional intelligence on job performance and life satisfaction for the research and development scientists in China. *Asia Pacific Journal of Management*, *25*(1), 51–69. https://doi.org/10.1007/s10490-007-9062-3

*Lawrence, A. D., Doverspike, D., & O'Connell, M. S. (2004). *An examination of the role job fit plays in selection* [Conference session]. Society for Industrial and Organizational Psychology, Chicago, IL, United States.

Lee, C. I. S., & Steel, P. (2019). Naturally occurring selection: Using applicant pool data to estimate job relevant range restriction. *Academy of Management Proceedings*, *2019*(1), Article 16220. https://doi.org/10.5465/AMBPP.2019.16220abstract

*Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*(2), 460–468. https://doi.org/10.1037/a0025741

Lounsbury, J. W., & Gibson, L. W. (2006). *Personal Style Inventory: A personality assessment tool for the work place*. Resource Associates.

*Lyons, T. J., Bayless, J. A., & Park, R. K. (2001). Relationship of cognitive, biographical, and personality measures with the training and job performance of detention enforcement officers in a federal government agency. *Applied HRM Research*, *6*(1), 67–70.

*Manley, G. G., & Benavidez, J. (2008). Using g alternatives to reduce subgroup differences when predicting US public-sector performance. *Equal Opportunities International*, *27*(4), 337–354. https://doi.org/10.1108/02610150810874304

*Mann, C. (2011). *Cognitive ability and job performance in a New Zealand service organisation: A thesis presented in partial fulfilment of the requirements for the degree Master of Science in Industrial/Organisational Psychology at Massey University, Manawatu, New Zealand* [Master's Thesis]. Massey University. https://mro.massey.ac.nz/handle/10179/2877

*Marcus, B., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance*, *20*(3), 275–285. https://doi.org/10.1080/08959280701333362

*Marcus, B., Goldenberg, J., Fine, S., Hummert, H., & Traum, A. (2020). Self-presentation in selection settings: The case of personality tests. *Journal of Business and Psychology*, *35*(5), 557–571. https://doi.org/10.1007/s10869-019-09642-x

*Marcus, B., Schuler, H., Quell, P., & Humpfner, G. (2002). Measuring counterproductivity: Development and initial validation of a German Self-Report Questionnaire. *International Journal of Selection and Assessment*, *10*(1–2), 18–35. https://doi.org/10.1111/1468-2389.00191

*Markou, P. (2015). *Fashion or fad? The incremental validity of conscientiousness and openness to experience over cognitive ability in predicting overall job performance* (Order No. 3712657) [Doctoral dissertation, The Chicago School of Professional Psychology]. ProQuest Dissertations & Theses Global.

*McCulloch, M. C., & Turban, D. B. (2007). Using person–organization fit to select employees for high-turnover jobs. *International Journal of Selection and Assessment*, *15*(1), 63–71. https://doi.org/10.1111/j.1468-2389.2007.00368.x

McDaniel, M. A. (2009, April). *Cumulative meta-analysis as a publication bias method* [Poster presentation]. Society for Industrial and Organizational Psychology, New Orleans, LA, United States.

*Morgan, K. E. (2014). *Rethinking intelligence tests: Using multidimensional item response theory to assess ability* (Order No. 3690335) [Doctoral dissertation, North Carolina State University]. ProQuest Dissertations & Theses Global.

*Mosley, D. C., Jr. (2002). *The influence of person-job fit, person-organization fit, and self-efficacy perceptions on work attitudes, job performance, and turnover* [Doctoral dissertation]. Mississippi State University.

*Motowidlo, S. J., Brownlee, A. L., & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment*, *16*(3), 272–280. https://doi.org/10.1111/j.1468-2389.2008.00433.x

*Mount, M. K., Oh, I.-S., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, *61*(1), 113–139. https://doi.org/10.1111/j.1744-6570.2008.00107.x

*Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed Biodata Scales over Gma and the five factor personality constructs. *Personnel Psychology*, *53*(2), 299–323. https://doi.org/10.1111/j.1744-6570.2000.tb00203.x

*Muhamad, M. (2011). *Testing a criterion-centric approach to validation for a leadership effectiveness framework in the context of south east Asia* (Order No. 10074728) [Doctoral dissertation, University of Surrey]. ProQuest Dissertations & Theses Global.

*Murensky, C. L. (2000). *The relationships between emotional intelligence, personality, critical thinking ability and organizational leadership performance at upper levels of management* (Order No. 9962991) [Doctoral dissertation, George Mason University]. ProQuest Dissertations & Theses Global.

*Mussel, P. (2013). Introducing the construct curiosity for predicting job performance. *Journal of Organizational Behavior*, *34*(4), 453–472. https://doi.org/10.1002/job.1809

*Mussel, P., Spengler, M., Litman, J. A., & Schuler, H. (2012). Development and validation of the German work-related curiosity scale. *European Journal of Psychological Assessment*, *28*(2), 109–117. https://doi.org/10.1027/1015-5759/a000098

National Academy of Sciences. (1999). *The changing nature of work: Implications for occupational analysis*. National Academy Press.

*Naude, M. N. (2017). *Cognitive ability testing for employee selection: Implications for age discrimination* [Doctoral Dissertation]. Colorado State University. https://mountainscholar.org/handle/10217/191413

*Nelson, L. C. (2003). *Working memory, general intelligence, and job performance* (Order No. 3076329) [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations & Theses Global.

Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive ability and job performance: Meta-analytic evidence for the validity of narrow cognitive abilities. *Journal of Business and Psychology*, *37*(6), 1119–1139. https://doi.org/10.1007/s10869-022-09796-1

*O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., III, & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, *15*(1), 19–29. https://doi.org/10.1111/j.1468-2389.2007.00364.x

Oh, I., Le, H., & Roth, P. L. (in press). Revisiting Sackett et al.'s (2022) rationale behind their recommendation against correcting for range restriction in concurrent validity studies. *Journal of Applied Psychology*.

*Oh, I.-S., Le, H., Whitman, D. S., Kim, K., Yoo, T.-Y., Hwang, J.-O., & Kim, C.-S. (2014). The incremental validity of honesty–humility over cognitive ability and the big five personality traits. *Human Performance*, *27*(3), 206–224. https://doi.org/10.1080/08959285.2014.913594

*Ono, M., Sachau, D. A., Deal, W. P., Englert, D. R., & Taylor, M. D. (2011). Cognitive ability, emotional intelligence, and the big five personality dimensions as predictors of criminal investigator performance. *Criminal Justice and Behavior*, *38*(5), 471–491. https://doi.org/10.1177/0093854811399406

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology*, *78*(4), 569–582. https://doi.org/10.1037/0021-9010.78.4.569

*Ovidiu, S. (2015). Improving personnel selection through frame of reference effect on personality inventory: Predictive and incremental validity over cognitive ability and job knowledge. *Procedia: Social and Behavioral Sciences*, *187*, 261–265. https://doi.org/10.1016/j.sbspro.2015.03.049

*Patton, C. B. (2015). *Crystallized intelligence and openness to experience: Drawing on intellectual-investment theories to predict job performance longitudinally* (Order No. 3664533) [Doctoral dissertation, George Mason University]. ProQuest Dissertations & Theses Global.

*Periman, W. C. (2016). *The relationship of working memory to job performance and innovation with stress and effort as moderators* (Order No. 10243112) [Doctoral dissertation, Oklahoma State University]. ProQuest Dissertations & Theses Global.

*Perry, S. J., Hunter, E. M., Witt, L. A., & Harris, K. J. (2010). P = *f* (conscientiousness × ability): Examining the facets of conscientiousness. *Human Performance*, *23*(4), 343–360. https://doi.org/10.1080/08959285.2010.501045

*Peterson, S. J., & Byron, K. (2008). Exploring the role of hope in job performance: Results from four studies. *Journal of Organizational Behavior*, *29*(6), 785–803. https://doi.org/10.1002/job.492

*Provines, J. L. (2006). *Investigation of police officer selection procedures* [Doctoral dissertation]. Wichita State University.

*Roland, D. K. (2010). *Relationships between personality dimensions assessed by the Navy Computer adaptive personality scales and supervisor ratings of job performance* (Order No. 3421373) [Doctoral dissertation, The University of Memphis]. ProQuest Dissertations & Theses Global.

*Roodt, G., & La Grange, L. (2001). Personality and cognitive ability as predictors of the job performance of insurance sales people. *SA Journal of Industrial Psychology*, *27*(3), 35–43. https://hdl.handle.net/10520/EJC88869

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*(1), 66–80. https://doi.org/10.1037/0021-9010.87.1.66

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). Wiley.

Sackett, P. R., Berry, C. M., Lievens, F., & Zhang, C. (2023). Correcting for range restriction in meta analysis: A reply to Oh et al. *Journal of Applied Psychology*, *108*(8), 1311–1315. https://doi.org/10.1037/apl0001116

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., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology*, *102*(10), 1435–1447. https://doi.org/10.1037/apl0000236

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*(3), 573–605. https://doi.org/10.1111/j.1744-6570.2003.tb00751.x

*Sanderson, K. (2012). *Time orientation in organizations: Polychronicity and multitasking* [Doctoral dissertation, Florida International University]. FIU Electronic Theses and Dissertations.

*Savoy, P. J. (2004). *Development and validation of a measure of self - directed learning competency* (Order No. 3133686) [Doctoral dissertation, Kent State University]. ProQuest Dissertations & Theses Global.

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *The Career Development International*, *22*(5), 469–476. https://doi.org/10.1108/CDI-08-2017-0136

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*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses 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

*Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry*, *36*(3), 620–661. https://doi.org/10.1111/j.1747-4469.2011.01245.x

*Sitser, T. (2014). *Predicting sales performance: Strengthening the personality—Job performance linkage* [Doctoral dissertation]. Erasmus University Rotterdam.

*Srikanth, P. B. (2020). The relative contribution of personality, cognitive ability and the density of work experience in predicting human resource competencies. *Personnel Review*, *49*(8), 1573–1590. https://doi.org/10.1108/PR-09-2018-0329

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

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*(1), 19–42. https://doi.org/10.1002/jrsm.1211

*Sturman, M. C. (2001). Utility analysis for multiple selection devices and multiple outcomes. *Journal of Human Resource Costing & Accounting*, *6*(2), 9–28. https://doi.org/10.1108/eb029072

*Tews, M. J., Michel, J. W., & Noe, R. A. (2011). Beyond objectivity: The performance impact of the perceived ability to learn and solve problems. *Journal of Vocational Behavior*, *79*(2), 484–495. https://doi.org/10.1016/j.jvb.2010.11.005

*Ugaz, A. (2011). *The relationship between personality, cognitive ability, and performance, in an aging workforce* (Order No. 3471033) [Doctoral dissertation, University of Houston]. ProQuest Dissertations & Theses Global.

*Uhrich, B. B., Heggestad, E. D., & Shanock, L. R. (2021). Smarts or trait emotional intelligence? The role of trait emotional intelligence in enhancing the relationship between cognitive ability and performance. *The Psychologist Manager Journal*, *24*(1), 23–47. https://doi.org/10.1037/mgr0000112

U.S. Employment Service. (1970). *Manual for the USES General Aptitude test Battery*. https://eric.ed.gov/?id=ED164579

*Valentine, A., Larson, E., & Yusko, K. (2022). Merck's general management acceleration program: Developing future leaders in pharmaceuticals. *The Industrial-Organizational Psychologist*, *59*(3). https://www.siop.org/Research-Publications/Items-of-Interest/ArtMID/19366/ArticleID/5537/preview/true/The-Bridge-Connecting-Science-and-Practice

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

*Van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, *42*(7), 1811–1835. https://doi.org/10.1177/0149206313515524

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

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

*Votruba, R. (2012). *Predictive validity of intelligence and personality to job outcomes over time* (Order No. 3494369) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations & Theses Global.

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, *40*(5), 663–705. https://doi.org/10.3102/00346543040005663

*Whelpley, C. E. (2014). *How to score situational judgment tests: A theoretical approach and empirical test* (Order No. 3666943) [Doctoral dissertation, Virginia Commonwealth University]. ProQuest Dissertations & Theses Global.

*Witt, L. A., & Burke, L. A. (2002). Selecting high-performing information technology professionals. *Journal of Organizational and End User Computing*, *14*(4), 37–50. https://doi.org/10.4018/joeuc.2002100103

Zhou, Y., Shen, W., Beatty, A. S., & Sackett, P. R. (2022). *An updated meta-analysis of the interrater reliability of supervisor ratings of job performance* [Paper presentation]. Society for Industrial and Organizational Psychology conference, Seattle, WA, United States.

(*Appendix follows*)

Appendix

Description of Studies Included in the Meta-Analysis

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Al Ali et al. (2012)	310	.54	0	0	0	1	1	0	1	2	0	Test assessing general cognitive ability	NR	NA	33-3051.00	Police and sheriff's patrol officers	4.54	4.02	2.98
Alexander (2007)	2,785	.15	1	1	0	1	0	0	0	0	0	Test assessing numerical ability and matrix manipulation	NR	NA	NA	NA	NA	NA	NA
Allen (2016)	89	.09	1	0	0	1	0	0	0	1	0	Two tests of verbal and inductive reasoning	NR	NA	NA	NA	NA	NA	NA
Allworth and Hesketh (2000)	202	.15	0	0	0	1	1	0	0	2	1	Raven's Progressive Matrices and Ball Numerical Reasoning Test	NR	NA	NA	NA	NA	NA	NA
Bedford (2011)	294	.07	1	0	1	1	1	0	0	0	1	Employee Aptitude Survey and Watson–Glaser	NR	NA	NA	NA	NA	NA	NA
Benavidez (2005)	529	.14	0	0	0	1	1	1	1	2	0	Test assessing memory and numerical, verbal, and logical ability	0.83	NA	33-3012.00	Correctional officers and jailers	4.01	3.01	2.15
Benjamin (2006)	64	.03	1	1	1	1	1	0	1	1	1	ACER Advanced Test B90	NR	1.03	NA	NA	NA	NA	NA
Bergman et al. (2006)	123	.26	0	0	1	1	1	0	1	0	1	Wonderlic	NR	NA	43-1011.00	First-line supervisors of office and administrative support workers	3.92	3.93	1.92
Bergner (2020)	123	.24	0	0	1	1	1	0	1	2	1	Wonderlic	NR	NA	11-3071.00	Transportation, storage, and distribution managers	4.30	4.44	2.48
Blickle et al. (2009)	83	−.05	0	0	1	1	1	0	0	2	1	Wonderlic	0.83	NA	NA	NA	NA	NA	NA
Bosco et al. (2015)	100	−.11	0	0	0	1	0	0	0	2	1	Wonderlic	0.89	NA	NA	NA	NA	NA	NA
Bosco et al. (2015)	159	.19	0	0	1	1	0	0	0	1	1	Raven's Advanced Progressive Matrices	0.86	NA	NA	NA	NA	NA	NA
Brantley (2000)	928	.14	0	0	0	1	1	1	1	0	0	Test assessing reasoning and applied math	NR	NA	53-2021.00	Air traffic controllers	4.77	2.89	1.23
Callans et al. (2016)	176	.22	0	0	0	1	1	0	0	2	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
Cavazotte et al. (2012)	134	.36	0	0	1	0	0	0	0	1	1	GMAT (Graduate Management Admission Test)	NR	NA	NA	NA	NA	NA	NA
Chan and Schmitt (2002)	160	.13	0	0	0	1	1	1	1	2	1	Graduate and Managerial Assessment–Verbal (GMA-V)	NR	NA	43-6014.00	Secretaries and administrative assistants, except legal, medical, and executive	3.03	2.33	1.28
Childers (2016)	168	.21	0	0	0	1	1	0	1	2	1	Wonderlic	NR	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45
Christiansen et al. (2010)	69	.30	0	0	0	1	1	0	0	1	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA

(*Appendix continues*)

Appendix (*continued*)

Author	N	r	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	u ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Cober et al. (2003)	156	.23	0	0	0	1	0	0	0	2	0	Test comparable to Raven's Progressive Matrices	0.73	NA	NA	NA	NA	NA	NA
Connolly (2001)	107	.14	0	0	0	1	1	0	1	2	1	Watson–Glaser	NR	NA	33-3051.00	Police and sheriff's patrol officers	4.54	4.02	2.98
Côté and Miners (2006)	175	.35	0	0	0	1	0	0	0	2	1	Culture Fair Intelligence Test	0.81	NA	NA	NA	NA	NA	NA
David and Witt (2009)	106	.21	0	0	0	1	0	0	1	0	1	Wonderlic	NR	NA	15-1251.00	Computer programmers	4.81	3.09	1.27
Dayan et al. (2002)	417	.12	1	1	0	1	1	0	1	2	0	One test equivalent to Raven's Progressive Matrices, one test assessing verbal ability	NR	0.58	33-3051.00	Police and sheriff's patrol officers	4.54	4.02	2.98
Dhliwayo and Coetzee (2020)	299	.35	0	0	1	1	1	1	0	2	1	GAMA (General Ability Measure for Adults)	0.76	NA	NA	NA	NA	NA	NA
Doucet et al. (2016)	70	.33	1	0	0	1	1	0	1	2	1	Wonderlic	NR	NA	43-4051.00	Customer service representatives	3.58	2.79	1.23
Ejiogu et al. (2006)	66	.39	0	0	0	1	1	0	1	0	1	Watson–Glaser	0.76	NA	33-3021.06	Intelligence analysts	5.32	4.11	0.97
Ejiogu et al. (2007)	87	.38	0	0	1	1	1	0	0	2	1	Watson–Glaser and Advanced Numerical Reasoning Appraisal	Watson–Glaser = 0.85; Advanced Numerical Reasoning Appraisal = 0.87; Composite = .92	NA	NA	NA	NA	NA	NA
Faura (2016)	197	.32	0	0	0	1	1	0	0	1	0	Test assessing numerical and verbal ability	0.92	NA	NA	NA	NA	NA	NA
Ferris et al. (2001)	106	.12	0	0	0	1	1	1	1	1	1	Wonderlic	NR	NA	15-1251.00	Computer programmers	4.81	3.09	1.27
Fertig (2009)	124	.19	0	0	1	1	1	0	1	1	0	Numerical reasoning test (NCC2)	NR	NA	11-3031.00	Financial managers	4.06	4.11	0.56
Fetzer (2004)	153	.00	0	0	1	1	1	0	1	1	1	Watson–Glaser	NR	NA	11-2022.00	Sales managers	4.16	4.97	1.19
Flannery (2020)	122	−.17	0	0	0	1	0	0	0	1	0	Test of matrix reasoning	NR	NA	NA	NA	NA	NA	NA
Fort (2010)	285	.16	0	0	0	1	1	0	1	0	1	EAS	NR	NA	43-6014.00	Secretaries and administrative assistants, except legal, medical, and executive	3.03	2.33	1.28
Goldstein et al. (2001)	633	.10	0	0	1	1	1	0	0	2	1	Wesman Personnel Classification Test	NR	NA	NA	NA	NA	NA	NA
Green and Macqueen (2008)	37	.37	1	1	1	1	1	0	0	0	1	Critical reasoning (VMG2), numerical critical reasoning (NMG2), and Raven's Standard Progressive Matrices (RPM)	NR	NA	NA	NA	NA	NA	NA

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Grim (2010)	100	.04	1	1	0	1	1	0	0	2	1	ASVAB	NR	1.04	NA	NA	NA	NA	NA
Gunter (2010)	90	.01	0	0	0	1	1	1	1	0	1	Wonderlic	NR	NA	43-4051.00	Customer service representatives	3.58	2.79	1.23
Harlaar (2013)	126	−.14	0	0	0	1	1	0	1	1	0	Connector Ability Test	Connector Ability–Figure Series = .78; Connector Ability–Matrices = .73; Connector Ability–Number Series = .88	NA	43-4051.00	Customer service representatives	3.58	2.79	1.23
Harris (2017)	300	.13	0	0	0	1	0	0	0	2	0	Composite assessing general cognitive ability	NR	NA	NA	NA	NA	NA	NA
Harris (2017)	1,410	.10	0	0	0	1	0	0	0	2	0	Test assessing memory and verbal ability	NR	NA	NA	NA	NA	NA	NA
Harzer et al. (2021)	169	.39	0	0	0	1	1	1	0	2	1	Culture Fair Intelligence Test	0.89	NA	NA	NA	NA	NA	NA
Hattrup et al. (2005)	121	.23	0	0	0	1	0	0	1	2	0	Test assessing numerical, analytical, and applied reasoning	0.78	NA	51-9124.00	Coating, painting, and spraying machine setters; operators; and tenders	3.66	2.44	3.65
Hattrup et al. (2005)	142	−.03	0	0	0	1	1	0	1	2	0	Test assessing numerical, analytical, and applied reasoning	0.47	NA	51-2092.00	Team assemblers	3.07	2.29	3.22
Hausdorf and Risavy (2015)	127	−.08	1	1	0	1	1	0	1	2	1	Wonderlic	0.76	NA	53-3052.00	Bus drivers, transit and intercity	3.07	2.33	2.75
Hawes (2000)	410	.05	0	0	0	1	1	0	1	2	0	Test of numerical and verbal ability	NR	NA	43-4051.00	Customer service representatives	3.58	2.79	1.23
Henderson (2010)	74	.47	1	1	0	1	1	0	1	0	0	Test assessing memory and numerical, verbal, logical, and spatial ability	0.93	0.68	33-2011.00	Firefighters	4.10	3.74	4.43
Hochwarter et al. (2000)	813	.18	0	0	0	1	0	0	0	0	1	Wonderlic and Watson–Glaser	NR	NA	NA	NA	NA	NA	NA
Hodge (2010)	91	.02	0	0	0	1	1	0	0	0	1	Watson–Glaser, Thurstone Test of Mental Alertness, Verbal Reasoning Test, and Basic Skills Test	NR	NA	NA	NA	NA	NA	NA
Huffcutt et al. (2017)	84	.12	0	0	0	1	1	0	0	0	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
Hunthausen (2000)	212	.38	0	0	1	1	1	1	1	2	1	Wonderlic	NR	NA	11-3071.00	Transportation, storage, and distribution managers	4.30	4.44	2.48
Hunthausen et al. (2003)	208	.24	0	0	1	1	1	0	1	2	1	Wonderlic	NR	NA	43-1011.00	First-line supervisors of office and administrative support workers	3.92	3.93	1.92
Iliescu et al. (2012)	223	.41	0	0	0	1	1	0	1	2	1	GAMA (General Ability Measure for Adults)	0.87	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45

(*Appendix continues*)

Appendix (*continued*)

Author	N	r	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	u ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Iliescu et al. (2012)	61	.35	0	0	1	1	1	0	1	2	1	GAMA (General Ability Measure for Adults)	0.87	NA	11-1011.00	Chief executive	4.94	5.13	0.94
Ion et al. (2016)	186	.35	1	0	0	1	1	0	1	0	1	GAMA (General Ability Measure for Adults)	0.87	NA	51-6021.00	Pressers, textile, garment, and related materials	2.49	1.57	2.88
Ion et al. (2016)	253	.37	0	0	0	1	1	0	1	0	1	GAMA (General Ability Measure for Adults)	0.83	NA	51-6021.00	Pressers, textile, garment, and related materials	2.49	1.57	2.88
Ispas et al. (2014)	129	.23	1	0	0	0	0	0	1	2	1	GAMA (General Ability Measure for Adults)	0.88	NA	41-4011.00	Sales representatives, wholesale and manufacturing, technical and scientific products	3.04	2.68	0.87
Jansen et al. (2013)	107	.18	0	0	0	1	1	0	0	0	1	Wonderlic	0.82	NA	NA	NA	NA	NA	NA
J. W. Johnson (2001)	1894	.20	0	0	0	1	0	0	0	0	0	Composite of reading comprehension, reasoning, numerical reasoning, and language usage	NR	NA	NA	NA	NA	NA	NA
Johnson et al. (2010)	32	.31	0	0	0	1	0	0	0	2	1	Wesman	NR	NA	NA	NA	NA	NA	NA
Klingner and Schuler (2004)	138	.26	0	0	0	1	1	0	1	2	1	Intelligence Structure Test (IST-70)	NR	NA	NA	NA	NA	NA	NA
Kluemper (2006)	66	.22	0	0	0	1	0	0	1	1	1	Wonderlic	NR	NA	39-9041.00	Residential advisors	3.61	3.67	1.95
Kluemper et al. (2013)	102	.03	0	0	0	1	0	0	0	2	1	ACT and SAT	NR	NA	NA	NA	NA	NA	NA
Kluemper et al. (2013)	85	.15	0	0	0	1	0	0	1	2	1	Wonderlic	NR	NA	33-3012.00	Correctional officers and jailers	4.01	3.01	2.15
Knapp and Tremble (2007)	414	.30	0	0	0	1	0	0	0	2	1	AFQT	NR	NA	NA	NA	NA	NA	NA
Knapp et al. (2003)	440	.08	1	1	0	1	1	0	0	0	1	ASVAB	NR	NA	NA	NA	NA	NA	NA
Knapp et al. (2003)	370	.11	1	1	0	1	1	0	0	0	1	ASVAB	NR	NA	NA	NA	NA	NA	NA
Konkin (2013)	161	−.09	0	0	1	1	1	0	0	1	1	Watson–Glaser	NR	NA	NA	NA	NA	NA	NA
Kostman (2004)	147	.32	0	0	0	0	0	0	1	2	1	Wonderlic	0.88	NA	41-9041.00	Telemarketers	2.16	1.89	0.65
Krajewski et al. (2006)	157	.23	1	1	1	1	1	0	1	0	1	EAS	NR	NA	11-3051.00	Industrial production managers	4.06	3.78	3.52
Krishnakumar (2008)	278	.13	0	0	0	1	1	1	0	1	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
Kuncel et al. (2014)	108	.37	0	0	0	1	1	0	0	1	1	Watson–Glaser and Advanced Numerical Reasoning Appraisal	Watson–Glaser = 0.86; Advanced Numerical Reasoning Appraisal = 0.88; Composite = .92	NA	NA	NA	NA	NA	NA
Law et al. (2008)	102	−.07	0	0	0	1	1	0	0	2	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/ task (1), only overall/ task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Lawrence et al. (2004)	130	.12	1	1	0	1	1	1	1	1	0	Test similar to Raven's Progressive Matrices	NR	0.92	53-7062.00	Laborers and freight, stock, and material movers, hand	2.44	1.56	3.10
Lawrence et al. (2004)	143	.07	1	1	0	1	1	1	1	2	0	Test similar to Raven's Progressive Matrices	NR	0.88	53-7065.00	Stockers and order fillers	2.53	2.20	2.21
Lievens and Sackett (2012)	103	.10	1	1	0	1	1	0	1	2	0	Four tests assessing scientific knowledge, one test assessing general cognitive ability	NR	0.55	29-1229.02	Hospitalist	5.12	4.15	1.59
Lounsbury and Gibson (2006)	188	.33	0	0	0	1	1	0	1	2	0	Composite scale of numerical, verbal, and abstract reasoning ability	0.84	NA	49-3023.00	Automotive service technicians and mechanics	3.75	2.67	4.28
Lounsbury and Gibson (2006)	105	.23	0	0	0	1	1	0	1	2	0	Composite scale of numerical, verbal, and abstract reasoning ability	0.84	NA	51-9197.00	Tire builders	3.18	2.33	3.61
Lounsbury and Gibson (2006)	154	.24	0	0	0	1	1	0	1	2	0	Composite scale of numerical, verbal, and abstract reasoning ability	0.84	NA	43-3071.00	Tellers	3.76	2.95	2.10
Lounsbury and Gibson (2006)	105	.28	0	0	0	1	1	0	1	2	0	Composite scale of numerical, verbal, and abstract reasoning ability	0.84	NA	43-3011.00	Bill and account collectors	2.79	2.15	0.51
Lyons et al. (2001)	381	.22	0	0	0	1	1	0	1	0	0	Test assessing numerical, verbal, and logical ability	Name and Number Comparison = 0.82; Logical Reasoning = 0.82; Following Policies and Procedures = 0.71	NA	33-3012.00	Correctional officer and jailer	4.01	3.01	2.15
Manley and Benavidez (2008)	529	.30	1	1	0	1	1	1	1	2	0	Test assessing memory and numerical, verbal, and logical ability	0.83	NA	33-3012.00	Correctional officer and jailer	4.01	3.01	2.15
Mann (2011)	43	.26	1	1	0	1	1	0	0	0	0	Internet Reasoning Test and Adaptive General Reasoning Test	NR	NA	NA	NA	NA	NA	NA
Marcus et al. (2002)	174	.21	0	0	0	1	0	0	0	0	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
Marcus et al. (2007)	119	.07	1	1	1	1	1	0	1	0	1	EAS	NR	0.92	11-1021.00	General and operations managers	4.05	4.13	2.44
Marcus et al. (2020)	344	.25	0	0	0	1	1	0	0	2	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
Markou (2015)	71	−.09	0	0	0	1	1	1	0	1	1	Wonderlic	NR	NA	NA	NA	NA	NA	NA
McCulloch and Turban (2007)	228	.23	0	0	0	1	1	0	1	1	0	Performance Skills Index	0.85	NA	43-2011.00	Switchboard operators, including answering service	3.09	2.23	1.43

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Morgan (2014)	285	.17	0	0	0	1	1	0	1	0	0	Test assessing numerical, verbal, and logical ability	NR	NA	41-2011.00	Cashiers	2.51	2.42	1.92
Morgan (2014)	132	.19	0	0	0	1	1	0	1	0	0	Test assessing numerical, verbal, and logical ability	NR	NA	43-9041.00	Insurance claims and policy processing clerks	3.33	2.09	0.66
Mosley (2002)	53	.17	1	0	0	1	0	0	1	1	1	Wonderlic	NR	NA	51-7011.00	Cabinetmakers and bench carpenters	3.29	2.05	3.85
Motowidlo et al. (2008)	140	.10	0	0	0	1	1	0	1	2	1	Wonderlic	0.87	NA	41-2031.00	Retail salespersons	2.83	2.35	1.99
Mount et al. (2000)	222	.18	0	0	0	1	1	0	1	2	1	Wonderlic	NR	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45
Mount et al. (2000)	146	.13	0	0	0	1	1	0	1	2	1	Wonderlic	NR	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45
Mount et al. (2008)	133	.21	0	0	0	1	0	0	1	2	1	Wonderlic	0.86	NA	53-7065.00	Stockers and order fillers	2.53	2.20	2.21
Muhamad (2011)	137	.32	1	0	1	1	1	0	0	0	0	Swift Aptitude Analysis	NR	NA	NA	NA	NA	NA	NA
Murensky (2000)	90	−.12	0	0	1	0	1	0	0	1	1	Watson–Glaser	NR	NA	NA	NA	NA	NA	NA
Mussel (2013)	224	.36	0	0	0	1	0	0	0	2	0	Test comparable to Wonderlic	0.91	NA	NA	NA	NA	NA	NA
Mussel et al. (2012)	219	.19	0	0	0	1	1	0	1	0	0	Test comparable to Wonderlic	0.86	NA	NA	NA	NA	NA	NA
Naude (2017)	214	.15	0	0	1	1	1	0	0	0	0	Test assessing numerical, deductive, and figural reasoning	0.73	NA	NA	NA	NA	NA	NA
Nelson (2003)	378	.06	0	0	0	1	1	0	1	2	0	Two tests assessing numerical and logical ability	Numerical = .89; Reasoning = .76; Composite = .87	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45
O'Connell et al. (2007)	1,140	.15	0	0	0	1	0	0	1	1	0	Composite of four reasoning subtests	0.79	NA	51-2092.00	Team assemblers	3.07	2.29	3.22
Oh et al. (2014)	217	.24	0	0	0	1	0	0	1	1	0	Korean Police Officer Aptitude Battery, comparable to the Wonderlic	0.73	NA	NA	NA	NA	NA	NA
Ono et al. (2011)	38	.44	1	0	0	1	1	0	1	2	1	Shipley Institute of Living Scale	NR	NA	33-3051.00	Police and sheriff's patrol officers	4.54	4.02	2.98
Ovidiu (2015)	36	.42	1	1	0	1	1	0	1	2	1	Raven's Advanced Progressive Matrices	NR	NA	NA	NA	NA	NA	NA
Patton (2015)	92	.09	0	0	0	1	1	0	0	0	1	Watson–Glaser	NR	NA	NA	NA	NA	NA	NA
Periman (2016)	214	.16	0	0	0	1	0	0	0	0	0	Test measuring basic vocabulary, grammar, geometry, logic, arithmetic, and problem solving	NR	NA	NA	NA	NA	NA	NA
Perry et al. (2010)	208	.13	0	0	0	1	0	0	1	0	1	Watson–Glaser	NR	NA	43-4051.00	Customer service representatives	3.58	2.79	1.23
Peterson and Byron (2008)	65	.26	1	0	1	0	0	0	1	2	1	Wechsler (WAIS-R)	0.90	NA	NA	NA	NA	NA	NA

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Peterson and Byron (2008)	163	.36	1	0	0	0	0	0	1	2	1	Wechsler (WAIS-R)	0.93	NA	41-2031.00	Retail salespersons	2.83	2.35	1.99
Peterson and Byron (2008)	79	.39	1	0	0	0	0	0	1	2	1	Wechsler (WAIS-R)	0.94	NA	13-2072.00	Loan officers	4.06	3.15	0.70
Provines (2006)	69	.00	1	1	0	1	1	0	1	0	1	Wonderlic and a test assessing verbal and spatial ability	NR	NA	33-3051.00	Police and sheriff's patrol officers	4.54	4.02	2.98
Roland (2010)	1,315	.08	1	1	0	1	1	1	1	1	1	AFQT	NR	1.78	NA	NA	NA	NA	NA
Roodt and La Grange (2001)	170	.16	0	0	0	1	0	0	1	0	0	Verbal Evaluation test (VCC3-SHL)	NR	NA	41-3021.00	Insurance sales agents	3.53	2.75	1.09
Sanderson (2012)	175	.25	0	0	0	1	1	0	0	1	0	Global Cognitive Index	NR	NA	NA	NA	NA	NA	NA
Sanderson (2012)	119	.16	0	0	0	1	1	0	0	2	0	Global Cognitive Index	NR	NA	NA	NA	NA	NA	NA
Savoy (2004)	54	.58	0	0	0	0	0	0	1	0	1	Wonderlic	0.93	NA	51-4041.00	Machinists	3.55	2.45	4.00
Shultz and Zedeck (2011)	1,148	.04	1	1	0	1	1	0	1	2	1	LSAT	NR	0.70	23-1011.00	Lawyers	4.65	3.91	1.07
Sitser (2014)	105	.07	0	0	0	1	1	0	1	1	0	Bridge Ability Suite	0.72	NA	41-3021.00	Insurance sales agents	3.53	2.75	1.09
Sitser (2014)	403	.09	0	0	0	1	1	0	1	1	0	Bridge Ability Suite	0.70	NA	41-3021.00	Insurance sales agents	3.53	2.75	1.09
Srikanth (2020)	140	.34	1	0	0	1	1	0	1	2	1	Wesman and Watson–Glaser	0.83	NA	13-1071.00	Human resources specialist	3.69	3.45	0.55
Sturman (2001)	296	.29	0	0	0	1	1	0	1	0	1	Wonderlic	NR	NA	43-6014.00	Secretaries and administrative assistants, except legal, medical, and executive	3.03	2.33	1.28
Tews et al. (2011)	265	.12	0	0	1	1	1	0	1	2	1	Wonderlic	NR	NA	11-9051.00	Food service managers	2.74	2.98	2.13
Ugaz (2011)	1,428	.12	1	1	0	1	0	0	0	0	0	Proprietary scale of mental alertness	NR	NA	NA	NA	NA	NA	NA
Uhrich et al. (2021)	102	.27	0	0	0	1	0	0	1	2	1	Wonderlic	NR	NA	13-1071.00	Human resources specialist	3.69	3.45	0.55
Valentine et al. (2022)	312	.16	1	1	1	1	0	0	0	0	0	Siena Reasoning Test (SRT)	NR	NA	NA	NA	NA	NA	NA
Van Iddekinge et al. (2016)	142	.20	1	0	0	1	1	1	0	1	1	ACT	NR	NA	NA	NA	NA	NA	NA
Votruba (2012)	9	−.08	0	0	0	0	0	0	1	0	1	Wonderlic	NR	NA	13-2052.00	Personal financial advisors	4.48	4.12	0.29
Votruba (2012)	19	.09	1	1	0	0	0	0	1	0	1	Wonderlic	NR	NA	13-2052.00	Personal financial advisors	4.48	4.12	0.29
Whelpley (2014)	1859	.11	0	0	1	1	0	0	1	2	0	Test assessing numerical and verbal ability	0.91	NA	11-3061.00	Purchasing managers	4.17	4.58	1.33
Whelpley (2014)	1,094	.06	0	0	1	1	0	0	1	2	0	Test assessing numerical and verbal ability	0.90	NA	11-3061.00	Purchasing managers	4.17	4.58	1.33
Witt and Burke (2002)	94	.36	0	0	0	1	0	0	1	1	1	Wonderlic	NR	NA	15-1232.00	Computer user support specialists	3.77	2.64	2.83

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Consultant study 1	74	.14	1	1	0	1	1	0	1	NA	1	Wonderlic	NR	NA	33-2011.00	Firefighters	4.10	3.74	4.43
Consultant study 2	268	.16	0	0	0	1	1	0	1	NA	0	Mechanical problem-solving test	NR	NA	49-3031.00 and 49-3042.00	Bus and truck mechanics and diesel engine specialists/mobile heavy equipment mechanics, except engines	3.63	2.78	4.60
Consultant study 3	38	.32	1	1	0	1	0	0	1	NA	1	Shipley	NR	NA	43-5031.00	Public safety telecommunicators	4.23	3.28	1.51
Consultant study 4	170	.14	0	0	0	1	1	0	1	NA	0	Test composed of logical series, analogies, and number series	0.76	NA	41-2031.00	Retail salespersons	2.83	2.35	1.99
Consultant study 5	293	.19	0	0	1	1	1	0	0	NA	0	Consultant measure of general ability	NR	NA	NA	NA	NA	NA	NA
Consultant study 6	213	.02	1	0	0	1	1	0	0	NA	1	Consultant measure of general ability	NR	NA	NA	NA	NA	NA	NA
Consultant study 7	339	.05	1	0	1	1	1	0	0	NA	1	Consultant measure of general ability	NR	NA	NA	NA	NA	NA	NA
Consultant study 8	397	.25	0	0	0	1	1	0	1	NA	0	Test measuring analytic thinking, multitasking, mathematical usage, and reading comprehension	0.89	NA	51-8013.00	Power plant operators	3.79	2.60	3.85
Consultant study 9	209	.17	0	0	0	1	0	0	1	NA	1	Basic Skills Test (BST)	0.72	NA	43-4111.00	Interviewers, except eligibility and loan	4.37	3.94	1.69
Consultant study 10	373	.07	1	0	0	1	1	0	1	NA	0	Test assessing reading, arithmetic, and mechanical problem solving	NR	NA	51-2092.00	Team assemblers	3.07	2.29	3.22
Consultant study 11	140	.25	0	0	0	1	1	0	1	NA	0	Test assessing reading, arithmetic, and mechanical problem solving	NR	NA	49-2022.00	Telecommunications equipment installers and repairers, except line installers	3.41	2.51	3.57
Consultant study 12	146	.00	0	0	1	1	1	0	1	NA	1	Watson–Glaser	NR	NA	11-2022.00	Sales managers	4.16	4.97	1.19
Consultant study 13	118	.10	0	0	0	1	1	0	0	NA	0	Test assessing reading, arithmetic, and mechanical problem solving	NR	NA	NA	NA	NA	NA	NA
Consultant study 14	94	−.16	0	0	0	1	1	0	1	NA	0	Test measuring adative reasoning	NR	NA	41-4011.00	Sales representatives, wholesale and manufacturing, technical and scientific products	3.04	2.68	0.87

(*Appendix continues*)

Appendix (*continued*)

Author	*N*	*r*	Predictive (1), concurrent (0)	Applicant (1), incumbent (0)	Manager (1), nonmanager (0)	Rating (1), objective (0)	Overall performance (1), task performance (0)	Both overall/task (1), only overall/task (0)	Single job (1), multijob (0)	Incremental did not look (0), looked; not found (1), looked; found (2)	Legacy test (1), other (0)	Cognitive test	Cognitive test reliability	*u* ratio	O*NET job code	O*NET job title	O*NET working with information (range 1.4–5.8)	O*NET leading, motivating, and coordinating (range 1.1–5.2)	O*NET manual and physical activities (range 0.12–5.4)
Consultant study 15	56	.36	0	0	0	1	1	0	1	NA	1	Basic Skills Test (BST) and EAS	Basic Skills Test 8–Following Written Directions = .75; Basic Skills Test 11–Classifying = .82; Employee Aptitude Survey 1–Verbal Comprehension = .85; Employe Aptitude Survey 2–Numerical Ability = .87; Employee Aptitude Survey 4–Visual Speed, and Accuracy = .91; Employee Aptitude Survey 7–Verbal Reasoning = .82	NA	43-9061.00	Office clerks, general	2.83	2.08	1.45
Consultant study 16	43	.26	0	0	1	1	1	0	1	NA	1	Basic Skills Test (BST) and EAS	NR	NA	33-3051.01	Police and sheriff's patrol officers	4.54	4.02	2.98
Consultant study 17	175	.16	0	0	1	1	1	0	1	NA	1	EAS and Professional Employment Test (PET)	Employee Aptitude Survey 7–Verbal Reasoning = .82; Professional Employment Test = .86	NA	41-1012.01	First-line supervisors of nonretail sales eorkers	4.31	4.16	2.05
Consultant study 18	226	.10	0	0	1	1	1	0	1	NA	1	EAS and Professional Employment Test (PET)	NR	NA	11-9131.00	Postmasters and mail superintendents	4.03	4.45	2.21
Consultant study 19	143	.07	1	1	0	1	1	0	1	NA	1	Professional Employment Test (PET)	NR	NA	13-1161.00	Market research analysts and marketing specialists	4.89	4.40	0.64
Consultant study 20	387	.11	0	0	0	1	1	0	1	NA	1	EAS and Professional Employment Test (PET)	NR	NA	15-1253.00	Software quality assurance analysts and testers	4.53	3.25	1.16
Consultant study 21	347	.10	0	0	0	1	1	0	1	NA	1	EAS and Professional Employment Test (PET)	NR	NA	15-1232.00	Computer user support specialists	3.77	2.64	2.83
Consultant study 22	165	.20	0	0	0	1	1	0	1	NA	1	EAS	NR	NA	49-2022.00	Telecommunications equipment installers and repairers, except line installers	3.41	2.51	3.57

Note. Cognitive test reliabilities were only taken from samples that report local reliability estimates; samples that only reported prior published test reliabilities, such as from test manuals, were not included. Five samples reported split-half reliabilities: Faura (2016); both samples from Iliescu et al. (2012); Ispas et al. (2014); Oh et al. (2014). ASVAB = Armed Services Vocational Aptitude Battery; NR= not reported; NA = not available; ACER = Australian Council for Educational Research; AFQT = Armed Forces Qualification Test; EAS = Employee Aptitude Survey; WAIS-R = Wechsler Adult Intelligence Test-Revised; LSAT = Law School Admissions Test.

Received June 10, 2022
Revision received September 28, 2023
Accepted October 2, 2023 ■