The Relationships Between Traditional Selection Assessments and Workplace Performance Criteria Specificity: A Comparative Meta-Analysis

The relationships between traditional selection assessments and workplace performance criteria specificity: A comparative meta-analysis

Céline Rojon, Edinburgh Business School, University of Edinburgh

Almuth McDowall, Department of Psychology, University of Surrey

Mark NK Saunders, Surrey Business School, University of Surrey

This is an accepted version of an article published by Taylor & Francis Group in Human Performance on 13 January 2015. Available online:
*http://www.tandfonline.com/10.1080/*08959285.2014.974757

To reference this paper:

Rojon, C., McDowall, A. and 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

Abstract

Individual workplace performance is a crucial construct in Work Psychology. However, understanding of its conceptualization remains limited, particularly regarding predictor-criterion linkages. This study examines to what extent operational validities differ when criteria are measured as overall job performance compared to specific dimensions as predicted by ability and personality measures respectively. Building on Bartram's work (2005), systematic review methodology is used to select studies for meta-analytic examination. We find validities for both traditional predictor types to be enhanced substantially when performance is assessed specifically rather than generically. Findings indicate assessment decisions can be facilitated through a thorough mapping and subsequent use of predictor measures using specific performance criteria. We discuss further implications, referring particularly to the development and operationalization of even more finely grained performance conceptualizations.

Introduction and Background

Workplace performance is a core topic in Industrial, Work and Organizational (IWO) Psychology (Campbell, 2010; Motowidlo, 2003; Arvey & Murphy, 1998), as well as related fields, such as Management and Organization Sciences (MOS). Research in this area has a long tradition (Austin & Villanova, 1992), performance assessment being "so important for work psychology that it is almost taken for granted" (Arnold et al., 2010, p. 241). For management of human resources, the concept of performance is of crucial importance, organizations implementing formal and systematic performance management systems outperforming other organizations by more than 50% regarding financial outcomes and by more than 40% regarding other outcomes such as customer satisfaction (Cascio, 2006). However, to ensure accuracy in the

measurement and prediction of workplace performance, further development of an evidence-based understanding of this construct remains vital (Viswesvaran & Ones, 2000).

This research focuses on refining our understanding of individual, rather than team or organizational level performance. In particular, it investigates operational validities of both personality and ability measures as predictors of individual workplace performance across differing levels of criterion specificity. Although individual workplace performance is arguably one of the most important dependent variables in IWO Psychology, knowledge about underlying constructs is as yet insufficient (e.g., Campbell, 2010) for advancing research and practice. Prior research has focused typically on the relationship between predictors and criteria. The predictor-side refers generally to assessments or indices including psychometric measures such as ability tests, personality questionnaires and biographical information or other measures such as structured interviews or role plays, which vary considerably in the amount of variance they can explain (e.g., Hunter & Schmidt, 1998). Extensive research has been conducted on such predictors, often through validation studies for specific psychometric predictor measures (Bartram, 2005).

The criterion-side is concerned with how performance is conceptualized and measured in practice, considering both subjective performance ratings by relevant stakeholders and objective measurement through organizational indices such as sales figures (Arvey & Murphy, 1998). Such criteria have attracted relatively less interest from scholars, Deadrick and Gardner (2008, p. 133) observing that "after more than 70 years of research, the 'criterion problem' persists, and the performance-criterion linkage remains one of the most neglected components of performance-related research". Hence, despite much being known about how to predict performance and what measures to use, our understanding regarding what is actually being predicted and how predictor- and criterion-sides relate to each other remains limited (Bartram, 2005; Bartram, Warr & Brown, 2010). For instance, whilst research has evaluated the operational validities of different types of predictors against each other (e.g., Hunter & Schmidt, 1998), there is little comparative work juxtaposing different types of criteria against one another. Thus, we argue there is an imbalance of evidence, knowledge being greater regarding the validity of different predictors than for an understanding of relevant criteria which these are purported to measure. In particular, the plethora of studies, reviews and meta-analyses that exist on the performance-criterion link indicate a need to draw together the extant evidence base in a systematic and rigorous fashion to formulate clear directions for research and practice.

Systematic review (SR) methodology, widely used in health-related sciences (Leucht, Kissling & Davis, 2009), is particularly useful for such 'stock taking exercises' in terms of "locating, appraising, synthesizing, and reporting 'best evidence'" (Briner, Denyer & Rousseau, 2009, p. 24). Whereas for IWO Psychology, SR methodology offers a relatively new approach (Briner & Rousseau, 2011), we note that as early as 2005, Hogh and Viitasara for instance published a SR on workplace violence in the *European Journal of Work and Organizational Psychology*, whilst other examples include Wildman and her colleagues (2012) on team knowledge, Plat, Frings-Dresen and Sluiter (2011) on health surveillance and Joyce, Pabayo, Critchley and Bambra (2010) on flexible working. The rigor and standardization of SR methodology allow for greater transparency, replicability and clarity of review findings (Briner, Denyer & Rousseau, 2009; Petticrew & Roberts, 2006; Denyer & Tranfield, 2009), providing

one of the main reasons why scholars (e.g., Briner & Denyer, 2012; Briner & Rousseau, 2011; Denyer & Neely, 2004) have recommended it be employed more in IWO Psychology and related fields such as MOS. In the Social Sciences and Medical Sciences, where SRs are widely used (e.g., Sheehan, Fenwick & Dowling, 2010), these regularly involve a statistical synthesis of primary studies' findings (Petticrew & Roberts, 2006; Green, 2005), Healthcare Sciences' Cochrane Collaboration (Higgins & Green, 2011) recommending that it can be useful to undertake a meta-analysis as part of an SR (Alderson & Green, 2002). In MOS also, a meta-analysis can be an integral component of a larger SR, for example to assess the effect of an intervention (Denyer, 2009; *cf.* Tranfield, Denyer & Smart, 2003).

Where existing 'regular' meta-analyses list the databases searched, the variables under consideration and relevant statistics (e.g., correlation coefficients), the principles of SR go further. Indeed, meta-analyses can be considered as one type of SR (e.g., Briner & Rousseau, 2011; Xiao & Nicholson, 2013), where the review and inclusion strategy, incorporating quality criteria for primary papers – of which a reviewer prescribed number have to be met – are detailed in advance in a transparent manner to provide a clear audit trail throughout the research process. Although it might be argued that concentrating on a relatively small number of included studies based on the application of inclusion/exclusion criteria may be overly selective, such careful evaluation of the primary literature is precisely the point when using SR methodology (Authors, 2011). Hence, whilst meta-analyses in IWO Psychology have traditionally not been conducted within the framework of SR methodology, this reviewing approach for meta-analysis ensures only robust studies measured against clear quality criteria are included, allowing the eventual conclusions drawn to be based on quality checked evidence. We provide further details regarding our meta-analytical strategies below.

To this extent, we undertook a SR of the criterion-side of individual workplace performance to investigate current knowledge and any gaps therein. We conducted a pre-review scoping study and consultation with ten expert stakeholders (Denyer & Tranfield, 2009; Petticrew & Roberts, 2006), who were Psychology and Management scholars with research foci in workplace performance and human resources practitioners in public and private sectors. Our expert stakeholders commented that a differentiated examination and measurement of the criterion-side was required, rather than conceptualizing it through assessments of overall job performance (OJP), operationalized typically via subjective supervisor ratings consisting oftentimes of combined or summed up scales. In other words, there is a need to establish specifically what types of predictors work with what types of criteria. Viswesvaran, Schmidt and Ones (2005) meta-analyzed performance dimensions' intercorrelations to indicate a case for a single general job performance factor, which they purported does not contradict the existence of several distinct performance dimensions. The authors note that performance dimensions are usually positively correlated, suggesting that their shared variance (60%) is likely to originate from a higher-level, more general factor and as such, both notions can coexist and it is useful to compare predictive validities at different levels of criterion specificity. Within this, scholars have highlighted the importance of matching predictors and criteria, both in terms of content and in terms of specificity/generality, this perspective also being known as 'construct correspondence' (e.g. Hough & Furnham, 2003; Ajzen & Fishbein, 1977; Fishbein & Ajzen, 1974). Yet, much of this work has been of a theoretical nature (e.g., Schmitt & Chan, 1998; Dunnette, 1963), emphasizing the necessity to empirically examine this issue. Therefore, using meta-analytical

procedures, we set out to investigate the review question: What are the relationships between overall versus criterion-specific measures of individual workplace performance and established predictors (i.e., ability and personality measures)?

Previous meta-analyses have investigated criterion-related validities of personality questionnaires (e.g., Barrick, Mount & Judge, 2001) and ability tests (e.g., Salgado & Anderson, 2003, examining validities for tests of general mental ability (GMA)) as traditional predictors of performance. These indicate both constructs can be good predictors of performance, taking into account potential moderating effects (e.g., participants' occupation; Vinchur, Schippmann, Switzer & Roth, 1998). Several meta-analyses focusing on the personality domain have examined predictive validities for more specific criterion constructs than OJP, such as organizational citizenship behavior (Chiaburu, Oh, Berry, Li & Gardner, 2011; Hurtz & Donovan, 2000), job dedication (Dudley, Orvis, Lebiecki & Cortina, 2006) or counterproductive work behavior (Berry, Ones & Sackett, 2007; Dudley et al., 2006; Salgado, 2002). Hogan and Holland (2003) investigated whether differentiating the criterion-domain into two performance dimensions (*getting along* and *getting ahead*) impacts on predictive validities of personality constructs, using the Hogan Personality Inventory for a series of meta-analyses. They argued that a priori alignment of predictor- and criterion-domains based on common meaning increases predictive validity of personality measures compared to atheoretical validation approaches. In line with expectations, these scholars found that predictive validities of personality constructs were higher when performance was assessed using these getting ahead and getting along dimensions separately compared to a combined, global measurement of performance.

Bartram (2005) explored the separate dimensions of performance further, differentiating the criterion-domain using an eight-fold taxonomy of competence, the *Great Eight*, these being hypothesized to relate to the Five Factor Model (FFM), motivation and general mental ability constructs. Employing meta-analytic procedures, he investigated links between personality and ability scales on the predictor-side in relation to these specific eight criteria, as well as to OJP. Within this, he used Warr's (1999) logical judgment method to ensure point-to-point alignment of predictors and criteria at the item level. His results corroborated the predictive validity of the criterion-centric approach, relationships between predicted associations being larger than those not predicted, with strongest relationships observed where both predictors and criteria had been mapped onto the same constructs. However, the study was limited in that it involved a restricted range of predictor instruments, using data solely from one family of personality measures and criteria which were aligned to these predictors. Consequently, there remains a need to further investigate these findings, using a wider range of primary studies to enable generalization. Building on Hogan and Holland's (2003) and Bartram's (2005) research, the aim of our study was to investigate criterion-related validities of both personality and ability measures as established predictors of a range of individual workplace performance criteria. In particular we focused our investigation on the validities of these two types of predictor measures across differing levels of criterion specificity: performance being operationalized as OJP or through separate performance dimensions, such as task proficiency. We examined the following theoretical proposition (*cf.* Hunter & Hirsh, 1987):

> Criterion-related validities for ability and personality measures increase when specific performance criteria are assessed compared to the criterion-side being measured as only one

general construct. Thus, the degree of specificity on the criterion-side acts as a moderator of the predictive validities of such established predictor measures.

Our approach therefore enables us to determine the extent to which linking specific predictor dimensions to specific criterion dimensions is likely to result in more precise performance prediction. In building upon previous studies examining predictor-criterion relationships of performance, our contribution is two-fold: (i) an investigation of criterion-related validities of both personality and ability measures as established predictors of performance, whereas most previous studies as outlined above focused on personality measures only; and (ii) a direct comparison of traditional predictor measures' operational validities at differing levels of criterion specificity, ranging from unspecific operationalization (OJP measures) through to specific criterion measurement using distinct performance dimensions.

Method

Literature search

Following established SR guidelines (Denyer & Tranfield, 2009; Petticrew & Roberts, 2006), we conducted a comprehensive literature search extending from 1950 to 2010, to retrieve published and unpublished (e.g., theses, dissertations) predictive and concurrent validation studies that had used personality and/or ability measures on the predictor-side and OJP or criterion-specific performance assessments on the criterion-side. Four sources of evidence were considered: (i) databases (namely PsycINFO, PsycBOOKS, Psychology and Behavioral Sciences Collection, Emerald Management e-Journals, Business Source Complete, International Bibliography of Social Sciences, Web of Science: Social Sciences Citation Index, Medline, British Library E-Theses, European E-Theses, Chartered Institute of Personnel Development database, Improvement and Development Agency (IDeA) database[1]); (ii) conference proceedings; (iii) potentially relevant journals inaccessible through the aforementioned databases; and (iv) direct requests for information, which were emailed to 18 established researchers in the field as identified by their publication record. For each database searched, we used four search strings in combined form, where the use of the asterisk, a wildcard, enabled searching on truncated word forms. For instance, the use of 'perform*' revealed all publications that included words such as 'performance', 'performativity', 'performing', 'perform'. Each search string represented a key concept of our study (e.g., 'performance' for the first search string), synomymical terms or similar concepts being included using the Boolean operator 'OR' to ensure that any relevant references would be found in the searches:

(1) perform* OR efficien* OR productiv* OR effective* (first search string)
(2) work* OR job OR individual OR task OR occupation* OR human OR employ* OR vocation* OR personnel (second search string)
(3) assess* OR apprais* OR evaluat* OR test OR rating OR review OR measure OR manage* (third search string)
(4) criteri* OR objective OR theory OR framework OR model OR standard (fourth search string).

Inclusion criteria

[1] This database is provided by the IDeA, a United Kingdom government agency. It was included here following recommendation by expert stakeholders consulted as part of determining review questions for the SR.

In line with SR methodology, all references from these searches (N = 59,465) were screened initially by title and abstract to exclude irrelevant publications, for instance those which did not address any of the constructs under investigation. This truncated the body of references to 314, which were then examined in detail, applying a priori defined quality criteria for inclusion/exclusion. As suggested (Briner et al., 2009), these criteria were informed by quality appraisal checklists developed and employed by academic journals in the field to assess the quality of submissions (e.g., *Academy of Management Review* (2013); *Journal of Occupational & Organizational Psychology* (Cassell, 2011); *Personnel Psychology* (Campion, 1993)), as well as by published advice (e.g., Denyer, 2009; Briner et al., 2009). We stipulated a publication would meet the inclusion/exclusion criteria if it (a) was relevant, addressing our review question (this being the base criterion) and (b) fulfilled at least five of eleven agreed quality criteria, thus judged to be of at least average overall quality (Table 1) (*cf.* Briner et al., 2009). No single quality criterion was used on its own to exclude a publication, but their application contributed to an integrated assessment of each study in turn. As such, no single criterion could disqualify potentially informative and relevant studies. For example the application of the criterion 'study is well informed by theory' (A) did not by default preclude the inclusion of technical reports, which are often not theory-driven, in the study pool. To illustrate, validation studies by Chung-Yan, Cronshaw and Hausdorf (2005) and Buttigieg (2006) were included in our meta-analysis despite not meeting criterion (A). Further, these two publications also did not fulfil criterion (F) relating to 'making a contribution'. Yet, by considering all quality criteria holistically, these studies were still included as they met other relevant criteria; for example: clearly articulating study purpose and aims (B) and providing a reasoned explanation for the data analysis methods used (D). Our point is further illustrated in the application of criterion (E), which refers to a study being 'published in a peer-reviewed journal': Whilst applying this criterion might be considered to disadvantage unpublished studies, this was not the case. Our use of the criteria in combination, requiring at least five of the eleven to be fulfilled, resulted in the inclusion of several studies that had not been published in peer-reviewed journals. These included unpublished doctoral theses (e.g., Garvin, 1995; Alexander, 2007) and a book chapter (Goffin, Rothstein and Johnston, 2000).

Some 61 publications (57 journal articles, three theses/dissertations, one book chapter) satisfied the criteria for the current meta-analysis and were included. These contained 67 primary studies conducted in academic and applied settings (N = 48,209). Studies were included even if their primary focus was not on the relationships between personality and/or ability and performance measures, as long as they provided correlation coefficients of interest to the meta-analysis and met the inclusion criteria outlined (e.g., a study by Côté & Miners, 2006). A sub-sample (10%) of all studies initially selected against the criteria for inclusion in the meta-analysis was checked by two of the three authors, interrater agreement was 100%. Our primary study sample size is in line with recent meta-analyses (e.g., Richter, Dawson & West, 2011; Whitman, Van Rooy & Viswesvaran, 2010; Mesmer-Magnus & DeChurch, 2009) and considerably higher than others (e.g., Hogan & Holland, 2003; Bartram, 2005).

Table 1 about here

Coding procedures

Each study was coded for the type and mode of predictor and criterion measures used. Initially, a subsample of 20 primary studies was chosen randomly to assess interrater agreement on the coding of the key variables of interest between the three authors. Agreement was 100% due to the unambiguous nature of the data, so coding and classification of the remaining studies was undertaken by the first author only. Ability and personality predictors were coded separately, further noting the personality or ability measure(s) each primary study had used. We also coded whether the criterion measured OJP, criterion-specific aspects of performance or both. Lastly, each study was coded with regards to their mode of criterion measurement, in other words whether performance had been measured objectively, subjectively or through a combination of both; this information being required to perform meta-analyses using the Hunter-Schmidt (2004) approach. Objective criterion measures were understood as counts of specified acts (e.g., production rate) or output (e.g., sales volume), usually maintained within organizational records. Performance ratings (e.g., supervisor ratings), assessment or development center evaluations and work sample measures were coded as subjective, being (even for standardized work samples) liable to human error and bias.

Database

Study sample sizes ranged from 38 to 8,274. Some 31 studies (46.3%) had used personality measures as predictors of workplace performance, 16 (23.9%) had used ability/GMA tests and the remaining 20 (29.8%) had used both personality and ability measures. In just over half of primary studies (n = 35, 52.3%), workplace performance had been assessed only in terms of OJP; 10 studies (14.9%) had evaluated criterion-specific aspects of performance only and the remaining 22 (32.8%) had used both OJP and criterion-specific measures. Further, nearly three quarters of primary studies had assessed individual performance subjectively (n = 48, 71.6%); 5 studies (7.5%) had assessed it objectively, and 14 studies (20.9%) both subjectively and objectively. Participants comprised five main occupational categories: managers (n = 14, 21.0%), military (n = 11, 16.4%), sales (n = 10, 14.9%), professionals (n = 8, 11.9%) and mixed occupations (n = 24, 35.8%). 54 (80.6%) studies had been conducted in the US, 10 (14.9%) in Europe, one in each of Asia and Australia (3%) and one (1.5%) across several cultures/countries simultaneously.

The choice of predictor and criterion variables

Tests of GMA were used as the predictor variable for ability-performance relationships (Figure 1, top left) rather than specific ability tests (e.g., verbal ability), the former being widely used measures for personnel selection worldwide (Salgado & Anderson, 2003; Bertua, Anderson & Salgado, 2005). Following Schmidt (2002; *cf.* Salgado, Anderson, Moscoso, Bertua, de Fruyt & Rolland, 2003), a GMA measure either assesses several specific aptitudes or includes a variety of items measuring specific abilities. Consequently, an ability composite was computed where different specific tests combined in a battery rather than an omnibus GMA test had been used.

To code predictor variables for personality-performance relationships, the dimensions of the FFM (Figure 1, top right) were used, in line with previous meta-analyses (Mol, Born, Willemsen & Van Der Molen, 2005; Salgado, 2003; Hurtz & Donovan, 2000; Mount, Barrick & Stewart, 1998; Vinchur et al., 1998; Salgado, 1998; Salgado, 1997; Robertson & Kinder, 1993; Barrick & Mount, 1991; Tett, Jackson & Rothstein, 1991). The FFM assumes five super-ordinate trait dimensions, namely Openness to Experience, Conscientiousness, Extraversion,

Agreeableness and Emotional Stability (Costa & McCrae, 1990), being employed very widely (Barrick et al., 2001). Personality scales not based explicitly on the FFM, or which had not been classified by the study authors themselves, were assigned using Hough and Ones' (2001) classification of personality measures. For the few cases of uncertainty, we drew on further descriptions of the FFM (Goldberg, 1990; Digman, 1990). These we also checked and coded independently by the three authors (100% agreement).

Three levels of granularity/specificity were distinguished on the criterion side, resulting in eleven criterion dimensions: Measures of OJP, operationalizing performance as one global construct, represent the broadest level of performance assessment (Figure 1, bottom right). At the medium grained level, Borman and Motowidlo's (1993; 1997) Task Performance/Contextual Performance distinction was used to provide a representation of the criterion with two components (Figure 1, bottom centre). At the finely grained level, Campbell and colleagues' performance taxonomy was utilized (Campbell, McHenry & Wise, 1990; Campbell, 1990; Campbell, McCloy, Oppler & Sager, 1993; Campbell, Houston & Oppler, 2001) (Figure 1, bottom left), its eight criterion-specific performance dimensions being: Job-Specific Task Proficiency (F1), Non-Job-Specific Task Proficiency (F2), Written and Oral Communication Task Proficiency (F3), Demonstrating Effort (F4), Maintaining Personal Discipline (F5), Facilitating Peer and Team Performance (F6), Supervision/Leadership (F7), Management/Administration (F8). Hence, using six predictor dimensions and these eleven criterion dimensions, validity coefficients for predictor-criterion relationships were obtained for a total of 66 combinations.

We gave particular consideration to the potential inclusion of previous meta-analyses. However, some of these had not stated the primary studies they had drawn upon explicitly (e.g., Robertson & Kinder, 1993; Barrick & Mount, 1991), whilst others omitted information regarding the criterion measures utilized. Previous relevant meta-analytic studies were therefore not considered further to eliminate the possibility of including the same study multiple times and also ensuring specificity of information regarding criterion measures. Nevertheless, we used previous meta-analyses investigating similar questions to inform our own meta-analysis, in terms of both content and process.

Figure 1 about here

As is customary in meta-analysis (e.g., Dudley et al., 2006; Salgado, 1998), composite correlation coefficients were obtained whenever more than one predictor measure had been used to assess the same constructs; such as in a study by Marcus, Goffin, Johnston and Rothstein (2007), where two different personality measures were employed to assess the FFM dimensions. Composite coefficients were also calculated where more than one criterion measure had been used; such as in research comparing the usefulness and psychometric properties of two methods of performance appraisal aimed at measuring the same criterion scales (Goffin, Gellatly, Paunonen, Jackson & Meyer, 1996).

Meta-analytic procedures

Psychometric meta-analysis procedures were employed following Hunter and Schmidt (2004, p. 461), using Schmidt and Le's (2004) software. This random-effects approach generally provides

the most accurate results when using *r* (correlation coefficient) in synthesizing studies (Brannick, Yang & Cafri, 2011) and has been widely employed by other researchers (e.g., Hurtz & Donovan, 2000; Barrick & Mount, 1991; Bartram, 2005; Salgado, 2003), providing "the basis for most of the meta-analyses for personnel selection predictors" (Robertson & Kinder, 1993, p. 233).

Following the Hunter-Schmidt approach, we considered criterion unreliability and predictor range restriction as artifacts, thereby allowing estimation of how much of the observed variance of findings across samples was due to artifactual error. Since insufficient information was provided to enable individually correcting meta-analytical estimates in many of the included primary studies, we employed artifact distributions to empirically adjust the true validity for artifactual error (Table 2) (Hunter & Schmidt, 2004; *cf.* Salgado, 2003). Values to correct for range restriction varied according to the type of predictor used (i.e., personality or ability), whilst values to correct for criterion unreliability were subdivided as to whether the criterion measurement was undertaken subjectively, objectively or in both ways (*cf.* Hogan & Holland, 2003; Hurtz & Donovan, 2000).

Table 2 about here

We report mean observed correlations ($\bar{r}_o$; uncorrected for artifacts), as well as mean true score correlations (ρ (rho)) and mean true predictive validities (MTV), which denote operational validities corrected for criterion unreliability and range restriction. We also present 80% credibility intervals (80% CV) of the true validity distribution to illustrate variability in estimated correlations across studies – upper and lower credibility intervals indicate the boundaries within which 80% of the values of the ρ distribution lie. The 80% lower credibility interval (CV_{10}) indicates that 90% of the true validity estimates are above that value; thus, if it is greater than zero, the validity is different from zero. Moreover, we report the percent variance accounted for by artifact corrections for each corrected correlation distribution (%VE). Validities are presented for categories in which two or more independent sample correlations were available.

Results

Table 3 presents the meta-analysis for each of the 66 predictor-criterion combinations. Our interpretation of these results draws partly on research into the distribution of effects across meta-analyses (Lipsey & Wilson, 2001), who argue an effect size of .10 can be interpreted as small/low, .25 as medium/moderate and .40 as large/high in terms of the importance/magnitude of the effect. Consideration of these results in relation to our theoretical proposition follows in the discussion section.

Table 3 about here

Relationships between ability/GMA and performance measures

GMA tests were found to be good predictors of OJP with an operational validity (MTV) of .27 (medium effect), predictive validity generally increasing at a more criterion-specific level. This held true particularly when using the eight factors, where operational validities were .72 for Job-Specific Task Proficiency (F1) and even .81 for Non-Job-Specific Task Proficiency (F2), corresponding to very large effects. This indicates that ability/GMA measures are valid

predictors of individual workplace performance, in particular for these criterion-specific dimensions. However, they did not predict the whole range of performance behaviors associated with the criterion-space, their predictive validity for three criterion-specific scales (F5/Maintaining Personal Discipline, F7/Supervision/Leadership, F8/Management/Administration) being low.

Relationships between personality (FFM) and performance measures

Results for the FFM dimensions as predictors of individual workplace performance followed a similar, but even more pronounced pattern compared to ability/GMA measures: All five dimensions displayed non-existent or low (.00 to .13) predictive validities for OJP. Yet, all showed some predictive validity at the two criterion-specific levels (i.e., medium to large effects).

In the case of *Conscientiousness*, predictive validity was low (.13) when OJP was assessed as the criterion. Higher operational validities for this dimension were observed when individual workplace performance was measured in more specific ways: For the prediction of Task Performance and Contextual Performance, moderate validities were found (.25 and .21 respectively). Equally, at the eight-factor level, Conscientiousness displayed moderate to high validities when used to predict F4/Demonstrating Effort (.23) and F5/Maintaining Personal Discipline (.31).

Predictive validity was low (.08) for *Extraversion* and OJP. At more specific criterion levels, moderate validities were observed, both at the Task Performance versus Contextual Performance level (.16 and .22 respectively), and the eight-factor level. As for Conscientiousness, two factors, F4/Demonstrating Effort (.22) and F5/Maintaining Personal Discipline (.20), were predicted to some extent by Extraversion, corresponding to a medium effect.

Predictive validity for *Emotional Stability* and OJP was low (.06). For Task Performance and Contextual Performance, Emotional Stability exhibited moderate operational validities of .28 and .26 respectively. Similarly, at the eight-factor level, Emotional Stability displayed moderate validities for the prediction of F3/Written & Oral Communication Task Proficiency (.24), F4/Demonstrating Effort (.22) and F5/Maintaining Personal Discipline (.20).

Results for *Openness to Experience* indicate this dimension's predictive validity was very low (.02) when the criterion was measured in general terms, as OJP. Operational validities increased, however, at the two criterion-specific levels, where medium effects were observed for Task Performance (.31) and Contextual Performance (.22). At the highest level of specificity, F5/Demonstrating Effort was predicted well by Openness to Experience, its validity of .40 corresponding to a large effect.

Predictive validity for *Agreeableness* and OJP was found to be non-existent (.00). Yet, for Task Performance and Contextual Performance, Agreeableness had a predictive validity of .14 and .31 respectively, indicating a medium effect. At the eight-factor level, the factors F3/Written & Oral Communication Task Proficiency (.32), F4/Demonstrating Effort (.20), F5/Maintaining Personal Discipline (.23), F7/Supervision/Leadership (.22) and

F8/Management/Administration (.24) were predicted to a moderate to large extent by Agreeableness.

Discussion

Adopting a criterion-centric approach (Bartram, 2005), the aim of this research was to investigate operational validities of both personality and ability measures as established predictors of a range of individual workplace performance criteria across differing levels of criterion specificity, examining frequently cited models of performance across various contexts. Previously, there has been a risk that because low operational validities were found for predictors of OJP (*cf.* Table 4), scholars might have drawn the conclusion that such predictors were not valid and, as a consequence, should not be used. Our findings suggest a different picture, emphasizing the importance of adopting a finely grained conceptualization and operationalization of the criterion-side in contrast to a general, overarching understanding of the construct: Addressing our theoretical proposition, taken together the operational validities suggest that prediction is improved where criterion measures are more specific.

It is evident that ability tests and personality assessments are generally predictive of performance, which is demonstrated both in our study and in previous meta-analytical research (e.g., Bartram, 2005; Barrick et al., 2001; Salgado & Anderson, 2003). In line with our proposition, our incremental contribution is that their respective predictive validity increases when individual workplace performance is measured through specific performance dimensions of interest as mapped point-to-point to relevant predictor dimensions. As such, our study helps to confirm that the issue does not lie in the validity of the predictors, but rather in the nature of measures of the criterion-side.

The more specific the match between predictor and criterion variables, the greater the precision of prediction: For example, distinguishing between Task Performance and Contextual Performance resulted in higher predictive validities for the majority of calculations. This applied also when performance was measured even more specifically, at the eight-factor level, where predictive validities were found to reach .81 for ability/GMA and .40 for personality dimensions, both corresponding to large effects (Lipsey & Wilson, 2001). Thus, as proposed, the specificity of the criterion measure moderates the relationships between predictors and criteria of performance.

Conceptualizing and measuring performance in specific terms is beneficial; establishing which variables are most predictive of relevant criteria and matching the constructs accordingly, when warranted, can enhance the criterion-related validities. Conscientiousness for example, which can be characterized by the adjectives efficient, organized, planful, reliable, responsible and thorough (McCrae & John, 1992, p. 178) was a good predictor of Demonstrating Effort (F4). This is plausible as this criterion dimension is "a direct reflection of the consistency of an individual's effort day by day" (Campbell, 1990, p. 709), which suggests a similarity in how these two constructs are conceptualized. As such, our findings further knowledge of the criterion-side and are important for both researchers and practitioners, who might benefit from more accurate performance predictions by adopting and adapting the suggestions put forward here.

Our validity coefficients are generally lower compared to those obtained by Barrick, Mount and Judge (2001) and Salgado and Anderson (2003) (Table 4) with regards to OJP. A likely reason for this is that our study divided criterion measures into OJP measures versus criterion-specific scales, a distinction not made by the two previous studies. Consequently, when the criterion is operationalized exclusively as OJP (as was the case at the lowest level of specificity in our research), operational validities of typical predictors may be lower than when the criterion includes both OJP indices and criterion-specific performance dimensions. A further potential reason for the differing findings may be that our meta-analysis included some data drawn from unpublished studies (e.g., theses) that possibly found lower coefficients for the predictor-criterion relationships than published studies might have observed (file drawer bias). We acknowledge also the possibility that operational validities observed here might to some extent vary compared to those reported in previous research partly as a result of the sampling frame employed, in other words our use of inclusion/exclusion criteria for the selection of studies for the meta-analysis. Yet, despite the validity coefficients being somewhat lower than those reported in previous meta-analytical studies, our results follow a similar pattern, whereby the best predictors of OJP are ability/GMA test scores, followed by the personality construct Conscientiousness. Similar to findings by Barrick and colleagues (2001), Openness to Experience and Agreeableness did not predict OJP.

Table 4 about here

Our theoretical contribution relates to how individual workplace performance is conceptualized. Performance frameworks differentiating only between two components, such as Borman and Motowidlo's (1993; 1997) Task Performance/Contextual Performance distinction, may be too blunt for facilitating assessment decisions. Addressing our theoretical proposition, both personality and ability predictor constructs relate most strongly to specific performance criteria, represented here by Campbell et al.'s eight performance factors (Campbell et al., 1990; Campbell, 1990; Campbell et al., 1993; Campbell et al., 2001). As such, our study corroborates and extends results of Bartram's research (2005), supporting the specific alignment of predictors to criteria (*cf.* Hogan & Holland, 2003), using more generic predictor and criterion models and a wider range of primary studies.

Limitations

To date, limited research has employed criterion-specific measures. Consequently, approximately a quarter of our analyses (26%) were based on less than five primary studies and a relatively small number of research participants ($N < 300$) (Table 3). The small number of studies is partly a result of having used SR methodology to identify and assess studies for inclusion in the meta-analysis. As aforementioned, we adopted this approach specifically because of its rigor and transparency compared to traditional reviewing methods, requiring active consideration of the quality of the primary studies. Within a meta-analysis, the studies included will always directly influence the resulting statistics, and we would argue that, as a consequence, attention needs to be devoted to ensuring the quality of these studies. Indeed, we believe positioning a meta-analysis within the wider SR framework, an approach that has a clear pedigree in other disciplines, represents a novel aspect of our paper in relation to IWO Psychology, and one that warrants further consideration within our academic community. However, we recognize that a different meta-analytical approach is likely to have yielded more

studies for inclusion, thus reducing the potential for second-order sampling error, whereby the meta-analysis outcome depends partly on study properties that vary randomly across samples (Hunter & Schmidt, 2004). Such second-order sampling error manifests itself in the predicted artifactual variance (%VE) being larger than 100%, which can be observed in 17% of our analyses. Yet, according to Hunter and Schmidt (2004), second-order sampling error has usually only a small effect on the validity estimates, indicating this is unlikely to have been a problem.

A further issue is that of validity generalization, in other words the degree to which evidence of validity obtained in one situation is generalizable to other situations without conducting further validation research. The credibility intervals (80% CV) give information about validity generalization: A lower 10% value (CV_{10}) above zero for 25 of the examined predictor-criterion relationships provides evidence that the validity coefficient can be generalized for these cases. For the remaining relationships, the 10% credibility value was below zero, indicating that validity cannot be generalized. It is important to note that at the coarsely grained level of performance assessment (i.e., OJP), validity can be generalized in merely 17% of cases. At the more finely grained levels, however, validity can be generalized in far more cases (42%), further supporting our argument that criterion-specific rather than overall measures of workplace performance are more adequate.

Conclusion

In 2005, Bartram noted that "differentiation of the criterion space would allow better prediction of job performance" (p. 1186) and that "we should be asking 'How can we best predict *Y*?' where *Y* is some meaningful and important aspect of workplace behavior" (p. 1185). Following his call, our research furthers our understanding of the criterion-side. Our use of SR methodology to add rigor and transparency to the meta-analytic approach is, we believe, a contribution in its own right as SR is currently underutilized in IWO Psychology. It offers a starting point for drawing together a well-established body of research, as we know much about the predictive validity of individual differences, particularly in a selection context (Bartram, 2005), whilst also enabling determination of a review question, which warranted answering: What are the relationships between overall versus criterion-specific measures of individual workplace performance and established predictors (i.e., ability and personality measures)?

Our findings indicate that the specificity/generality of the criterion measure acts as a moderator of the predictive validities – higher predictive validities for ability/GMA and personality measures were observed when these were mapped onto specific criterion dimensions in comparison to Task Performance versus Contextual Performance or a generic performance construct (OJP). This calls into question a dual categorical distinction as in Hogan and Holland's (2003) and Borman and Motowidlo's work (1993; 1997), adding weight to Bartram's (2005) argument – more is better. Whilst we have shown that a criterion-specific mapping is important for any predictor measure, this appears particularly crucial for personality assessments, where different personality scales tap into differentiated aspects of performance, therefore making them less suitable for contexts where only overall measures of performance are available. For ability, the best prediction achieved here explained 40% of variance in workplace performance using eight dimensions, whilst for personality this amounted to 20%. As such, we do not claim that the entirety of the criterion-side can be predicted by measures of ability/GMA or personality, even if

aligned carefully with specific criterion dimensions. Rather, we acknowledge there may be multiple additional constructs that might also be important to predict performance more precisely, an example being motivation (measures of this construct, specifically relating to need for power and control and need for achievement, are included in Bartram's (2005) Great Eight predictor-criterion mapping). Such alternative predictor constructs, as well as alternative criterion conceptualizations, should be examined as part of further research into the criterion-side. Future studies might also explore different levels of specificity and the impact thereof on operational validities both for criteria and predictors of performance, as well as the nature of point-to-point relationships. In particular, it would be worthwhile to investigate the extent to which performance predictors as operationalized through traditional versus contemporary measures improve operational validities, and also to what extent models of the predictor-side may concur with, or indeed deviate from, models of the criterion-side, given that Bartram's research (2005) elicited two- versus three-factor models respectively.

Our study offers one of the first direct comparisons of operational validities for both ability and personality predictors of OJP and criterion-specific performance measurements at differing levels of specificity. Besides investigating alternative predictor and criterion constructs, future research could build on our contribution by locating additional studies that have measured the criterion-side in more specific ways to corroborate the findings of the current meta-analysis and to avoid the potential danger of second-order sampling error. Through our SR approach we were able only to locate a relatively modest number of studies where criterion-specific measures of performance had been employed. However, as researchers become more aware of the increased value of using more specific criterion operationalizations, we believe the number of studies in this area will rise and allow more research into the criterion-related validities of performance predictors. Finally, our analysis suggests performance should be conceptualized specifically rather than generally, using frameworks more finely grained than dual categorical distinctions to increase the certainty with which inferences about associations between relevant predictors and criteria can be drawn. Thus, whilst we do not preclude the existence of a single general performance factor as suggested by Viswesvaran and colleagues (2005), taking a criterion-centric approach (Bartram, 2005) increases the likelihood of inferences about validity being accurate. Nevertheless, there is still a need to determine whether or not an even more specific and detailed conceptualization of performance further enhances operational validities of predictor measures. We offer this as a challenge for future research.

References

Studies included in the meta-analysis are denoted with an asterisk.

Academy of Management Review (2013). *Reviewer resources: Reviewer guidelines.* Briarcliff Manor, NY: Academy of Management (AOM). Retrieved from http://aom.org/Publications/AMR/Reviewer-Resources.aspx

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888-918. doi: 10.1037/0033-2909.84.5.888

Alderson, P. & Green, S. (2002). *Cochrane Collaboration open learning material for reviewers (version 1.1)*. Oxford, England: The Cochrane Collaboration. Retrieved from The Cochrane Collaboration website: http://www.cochrane-net.org/openlearning/PDF/Openlearning-full.pdf

*Alexander, S. G. (2007). *Predicting long term job performance using a cognitive ability test* (Unpublished doctoral dissertation). University of North Texas, Denton, TX.

*Allworth, E. & Hesketh, B. (1999). Construct-oriented biodata: Capturing change-related and contextually relevant future performance. *International Journal of Selection and Assessment, 7*(2), 97-111.

Arnold, J., Randall, R., Patterson, F., Silvester, J., Robertson, I., Cooper, C. & Burnes, B. (2010). *Work psychology: Understanding human behaviour in the workplace* (2nd ed.). Harlow, England: Pearson.

Arvey, R. D. & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168.

Austin, J. T. & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*(6), 836-874.

Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

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.

*Barrick, M. R., Mount, M. K. & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology, 78*(5), 715-722.

*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. doi: 10.1037//0021-9010.87.1.43

Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185-1203. doi: 10.1037/0021-9010.90.6.1185

Bartram, D., Warr, P. & Brown, A. (2010). Let's focus on two-stage alignment not just on overall performance. *Industrial and Organizational Psychology, 3*, 335-339.

*Bergman, M. E., Donovan, M. A., Drasgow, F., Overton, R. C. & Henning, J. B. (2008). Test of Motowidlo et al.'s (1997) theory of individual differences in task and contextual performance. *Human Performance, 21*, 227-253. doi: 10.1080/08959280802137606

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. doi: 10.1037/0021-9010.92.2.410

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*, 387-409.

*Bilgiç, R. & Sümer, H. C. (2009). Predicting military performance from specific personality measures: A validity study. *International Journal of Selection and Assessment, 17*(2), 231-238.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis (statistics in practice).* Oxford, England: Wiley Blackwell.

Borman, W. C. & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Bormann (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.

Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*(2), 99-109.

Brannick, M. T., Yang, L.-Q. & Cafri, G. (2011). Comparison of weights for meta-analysis of *r* and *d* under realistic conditions. *Organizational Research Methods, 14*(4), 587-607. doi: 10.1177/1094428110368725

Briner, R. B., Denyer, D. & Rousseau, D. M. (2009). Evidence-based management: Concept cleanup time? *Academy of Management Perspectives, 23*(4), 19-32.

Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In D. M. Rousseau (ed.), *The Oxford* handbook of evidence-based management *(p. 112-129)*. New York, NY: Oxford University Press.

Briner, R. B. & Rousseau, D. M. (2011). Evidence-based I-O Psychology: Not there yet. *Industrial and Organizational Psychology, 4*(1), 2-11.

*Buttigieg, S. (2006). Relationship of a biodata instrument and a Big 5 personality measure with the job performance of entry-level production workers. *Applied H.R.M. Research, 11*(1), 65-68.

*Caligiuri, P. M. (2000). The Big Five personality characteristics as predictors of expatriate's desire to terminate the assignment and supervisor-rated performance. *Personnel Psychology, 53*(1), 67-88.

Campbell, J. P. (1990). Modeling the performance prediction problem in Industrial and Organizational Psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (Vol. 1, 2nd ed., pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (2010). Individual occupational performance: The blood supply of our work life. In M. A. Gernsbacher, R. W. Pew & L. M. Hough (Eds.), *Psychology and the real world: essays illustrating fundamental contributions to society* (pp. 245-254). New York, NY: Worth Publishers.

Campbell, J. P., Houston, M. A. & Oppler, S. H. (2001). Modeling performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 307-333). Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, J. P., McCloy, R. A., Oppler, S. H. & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass.

Campbell, J. P., McHenry, J. J. & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*(2), 313-333.

Campion, M. A. (1993). Article review checklist: A criterion checklist for reviewing research articles in Applied Psychology. *Personnel Psychology, 46*(3), 705-718. doi: 10.1111/j.1744-6570.1993.tb00896.x

Cascio, W. F. (2006). Global performance management systems. In I. Bjorkman & G. Stahl (Eds.), *Handbook of research in international human resources management* (pp. 176-196). London, England: Edward Elgar Ltd.

Cassell, C. (2011). Criteria for evaluating papers using qualitative research methods. *Journal of Occupational and Organizational Psychology (JOOP),* [Online], Retrieved from Journal of Occupational and Organizational Psychology website: http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%292044-8325/homepage/qualitative_guidelines.htm

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 behavior: A meta-analysis. *Journal of Applied Psychology, 96*(6), 1140-1166. doi: 10.1037/a0024004

*Chung-Yan, G. A., Cronshaw, S. F. & Hausdorf, P. A. (2005). Information exchange article: A criterion-related validation study of transit operators. *International Journal of Selection and Assessment, 13*(2), 172-177.

*Conway, J. M. (2000). Managerial performance development constructs and personality correlates. *Human Performance, 13*(1), 23-46.

*Conte, J. M. & Gintoft, J. N. (2005). Polychronicity, Big Five personality dimensions, and sales performance. *Human Performance, 18*(4), 427-444.

*Cook, M., Young, A., Taylor, D. & Bedford, A. P. (2000). Personality and self-rated work performance. *European Journal of Psychological Assessment, 16*(3), 202-208.

*Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G. & Smith, R. G. (1992). The 'Big Five' personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology, 45*(1), 119-140.

Costa Jr., P. T. & McCrae, R. R. (1990). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

*Côté, S. & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly, 51*(1), 1-28.

*Day, D. V. & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*(1), 25-36.

*Deadrick, D. L. & Gardner, D. G. (2008). Maximal and typical measures of job performance: An analysis of performance variability over time. *Human Resource Management Review, 18*(3), 133-145. doi:10.1016/j.hrmr.2008.07.008

Denyer, D. (2009). *Reviewing the literature systematically.* Cranfield, England: Advanced Institute of Management Research (AIM). Retrieved from Advanced Institute of Management website: http://aimexpertresearcher.org/

Denyer, D. & Neely, A. (2004). Introduction to special issue: Innovation and productivity performance in the UK. *International Journal of Management Reviews, 5/6*(3&4), 131-135. doi: 10.1111/j.1460-8545.2004.00100.x

Denyer, D. & Tranfield, D. (2009). Producing a systematic review. In D. Buchanan & A. Bryman (Eds.). *The SAGE handbook of organizational research methods* (pp. 671-689). London, England: Sage.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

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*, 40-57. doi: 10.1037/0021-9010.91.1.40

Dunnette, M. D. (1963). A note on *the* criterion. *Journal of Applied Psychology, 47*(4), 251-254.

*Emmerich, W., Rock, D. A. & Trapani, C. S. (2006). Personality in relation to occupational outcomes among established teachers. *Journal of Research in Personality, 40*(5), 501-528. doi: 10.1016/j.jrp.2005.04.001

Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review, 81,* 59–74. doi: 10.1037/h0035872

*Ford, J. K. & Kraiger, K. (1993). Police officer selection validation project: The multijurisdictional police officer examination. *Journal of Business and Psychology, 7*(4), 421-429.

*Fritzsche, B. A., Powell, A. B. & Hoffman, R. (1999). Person-environment congruence as a predictor of customer service performance. *Journal of Vocational Behavior, 54*(1), 59-70.

*Furnham, A., Jackson, C. J. & Miller, T. (1999). Personality, learning style and work performance. *Personality and Individual Differences, 27*(6), 1113-1122.

*Garvin, J. D. (1995). *The relationship of perceived instructor performance ratings and personality trait characteristics of U.S. Air Force instructor pilots* (Unpublished doctoral dissertation). Graduate Faculty, Texas Tech University, Lubbock, TX.

*Gellatly, I. R., Paunonen, S. V., Meyer, J. P., Jackson, D. N. & Goffin, R. D. (1991). Personality, vocational interest, and cognitive predictors of managerial job performance and satisfaction. *Personality and Individual Differences, 12*(3), 221-231.

*Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N. & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The Behavioral Observation Scale and the Relative Percentile Method. *Journal of Business & Psychology, 11*(1), 23-33.

*Goffin, R. D., Rothstein, M. G. & Johnston, N. G. (2000). Predicting job performance using personality constructs: Are personality tests created equal? In R. D. Goffin & E. Helmes, *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 249-264). New York, NY: Kluwer Academic/Plenum Publishers.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.

Green, S. (2005). Systematic reviews and meta-analysis. *Singapore Medical Journal, 46*(6), 270-270.

*Hattrup, K., O'Connell, M. S. & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11*(4), 305-319.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions.* Version 5.1.0 (updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org.

*Hogan, J., Hogan, R. & Murtha, T. (1992). Validation of a personality measure of managerial performance. *Journal of Business and Psychology, 7*(2), 225-237.

Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance ratings: A socioanalytic perspective. *Journal of Applied Psychology, 88*(1), 100-112. doi: 10.1037/0021-9010.88.1.100

Hogh, A., & Viitasara, E. (2005). A systematic review of longitudinal studies of nonfatal workplace violence. *European Journal of Work and Organizational Psychology, 14*(3), 291-313. doi: 10.1080/13594320500162059

*Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D. & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*(5), 581-595.

Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In W. Borman, D. Ilgen, & R. Klimoski (Eds.), *Handbook of Psychology* (pp. 131-169). New York, NY: Wiley.

Hough, L. M. & Ones, D. O. (2001). The structure, measurement, validity, and use of personality variables in Industrial, Work, and Organizational Psychology. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology: Volume 1 Personnel Psychology* (pp. 233-277). London, England: Sage.

Hunter, J. E. & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper & I. T. Robertson (Eds.), *International review of Industrial and Organizational Psychology 1987* (pp. 321-357). New York, NY: Wiley.

Hunter, J. E. & Schmidt, F. L. (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.

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

Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*(6), 869-879. doi: 10.1037//002I-9010.85.6.869

*Inceoglu, I. & Bartram, D. (2007). Die Validität von Persönlichkeitsfragebögen: Zur Bedeutung des verwendeten Kriteriums. *Zeitschrift für Personalpsychologie, 6*(4), 160-173. doi: 10.1026/1617-6391.6.4.160

*Jenkins, M. & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology, 19*(2), 255-269.

Joyce, K., Pabayo, R., Critchley, J. A., & Bambra, C. (2010). Flexible working conditions and their effects on employee health and wellbeing. *Cochrane Database of Systematic Reviews*, Art. No. CD008009.

*Kanfer, R., Wolf, M. B., Kantrowitz, T. M. & Ackerman, P. L. (2010). Ability and trait complex predictors of academic and job performance: A person-situation approach. *Applied Psychology: An International Review, 59*(1), 40-69. doi: 10.1111/j.1464-0597.2009.00415.x

Kelly, A., Morris, H., Rowlinson, M. & Harvey, C. (Eds.) (2010). *Academic journal quality guide 2009. Version 3. Subject area listing.* London, England: The Association of Business Schools. Retrieved from The Association of Business Schools website: http://www.the-abs.org.uk/?id=257

*Lance, C. E. & Bennett Jr., W. (2000). Replication and extension of models of supervisory job performance ratings. *Human Performance, 13*(2), 139-158.

Leucht, S., Kissling, W., & Davis, J. M. (2009). How to read and understand and use systematic reviews and meta-analyses. *Acta Psychiatrica Scandinavica, 119*, 443–450. doi: 10.1111/j.1600-0447.2009.01388.x

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series (Vol. 49).* Thousand Oaks, CA: Sage Publications.

*Loveland, J. M., Gibson, L. W., Lounsbury, J. W. & Huffstetler, B. C. (2005). Broad and narrow personality traits in relation to the job performance of camp counselors. *Child & Youth Care Forum, 34*(3), 241-255. doi: 10.1007/s10566-005-3471-6

*Lunenburg, F. C. & Columba, L. (1992). The 16PF as a predictor of principal performance: An integration of quantitative and qualitative research methods. *Education, 113*(1), 68-73.

*Mabon, H. (1998). Utility aspects of personality and performance. *Human Performance, 11*(2-3), 289-304.

*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.

McCrae, R. R. & John, O. P (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175-215. doi: 10.1111/j.1467-6494.1992.tb00970.x

*McHenry, J. J., Hough, L. M., Toquam, J. L. Hanson, M. A. & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*(2), 335-354.

Meadows, A. J. (1974). *Communication in science*. London, England: Butterworths.

Mesmer-Magnus, J. R. & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535-546. doi: 10.1037/a0013773

*Minbashian, A., Bright, J. E. H. & Bird, K. D. (2009). Complexity in the relationships among the subdimensions of extraversion and job performance in managerial occupations. *Journal of Occupational and Organizational Psychology, 82*(3), 537-549. doi: l0.i348/0963l7908X37l097

*Miner, J. B. (1962). Personality and ability factors in sales performance. *Journal of Applied Psychology, 46*(1), 6-13.

Mol, S. T., Born, M. P., Willemsen, M. E. & Van Der Molen, H. T. (2005). Predicting expatriate job performance for selection purposes: A quantitative review. *Journal of Cross-Cultural Psychology, 36*(5), 590-620. doi: 10.1177/0022022105278544

Motowidlo, S. J. (2003). Job performance. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology* (Vol. 12, pp. 39-53). New York, NY: Wiley.

Mount, M. K., Barrick, M. R. & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*(2/3), 145-165.

*Muchinsky, P. M. & Hoyt, D. P. (1974). Predicting vocational performance of engineers from selected vocational interest, personality, and scholastic aptitude variables. *Journal of Vocational Behavior, 5*(1), 115-123.

*Nikolaou, I. (2003). Fitting the person to the organisation: examining the personality-job performance relationship from a new perspective. *Journal of Managerial Psychology, 18*(7), 639-648. doi: 10.1108/02683940310502368

*Oh, I.-S. & Berry, C. M. (2009). The five-factor model of personality and managerial performance: Validity gains through the use of 360 degree performance ratings. *Journal of Applied Psychology, 94*(6), 1498-1513. doi: 10.1037/a0017221

*Olea, M. M. & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology, 79*(6), 845-851.

*Peacock, A. C. & O'Shea, B. (1984). Occupational therapists: Personality and job performance. *The American Journal Of Occupational Therapy, 38*(8), 517-521.

*Pettersen, N. & Tziner, A. (1995). The cognitive ability test as a predictor of job performance: Is its validity affected by job complexity and tenure within the organization? *International Journal of Selection and Assessment, 3*(4), 237-241.

Petticrew, M. & Roberts, H. (2006). *Systematic reviews in the Social Sciences. A practical guide.* Oxford, England: Blackwell Publishing.

*Piedmont, R. L. & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO Personality Inventory. *Journal of Psychology: Interdisciplinary and Applied, 128*(3), 255-265.

*Plag, J. A. & Goffman, J. M. (1967). The Armed Forces Qualification Test: Its validity in predicting military effectiveness for naval enlistees. *Personnel Psychology, 20*(3), 323-340.

Plat, M. J., Frings-Dresen, M. H. W., & Sluiter, J. K. (2011). A systematic review of job-specific workers' health surveillance activities for fire-fighting, ambulance, police and military personnel. *International Archives of Occupational & Environmental Health, 84*, 839-857. doi: 10.1007/s00420-011-0614-y

*Prien, E. P. (1970). Measuring performance criteria of bank tellers. *Journal of Industrial Psychology, 5*(1), 29-36.

*Prien, E. P. & Cassel, R. H. (1973). Predicting performance criteria of institutional aides. *American Journal of Mental Deficiency, 78*(1), 33-40.

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

Richter, A. W., Dawson, J. F. & West, M. A. (2011). The effectiveness of teams in organizations: A meta-analysis. *The International Journal of Human Resource Management, 22*(13), 2749-2769. doi: 10.1080/09585192.2011.573971

*Robertson, I. T., Baron, H., Gibbons, P., MacIver, R. & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology, 73*, 171-180.

Robertson, I. T. & Kinder, A. (1993). Personality and job competence: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology, 66*, 222-244.

*Rust, J. (1999). Discriminant validity of the 'Big Five' personality traits in employment settings. *Social Behavior and Personality, 27*(1), 99-108.

*Sackett, P. R., Gruys, M. L. & Ellingson, J. E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology, 83*(4), 545-556.

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.

Salgado, J. F. (1998). Big Five personality dimensions and job performance in Army and civil occupations: A European perspective. *Human Performance, 11*(2-3), 271-288.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*(1/2), 117-125.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76*, 323-346.

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

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology, 88*(6), 1068-1081. doi: 10.1037/0021-9010.88.6.1068

*Salgado, J. F. & Rumbo, A. (1997). Personality and job performance in financial services managers. *International Journal of Selection and Assessment, 5*(2), 91-100.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in Psychology. *American Psychologist, 47*(10), 1173-1181.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15,* 187–211. doi: 10.1080/08959285.2002.9668091

Schmidt, F. L., & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods [Computer software]. University of Iowa, Department of Management & Organizations, Iowa City, IA.

Schmitt, N. & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.

Sheehan, C., Fenwick, M. & Dowling, P. J. (2010). An investigation of paradigm choice in Australian international human resource management research. *The International Journal of Human Resource Management, 21*(11), 1816-1836. doi: 10.1080/09585192.2010.505081

*Small, R. J. & Rosenberg, L. J. (1977). Determining job performance in the industrial sales force. *Industrial Marketing Management, 6*(1), 99-102.

*Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84*(6), 959-968.

*Stokes, G. S., Toth, C. S., Searcy, C. A., Stroupe, J. P. & Carter, G. W. (1999). Construct/rational biodata dimensions to predict salesperson performance: Report on the US Department of labor sales study. *Human Resource Management Review, 9*(2), 185-218.

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.

*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. doi: 10.1002/job.191

Thomson Reuters (2010). *Journal Citation Reports® Social Sciences edition 2008*. London, England: Thomson Reuters. Retrieved from Web of Knowledge website: http://apps.isiknowledge.com/additional_ resources.do?highlighted_tab=additional_ resources&product=UA&SID=W1iIe4GeCHKpbH2cbNg&cacheurl=no/JCR?PointOfEntry=Home&SID=Q27iIfaEhdIPf63KEJ1

Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*, 207-222.

*Tyler, G. P. & Newcombe, P. A. (2006). Relationship between work performance and personality traits in Hong Kong organizational settings. *International Journal of Selection and Assessment, 14*(1), 37-50.

*Van Scotter, J. R. (1994). *Evidence for the usefulness of task performance, job dedication and interpersonal facilitation as components of performance* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S. & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83*(4), 586-597.

Viswesvaran, C. & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*(4), 216-226.

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.

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

Warr, P. (1999). Logical and judgmental moderators of the criterion-related validity of personality scales. *Journal of Occupational and Organizational Psychology, 72*, 187-204.

Whitman, D. S., Van Rooy, D. L. & Viswesvaran, C. (2010). Satisfaction, citizenship behaviors, and performance in work units: A meta-analysis of collective construct relations. *Personnel Psychology, 63*(1)*,* 41-81. doi: 10.1111/j.1744-6570.2009.01162.x

Wildman, J. L., Thayer, A. L., Pavlas, D., Salas, E., Stewart, J. E., & Howse, W. R. (2012). Team knowledge research: Emerging trends and critical needs. *Human Factors, 54*(1), 84-111. doi: 10.1177/0018720811425365

*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.

Xiao, S. H., & Nicholson, M. (2013). A multidisciplinary cognitive behavioural framework of impulse buying: A systematic review of the literature. *International Journal of Management Reviews, 15*, 333-356. doi: 10.1177/0018720811425365

*Young, B. S., Arthur, Jr., W. & Finch, J. (2000). Predictors of managerial performance: More than cognitive ability. *Journal of Business and Psychology, 15*(1), 53-72.

Table 1
Sample inclusion/exclusion criteria for study selection

Sample criterion	Explanation/justification for sample criterion
(A) Study is well informed by theory	Explanation of theoretical framework for study; integration of existing body of evidence and theory (Cassell, 2011; Denyer, 2009). The study is framed within appropriate literature. The specific direction taken by the study is justified (Campion, 1993).
(B) Purpose and aims of the study are described clearly	Clear articulation of the research questions and objectives addressed (Denyer, 2009).
(C) Rationale for the sample used is clear	Explanation of and reason for the sampling frame employed and also for potential exclusion of cases (Cassell, 2011; Denyer, 2009). The sample used is appropriate for the study's purpose (Campion, 1993).
(D) Data analysis is appropriate	Clear explanation of methods of data analysis chosen; justification for methods used (Cassell, 2011; Denyer, 2009). The analytical strategies are fit for purpose (Campion, 1993).
(E) Study is published in a peer-reviewed journal[a]	Studies published in peer-reviewed journals are subjected to a process of impartial scrutiny prior to publication, offering quality control (e.g. Meadows, 1974).
(F) Study makes a contribution	Creation, extension or advancement of knowledge in a meaningful way; clear guidance for future research is offered (Academy of Management Review, 2013).

Note. [a]Whilst applying this criterion in particular might seem to disadvantage unpublished studies (e.g. theses), this is not the case, as it is merely one out of a total of eleven quality criteria used for the selection of studies.

Table 2
Mean values used to correct for artifacts

	Type of predictor	
Type of artifact	**Personality**	**Ability**
Range restriction	.94 (*SD* = .04) (Barrick & Mount, 1991; Mount et al., 1998)	.60 (*SD* = .24) (Bertua et al., 2005)
Criterion reliability		
Subjective measures	.52 (*SD* = .10) (Viswesvaran, Ones & Schmidt, 1996)	
Objective measures	Perfect reliability (1.00) is assumed in line with Salgado (1997), Hogan and Holland (2003) and Vinchur et al. (1998)	
Combined use of objective and subjective measures	.59 (*SD* = .19) (Hurtz & Donovan, 2000; Dudley et al., 2006)	

Table 3
Meta-analysis results for overall versus criterion-specific performance dimensions for ability (GMA) and FFM personality measures

								80% CV		
Category	***k***	***N***	$\bar{r}_o$	**ρ (rho)**	***SD*ρ**	**MTV**	SD_{MTV}	CV_{10}	CV_{90}	**%VE**
Ability (GMA)										
Overall job performance	29	13,517	.12	.28	.19	.27	.19	.03	.51	23.67
Task Performance	6	1,148	.10	.23	.24	.22	.23	-.07	.52	31.83
Contextual Performance	7	1,365	.13	.30	.28	.29	.28	-.06	.64	23.65
Job-Specific Task Proficiency (F1)	7	13,562	.40	.79	.10	.76	.10	.64	.89	6.15
Non-Job-Specific Task Proficiency (F2)	6	13,279	.45	.84	.08	.81	.08	.71	.91	7.00
Written and Oral Communication Task Proficiency (F3)	4	789	.07	.17	.24	.16	.23	-.14	.46	34.70
Demonstrating Effort (F4)	7	13,236	.17	.43	0	.41	0	.41	.41	151.50
Maintaining Personal Discipline (F5)	6	12,997	.02	.06	.07	.06	.06	-.02	.14	41.62
Facilitating Peer and Team Performance (F6)	4	925	.14	.28	.49	.27	.47	-.33	.89	6.25
Supervision/Leadership (F7)	4	959	.03	.08	.22	.08	.21	-.19	.34	36.65
Management/Administration (F8)	4	788	.04	.10	.23	.10	.22	-.18	.39	37.75
Openness to Experience										
Overall job performance	22	5,791	.01	.02	.10	.02	.09	-.09	.13	53.02
Task Performance	4	784	.21	.35	.24	.31	.21	.04	.58	18.15
Contextual Performance	4	784	.14	.24	.23	.22	.20	-.04	.48	20.46
Job-Specific Task Proficiency (F1)	5	642	-.05	-.08	0	-.08	0	-.08	-.08	100.37
Non-Job-Specific Task Proficiency (F2)	7	1,399	.04	.07	.18	.06	.16	-.15	.27	29.92
Written and Oral Communication Task Proficiency (F3) [a]										
Demonstrating Effort (F4)	4	731	-.07	-.12	0	-.11	0	-.11	-.11	194.08
Maintaining Personal Discipline (F5)	3	413	.27	.45	.13	.40	.12	.24	.56	47.87
Facilitating Peer and Team Performance (F6)	5	836	-.04	-.07	.05	-.06	.04	-.11	-.01	88.69
Supervision/Leadership (F7)	4	933	-.04	-.07	0	-.06	0	-.06	-.06	263.73
Management/Administration (F8)	2	198	-.12	-.20	0	-.18	0	-.18	-.18	871.52
Conscientiousness										
Overall job performance	36	9,205	.09	.14	.14	.13	.13	-.04	.29	33.44
Task Performance	6	1,230	.18	.29	.18	.25	.16	.05	.46	28.38
Contextual Performance	7	1,353	.14	.24	.22	.21	.20	-.04	.47	21.49
Job-Specific Task Proficiency (F1)	10	10,195	.04	.07	.08	.06	.07	-.03	.15	31.76
Non-Job-Specific Task Proficiency (F2)	13	12,806	.05	.08	.07	.07	.06	-.01	.15	35.38
Written and Oral Communication Task Proficiency (F3)	4	654	.10	.18	0	.16	0	.16	.16	201.20
Demonstrating Effort (F4)	10	12,236	.15	.26	.14	.23	.12	.07	.39	10.00

								80% CV		
Category	***k***	***N***	$\bar{r}_o$	**ρ (rho)**	***SD*ρ**	**MTV**	SD_{MTV}	CV_{10}	CV_{90}	**%VE**
Maintaining Personal Discipline (F5)	6	9,434	.21	.35	0	.31	0	.31	.31	190.92
Facilitating Peer and Team Performance (F6)	9	3,380	.05	.08	.06	.07	.05	.01	.13	70.85
Supervision/Leadership (F7)	9	4,247	.03	.05	.15	.04	.13	-.13	.21	20.94
Management/Administration (F8)	6	850	.01	.02	.09	.02	.08	-.08	.12	73.00
Extraversion										
Overall job performance	33	8,608	.05	.09	.15	.08	.13	-.09	.25	32.07
Task Performance	5	1,163	.11	.19	.29	.16	.26	-.16	.49	12.21
Contextual Performance	6	1,286	.15	.25	.32	.22	.29	-.14	.59	10.39
Job-Specific Task Proficiency (F1)	8	9,663	.03	.05	.07	.05	.07	-.04	.13	30.16
Non-Job-Specific Task Proficiency (F2)	11	12,289	.02	.03	.05	.03	.05	-.03	.09	49.00
Written and Oral Communication Task Proficiency (F3)	3	340	.10	.17	.09	.15	.08	.05	.25	76.97
Demonstrating Effort (F4)	10	12,021	.14	.24	.12	.22	.11	.07	.36	12.74
Maintaining Personal Discipline (F5)	6	9,434	.13	.22	.07	.20	.06	.12	.27	26.81
Facilitating Peer and Team Performance (F6)	8	3,066	.06	.10	.14	.09	.12	-.06	.25	28.30
Supervision/Leadership (F7)	7	3,712	.09	.15	.15	.14	.13	-.03	.30	19.67
Management/Administration (F8)	4	317	-.06	-.10	0	-.09	0	-.09	-.09	1,136.19
Agreeableness										
Overall job performance	25	7,184	.00	.01	.17	.00	.15	-.19	.20	24.70
Task Performance	5	1,163	.09	.16	.16	.14	.14	-.04	.32	32.30
Contextual Performance	5	1,163	.21	.35	.20	.31	.18	.08	.54	20.74
Job-Specific Task Proficiency (F1)	7	8,774	.01	.02	.01	.02	.01	.00	.03	91.60
Non-Job-Specific Task Proficiency (F2)	9	9,570	.01	.02	.07	.02	.06	-.06	.10	36.31
Written and Oral Communication Task Proficiency (F3)	3	340	.21	.36	0	.32	0	.32	.32	208.92
Demonstrating Effort (F4)	7	9,123	.13	.22	.10	.20	.09	.08	.32	16.58
Maintaining Personal Discipline (F5)	5	8,545	.16	.26	.13	.23	.12	.08	.38	8.29
Facilitating Peer and Team Performance (F6)	6	1,057	.07	.13	.21	.11	.18	-.12	.35	27.54
Supervision/Leadership (F7)	5	993	.14	.24	.19	.22	.17	-.00	.44	26.31
Management/Administration (F8)	5	652	.16	.27	.15	.24	.14	.06	.41	46.16
Emotional Stability										
Overall job performance	29	7,099	.04	.07	.13	.06	.11	-.08	.20	41.87
Task Performance	4	733	.19	.32	.25	.28	.22	.00	.57	18.16
Contextual Performance	5	856	.17	.29	.31	.26	.28	-.09	.66	13.09
Job-Specific Task Proficiency (F1)	7	8,774	.01	.02	.02	.02	.01	-.00	.04	89.03
Non-Job-Specific Task Proficiency (F2)	8	9,471	.02	.03	.07	.02	.06	-.05	.10	34.90
Written and Oral Communication Task Proficiency (F3)	2	119	.16	.27	0	.24	0	.24	.24	579.49

								80% CV		
Category	***k***	***N***	$\bar{r}_o$	**ρ (rho)**	***SD*ρ**	**MTV**	SD_{MTV}	CV_{10}	CV_{90}	**%VE**
Demonstrating Effort (F4)	5	8,803	.15	.25	.12	.22	.10	.09	.36	10.22
Maintaining Personal Discipline (F5)	5	8,545	.13	.22	.05	.20	.05	.14	.26	35.17
Facilitating Peer and Team Performance (F6)	5	836	-.07	-.13	0	-.11	0	-.11	-.11	195.80
Supervision/Leadership (F7)	5	993	-.03	-.05	.16	-.05	.14	-.22	.13	37.66
Management/Administration (F8)	4	317	-.05	-.08	.13	-.07	.11	-.22	.07	69.25

Note. k = number of correlations; N = total sample size across meta-analyzed correlations; $\bar{r}_o$ = sample-size weighted mean observed correlation (uncorrected for artifacts); ρ (rho) = mean true score correlation; *SD*ρ = standard deviation of true score correlation; MTV = mean true predictive validity; SD_{MTV} = mean true standard deviation; 80% CV = 80% credibility intervals (CV_{10} = lower-bound 80% credibility interval; CV_{90} = upper-bound 80% credibility interval); %VE = percent variance accounted for by artifact corrections for each corrected correlation distribution.

[a]It was not possible to compute results for this combination, as none of the studies had assessed how well Openness to Experience can be a predictor of Factor 3

Table 4
Comparison of current meta-analysis results with findings from previous studies

Predictor construct	Predictive validity in current study	Predictive validity in previous studies
Ability/GMA	.27	.44 (US) (Salgado & Anderson, 2003)
		.56-.68 (European Community) (Salgado & Anderson, 2003)
Openness to Experience	.02	.07 (Barrick et al., 2001)
Conscientiousness	.13	.27 (Barrick et al., 2001)
Extraversion	.08	.15 (Barrick et al., 2001)
Agreeableness	.00	.13 (Barrick et al., 2001)
Emotional Stability	.06	.13 (Barrick et al., 2001)

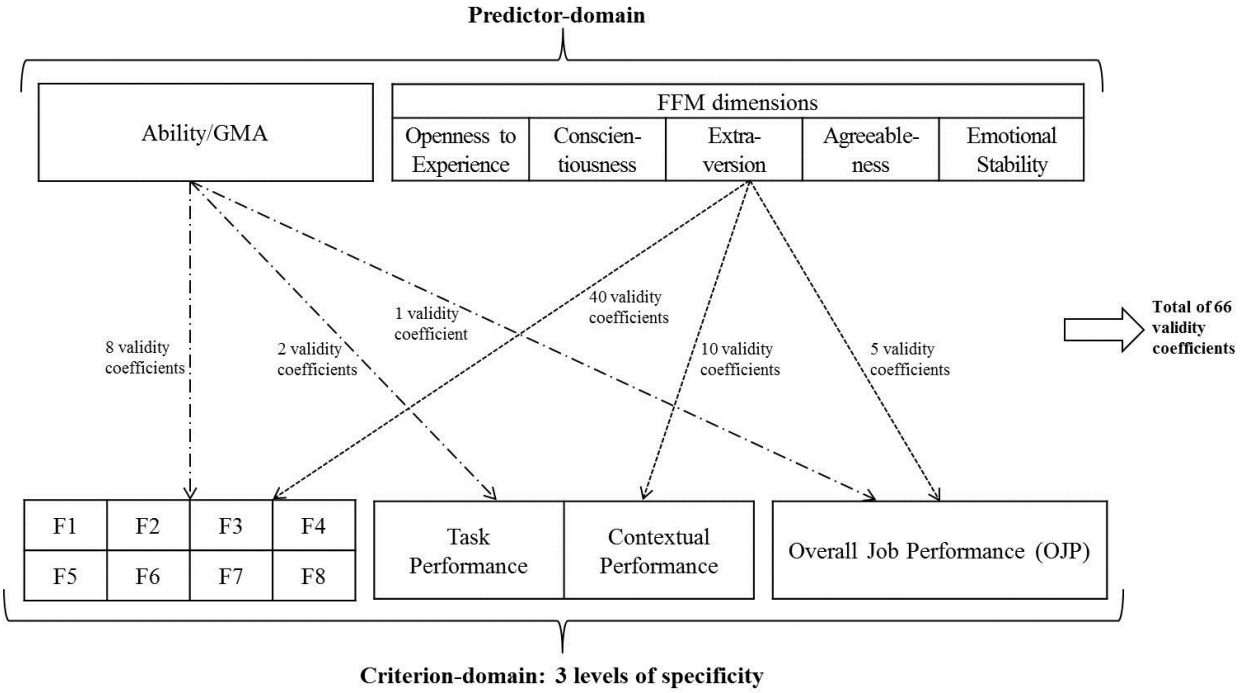

Figure 1
Overview of meta-analysis

Note. F1 to F8 represent the eight factors from Campbell et al.'s performance taxonomy, that is Job-Specific Task Proficiency (F1), Non-Job-Specific Task Proficiency (F2), Written and Oral Communication Task Proficiency (F3), Demonstrating Effort (F4), Maintaining Personal Discipline (F5), Facilitating Peer and Team Performance (F6), Supervision/Leadership (F7), Management/Administration (F8)