

Watrin, Luc; Weihrauch, Lucas; Wilhelm, Oliver

**Article — Published Version**

## The criterion-related validity of conscientiousness in personnel selection: A meta-analytic reality check

International Journal of Selection and Assessment

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Watrin, Luc; Weihrauch, Lucas; Wilhelm, Oliver (2022) : The criterion-related validity of conscientiousness in personnel selection: A meta-analytic reality check, International Journal of Selection and Assessment, ISSN 1468-2389, Wiley, Hoboken, NJ, Vol. 31, Iss. 2, pp. 286-301,  
<https://doi.org/10.1111/ijsa.12413>

This Version is available at:

<https://hdl.handle.net/10419/288027>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc/4.0/>

# The criterion-related validity of conscientiousness in personnel selection: A meta-analytic reality check

Luc Watrin  | Lucas Weihrauch  | Oliver Wilhelm 

Department of Individual Differences and Psychological Assessment, Institute of Psychology and Education, Ulm University, Ulm, Germany

## Correspondence

Luc Watrin, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany.

Email: [luc.watrin@uni-ulm.de](mailto:luc.watrin@uni-ulm.de)

## Abstract

A key finding in personnel selection is the positive correlation between conscientiousness and job performance. Evidence predominantly stems from concurrent validation studies with incumbent samples but is readily generalized to predictive settings with job applicants. This is problematic because the extent to which faking and changes in personality affect the measurement likely vary across samples and study designs. Therefore, we meta-analytically investigated the relation between conscientiousness and job performance, examining the moderating effects of sample type (incumbent vs. applicant) and validation design (concurrent vs. predictive). The overall correlation of conscientiousness and job performance was in line with previous meta-analyses ( $\bar{r} = .17, k = 102, n = 23,305$ ). In our analyses, the correlation did not differ across validation designs (concurrent:  $\bar{r} = .18, k = 78, n = 19,132$ ; predictive:  $\bar{r} = .15, k = 24, n = 4173$ ), sample types (incumbents:  $\bar{r} = .18, k = 92, n = 20,808$ ; applicants:  $\bar{r} = .14, k = 10, n = 2497$ ), or their interaction. Critically, however, our review revealed that only a small minority of studies (~12%) were conducted with real applicants in predictive designs. Thus, barely a fraction of research is conducted under realistic conditions. Therefore, it remains an open question if self-report measures of conscientiousness retain their predictive validity in applied settings that entail faked responses. We conclude with a call for more multivariate research on the validity of selection procedures in predictive settings with actual applicants.

## KEYWORDS

conscientiousness, criterion-related validity, faking, job performance, meta-analysis

## Practitioner points

- Research on the predictive validity of conscientiousness is almost exclusively conducted with incumbents and criterion data gathered at the same time as test data.
- Such studies likely underestimate the detrimental effects of faking and personality change across time.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *International Journal of Selection and Assessment* published by John Wiley & Sons Ltd.

- Predictive studies with real applicant samples are scarce.
- Self-report personality measures should be used with caution if faking was expected.
- In general, a stronger emphasis on incremental validity instead of individual predictors is desirable.

## 1 | INTRODUCTION

The application of personality tests in personnel selection is presumably legitimized by several meta-analyses that consistently report small to moderate associations between self-report measures of personality and measures of job performance (e.g., Barrick et al., 2001; He et al., 2019; Judge et al., 2013; Salgado, 1997; Shaffer & Postlethwaite, 2012). The largest and most generalizable correlations have been reported for the personality trait conscientiousness, with raw correlations around  $\bar{r} = .15$  and disattenuated correlations around  $\rho = .24$  (e.g., He et al., 2019). To provide unbiased and generalizable estimates, however, meta-analyses must ensure that primary studies are representative of the settings towards which generalization is sought (Cooper et al., 2019). Morgeson et al. (2007b, p. 1045) rightly stated that “only studies that use a predictive model with actual job applicants should be used to support the use of personality in personnel selection.” However, reading the pertinent literature and reference sections of published meta-analyses (e.g., Judge et al., 2013; Salgado, 1997; Tett et al., 1991) conveys the impression that research primarily investigates job incumbents (or students) in concurrent designs. We assume this is not because researchers are unaware of the inadequacies of such studies but because predictive studies are often costly, and access to actual applicant samples can be difficult.

Sample and design characteristics have been discussed as important moderators of validity.<sup>1</sup> However, the evidence has remained scarce and inconclusive due to an insufficient number of primary studies to analyze (Van Iddekinge & Ployhart, 2008; see also Table 1). In the present study, we revisited this question based on a broader database covering the last 40 years. Thereby, we were less interested in the effects of sample type or study design per se. Instead, we argue that different processes operate in different samples and designs, potentially influencing personality measurement and thus criterion validity. As we will elaborate more comprehensively in the following sections, we argue that both faking (Ziegler et al., 2011) and personality change (Roberts et al., 2006; Wille et al., 2012) are more prevalent in applicant samples with predictive validation designs than in concurrent studies with incumbents.

In the present study, we first describe our view of a typical personnel selection process and derive requirements for primary studies. Next, we review the prevalence of sample types and study designs in primary studies on the validity of conscientiousness to predict job performance over the last 40 years. Finally, we

**TABLE 1** Overview of first- and second-order meta-analyses addressing design or sample characteristics of primary studies

	Concurrent/ predictive	Incumbent/ applicant
Conscientiousness and overall job performance		
Barrick and Mount (1991)	–	–
Barrick et al. (2001) <sup>a</sup>	–	–
Darr (2011)	–	Coded but not reported
Dudley et al. (2006)	–	Coded but not reported
He et al. (2019) <sup>a</sup>	–	–
Hogan and Holland (2003)	41/2	–
Hough (1992)	–	–
Hurtz and Donovan (2000)	–	–
Oh et al. (2011)	14/2	16/0
Judge et al. (2013)	–	–
Rojon et al. (2015)	–	–
Salgado (1997)	–	–
Salgado (2003)	–	–
Shaffer and Postlethwaite (2012)	112/7	117/3
Shaffer and Postlethwaite (2013)	–	–
van Aarde et al. (2017)	–	–
Wilmot and Ones (2019) <sup>a</sup>	–	–
Other/multiple constructs or alternative job-related criteria		
Bartram (2005)	21/7	–
Berry et al. (2007)	–	–
Chiaburu et al. (2011)	–	–
Huang et al. (2014)	64/1	–
Ilies et al. (2009)	–	–
Judge et al. (2002)	–	–
Lee et al. (2019)	–	–
Ones et al. (1993)	135/79	135/43
Pletzer et al. (2019)	–	–
Salgado (2002)	–	–

(Continues)

TABLE 1 (Continued)

	Concurrent/ predictive	Incumbent/ applicant
Salgado and Táuriz (2014)	–	–
Schmitt et al. (1984)	153/213	–
Tett et al. (1991)	–	81/12
Van Iddekinge et al. (2012)	38/32	47/24
Woo et al. (2014)	–	–

Note: –, design or sample characteristics were not considered.

<sup>a</sup>Second-order meta-analysis.

meta-analytically investigate the validity of conscientiousness to predict job performance. Thereby, we specifically focus on the moderating effects of sample type, study design, and, most importantly, the interaction of both moderators.

### 1.1 | Characteristics of a common personnel selection process

In our view, a typical personnel selection process has several key characteristics. First, there are more applicants than positions to be filled, making a selection based on some attributes of applicants inevitable in the first place. Second, the selection decision depends on the performance of applicants in the selection procedure (e.g., personality test). Consequently, applicants are highly motivated to perform well in the procedure and portray what they think is a favorable picture of themselves to get the job. While there are certainly exceptions to this (e.g., applications enforced by employment agencies threatening to withdraw welfare), we argue that it is prudent to assume that applicants want to receive a job offer in the vast majority of personnel selection procedures. Third, a time lag between the time of the selection process and the collection of criterion data (e.g., supervisory ratings of job performance) is inevitable. Obviously, the newly hired employee (or the old employee in a new position) must have had the opportunity to demonstrate observable behavior (e.g., sell, build, or invent things), which can serve as an indicator of job performance. In sum, then, selection processes inevitably involve real applicants and predictive validation designs. In the following, we review the prevalence of sample types (incumbents vs. applicants) and study designs (concurrent vs. predictive) in previously published meta-analyses.

### 1.2 | Incumbent versus applicant samples

Incumbents and applicants likely differ in key characteristics. Relative to job incumbents, job applicants put more effort and motivation into the tests (Arvey et al., 1990) and distort their responses to portray a

favorable image of themselves (Griffith & Converse, 2011). This is conclusive because, for job applicants, the testing process usually constitutes a high-stakes situation with far-reaching consequences for their personal and vocational life. In turn, for job incumbents with secure jobs, the consequences of poor performance are relatively less essential. This intentional distortion to portray a favorable picture of oneself to achieve personal goals has been termed “faking” (Ziegler et al., 2011). According to several studies, around 30%–50% of applicants fake personality tests (e.g., Donovan et al., 2014; Griffith et al., 2007). Therefore, the occurrence of faking must be considered the rule rather than the exception in applicant samples.

On the one hand, some authors argued that faking does not affect the validity of personality measures (e.g., Hough, 1998b; Komar et al., 2008; Tett & Simonet, 2021; Weekley et al., 2004). On the other hand, extensive empirical evidence suggests that faking affects the mean structure, the covariance structure and criterion validity of self-report measures of personality (e.g., Christiansen et al., 2021; Donovan et al., 2014; Geiger et al., 2018; Krammer et al., 2017; MacCann et al., 2017; Pauls & Crost, 2005; Schmit & Ryan, 1993). Mean shifts are commonly observed as a consequence of faking (Birkeland et al., 2006). More importantly, faking will alter the rank order of participants and thus the construct validity of self-reports and also the criterion validity of personnel selection decisions based upon such reports (e.g., Anglim et al., 2018; Donovan et al., 2014; Jeong et al., 2017). Even if faking only alters scores slightly on average, this can substantially affect top-down selection decisions (Donovan et al., 2014; Pavlov et al., 2018); if only three out of 100 applicants fake and thus prevail in the process for three vacancies, the proportion of persons hired based on invalid personality scores is 100%. Changes in rank orders occur because persons fake to a different extent because they differ in the extent to which they are willing and able to fake (Boss et al., 2015; Geiger et al., 2018, 2021; Krammer, 2020; Pavlov et al., 2018). In instructed faking studies, correlations below .50 have been reported between honest and faked personality scores (e.g., Galić & Jerneić, 2013; Ng et al., 2020; Pavlov et al., 2018) and faking seemingly affects the measurement invariance of personality tests (e.g., Krammer et al., 2017). However, changes in rank order seem to be less pronounced in applicant samples than in instructed faking studies (Hu & Connelly, 2021). If within-person correlations are weak, honest and faked personality scores may represent a jingle fallacy (Kelley, 1927), in that they are assigned the same name (e.g., conscientiousness) but what is actually measured is partly or fundamentally different. If personality, as measured in low-stakes settings, differs markedly from personality as measured in high-stakes settings, the question arises as to which kind of construct is measured in primary studies of seminal meta-analyses relating personality scores to job performance.

The type of sample has rarely been investigated or reviewed as a potential moderator of validity. In a meta-analysis, Tett et al. (1991) initially found higher validities for incumbents than for recruits, but this effect was due to a single study with an outsized influence on the effect (the sample had over 4000 military recruits). After the removal of this study, only 11 studies with 814 subjects remained, and the

difference between the sample types was found to be nonsignificant. Darr (2011), as well as Shaffer and Postlethwaite (2012), coded sample type but were unable to perform moderator analyses because the number of applicant samples was far too low (e.g., three applicant studies in Shaffer & Postlethwaite, 2012). Van Iddekinge et al. (2012) investigated the validity of self-report measures of integrity. They found descriptively lower validity estimates in applicant ( $\rho = .15$ ) than in incumbent samples ( $\rho = .20$ ), but the difference was not significant. As our broader review in Table 1 shows, most published meta-analyses on the criterion-related validity of personality measures do not consider whether the included samples comprise applicants or incumbents.

In sum, the difference in the predictive validity of personality in applicant versus incumbent samples is inconclusive, but points towards slightly lower validity estimates in applicant samples than in incumbent samples. This conclusion supposedly applies to self-report measures of personality more broadly and conscientiousness more specifically. On a side note, the lack of attention to design and sample characteristics is also prevalent in investigations of the validity of cognitive ability (Schmitt & Sinha, 2011). However, faking (good) is less of an issue in tests of maximum performance (but see Steger et al., 2018 for evidence concerning problems with unprocured online testing).

### 1.3 | Concurrent versus predictive validation designs

In concurrent validation designs, the predictor and criteria data are collected at the same time, whereas in predictive designs, criteria are collected at a later time point. Evidently, the validation design and sample are not independent—concurrent validation studies cannot be conducted with job applicants as their criterion data are not available at the time of testing. Concurrent designs are cross-sectional in nature and therefore incompatible with the overarching predictive purposes of personality testing in personnel selection, which is to predict an applicant's future performance based on the current data.

Predictive designs, in turn, exist in different variants, but all have in common that some time passes between assessing predictors and gathering criterion data (Guion & Cranny, 1982). Therefore, predictive studies can comprise both job incumbents and job applicants, but concurrent studies can only comprise job incumbents. The temporal distance in longitudinal studies can change the association of initial personality assessment and subsequent criterion measure. Personality traits have long been viewed as invariant over time. However, personality research has increasingly moved to acknowledge systematic age-related changes in personality (Roberts et al., 2006) and particularly following significant life events (Bleidorn et al., 2018; Woods et al., 2013). Besides private events (e.g., marriage and parenthood), work-related events such as entering the work force or taking on a new job role (Wille et al., 2012), have been shown to change personality traits. Thus, the trait levels at the time of hiring can

differ from the trait levels that shape behavior in the further career. If that's the case, the predictive power of initial personality measurement should decrease as a function of time since recruitment. Further, in predictive applicant samples with selection, job performance criteria can only be assessed for the (usually small) subsample of hired individuals, requiring corrections for range restriction to obtain unbiased estimates of predictive validity.

The type of design has been discussed as an important characteristic of validation studies (Van Iddekinge & Ployhart, 2008). In a 40-year-old meta-analysis, Schmitt and colleagues (1984) found minimally lower criterion-related validities for aggregates of personality scales in predictive ( $r = .30$ ) relative to concurrent designs ( $r = .34$ ), and predictive designs with selection ( $r = .26$ ). Hough (1998a) reported a predictive raw correlation of dependability and job proficiency that was .05 lower than the concurrent validity estimate, although this difference is unlikely to be significant. As Van Iddekinge and Ployhart (2008) rightly point out, this might not seem much but given that observed validity estimates are generally low, this amounts to a substantial decrease in explained variance. Because Hough (1998a) based her analyses solely on military samples, the generalizability of the results might be limited. In their meta-analysis on the validity of integrity tests, Van Iddekinge et al. (2012) did not find significant differences between predictive ( $\rho = .17$ ) and concurrent ( $\rho = .19$ ) studies.

As our review illustrates, several other meta-analyses have coded the study design of primary studies (Table 1). However, they have not reported moderator analyses due to the literature's low prevalence of predictive validation studies. Concerning sample type as moderator, the effect of the study design is inconclusive. If anything, the extant literature suggests slightly lower validity estimates in predictive than in concurrent designs.

### 1.4 | The present study

Taken together, there is a discrepancy between the alleged awareness that only predictive studies with real applicants are suitable to investigate the validity of self-report measures of personality in personnel selection and common practice. In fact, barely any previous meta-analyses had a sufficient number of samples to investigate validity under what we would call realistic conditions. The present study aims to fill this gap.

## 2 | METHODS

### 2.1 | Literature search

The literature search was conducted in July 2020. We searched OVID (PsyArticles, PsyINFO, and PSYNDEX), Sage Journals (Social Sciences and Humanities), ScienceDirect, Web of Science (Core Collection without Chemical Indexes) and the reference sections from the meta-

analyses reported in Table 1 for journal articles written in English and published between 1980 and 2020. To be inclusive, we first developed a broad taxonomy of conscientiousness (see Appendix A). Terms from this taxonomy were subsequently searched for in combination with "job performance," "work performance," "performance rating," or "overall performance" (see OS1). We searched in titles, abstracts, and keywords. After the removal of duplicates, the initial search led to 10,713 articles. Figure B1 (Appendix B) provides a PRISMA chart of the literature search.

## 2.2 | Inclusion and exclusion criteria

Concerning the predictor, we included studies that reported at least one self-report measure of conscientiousness or a facet thereof. We restricted the analysis to conscientiousness because it is supposedly the most important personality factor in personnel selection and it is presumably the most studied personality factor with the richest database. We excluded forced-choice measures from the analysis for methodological reasons (Brown & Maydeu-Olivares, 2013; Bürkner et al., 2019). Regarding the criterion, studies had to report a measure of overall job performance provided by a supervisor. Other sources (e.g., peers) or types of performance (e.g., OCB) were not included. Studies had to include some estimate of the association between the predictor and the criterion that could be transformed to a bivariate correlation. We only included studies that reported individual-participant data and excluded studies that reported group-level or unit-level analyses. Concerning sample and design, we included concurrent and predictive studies with job applicants or job incumbents conducted in a field (i.e., organizational) setting. Simulation studies or studies with student samples were not considered. The comprehensive coding manual is provided in OS2.

## 2.3 | Data extraction

Data screening and extraction were performed in three steps. In a preliminary step, one rater assessed all articles for general viability. This was necessary as the breadth of the search terms led to a multitude of articles from other fields within (e.g., psychotherapy) and outside of psychology (e.g., chemistry). Next, three additional raters underwent training and subsequently helped to screen the remaining articles based on their abstract. Finally, 360 studies were considered for full-text screening, of which 92 articles were included in the final analysis. All articles were double-coded by four trained psychology students; one rater coded all studies, and three other raters coded a subset. The proportion agreement for the main categorical variables ranged between 0.92 and 1.00. The intra-class correlation coefficient for the main continuous variables was between 0.92 and 1.00 (see OS3 for comprehensive tables of reliability estimates for all variables). Discrepancies were discussed and resolved before further analysis.

## 2.4 | Data analytic strategy

Unreliability in the measurement, as well as range restriction, can attenuate the observed predictive validity estimates. Comprehensive information about reliability estimates in the initial sample, the selected sample, and selection ratios are necessary to adequately correct for such attenuating factors (Sackett et al., 2021). Unfortunately, most primary studies did not report sufficient information to perform suitable corrections for either reliability or range restriction. In fact, only three studies reported selection ratios. Therefore, as we concur with Sackett et al. (2021) evaluation that common correction methods tend to inflate validity estimates if based on insufficient data, and their principle of conservative estimation, we chose only to report raw estimates.

When studies reported multiple measures of the same construct (e.g., two measures of conscientiousness which could be classified under the same facet of conscientiousness), we computed composite scores (Schmidt & Hunter, 2015; as implemented in the `composite_r_scalar` function of the `psychmeta` package). Correlations were then transformed to Fisher's  $z$  values, which were the basis for the random-effects meta-analyses. For ease of interpretation, final estimates were converted back into correlation coefficients (Cooper et al., 2019). We quantified heterogeneity and uncertainty in effect sizes using  $\tau^2$  using the restricted maximum likelihood estimator (REML), Higgins  $I^2$ , and prediction intervals (Cooper et al., 2019; IntHout et al., 2016). When testing individual coefficients, we used the method proposed by Knapp and Hartung (2003) to adjust standard errors. To identify outliers and influential studies, we used Baujat plots (Baujat et al., 2002) and influence plots (Viechtbauer & Cheung, 2010). Publication bias was investigated via funnel plots and Egger's regression test (Sterne & Egger, 2005).

All analyses were performed in R (version 4.0.4; R Core Team, 2020). Data handling and visualization were performed with packages of the `tidyverse` (version 1.3.1, Wickham et al., 2019). Descriptive and psychometric statistics were computed with the `summarytools` package (version 0.9.9; Comtois, 2021) and the `psych` package (version 2.0.12; Revelle, 2020), interrater reliability was computed with the `psych` package and the `irr` package (version 0.84.1; Gamer et al., 2019), composite scores were computed with the `psychmeta` package (version 2.4.2; Dahlke & Wiernik, 2019), and the meta-analysis was performed with the `metafor` package (version 2.4-0; Viechtbauer, 2010). All data, syntax, and materials are available at <https://osf.io/87gyr/>.

## 3 | RESULTS

We identified 132 effect sizes from 115 unique samples reported in 91 articles with a total of 32,499 participants. As Table 2 illustrates, the vast majority of correlations were for the association of a global measure of conscientiousness with job performance in concurrent studies with job incumbents. Given the low number of studies for

facets of conscientiousness and levels of the main moderators, we restrained all further analyses to the global measure of conscientiousness. Two studies were excluded due to the outlier/influence analysis (we report all results, including the two studies, in OS4). Neither the funnel plots nor Egger's regression test indicated meaningful publication bias (see OS5).

**TABLE 2** Frequencies of central variables

	<i>k</i>	%
Conscientiousness		
Global	104	78.79
Orderliness	6	4.55
Industriousness	15	11.36
Self-Control	3	2.27
Responsibility	4	3.03
Study design × sample type		
Concurrent	97	73.48
Incumbent	93	70.40
Internal applicant	4	3.03
External applicant	0	0
Predictive	35	26.52
Incumbent	20	15.20
Internal applicant	1	1.33
External applicant	14	10.60

Note: *k*, number of correlations.

### 3.1 | Overall effect

In the first step, we estimated the overall association of conscientiousness with job performance as a replication of previous meta-analyses. The raw estimate  $\bar{r} = .17$  [.15, .19] was in line with previously reported validity estimates (e.g., Barrick et al., 2001; He et al., 2019; Judge et al., 2013; Shaffer & Postlethwaite, 2012; Wilmot & Ones, 2019). All indices of heterogeneity indicated a fair amount of heterogeneity among the true effects ( $Q = 216.6$ ,  $df = 101$ ,  $p < .001$ ;  $\tau = .07$ ;  $I^2 = 52.61$ ). Accordingly, we investigated sample, design, and the interaction of both as sources of heterogeneity in a moderation analysis (Table 3).

### 3.2 | Moderator analyses

#### 3.2.1 | Sample type

To test the moderating effect of sample type, we estimated a meta-regression where sample type was dummy-coded and the applicant samples served as the reference group. Estimates were  $\bar{r} = .14$  [−.01, .29] for applicants and  $\bar{r} = .18$  [.04, .31] for incumbents, respectively. An omnibus test indicated that the small difference between effect sizes for the applicant and incumbent samples was not significant,  $F(1, 100) = 1.03$ ,  $p = .31$ .

#### 3.2.2 | Study design

The moderating effect of the study design was tested analogously. The concurrent designs served as the reference group. Estimates were

**TABLE 3** Meta-analysis and moderators of the validity of conscientiousness to predict overall job performance

Model	<i>k</i>	<i>n</i>	$\bar{r}$	95% CI		95% PI		<i>Q</i>	$\tau$	$I^2$
				LB	UB	LB	UB			
Overall	102	23,305	.17	.15	.19	.03	.31	216.6**	.07	52.61
Sample								210.1**	.07	52.06
Applicants	10	2497	.14	.08	.20	−.01	.29			
Incumbents	92	20,808	.18	.16	.20	.04	.31			
Design								212.9**	.07	52.43
Predictive	24	4173	.15	.11	.20	.01	.29			
Concurrent	78	19,132	.18	.16	.20	.04	.31			
Sample × Design								209.9**	.07	53.06
Appl./Pred.	9	2497	.16	.10	.22	.01	.31			
Incumb./Pred.	15	1902	.14	.08	.20	−.01	.29			
Incumb./Concur.	77	19,336	.18	.16	.20	.04	.31			

Abbreviations: 95% CI, 95% confidence interval;  $I^2$ , ratio of true heterogeneity to total variation; *k*, number of correlations; LB, lower bound; *n*, total sample size; 95% PI, 95% prediction interval; *Q*, homogeneity statistic;  $\bar{r}$ , mean estimate of uncorrected correlations;  $\tau^2$ , variance of the true effect sizes; UB, upper bound.

\*\* $p < .001$ .



$\bar{r} = .18$  [.04, 0.31] for concurrent designs and  $\bar{r} = .15$  [.01, 0.29] for predictive designs, respectively. An omnibus test indicated that the small descriptive difference between effect sizes between the applicant and incumbent samples was not significant,  $F(1, 100) = .97, p = .33$ .

### 3.2.3 | Sample type $\times$ study design

Ultimately, we tested the interaction of sample type and study design to investigate one of the main questions of this study: do estimates of validity generalize to predictive studies with applicant samples, as are encountered in real-life selection procedures? Thereto, we performed a moderator analysis with three groups: concurrent/incumbent, predictive/incumbent, and predictive/applicant. The effect sizes were  $\bar{r} = .18$  [.04, 0.31] for concurrent designs with incumbents,  $\bar{r} = .14$  [-.01, .29] for predictive designs with incumbents and  $\bar{r} = .16$  [.01, .31] for predictive designs with applicants, respectively. An omnibus test again indicated no significant differences between the groups,  $F(2, 98) = .58, p = .56$ .

## 4 | DISCUSSION

The current study reviewed the prevalence of key design and sample characteristics of studies investigating the validity of self-report measures of conscientiousness to predict job performance and to quantify their impact on the validity estimates. Overall, the correlation of conscientiousness with job performance was in line with previously reported estimates (e.g., He et al., 2019) and moderator analyses did not reveal differences across types of samples (incumbent vs. applicant) or validation designs (concurrent vs. predictive). However, the overwhelming majority of research is conducted in concurrent designs with job incumbents. Only 12% of the studies published in the last 40 years investigated job applicants in predictive designs. Thus, common research practice is at odds with requests to conduct research with samples that match the population to which results should apply (Morgeson et al., 2007b). We, therefore, caution against interpreting the present findings lightheartedly - based on available evidence, questions concerning the predictive validity of self-report questionnaires of conscientiousness cannot be answered reliably. Therefore, it remains an open question whether or not self-report measures of personality retain their predictive validity in applied settings that entail faked responses. In the following, we discuss our concerns regarding faking in high-stakes personality testing, suggest a stronger emphasis on incremental predictive validity, and propose guidelines for future research.

### 4.1 | Faking as key issue in validity of self-reports

At first sight, the present results are encouraging. Contrary to formerly raised concerns (e.g., Van Iddekinge & Ployhart, 2008), the validity of self-report measures of conscientiousness to predict job

performance does not seem to differ meaningfully in applicant versus incumbent samples or concurrent versus predictive validation designs. Yet, these results need to be reconciled with the extensive evidence that applicants do fake to a substantial degree (Griffith & Converse, 2011), that individuals differ in the extent they are willing and able to fake successfully (e.g., Geiger et al., 2018, 2021; Kleinmann et al., 2011; Pavlov et al., 2018) and that, as a consequence, faking substantially alters the rank order of participants (e.g., Griffith et al., 2007; Krammer, 2020). Changes in rank orders imply that the personality constructs assessed via self-report measures are not measuring the same underlying disposition and are not measurement invariant between honest and faking conditions (e.g., Krammer et al., 2017). This affects the construct validity of personality tests—what is measured under faking might differ fundamentally compared to the construct measured under honest conditions. We acknowledge that the amount of faking observed in applied contexts is likely lower or more subtle than in laboratory studies (Birkeland et al., 2006; Hu & Connelly, 2021), but it would be naive to assume that faking is not highly prevalent, particularly in high-stakes contexts with well-educated and prepared applicants.

The variance captured by faked self-report measures of personality still predicts job performance, but we question if that is still predominantly the same variance captured in low-stakes settings. Recent research suggests that faked personality scores reflect individual differences in the ability to fake successfully (i.e., achieve high scores in relevant personality traits) to some degree, which has been explained with the ability to identify criteria, i.e., the ability to identify the targeted selection criteria (e.g., Klehe et al., 2012). The ability to fake successfully has been linked to individual differences in fluid intelligence, crystallized intelligence, and interpersonal abilities (Geiger et al., 2018, 2021). Critically, all these abilities are correlated with job performance (Schmidt & Hunter, 1998). This perspective would reconcile findings that faking fundamentally affects self-report measures of personality but doesn't invariably lead to decrements in criterion validity.

While faking seems to be ignored or deemed irrelevant in validation studies, there is a vast body of research that illustrates its prevalence and investigates methods to prevent it (Ziegler et al., 2011). Among them, the forced-choice (FC) format enjoys continued popularity due to its presumed resistance against faking. Yet, we decided to omit forced-choice measures from the current review. We did so because FC response formats function fundamentally differently than the widely applied single stimulus (e.g., Likert) measures. Due to their relative nature, conventionally scored FC measures result in (quasi-) ipsative scores which prohibit interindividual comparisons (Brown et al., 2013; Hicks, 1970). There is an ongoing controversy and active research on how to best construct FC measures (e.g., Watrin et al., 2019), how to make FC measures faking resistant (e.g., Pavlov et al., 2021) and under which conditions scores from FC measures are valid (e.g., Bürkner, 2022). Taken together, FC format might have the potential to reduce detrimental effects of faking. However, a cascade of questions concerning the validity of FC personality tests in general and in the selection more specifically should be studied in a separate meta-analysis first.



## 4.2 | Call for a stronger focus on incremental validity

Because faked personality scores might capture variance other than honest personality, it is crucial to investigate the validity jointly with competing constructs to obtain meaningful estimates of predictive and, more importantly, incremental predictive validity. Personnel selection is rarely based on a single source of information. Instead, the aim is to obtain relevant and complementary pieces of valid information about the applicant and to combine this information in a way that maximizes the probability of valid inferences about future job performance. Comprehensive meta-analyses provide evidence on which dispositions are suitable for this purpose - first and foremost tests of cognitive abilities (e.g., Schmidt & Hunter, 1998). Conscientiousness has been shown to provide a small but significant amount of incremental validity above cognitive ability (Schmidt & Hunter, 1998). However, this estimate hinges upon the low correlation of conscientiousness and cognitive ability, which was estimated in samples where faking was unlikely to be an issue (e.g., Judge & Jackson, Shaw, et al., 2007; Table 3). Given the recent evidence that conscientiousness and cognitive ability are more strongly correlated under faking (Schilling et al., 2020), and that the construct validity of self-report measures of conscientiousness are potentially affected by faking, we need improved meta-analytic estimates of the correlation of both constructs and estimates of incremental validity under faking. However, this would presuppose enough primary studies in which both conscientiousness and cognitive ability were measured in high-stakes settings with applicants in predictive designs. Such studies would be a subset of all predictive studies with applicants, but none was present in our review, which shows a clear need for research.

## 4.3 | Limitations and future directions

The generalizability of the current results is limited due to the small number of studies with real applicants and predictive validation designs. Either such studies are rarely conducted, or their results are not being published. The former is not very likely given the prominence of conscientiousness in I&O psychological research and the ease with which such questionnaires can be applied in a wider variety of personnel selection settings. Companies offering personality tests for personnel selection can pursue a number of incentives by virtue of gathering evidence in support of their products. It is therefore likely, that either publishing results does not rank highly in all companies' reward agenda or results are actively withheld from going public. Personal communications with practitioners and anecdotal reports, such as in Van Iddekinge et al. (2012), suggest that the latter is certainly prevalent. As a consequence, there is reason for concern that unpublished results would depress the validity of the results collected here.

Given the far-reaching personal and economic consequences of personnel selection decisions, we call for more publications on the validity of self-report measures of personality under realistic conditions. Such publications should meet several criteria. First, they should comprise real job applicants who respond to personality tests in a high-stakes situation. Second, job performance of recruited applicants should be evaluated at a later stage based on sufficient valid performance data. Third, preferably personality and performance data should also be gathered from job incumbents. Using a matching procedure—as proposed by Jeong et al. (2017), for example—allows to compare validity estimates for applicants and incumbents. Fourth, available additional measures should be included for evaluating incremental validity. Fifth, commitments to quality standards for ensuring study quality should be default in such studies (e.g., DIN33430; ISO10667). Additionally, validation studies should be preregistered. In all likelihood, just like in other disciplines in (e.g., see Dechartres et al., 2016; for the systematic influence of mandatory preregistration on results in medicine) a bias of results in studies with conflicts of interests is to be expected. Sixth, such publications should adhere to open sciences practices. Sophisticated methods of de-identification and synthetic data procedures have been proposed that address legitimate privacy concerns while ensuring the ability to verify results (Grund et al., 2022; Walsh et al., 2018). If more studies meeting these requirements were available, a more dependable verdict concerning the predictive validity of self-report measures of personality and its moderators were possible.

## 5 | CONCLUSION

In sum, the current meta-analysis leaves us with an ambivalent view on the predictive validity of self-reported measures of conscientiousness. On the one hand, validity estimates were comparable between presumably low-stakes and high-stakes settings. On the other hand, their magnitude must still be considered low (Morgeson et al., 2007a) and the number of studies comprising real job applicants and predictive validation designs is frustratingly low. Research is predominantly conducted in settings that do not allow investigating the greatest threat to predictive validity: faking. Thus, the present results are preliminary and call for more multivariate studies including competing methods of selection in real high-stakes personnel selection processes to answer two important questions: how and to which degree does faking affect the predictive validity of self-report measures of personality in applied contexts and, more importantly, how does faking affect the incremental validity above and beyond established ability constructs with significantly higher criterion-related validity?

## ACKNOWLEDGMENTS

This research received no specific grant from any funding agency. We thank Anna Kaisinger, Raimund Krämer, Natasa Subotin, and Tim Trautwein for their help with coding studies and Sally Olderbak for

her methodological advice. Open Access funding enabled and organized by Projekt DEAL.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at <https://osf.io/87gyr/>

## ORCID

Luc Watrin  <http://orcid.org/0000-0003-4343-3781>

Lucas Weihrauch  <http://orcid.org/0000-0002-4665-8567>

Oliver Wilhelm  <http://orcid.org/0000-0001-7980-1166>

## ENDNOTE

<sup>1</sup> If not stated otherwise, we will use the term validity to refer to criterion-related validity throughout this article.

## REFERENCES

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

- \*Abbas, M., & Raja, U. (2019). Challenge-hindrance stressors and job outcomes: The moderating role of conscientiousness. *Journal of Business and Psychology*, 34(2), 189–201. <https://doi.org/10.1007/s10869-018-9535-z>
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education*, 23(2), 311–321. <https://doi.org/10.1007/s10459-017-9796-8>
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695–716. <https://doi.org/10.1111/j.1744-6570.1990.tb00679.x>
- \*Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology*, 17(1), 87–105. <https://doi.org/10.1023/A:1016200317002>
- \*Barling, J., & Boswell, R. (1995). Work performance and the achievement–strivings and impatience–irritability dimensions of type A behaviour. *Applied Psychology*, 44(2), 143–153. <https://doi.org/10.1111/j.1464-0597.1995.tb01071.x>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- \*Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78(1), 111–118. <https://doi.org/10.1037/0021-9010.78.1.111>
- \*Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81(3), 261–272. <https://doi.org/10.1037/0021-9010.81.3.261>
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1–2), 9–30. <https://doi.org/10.1111/1468-2389.00160>
- \*Barrick, M. R., Mount, M. K., & Strauss, J. P. (1994). Antecedents of involuntary turnover due to a reduction in force. *Personnel Psychology*, 47(3), 515–535. <https://doi.org/10.1111/j.1744-6570.1994.tb01735.x>
- \*Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87(1), 43–51. <https://doi.org/10.1037/0021-9010.87.1.43>
- \*Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for retention and performance. *Human Resource Management*, 48(2), 183–206. <https://doi.org/10.1002/hrm.20275>
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90(6), 1185–1203. <https://doi.org/10.1037/0021-9010.90.6.1185>
- Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, 21, 2641–2652. <https://doi.org/10.1002/sim.1221>
- \*Van den berg, P. T., & Feij, J. A. (2003). Complex relationships among personality traits, job characteristics, and work behaviors. *International Journal of Selection and Assessment*, 11(4), 326–339. <https://doi.org/10.1111/j.0965-075X.2003.00255.x>
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92(2), 410–424. <https://doi.org/10.1037/0021-9010.92.2.410>
- \*Bing, M. N., & Burroughs, S. M. (2001). The predictive and interactive effects of equity sensitivity in teamwork-oriented organizations. *Journal of Organizational Behavior*, 22(3), 271–290. <https://doi.org/10.1002/job.68>
- \*Bing, M. N., & Lounsbury, J. W. (2000). Openness and job performance in U.S.-based Japanese manufacturing companies. *Journal of Business and Psychology*, 14(3), 515–522. <https://doi.org/10.1023/A:1022940519157>
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Bleidorn, W., Hopwood, C. J., & Lucas, R. E. (2018). Life events and personality trait change: Life events and trait change. *Journal of Personality*, 86(1), 83–96. <https://doi.org/10.1111/jopy.12286>
- \*Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology*, 76(6), 863–872. <https://doi.org/10.1037/0021-9010.76.6.863>
- Boss, P., König, C. J., & Melchers, K. G. (2015). Faking good and faking bad among military conscripts. *Human Performance*, 28(1), 26–39. <https://doi.org/10.1080/08959285.2014.974758>
- Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, 18(1), 36–52. <https://doi.org/10.1037/a0030641>
- Bürkner, P.-C. (2022). On the information obtainable from comparative judgments. *Psychometrika*, 87, 1439–1472. <https://doi.org/10.1007/s11336-022-09843-z>
- Bürkner, P.-C., Schulte, N., & Holling, H. (2019). On the statistical and practical limitations of Thurstonian IRT models. *Educational and Psychological Measurement*, 79(5), 827–854. <https://doi.org/10.1177/0013164419832063>
- \*Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M.-C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, 99(4), 564–586. <https://doi.org/10.1037/a0034688>
- \*Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15(3), 233–254. [https://doi.org/10.1207/S15327043HUP1503\\_01](https://doi.org/10.1207/S15327043HUP1503_01)

- Chiaburu, D. S., Oh, I.-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 96(6), 1140–1166. <https://doi.org/10.1037/a0024004>
- Christiansen, N. D., Robie, C., Burns, G. N., Loy, R. W., Speer, A. B., & Jacobs, R. R. (2021). Effects of applicant response distortion on the relationship between personality trait scores and cognitive ability. *Personality and Individual Differences*, 171, 110542. <https://doi.org/10.1016/j.paid.2020.110542>
- \*Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86(3), 410–417. <https://doi.org/10.1037/0021-9010.86.3.410>
- \*Colbert, A. E., & Witt, L. A. (2009). The role of goal-focused leadership in enabling the expression of conscientiousness. *Journal of Applied Psychology*, 94(3), 790–796. <https://doi.org/10.1037/a0014187>
- Comtois, D. (2021). *Summarytools: Tools to quickly and neatly summarize data. R package version 0.9.9*. <https://CRAN.R-project.org/package=summarytools>
- \*Conte, J. M., & Gintoft, J. N. (2005). Polychronicity, Big Five personality dimensions, and sales performance. *Human Performance*, 18(4), 427–444. [https://doi.org/10.1207/s15327043hup1804\\_8](https://doi.org/10.1207/s15327043hup1804_8)
- \*Conte, J. M., & Jacobs, R. R. (2003). Validity evidence linking polychronicity and Big Five personality dimensions to absence, lateness, and supervisory performance ratings. *Human Performance*, 16(2), 107–129. [https://doi.org/10.1207/S15327043HUP1602\\_1](https://doi.org/10.1207/S15327043HUP1602_1)
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *Handbook of Research Synthesis and Meta-analysis* (3rd ed.). Russell Sage Foundation.
- Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. University of Illinois Press.
- \*Cucina, J. M., Vasilopoulos, N. L., Su, C., Busciglio, H. H., Cozma, I., DeCostanza, A. H., Martin, N. R., & Shaw, M. N. (2019). The effects of empirical keying of personality measures on faking and criterion-related validity. *Journal of Business and Psychology*, 34(3), 337–356. <https://doi.org/10.1007/s10869-018-9544-y>
- 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>
- Darr, W. (2011). Military personality research: A meta-analysis of the self description inventory. *Military Psychology*, 23, 272–296. <https://doi.org/10.1080/08995605.2011.570583>
- Davison, H. K., Kluemper, D. H., Tao, S., Stewart, D. W., & Bing, M. (2021). The effects of faking on the relationship between cognitive ability and conscientiousness: A cautionary note. *International Journal of Selection and Assessment*, 1–16. <https://doi.org/10.1111/ijsa.12319>
- \*Day, D. V., & Bedeian, A. G. (1991). Predicting job performance across organizations: The interaction of work orientation and psychological climate. *Journal of Management*, 17(3), 589–600. <https://doi.org/10.1177/014920639101700304>
- Dechartres, A., Ravaud, P., Atal, I., Riveros, C., & Boutron, I. (2016). Association between trial registration and treatment effect estimates: A meta-epidemiological study. *BMC Medicine*, 14(1), 100. <https://doi.org/10.1186/s12916-016-0639-x>
- \*DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment*, 15(1), 30–39. <https://doi.org/10.1111/j.1468-2389.2007.00365.x>
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, 29(3), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40–57. <https://doi.org/10.1037/0021-9010.91.1.40>
- \*Fallon, J. D., Avis, J. M., Kudisch, J. D., Gornet, T. P., & Frost, A. (2000). Conscientiousness as a predictor of productive and counter-productive behaviors. *Journal of Business and Psychology*, 15(2), 339–349. <https://doi.org/10.1023/A:1007880203956>
- \*Farh, C. I. C. C., Seo, M.-G., & Tesluk, P. E. (2012). Emotional intelligence, teamwork effectiveness, and job performance: The moderating role of job context. *Journal of Applied Psychology*, 97(4), 890–900. <https://doi.org/10.1037/a0027377>
- \*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>
- \*Forero, C. G., Gallardo-Pujol, D., Maydeu-Olivares, A., & Andrés-Pueyo, A. (2009). A longitudinal model for predicting performance of police officers using personality and behavioral data. *Criminal Justice and Behavior*, 36(6), 591–606. <https://doi.org/10.1177/0093854809333406>
- Galić, Z., & Jerneić, Ž. (2013). Measuring faking on five-factor personality questionnaires: The usefulness of communal and agentic management scales. *Journal of Personnel Psychology*, 12(3), 115–123. <https://doi.org/10.1027/1866-5888/a000087>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. <https://CRAN.R-project.org/package=irr>
- Geiger, M., Bärwaldt, R., & Wilhelm, O. (2021). The good, the bad, and the clever: Faking ability as a socio-emotional ability. *Journal of Intelligence*, 9(1), 13. <https://doi.org/10.3390/jintelligence9010013>
- Geiger, M., Olderbak, S., Sauter, R., & Wilhelm, O. (2018). The “g” in faking: Doublethink the validity of personality self-report measures for applicant selection. *Frontiers in Psychology*, 9, 1–15. <https://doi.org/10.3389/fpsyg.2018.02153>
- \*Glaser, L., Stam, W., & Takeuchi, R. (2016). Managing the risks of proactivity: A multilevel study of initiative and performance in the middle management context. *Academy of Management Journal*, 59(4), 1339–1360. <https://doi.org/10.5465/amj.2014.0177>
- \*Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology*, 81(6), 746–756. <https://doi.org/10.1037/0021-9010.81.6.746>
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341–355. <https://doi.org/10.1108/00483480710731310>
- Griffith, R. L., & Converse, P. D. (2011). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment*. Oxford University Press.
- Grund, S., Lüdtke, O., & Robitzsch, A. (2022). Using synthetic data to improve the reproducibility of statistical results in psychological research. *Psychological Methods*. <https://doi.org/10.1037/met0000526>
- 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>
- \*Gupta, N., Ganster, D. C., & Kepes, S. (2013). Assessing the validity of sales self-efficacy: A cautionary tale. *Journal of Applied Psychology*, 98(4), 690–700. <https://doi.org/10.1037/a0032232>
- \*Hakstian, A. R., Scratchley, L. S., MacLeod, A. A., Tweed, R. G., & Siddarth, S. (1997). Selection of telemarketing employees by standardized assessment procedures. *Psychology & Marketing*, 14(7), 703–726. [https://doi.org/10.1002/\(SICI\)1520-6793\(199710\)14:7%3C703::AID-MAR5%3E3.0.CO;2-K](https://doi.org/10.1002/(SICI)1520-6793(199710)14:7%3C703::AID-MAR5%3E3.0.CO;2-K)

- \*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: Predictive validity of transit operator performance. *International Journal of Selection and Assessment*, 23(2), 191–195. <https://doi.org/10.1111/ijsa.12107>
- \*He, H., Wang, W., Zhu, W., & Harris, L. (2015). Service workers' job performance: The roles of personality traits, organizational identification, and customer orientation. *European Journal of Marketing*, 49(11/12), 1751–1776. <https://doi.org/10.1108/EJM-03-2014-0132>
- He, Y., Donnellan, M. B., & Mendoza, A. M. (2019). Five-factor personality domains and job performance: A second order meta-analysis. *Journal of Research in Personality*, 82, 103848. <https://doi.org/10.1016/j.jrp.2019.103848>
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74(3), 167–184. <https://doi.org/10.1037/h0029780>
- \*Ho, J., & Nesbit, P. L. (2014). Self-Leadership in a Chinese context: Work outcomes and the moderating role of job autonomy. *Group & Organization Management*, 39(4), 389–415. <https://doi.org/10.1177/1059601114539389>
- \*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>
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100–112. <https://doi.org/10.1037/0021-9010.88.1.100>
- \*Holtrop, D., Born, M. P., & de Vries, R. E. (2014). Predicting performance with contextualized inventories, no frame-of-reference effect? *International Journal of Selection and Assessment*, 22(2), 219–223. <https://doi.org/10.1111/ijsa.12071>
- Hough, L. M. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance*, 5(1–2), 139–155. <https://doi.org/10.1080/08959285.1992.9667929>
- Hough, L. M. (1998a). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection*. Erlbaum.
- Hough, L. M. (1998b). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11(2–3), 209–244. <https://doi.org/10.1080/08959285.1998.9668032>
- Hu, J., & Connelly, B. S. (2021). Faking by actual applicants on personality tests: A meta-analysis of within-subjects studies. *International Journal of Selection and Assessment*, 1–15. <https://doi.org/10.1111/ijsa.12338>
- Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology*, 99(1), 162–179. <https://doi.org/10.1037/a0034285>
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85(6), 869–879. <https://doi.org/10.1037/0021-9010.85.6.869>
- \*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>
- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, 61(4), 871–925. <https://doi.org/10.1111/j.1744-6570.2008.00133.x/full>
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499–530. <https://doi.org/10.1037/a0021196>
- Ilies, R., Fulmer, I. S., Spitzmuller, M., & Johnson, M. D. (2009). Personality and citizenship behavior: The mediating role of job satisfaction. *Journal of Applied Psychology*, 94(4), 945–959. <https://doi.org/10.1037/a0013329>
- \*Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2013). Examining the psychometric properties of the Mayer-Salovey-Caruso Emotional Intelligence Test. *European Journal of Psychological Assessment*, 29(2), 121–128. <https://doi.org/10.1027/1015-5759/a000132>
- Int'Hout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7), e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
- \*Jackson, C. J., Hobman, E. V., Jimmieson, N. L., & Martin, R. (2009). Comparing different approach and avoidance models of learning and personality in the prediction of work, university, and leadership outcomes. *British Journal of Psychology*, 100(2), 283–312. <https://doi.org/10.1348/000712608X322900>
- Jeong, Y. R., Christiansen, N. D., Robie, C., Kung, M.-C., & Kinney, T. B. (2017). Comparing applicants and incumbents: Effects of response distortion on mean scores and validity of personality measures: RA JEONG et al. *International Journal of Selection and Assessment*, 25(3), 311–315. <https://doi.org/10.1111/ijsa.12182>
- \*Johnson, M. K., Rowatt, W. C., & Petrini, L. (2011). A new trait on the market: Honesty-humility as a unique predictor of job performance ratings. *Personality and Individual Differences*, 50(6), 857–862. <https://doi.org/10.1016/j.paid.2011.01.011>
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4), 765–780. <https://doi.org/10.1037/0021-9010.87.4.765>
- \*Judge, T. A., & Erez, A. (2007). Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *Personnel Psychology*, 60(3), 573–596. <https://doi.org/10.1111/j.1744-6570.2007.00084.x>
- Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, 92(1), 107–127. <https://doi.org/10.1037/0021-9010.92.1.107>
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98(6), 875–925. <https://doi.org/10.1037/a0033901>
- Kelley, T. L. (1927). *Interpretation of educational measurements*. World Press.
- Klehe, U.-C., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P. A., & Lievens, F. (2012). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance*, 25(4), 273–302. <https://doi.org/10.1080/08959285.2012.703733>
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review*, 1(2), 128–146. <https://doi.org/10.1177/2041386610387000>
- 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>



- Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, 93(1), 140–154. <https://doi.org/10.1037/0021-9010.93.1.140>
- Krammer, G. (2020). Applicant faking of personality inventories in college admission: Applicants' shift from honest responses is unsystematic and related to the perceived relevance for the profession. *Journal of Personality Assessment*, 102(6), 758–769. <https://doi.org/10.1080/00223891.2019.1644342>
- Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523. <https://doi.org/10.1080/00223891.2017.1285781>
- \*LaHuis, D. M., Martin, N. R., & Avis, J. M. (2005). Investigating nonlinear conscientiousness-job performance relations for clerical employees. *Human Performance*, 18(3), 199–212. [https://doi.org/10.1207/s15327043hup1803\\_1](https://doi.org/10.1207/s15327043hup1803_1)
- \*Lanyon, R. I., Goodstein, L. D., & Wershba, R. (2014). Good impression' as a moderator in employment-related assessment: 'good impression' as a moderator in pre-employment. *International Journal of Selection and Assessment*, 22(1), 52–61. <https://doi.org/10.1111/ijsa.12056>
- Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, 104(12), 1535–1546. <https://doi.org/10.1037/apl0000421>
- \*Lee, Y., Stettler, A., & Antonakis, J. (2011). Incremental validity and indirect effect of ethical development on work performance. *Personality and Individual Differences*, 50(7), 1110–1115. <https://doi.org/10.1016/j.paid.2011.01.036>
- \*Lin, W., Ma, J., Wang, L., & Wang, M. (2015). A double-edged sword: The moderating role of conscientiousness in the relationships between work stressors, psychological strain, and job performance. *Journal of Organizational Behavior*, 36(1), 94–111. <https://doi.org/10.1002/job.1949>
- \*Lounsbury, J. W., Gibson, L. W., & Hamrick, F. L. (2003). The development and validation of a personological measure of work drive. *Journal of Business and Psychology*, 18(4), 427–451. <https://doi.org/10.1023/B:JOBU.0000028445.29004.d1>
- \*Luuk, K., Luuk, A., & Aluoja, A. (2009). Predicting professional success of air traffic control personnel from their personality profile at admission to ab initio training. *The International Journal of Aviation Psychology*, 19(3), 235–251. <https://doi.org/10.1080/10508410902983896>
- MacCann, C., Pearce, N., & Jiang, Y. (2017). The general factor of personality is stronger and more strongly correlated with cognitive ability under instructed faking. *Journal of Individual Differences*, 38(1), 46–54. <https://doi.org/10.1027/1614-0001/a000221>
- \*Minbashian, A., Bright, J. E. H., & Bird, K. D. (2010). A comparison of artificial neural networks and multiple regression in the context of research on personality and work performance. *Organizational Research Methods*, 13(3), 540–561. <https://doi.org/10.1177/1094428109335658>
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683–729. <https://doi.org/10.1111/j.1744-6570.2007.00089.x>
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60(4), 1029–1049. <https://doi.org/10.1111/j.1744-6570.2007.00100.x>
- \*Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology*, 79(2), 272–280. <https://doi.org/10.1037/0021-9010.79.2.272>
- \*Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management*, 25(5), 707–721. <https://doi.org/10.1177/014920639902500505>
- \*Muchinsky, P. M. (1993). Validation of personality constructs for the selection of insurance industry employees. *Journal of Business and Psychology*, 7(4), 475–482. <https://doi.org/10.1007/BF01013760>
- Ng, V., Lee, P., Ho, M.-H. R., Kuykendall, L., Stark, S., & Tay, L. (2020). The development and validation of a multidimensional forced-choice format character measure: Testing the Thurstonian IRT approach. *Journal of Personality Assessment*, 1–14. <https://doi.org/10.1080/00223891.2020.1739056>
- Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96(4), 762–773. <https://doi.org/10.1037/a0021832>
- \*O'Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2010). Test-taking motivation and personality test validity. *Journal of Personnel Psychology*, 9(3), 117–125. <https://doi.org/10.1027/1866-5888/a000012>
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78(4), 679–703. <https://doi.org/10.1037/0021-9010.78.4.679>
- \*Ones, D. S., Viswesvaran, C., & Dilchert, S. (2017). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson, & O. Voskuil (Eds.), *The Blackwell handbook of personnel selection* (pp. 143–173). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405164221.ch7>
- \*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>
- Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences*, 39(2), 297–308. <https://doi.org/10.1016/j.paid.2005.01.003>
- Pavlov, G., Maydeu-Olivares, A., & Fairchild, A. J. (2018). Effects of applicant faking on forced-choice and Likert scores. *Organizational Research Methods*, 22(3), 710–739. <https://doi.org/10.1177/1094428117753683>
- Pavlov, G., Shi, D., Maydeu-Olivares, A., & Fairchild, A. (2021). Item desirability matching in forced-choice test construction. *Personality and Individual Differences*, 183, 111114. <https://doi.org/10.1016/j.paid.2021.111114>
- \*Peng, J.-C., & Tseng, M.-M. (2019). Antecedent and consequence of nurse engagement. *The Journal of Psychology*, 153(3), 342–359. <https://doi.org/10.1080/00223980.2018.1536639>
- \*Piedmont, R. L., & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO personality inventory. *The Journal of Psychology*, 128(3), 255–265. <https://doi.org/10.1080/00223980.1994.9712728>
- Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & de Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, 112, 369–383. <https://doi.org/10.1016/j.jvb.2019.04.004>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- \*Renn, R. W., Steinbauer, R., & Fenner, G. (2014). Employee behavioral activation and behavioral inhibition systems, manager ratings of employee job performance, and employee withdrawal. *Human Performance*, 27(4), 347–371. <https://doi.org/10.1080/08959285.2014.929694>
- Revelle, W. (2020). Psych: Procedures for psychological, psychometric, and personality research (R package Version 2.0.12). <https://cran.r-project.org/web/packages/psych/psych.pdf>
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. <https://doi.org/10.1037/0033-2909.132.1.1>
- \*Robie, C., & Ryan, A. M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment*, 7(3), 157–169. <https://doi.org/10.1111/1468-2389.00115>
- \*Robson, S. M., Abraham, J. D., & Weiner, J. (2010). Characteristics of successful direct support professionals: An examination of personality and cognitive ability requirements. *International Journal of Selection and Assessment*, 18(2), 215–219. <https://doi.org/10.1111/j.1468-2389.2010.00503.x>
- \*Rodrigues, N., & Rebelo, T. M. M. S. D. (2013). Incremental validity of proactive personality over the Big Five for predicting job performance of software engineers in an innovative context. *Journal of Work and Organizational Psychology*, 29(1), 21–27. <https://doi.org/10.5093/tr2013a4>
- Rojon, C., McDowall, A., & Saunders, M. N. K. (2015). The relationships between traditional selection assessments and workplace performance criteria specificity: A comparative meta-analysis. *Human Performance*, 28(1), 1–25. <https://doi.org/10.1080/08959285.2014.974757>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000994>
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82(1), 30–43. <https://doi.org/10.1037/0021-9010.82.1.30>
- Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10(1 & 2), 117–125. <https://doi.org/10.1111/1468-2389.00198>
- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76(3), 323–346. <https://doi.org/10.1348/096317903769647201>
- \*Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, 21(1), 74–84. <https://doi.org/10.1111/ijsa.12018>
- \*Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2015). Validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology*, 24(3), 325–349. <https://doi.org/10.1080/1359432X.2014.903241>
- \*Salgado, J. F., & Rumbos, A. (1997). Personality and job performance in financial services managers. *International Journal of Selection and Assessment*, 5(2), 91–100. <https://doi.org/10.1111/1468-2389.00049>
- Salgado, J. F., & Táuriz, G. (2014). The Five-Factor model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, 23(1), 3–30. <https://doi.org/10.1080/1359432X.2012.716198>
- Schilling, M., Becker, N., Grabenhorst, M. M., & König, C. J. (2020). The relationship between cognitive ability and personality scores in selection situations: A meta-analysis. *International Journal of Selection and Assessment*, 1–18. <https://doi.org/10.1111/ijsa.12314>
- 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>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting for error and bias in research findings* (3rd ed.). Sage Publications.
- \*Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61(4), 827–868. <https://doi.org/10.1111/j.1744-6570.2008.00132.x>
- Schmitt, M. J., & Ryan, A. M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78(6), 966–974. <https://doi.org/10.1037/0021-9010.78.6.966>
- 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>
- Schmitt, N., & Sinha, R. (2011). Validation support for selection procedures. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Selecting and developing members for the organization* (Vol. 2, pp. 399–420). American Psychological Association. <https://doi.org/10.1037/12170-013>
- \*Scotter, V., & R., J. (1994). *Evidence for the usefulness of task performance, job dedication, and interpersonal facilitation as components of overall performance*. Air Force Institute of Technology. <https://apps.dtic.mil/sti/citations/ADA283491>
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65(3), 445–494. <https://doi.org/10.1111/j.1744-6570.2012.01250.x>
- Shaffer, J. A., & Postlethwaite, B. E. (2013). The validity of conscientiousness for predicting job performance: A meta-analytic test of two hypotheses. *International Journal of Selection and Assessment*, 21(2), 183–199. <https://doi.org/10.1111/ijsa.12028>
- \*Siers, B. P., & Christiansen, N. D. (2013). On the validity of implicit association measures of personality traits. *Personality and Individual Differences*, 54(3), 361–366. <https://doi.org/10.1016/j.paid.2012.10.004>
- \*Sitser, T., van der Linden, D., & Born, M. P. (2013). Predicting sales performance criteria with personality measures: The use of the general factor of personality, the Big Five and narrow traits. *Human Performance*, 26(2), 126–149. <https://doi.org/10.1080/08959285.2013.765877>
- \*Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9(2), 219–242. [https://doi.org/10.1016/S1053-4822\(99\)00019-4](https://doi.org/10.1016/S1053-4822(99)00019-4)
- \*Speer, A. B., Robie, C., & Christiansen, N. D. (2016). Effects of item type and estimation method on the accuracy of estimated personality trait scores: Polytomous item response theory models versus summated scoring. *Personality and Individual Differences*, 102, 41–45. <https://doi.org/10.1016/j.paid.2016.06.058>
- Steger, D., Schroeders, U., & Gnams, T. (2018). A meta-analysis of test scores in proctored and unproctored ability assessments. *European Journal of Psychological Assessment*, 1–11. <https://doi.org/10.1027/1015-5759/a000494>



- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Wiley.
- \*Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, 74(5), 637–657. <https://doi.org/10.1348/096317901167569>
- \*Sy, T., Tram, S., & O'Hara, L. A. (2006). Relation of employee and manager emotional intelligence to job satisfaction and performance. *Journal of Vocational Behavior*, 68(3), 461–473. <https://doi.org/10.1016/j.jvb.2005.10.003>
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44(4), 703–742. <https://doi.org/10.1111/j.1744-6570.1991.tb00696.x>
- Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A "multisaturation" perspective on faking as performance. *Human Performance*, 24(4), 302–321. <https://doi.org/10.1080/08959285.2011.597472>
- Tett, R., & Simonet, D. (2021). Applicant faking on personality tests: Good or bad and why should we care? *Personnel Assessment and Decisions*, 7(1). <https://doi.org/10.25035/pad.2021.01.002>
- \*Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior*, 24(3), 335–356. <https://doi.org/10.1002/job.191>
- \*Tracey, J. B., Sturman, M. C., & Tews, M. J. (2007). Ability versus personality: Factors that predict employee job performance. *Cornell Hotel and Restaurant Administration Quarterly*, 48(3), 313–322. <https://doi.org/10.1177/0010880407302048>
- \*Tziner, A., Meir, E. I., & Segal, H. (2002). Occupational congruence and personal task-related attributes: How do they relate to work performance? *Journal of Career Assessment*, 10(4), 401–412. <https://doi.org/10.1177/1069072702238403>
- van Aarde, N., Meiring, D., & Wiernik, B. M. (2017). The validity of the Big Five personality traits for job performance: Meta-analyses of South African studies. *International Journal of Selection and Assessment*, 25(3), 223–239. <https://doi.org/10.1111/ijsa.12175>
- \*Vecchione, M., Alessandri, G., & Barbaranelli, C. (2013). Measurement and application of egoistic and moralistic self-enhancement. *International Journal of Selection and Assessment*, 21(2), 170–182. <https://doi.org/10.1111/ijsa.12027>
- \*Vecchione, M., Dentale, F., Alessandri, G., Imbesi, M. T., Barbaranelli, C., & Schnabel, K. (2017). On the applicability of the Big Five implicit association test in organizational settings. *Current Psychology*, 36(3), 665–674. <https://doi.org/10.1007/s12144-016-9455-x>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Walsh, C. G., Xia, W., Li, M., Denny, J. C., Harris, P. A., & Malin, B. A. (2018). Enabling open-science initiatives in clinical psychology and psychiatry without sacrificing patients' privacy: Current practices and future challenges. *Advances in Methods and Practices in Psychological Science*, 1(1), 104–114. <https://doi.org/10.1177/2515245917749652>
- Watrin, L., Geiger, M., Spengler, M., & Wilhelm, O. (2019). Forced-choice versus likert responses on an occupational Big Five questionnaire. *Journal of Individual Differences*, 1–16. <https://doi.org/10.1027/1614-0001/a000285>
- Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance*, 17(4), 433–461. [https://doi.org/10.1207/s15327043hup1704\\_5](https://doi.org/10.1207/s15327043hup1704_5)
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- \*Wihler, A., Blickle, G., Ellen, B. P., Hochwarter, W. A., & Ferris, G. R. (2017). Personal initiative and job performance evaluations: Role of political skill in opportunity recognition and capitalization. *Journal of Management*, 43(5), 1388–1420. <https://doi.org/10.1177/0149206314552451>
- Wille, B., Beyers, W., & De Fruyt, F. (2012). A transactional approach to person–environment fit: Reciprocal relations between personality development and career role growth across young to middle adulthood. *Journal of Vocational Behavior*, 81(3), 307–321. <https://doi.org/10.1016/j.jvb.2012.06.004>
- Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences*, 116(46), 23004–23010. <https://doi.org/10.1073/pnas.1908430116>
- \*Witt, L. A. (2002). The interactive effects of extraversion and conscientiousness on performance. *Journal of Management*, 28(6), 835–851. <https://doi.org/10.1177/014920630202800607>
- \*Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology*, 87(1), 164–169. <https://doi.org/10.1037/0021-9010.87.1.164>
- \*Witt, L. A., & Carlson, D. S. (2006). The work–family interface and job performance: Moderating effects of conscientiousness and perceived organizational support. *Journal of Occupational Health Psychology*, 11(4), 343–357. <https://doi.org/10.1037/1076-8998.11.4.343>
- Woo, S. E., Chernyshenko, O. S., Stark, S. E., & Conz, G. (2014). Validity of six openness facets in predicting work behaviors: A meta-analysis. *Journal of Personality Assessment*, 96(1), 76–86. <https://doi.org/10.1080/00223891.2013.806329>
- Woods, S. A., Lievens, F., De Fruyt, F., & Wille, B. (2013). Personality across working life: The longitudinal and reciprocal influences of personality on work. *Journal of Organizational Behavior*, 34(S1), S7–S25. <https://doi.org/10.1002/job.1863>
- \*Wright, P. M., Kacmar, K. M., McMahan, G. C., & Deleeuw, K. (1995). P=f (M X A): Cognitive ability as a moderator of the relationship between personality and job performance. *Journal of Management*, 21(6), 1129–1139. <https://doi.org/10.1177/014920639502100606>
- \*Zhang, S., Zhou, M., Zhang, J., & Chen, S. (2012). The nonlinear effects of conscientiousness on overall job performance and performance dimensions in the Chinese context. *Asian Journal of Social Psychology*, 15(4), 231–237. <https://doi.org/10.1111/j.1467-839X.2012.01375.x>
- Ziegler, M., MacCann, C., & Roberts, R. D. (Eds.). (2011). *New perspectives on faking in personality assessment*. Oxford University Press.

**How to cite this article:** Watrin, L., Weihrauch, L., & Wilhelm, O. (2023). The criterion-related validity of conscientiousness in personnel selection: A meta-analytic reality check. *International Journal of Selection and Assessment*, 31, 286–301. <https://doi.org/10.1111/ijsa.12413>

## APPENDIX A

Table A1.

**TABLE A1** Search terms used for the meta-analysis

Conscientiousness	AND	Performance
Personality, Five-factor model, 5-factor model, FFM, Big5, Big Five		Job Performance,
Conscientiousness		Work Performance,
<i>Orderliness</i> , Order, Organization, Task planning, Planfulness, Tidiness, Cleanliness, Neatness, Punctuality, Perfectionism, Diligence, Meticulousness, Methodicalness, Superficiality, Discipline, Formalness, Conventionality, Traditionalism		Performance Rating,
<i>Industriousness</i> , Achievement-Striving, Achievement Motivation, Achievement, Action Orientation, Activity, Autonomy, Competence, Decisiveness, Endurance, Efficiency, Goal-Striving, Initiative, Laziness, Perseverance, Persistence, Procrastination, Purposefulness, Rationality, Self-Efficiency		Overall Performance
<i>Self-Control</i> , Careless, Cautiousness, <sup>a</sup> Control, Constraint, Deliberation, Impulse Control, Impulsivity, Impulsiveness, Self-Discipline		
<i>Responsibility</i> , Caution, <sup>a</sup> Compliance, Dutifulness, Dependability, Prudence, Reliability		

Note: Detailed search strings for the respective databases are provided in OS1.

<sup>a</sup>This duplication did not affect the present results because none of the constructs in the primary studies were coded as "Cautiousness" or "Caution."

## APPENDIX B

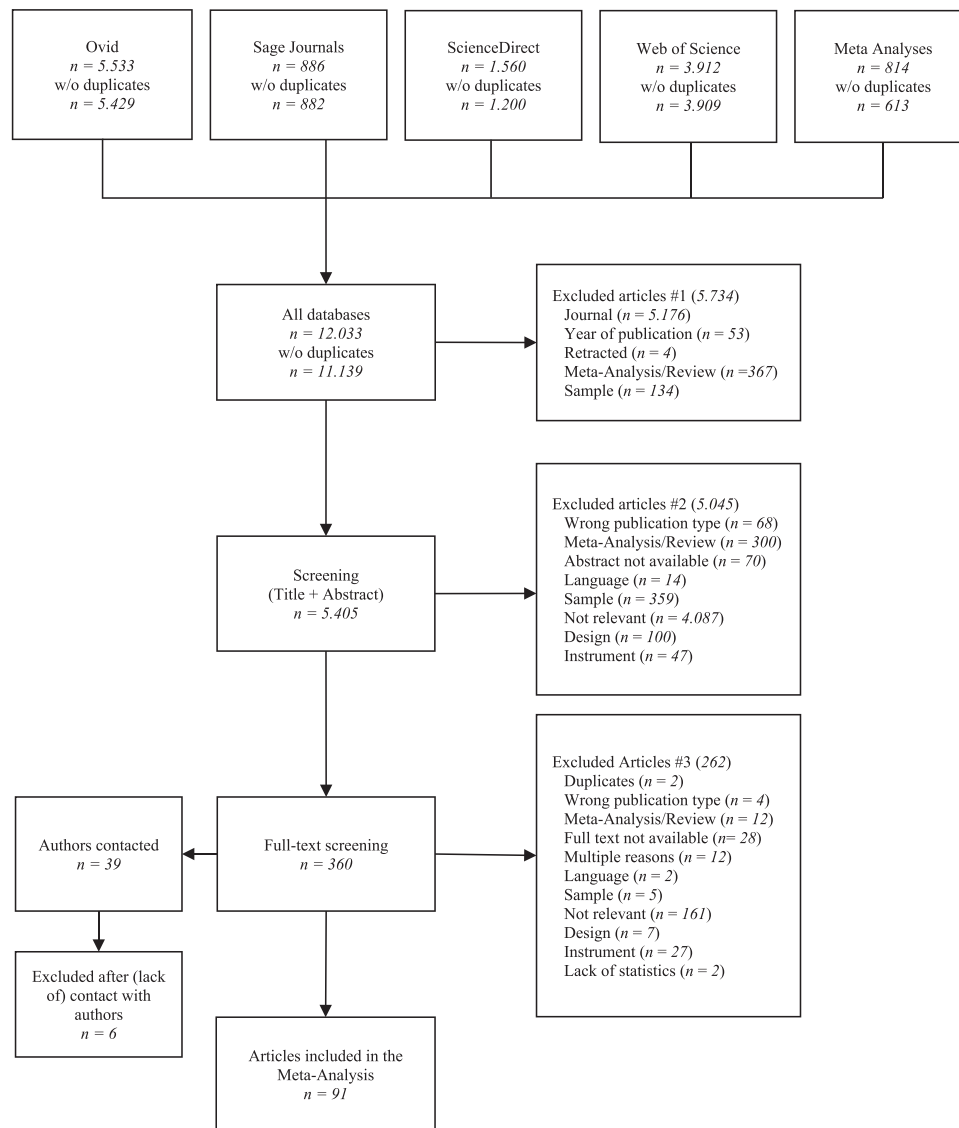


FIGURE B1 Overview of the literature search process.