

Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next?

Murray R. Barrick, Michael K. Mount and Timothy A. Judge*

As we begin the new millennium, it is an appropriate time to examine what we have learned about personality-performance relationships over the past century and to embark on new directions for research. In this study we quantitatively summarize the results of 15 prior meta-analytic studies that have investigated the relationship between the Five Factor Model (FFM) personality traits and job performance. Results support the previous findings that conscientiousness is a valid predictor across performance measures in all occupations studied. Emotional stability was also found to be a generalizable predictor when overall work performance was the criterion, but its relationship to specific performance criteria and occupations was less consistent than was conscientiousness. Though the other three Big Five traits (extraversion, openness and agreeableness) did not predict overall work performance, they did predict success in specific occupations or relate to specific criteria. The studies upon which these results are based comprise most of the research that has been conducted on this topic in the past century. Consequently, we call for a moratorium on meta-analytic studies of the type reviewed in our study and recommend that researchers embark on a new research agenda designed to further our understanding of personality-performance linkages.

Introduction

The relationship between personality and job performance has been a frequently studied research topic in industrial-organizational psychology in the past century. Generally speaking, the research can be categorized into two distinct phases. The first phase spans a relatively long time period and includes studies conducted from the early 1900s through the mid-1980s. Research conducted during this time period was characterized by primary studies in which researchers investigated the relationships of individual scales from numerous personality inventories to various aspects of job performance. The overall conclusion from this body of research was that personality and job performance were not related in any meaningful way across traits and across situations. In fact, some have sarcastically referred to this as the time when we had no personalities. As Guion and Gottier (1965, p. 159) noted in their influential review, [T]here is no generalizable evidence that personality measures can be recommended as good or practical tools for

employee selection.' For the most part, this conclusion went unchallenged for 25 years.

There are several possible explanations for these pessimistic conclusions. First, no classification system was used to reduce the thousands of personality traits into a smaller, more manageable number. Second, there was lack of clarity about the traits being measured. For example, in some cases researchers were using the same name to refer to traits with different meanings and in others were using different names for traits with the same meaning. A related problem was that researchers did not distinguish between measurement of personality at the construct level and measurement at the inventory scale level. Researchers implicitly treated each individual personality scale as if it measured a distinct construct, rather than recognizing that each scale from a personality inventory assessed only one aspect or facet of a larger construct. Further, much of the research at this time was characterized by a 'shotgun' approach in which the relationship of all personality scales on personality inventories was correlated with all the criteria investigated

* Address for correspondence: Murray R. Barrick, Department of Management, Broad Graduate College of Business, Michigan State University, East Lansing, MI 48824-1122. E-mail: Barrick@pilot.msu.edu

in the study. Not surprisingly, researchers found that many of the correlations were near zero. (Of course, this is exactly what one would expect when the presumed personality-performance linkages had not been established theoretically or through job analysis.) Finally, the reviews of the literature at this time were largely narrative rather than quantitative, and did not correct for study artifacts that led to downwardly biased validity estimates. Understandably, these problems made it difficult to identify consistent relationships among personality traits and criteria and, consequently, little advancement was made in understanding personality performance relationships.

Currently, the area is experiencing something of a renaissance. The second phase, which covers the period from the mid-1980s to the present, is characterized by the use of the FFM, or some variant, to classify personality scales. Most primary studies conducted since 1990 have used instruments that assess personality traits at the FFM level, or have used the FFM to classify individual scales from personality inventories. The second distinguishing characteristic is the use of meta-analytic methods to summarize results quantitatively across studies. To date, there have been 15 meta-analytic studies of personality-performance relationships (11 published articles and 4 conference presentations). Taken together, the results of both the primary studies using FFM constructs and meta-analytic studies using FFM constructs appear to have led to more optimistic conclusions than those from the prior era, and have helped increase our understanding of personality-performance relationships. Thus, in contrast to the previous era, it appears that we do have a personality and that at least some aspects of it are meaningfully related to performance. As Guion (1998, p.145) recently noted, 'Meta-analyses have provided grounds for optimism.'

As we begin a new millennium, we believe that it is an appropriate time to take stock of what we have learned. Therefore there are two major aims of this study. The first is to quantitatively review previous meta-analytic studies about personality-performance linkages. The second major aim is to suggest directions for future research that can further our knowledge of personality-performance relations.

Although recent research has provided grounds for optimism, a close examination of the findings from the extant quantitative reviews reveals some discrepancies in the results. For example, Barrick and Mount (1991) found that conscientiousness was the only FFM trait to display non-zero correlations with job performance across different occupational groups and criterion types. In contrast, Tett, Rothstein and Jackson (1991) found that only emotional

stability displayed non-zero correlations with performance, and two other Big Five traits – agreeableness and openness – displayed higher correlations with performance than conscientiousness. More recently, Salgado (1997) and Anderson and Viswesvaran (1998) found that two traits from the five-factor model – emotional stability and conscientiousness – displayed non-zero correlations with job performance. Other meta-analyses have also been conducted, with as much variance in the findings as those reported above (Hough 1992; Salgado 1998). Referring only to the first two studies, Goldberg (1993) has described the differences in findings based on a similar body of knowledge as 'befuddling' (p. 31).

Meta-analysis has effectively demonstrated that differences in correlations across primary studies are often more a function of small sample sizes than meaningful differences in the nature of the relationship between two variables across settings. Similarly, some of the differences in true score correlations reported in prior meta-analyses may be due to these estimates being based on only a few studies or relatively small samples. For example, the widely cited meta-analytic result for agreeableness in the Tett *et al.* (1991) study ($\rho = .33$) is based on only four primary studies and a total sample size of 280.¹ When meta-analyses are based on a small number of studies, the average sampling error effects will still be largely unaccounted for, thus biasing the meta-analytic mean and standard deviation estimates. Hunter and Schmidt (1990) referred to this problem as second-order sampling error. Therefore, we examine whether the differences across meta-analyses are an artifact of small samples in some of these analyses. To further minimize artifactual differences across meta-analyses investigated in this analysis, we corrected for statistical artifacts in similar ways across all meta-analyses. Thus, the primary aim of this study is to clarify what we have learned over the last century of research in this area by conducting a second-order meta-analysis of existing meta-analyses.

Having provided the context for our study, in the next two sections we provide a brief review of the five-factor model of personality, and review linkages between the Big Five traits and job performance.

Five-Factor Model of Personality

Recent meta-analyses have utilized a construct-oriented approach to study the relationship between specific personality traits and performance in various jobs. The FFM or 'Big Five' has been the most frequently used taxonomy in these meta-analyses. Although the five-factor

model enjoys widespread support, critics have challenged the model on numerous grounds (Block 1995; Eysenck 1992). In response to these concerns, a number of theoretical and empirical developments supporting the Big Five model have emerged in the past few years. This evidence includes: (a) demonstrations of the genetic influences on measures constituting the five factor model, with (uncorrected) heritability estimates ranging from .39 for agreeableness to .49 for extraversion (Bouchard 1997); (b) the stability of the Big Five model across the lifespan (Conley 1984; Costa and McCrae 1988); and (c) the replicability of the five factor structure across different theoretical frameworks, using different assessment approaches including questionnaires and lexical data, in different cultures, with different languages, and using ratings from different sources (e.g., Digman and Shmelyov 1996). While there is not universal agreement on the Big Five model, it is a useful taxonomy and currently the one considered most useful in personality research.

Although there is some disagreement about the names and content of these five personality dimensions, they generally can be defined as follows. Extraversion consists of sociability, dominance, ambition, positive emotionality and excitement-seeking. Cooperation, trustfulness, compliance and affability define agreeableness. Emotional stability is defined by the lack of anxiety, hostility, depression and personal insecurity. Conscientiousness is associated with dependability, achievement striving, and planfulness. Finally, intellectance, creativity, unconventionality and broad-mindedness define openness to experience. Taken together, the five-factor model has provided a comprehensive yet parsimonious theoretical framework to systematically examine the relationship between specific personality traits and job performance.

Relations between the FFM Traits and Job Performance

Adopting the FFM taxonomy has enabled researchers to develop specific hypotheses about the predictive validity of personality constructs at work. Prior meta-analytic evidence suggests some FFM traits are related to overall job performance in virtually all jobs, whereas other traits are related to performance in only a few jobs. For example, while agreeableness may be a useful predictor of service orientation and teamwork, extraversion and openness to experience appear to be related to training proficiency. In this section, we review prior meta-analytic findings regarding the relationship between the FFM traits and job performance.

Most meta-analyses have suggested that two of the FFM – conscientiousness and emotional stability – are positively correlated with job performance in virtually all jobs (Anderson and Viswesvaran 1998; Barrick and Mount 1991; Salgado 1997; Tett *et al.* 1991). Of these two traits, most meta-analyses have suggested that conscientiousness is somewhat more strongly related to overall job performance than is emotional stability. Indeed, it is hard to conceive of a job where it is beneficial to be careless, irresponsible, lazy, impulsive and low in achievement striving (low conscientiousness). Therefore, employees with high scores on conscientiousness should also obtain higher performance at work. Similarly, being anxious, hostile, personally insecure and depressed (low emotional stability) is unlikely to lead to high performance in any job. Thus, we expect that conscientiousness and emotional stability will be positively related to overall performance across jobs.

These two personality dimensions are also expected to be related to some specific dimensions of performance. First, both conscientiousness and emotional stability are expected to influence success in teamwork (Hough 1992; Mount, Barrick and Stewart 1998). In jobs involving considerable interpersonal interaction, being more dependable, thorough, persistent and hard working (high in conscientiousness), as well as being calm, secure and not depressed or hostile (high in emotional stability), should result in more effective interactions with co-workers or customers. Second, employees who approach training in a careful, thorough, persistent manner (high in conscientiousness) are more likely to benefit from training (Barrick and Mount 1991). Based on these findings, we believe that emotional stability and conscientiousness will be positively related to measures of teamwork performance, and that conscientiousness will be positively related to performance in training.

The other three FFM dimensions are expected to be valid predictors of performance, but only in some occupational groups or for specific criteria. For example, extraversion has been found to be related to job performance in occupations where interactions with others are a significant portion of the job (Barrick and Mount 1991; Mount *et al.* 1998). In such jobs, such as sales and management, being sociable, gregarious, assertive, energetic and ambitious is likely to contribute to success on the job. Furthermore, if working in a team comprises an important component of the work, higher scores on extraversion would be expected to be related to more effective teamwork. Finally, although it has not been frequently examined, there is some meta-analytic evidence (Barrick and Mount 1991; Hough 1992) which shows that higher scores on extraversion are associated with greater training proficiency.

One explanation for this finding is that highly extraverted trainees are more active during training and ask more questions, which enables them to learn more efficiently. Based on these arguments, we expect that extraversion will be positively related to performance in jobs with an important interpersonal component, especially those involving mentoring, leading, persuading – such as management and sales positions – and that extraverted employees will receive higher ratings on teamwork and be more successful in training.

The final two dimensions, agreeableness and openness to experience, are generally expected to have weak relationships with overall job performance. The one situation in which agreeableness appears to have high predictive validity is in jobs that involve considerable interpersonal interaction, particularly when the interaction involves helping, cooperating and nurturing others. In fact, in those settings, agreeableness may be the single best personality predictor (Barrick, Stewart, Neubert and Mount 1998; Mount *et al.* 1998). Thus, employees who are more argumentative, inflexible, uncooperative, uncaring, intolerant and disagreeable (low in agreeableness) are likely to have lower ratings of teamwork. Turning to openness to experience, training proficiency is the one criterion that has consistently been associated with openness (Barrick and Mount 1991; Salgado 1997). It appears that employees who are intellectual, curious, imaginative, and have broad interests are more likely to benefit from the training. These employees are likely to be 'training ready' or more willing to engage in learning experiences. Taken together, we expect that agreeableness will be positively related to ratings of teamwork, while openness to experience will be positively related to training performance.

Method

Literature Review

All prior meta-analyses of the relationship between personality (categorized using the Big Five model or some variant) and job performance conducted during the 1990s and published or presented at a conference served as the source of data for this study. Three separate methods were used to locate appropriate meta-analyses for use in the present study. First, a computer-based literature search was conducted using PsycLIT. Second, a manual article-by-article search was conducted in the following journals for the period of time from 1990 to 1998: *Journal of Applied Psychology*, *Personnel Psychology*, *Academy of Management Journal*, *Organizational Behavior and Human Decision Processes*, and *Journal of*

Management. Third, conference programs from the last four annual conferences (1995–98) of both the Society for Industrial and Organizational Psychology (SIOP) and the Academy of Management were searched for potential meta-analyses to be included in the present review.

All told, 15 meta-analyses (11 published articles and 4 paper presentations) reporting usable statistics were identified. The Appendix provides the meta-analyses and identifies the criteria and occupations for which data are reported. Data from 11 of these meta-analyses were included in this study. To be included, the meta-analysis had to use only actual workers or applicants as subjects. Second, the meta-analysis had to include a measure of personality that had been categorized or could be categorized according to the Big Five model. Third, the data for a specific criterion or occupational group had to be assessed in at least two separate meta-analyses. Thus, the meta-analysis by Organ and Ryan (1995), for example, was excluded, as it was the only known meta-analysis to report findings for organizational citizenship. Fourth, a meta-analysis could not report data that was largely or fully incorporated into another, larger meta-analysis. For example, the meta-analytic data reported by Hough *et al.* (1990) are included in Hough (1992). In contrast, meta-analyses were included (e.g. Hough [1992] and Barrick and Mount [1991]), if there was only partial overlap in the studies included.

However, using meta-analyses that included some of the same studies will 'double count' some primary studies. This non-independent sampling procedure will produce inflated counts of the number of studies and the total sample size, as well as bias the true score standard deviation estimates (it would not, however, bias estimates of the mean correlation). To address this issue, two sets of analyses were conducted in relating the Big Five traits to an overall measure of work performance. One analysis was based on a set of meta-analyses with no overlap in studies (independent analyses). The other analysis, in addition to including those analyses used in the independent analysis, included other studies with unknown or even substantial overlap in primary studies (non-independent analyses). The first set of analyses (i.e., the independent analyses) provides bias-free estimates of the population effects. It will be the focus of our discussion. The second set of analyses (i.e., non-independent analyses) reports a comprehensive estimate based on all available meta-analyses, but whose variability estimates may be biased by including the same set of primary studies in multiple meta-analyses. These results are included for comparative purposes only for the overall work performance and are of secondary interest in the study.

As noted earlier, the Appendix details which prior meta-analysis is included in each second-order meta-analysis conducted to this study for each criterion type, occupational group, and across levels of analysis. The Appendix also reports which meta-analysis was included in the independent analysis or in both the non-independent and independent analyses.

Analyses

The purpose of this study was to summarize the cumulative knowledge that has accrued over the past century pertaining to personality and performance relations. The quantitative analyses we conducted are referred to as a second-order meta-analysis (Hunter and Schmidt 1990). It is standard practice to calculate meta-analytic estimates using relative sample weights (Hunter and Schmidt 1990). Consequently, all the analyses reported in this study weight all values by their respective sample sizes. Therefore, our second-order ρ and $SD\rho$ estimates reflect a weighted average of all of the meta-analytically derived ρ and $SD\rho$ estimates reported in prior meta-analyses. We believe our second-order sample size weighted ρ and $SD\rho$ estimates (labeled $\bar{\rho}^{sw}$ and $SD\bar{\rho}^{sw}$ in Tables 1 to 5) provide a better approximation of the actual population ρ and actual residual variance than those reported in any single previous meta-analysis. To estimate the validity of each Big Five dimension at the construct level, the second-order sample size weighted $\bar{\rho}$ was adjusted for imperfect construct assessment. This occurs because prior meta-analyses probably understated the magnitude of validities by examining facets of the trait, rather than comprehensive measures of the Big Five construct (Barrick and Mount 1991; Mount and Barrick 1995; Salgado 1997). As suggested by Mount and Barrick (1995), these $\bar{\rho}^{sw}$ estimates were adjusted using the composite score formula of Hunter and Schmidt (1990). This formula combines the lower level of the elements or facets as if the raw scores were summed to an overall construct. In this study, we used the average correlation between Big Five predictor constructs reported by Ones (1993). The sample size for these analyses are based on N as reported in each table. This results in $\bar{\rho}^{FFM}$, which is a construct valid second-order sample size weighted estimate of the population effect size.

Furthermore, rather than conduct a chi-square test for homogeneity of predictive validity coefficients, we believe our weighted average $SD\rho$ estimates coupled with the sample weighted reciprocal average percentage of variance accounted for by statistical artifacts across meta-analyses should be taken at face value as an indicator of the variation in effect

sizes across studies. However, to provide additional evidence about the amount of variance expected across studies, we also report the standard deviation of prior meta-analytic ρ 's (SD of ρ).

Finally, because our interest was in the true (theoretical) relationship between the Big Five and job performance, all the meta-analytic estimates were fully corrected for measurement error in the predictor and criterion, as well as for range restriction (Hunter and Schmidt 1990). In those meta-analyses where the ρ and $SD\rho$ estimates were not fully corrected for these artifacts, sample weighted average artifact estimates were used to correct these values. In addition, a few prior meta-analyses (e.g. Hough *et al.* 1990; Hough 1992) did not report variance estimates. In these cases, the average sample-weighted variance estimates from other meta-analyses were used to estimate the studies expected variance estimate. Thus, all meta-analytic estimates were corrected for statistical artifacts in similar ways across all meta-analyses.

The complete results of the second-order meta-analytic findings for each of the Big Five traits are presented in Tables 1 to 5 (each table reports results for a Big Five trait). Analyses are conducted across a number of performance criteria, including overall work performance, supervisory ratings of performance, objective indicators of performance, training and teamwork. Overall work performance consists of measures of overall job performance, which includes ratings as well as productivity data. Objective indicators of performance include productivity data, turnover, promotions and salary measures. Additional analyses are conducted across specific occupational groups (sales, managers, professionals, police and skilled or semi-skilled labor). For each of these categories, analyses were conducted using only independent samples. Although overall analysis results suggested that violating the assumption of independence across primary studies had a minimal biasing effect on the true score correlation, we took the conservative approach in reporting criteria and occupational validities based solely on independent studies.

The second column of Tables 1–5 provides the number of prior meta-analyses each analysis is based on, while the third and fourth summarizes the number of studies (K) and total sample size (N) of studies reported in the prior meta-analyses. The next three columns report the second-order sample size weighted effect size (uncorrected [observed r^{sw}]), the sample weighted, corrected ($\bar{\rho}^{sw}$) estimate, and the construct valid population estimate ($\bar{\rho}^{FFM}$). Columns 8–10 report the variability in effect sizes ($SD\bar{\rho}^{sw}$, standard deviation of prior meta-analytic ρ 's, and the percentage variance

Table 1: Summary of second-order meta-analytic results for extraversion across criteria and occupational groups

Criteria	No. MAs	K	N	Obs r^{sw}	ρ^{sw}	ρ^{FFM}	$SD\rho^{sw}$	Std Dev of ρ 's	% Var Accted ^{sw}	CV _L	CV _U
Work performance											
Non-independent	8	559	82,032	.08	.12	.15	.15	.04	21	-.07	.32
Independent	5	222	39,432	.06	.12	.15	.12	.04	43	-.03	.27
Specific criteria											
Supervisor ratings	4	164	23,785	.07	.11	.13	.14	.03	26	-.06	.29
Objective performance	2	37	7,101	.06	.11	.13	.17	.02	36	-.12	.33
Training performance	2	21	3,484	.13	.23	.28	.12	.17	112	.07	.39
Teamwork	2	48	3,719	.08	.13	.16	.00	.01	171	.13	.13
Specific occupations											
Sales performance	3	35	3,806	.07	.09	.11	.16	.18	54	-.11	.29
Managerial performance	3	67	12,602	.10	.17	.21	.12	.08	52	.01	.32
Professionals	1	4	476	-.05	-.09	-.11	.05		92	-.15	-.03
Police	2	20	2,074	.06	.10	.12	.00	.03	129	.10	.10
Skilled or semi-skilled	3	44	6,830	.03	.05	.06	.07	.06	71	-.05	.14

Note: ^{sw} Sample Weighted Estimates. CV_U and CV_L are the upper and lower limits of a 90% credibility value for ρ^{sw} . ρ^{sw} = estimated sample weighted true correlation at the scale level; ρ^{FFM} = estimated true correlation at the construct level. No. MAs = number of independent meta-analyses included in the analysis; K and N = number of studies and subjects reported across meta-analyses, respectively.

Table 2: Summary of second-order meta-analytic results for emotional stability across criteria and occupational groups

Criteria	No. MAs	K	N	Obs r^{sw}	ρ^{sw}	ρ^{FFM}	$SD\rho^{sw}$	Std Dev of ρ 's	% Var Accted ^{sw}	CV _L	CV _U
Work performance											
Non-independent	8	453	73,047	.09	.14	.15	.09	.04	53	.02	.26
Independent	5	224	38,817	.06	.12	.13	.08	.03	66	.01	.22
Specific criteria											
Supervisor ratings	4	167	23,687	.07	.12	.13	.07	.03	75	.03	.20
Objective performance	2	32	6,219	.05	.09	.10	.15	.01	45	-.10	.27
Training performance	2	25	3,753	.05	.08	.09	.00	.08	127	.08	.08
Teamwork	2	41	3,558	.13	.20	.22	.00	.01	542	.20	.20
Specific occupations											
Sales performance	3	30	3,664	.03	.05	.05	.15	.07	115	-.14	.25
Managerial performance	3	63	11,591	.05	.08	.09	.09	.01	66	-.03	.19
Professionals	2	8	926	.04	.06	.06	.08	.30	73	-.05	.16
Police	2	22	2,275	.07	.11	.12	.00	.04	368	.11	.11

Note: ^{sw} Sample Weighted Estimates. CV_U and CV_L are the upper and lower limits of a 90% credibility value for ρ^{sw} . ρ^{sw} = estimated sample weighted true correlation at the scale level; ρ^{FFM} = estimated true correlation at the construct level. No. MAs = number of independent meta-analyses included in the analysis; K and N = number of studies and subjects reported across meta-analyses, respectively.

Table 3: Summary of second-order meta-analytic results for agreeableness across criteria and occupational groups

Criteria	No. MAs	K	N	Obs r^{sw}	ρ^{sw}	ρ^{FFM}	$SD\rho^{sw}$	Std Dev of ρ 's	% Var Accted ^{sw}	CV _L	CV _U
<i>Work performance</i>											
Non-independent	8	308	52,633	.06	.09	.11	.09	.10	46	-.02	.21
Independent	5	206	36,210	.06	.10	.13	.09	.08	47	-.01	.22
<i>Specific criteria</i>											
Supervisor ratings	4	151	22,193	.06	.10	.13	.08	.09	58	.00	.20
Objective performance	2	28	4,969	.07	.13	.17	.10	.09	79	.00	.25
Training performance	2	24	4,100	.07	.11	.14	.01	.06	129	.10	.12
Teamwork	2	17	1,820	.17	.27	.34	.00	.00	272	.27	.27
<i>Specific occupations</i>											
Sales performance	3	27	3,551	.01	.01	.01	.21	.02	34	-.26	.28
Managerial performance	3	55	9,864	.04	.08	.10	.03	.10	95	.04	.11
Professionals	2	10	965	.03	.05	.06	.03	.05	121	.01	.09
Police	2	18	2,015	.06	.10	.13	.04	.01	99	.05	.15
Skilled or semi-skilled	3	44	7,194	.05	.08	.10	.11	.05	77	-.06	.22

Note: ^{sw} Sample Weighted Estimates. CV_U and CV_L are the upper and lower limits of a 90% credibility value for ρ^{sw} . ρ^{sw} = estimated sample weighted true correlation at the scale level; ρ^{FFM} = estimated true correlation at the construct level. No. MAs = number of independent meta-analyses included in the analysis; K and N = number of studies and subjects reported across meta-analyses, respectively.

Table 4: Summary of second-order meta-analytic results for conscientiousness across criteria and occupational groups

Criteria	No. MAs	K	N	Obs r^{sw}	ρ^{sw}	ρ^{FFM}	$SD\rho^{sw}$	Std Dev of ρ 's	% Var Accted ^{sw}	CV_L	CV_U
<i>Work performance</i>											
Non-independent	8	442	79,578	.12	.20	.24	.11	.05	18	.06	.34
Independent	5	239	48,100	.12	.23	.27	.10	.05	30	.10	.35
<i>Specific criteria</i>											
Supervisor ratings	4	185	33,312	.15	.26	.31	.11	.06	17	.11	.40
Objective performance	2	35	6,905	.10	.19	.23	.09	.09	76	.07	.30
Training performance	2	20	3,909	.13	.23	.27	.14	.01	79	.05	.41
Teamwork	2	38	3,064	.15	.23	.27	.00	.04	315	.23	.23
<i>Specific occupations</i>											
Sales performance	3	36	4,141	.11	.21	.25	.00	.07	491	.21	.21
Managerial performance	3	60	11,325	.12	.21	.25	.10	.06	65	.08	.33
Professionals	1	6	767	.11	.20	.24	.00		106	.20	.20
Police	2	22	2,369	.13	.22	.26	.17	.01	103	.00	.44
Skilled or semi-skilled	3	44	7,682	.12	.19	.23	.08	.04	60	.09	.29

Note: ^{sw} Sample Weighted Estimates. CV_U and CV_L are the upper and lower limits of a 90% credibility value for ρ^{sw} . ρ^{sw} = estimated sample weighted true correlation at the scale level; ρ^{FFM} = estimated true correlation at the construct level. No. MAs = number of independent meta-analyses included in the analysis; K and N = number of studies and subjects reported across meta-analyses, respectively.

Table 5: Summary of second-order meta-analytic results for openness to experience across criteria and occupational groups

Criteria	No. MAs	K	N	Obs r^{sw}	ρ^{sw}	ρ^{FFM}	$SD\rho^{sw}$	Std Dev of ρ 's	% Var Accted ^{sw}	CV_L	CV_U
<i>Work performance</i>											
Non-independent	7	218	38,786	.03	.05	.07	.14	.09	37	-.12	.23
Independent	4	143	23,225	.03	.05	.07	.11	.03	53	-.09	.19
<i>Specific criteria</i>											
Supervisor ratings	4	116	18,535	.03	.05	.07	.12	.04	41	-.11	.21
Objective performance	2	25	4,401	.02	.02	.03	.13	.04	53	-.15	.18
Training performance	2	18	3,177	.14	.24	.33	.14	.06	66	.06	.41
Teamwork	2	10	2,079	.08	.12	.16	.00	.05	272	.12	.12
<i>Specific occupations</i>											
Sales performance	2	17	2,168	-.01	-.02	-.03	.16	.01	46	-.22	.19
Managerial performance	3	44	8,678	.05	.07	.10	.15	.03	41	-.12	.27
Professionals	1	4	476	-.05	-.08	-.11	.04		94	-.13	-.03
Police	2	16	1,688	.02	.02	.03	.00	.08	217	.02	.02
Skilled or semi-skilled	3	32	6,055	.03	.04	.05	.09	.05	67	-.08	.15

Note: ^{sw} Sample Weighted Estimates. CV_U and CV_L are the upper and lower limits of a 90% credibility value for ρ^{sw} . ρ^{sw} = estimated sample weighted true correlation at the scale level; ρ^{FFM} = estimated true correlation at the construct level. No. MAs = number of independent meta-analyses included in the analysis; K and N = number of studies and subjects reported across meta-analyses, respectively.

accounted^{sw} for by statistical artifacts). Finally, the last two columns report the lower and upper bound of the 90% credibility values for the second-order sample's weighted ρ and $SD\rho$ estimates.

Results

A main aim of this study was to summarize the results of earlier meta-analytic findings for the five-factor model of personality. Figure 1 reports the true score correlations (and 90% credibility values) for eight separate meta-analyses between specific FFM traits and overall measures of work performance, as well as the second-order meta-analytic result (ρ) from this study. True score correlations across meta-analyses of the same literature should be consistent and, generally speaking this is what we found. Figure 1 shows the relatively small differences reported across study ρ estimates. In fact, only 3 of 40 study ρ estimates exceeded the 90% credibility value for the second-order estimate ($\bar{\rho}^{sw}$). Tett *et al.*'s (1991) ρ estimates for agreeableness, emotional stability and openness to experience (.34, .23 and .28, respectively), exceeded the 90% credibility values (90% $CV_u = .27, .22$ and $.19$, respectively). The fact that all other single study ρ estimates were within the 90% credibility value suggests that there are not large differences in the magnitude of effect sizes (study ρ estimates) reported across meta-analyses. This finding demonstrates there is more similarity in results across a number of meta-analyses of the same domain, the FFM personality dimensions and job performance, than has heretofore been recognized. Next, we examine the results across specific criteria and occupational groups. Examination of these findings will inform the debate about the intricate relationship that exists between specific FFM personality traits and job performance.

Second-Order Meta-Analytic Results for Each Big Five Dimension

Extraversion. Whether based on independent or non-independent samples, the relationship between extraversion and job performance for the work performance criterion was estimated to be .15 ($\bar{\rho}^{FFM}$). However, based on the lower bound 90% credibility value, this correlation could not be distinguished from zero. Based on previous findings, it was expected that higher scores on extraversion would be related to two specific criteria, higher training proficiency and teamwork. It was also expected that higher scores would predict successful work performance in two occupations, sales and managerial jobs. As

shown in Table 1, the results provided only mixed support for these expectations. The teamwork-extraversion relationship was supported ($\bar{\rho}^{FFM} = .16$, $K = 48$, $N = 3,719$), as was the relationship between extraversion and training performance ($\bar{\rho}^{FFM} = .28$, $K = 21$, $N = 3,484$). Extraversion also showed non-zero relationships with managerial performance ($\bar{\rho}^{FFM} = .21$, $K = 67$, $N = 12,602$) and police officer performance ($\bar{\rho}^{FFM} = .12$, $K = 20$, $N = 2,074$), but not for sales. Finally, it should be noted that the 90% credibility values across criteria and occupational groups provide some indication about the magnitude of likely moderator variables. In almost all cases, true score correlations may be expected to range up to the low to mid .30's, based on the upper bound of the 90% credibility values. Such validities are considerably larger than the average true score correlation estimates in the low to mid .10's.

Emotional stability. Table 2 summarizes the findings for emotional stability. As expected, emotional stability was found to be a valid predictor of work performance across jobs, with independent analyses ($\bar{\rho}^{FFM} = .13$, $K = 224$, $N = 38,817$) reporting very similar results to non-independent analyses ($\bar{\rho}^{FFM} = .15$, $K = 453$, $N = 73,047$). More importantly, the 90% credibility values for both analyses were found to exclude zero (for non-independent analyses, 90% C.I. for $\bar{\rho}^{sw} = .02 < .14 < .26$). In addition, as expected, emotional stability was a valid predictor of teamwork ($\bar{\rho}^{FFM} = .22$, $K = 41$, $N = 3,558$). Considering the specific occupational breakdowns, emotional stability was related to performance in some occupations (police, skilled or semi-skilled), but not others. The 90% credibility values reported across criteria and occupational groups also indicated that any potential moderator would be fairly modest in magnitude, as the upper bound of the 90% credibility values rarely exceeded a true score correlation in the mid .20's.

Agreeableness. Table 3 reports the second-order meta-analytic results for agreeableness. As expected, agreeableness displayed a weak relationship with the work performance criterion that was indistinguishable from zero. While agreeableness was found to predict teamwork ($\bar{\rho}^{FFM} = .34$, $K = 17$, $N = 1,820$), the number of studies and total sample size of this analysis were not large. Furthermore, as expected, agreeableness was not strongly related to any other criterion or occupational group. More importantly, the magnitude of a potential moderator variable rarely exceeded the mid .20's, based on an examination of the upper bounds of the 90% credibility values across criteria and occupational groups.

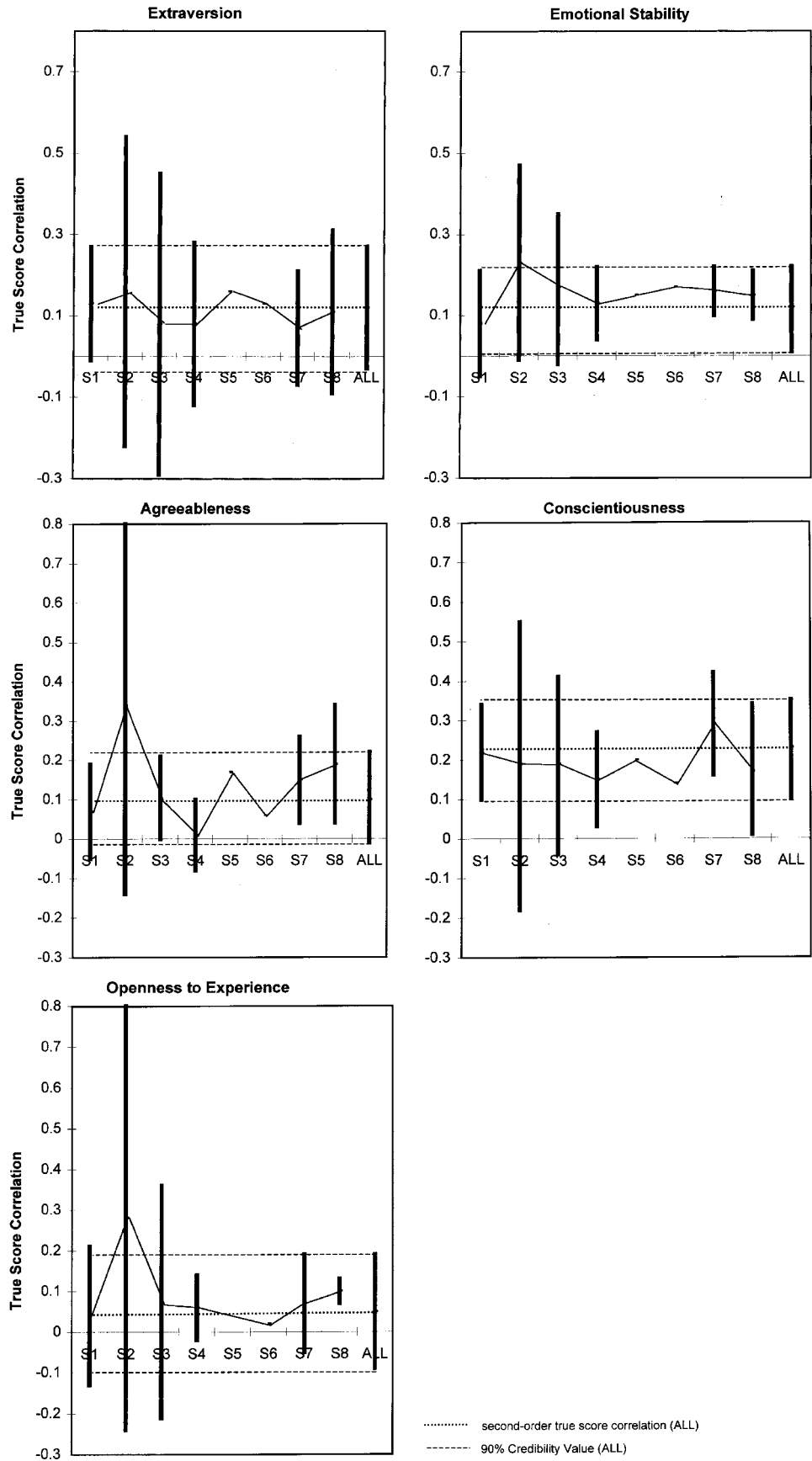


Figure 1. Meta-Analytic and Second-Order Meta-Analytic Estimates with Work Performance.

Conscientiousness. Table 4 demonstrates that conscientiousness is a valid predictor of performance across all criterion types and occupational groups. In all of these analyses, the 90% credibility value excluded zero. Thus, as expected, conscientiousness is related to work performance across all jobs for both independent analyses ($\bar{\rho}^{\text{FFM}} = .27$, $K = 239$, $N = 48,100$) and non-independent analyses ($\bar{\rho}^{\text{FFM}} = .24$, $K = 442$, $N = 79,578$), as well as for teamwork ($\bar{\rho}^{\text{FFM}} = .27$, $K = 38$, $N = 3,064$) and training performance ($\bar{\rho}^{\text{FFM}} = .27$, $K = 20$, $N = 3,909$). Furthermore, the magnitude of the true score correlations for this trait were consistently the highest among the Big Five, with the average true score correlation estimates ranging from the mid .20s to low .30s. Examination of the 90% credibility values demonstrated that the upper bound of these validity estimates generally was in the upper .30s. Thus, conscientiousness appears to consistently predict success in virtually all jobs moderately well, and may predict success strongly when moderator effects are accounted for in certain situations.

Openness to experience. Table 5 reports the second-order meta-analytic results for the final Big Five dimension, openness to experience. First, as expected, openness to experience was not relevant to many work criteria. In fact, along with agreeableness, this dimension consistently reported the lowest average true score correlations across criteria and occupational groups. More importantly, the magnitude of the upper bound of the 90% credibility values suggests the search for potential moderators would not be very useful, as the upper bound on true score correlations across criteria and occupational groups typically was in the low .20s. One exception to this finding, as expected, was reported for training proficiency, where the independent analyses suggest moderate effects ($\bar{\rho}^{\text{FFM}} = .33$, $K = 18$, $N = 3,177$).

Discussion

Understanding the relationship of personality to job performance is a fundamental concern of industrial-organizational psychologists throughout the world. The purpose of this study was to take stock of what we have learned after a century of personality research by summarizing prior meta-analytic studies, and to suggest new directions for future research. We did not expect to discover dramatically new relationships between personality traits and performance in different occupations. Rather, our goal was to resolve discrepancies in previous meta-analyses (to the extent that they existed) in order to

clearly understand what we now know about personality-performance relationships.

What Have We Learned about the FFM and Job Performance?

As discussed earlier, the first phase of research in this area covered the period prior to the mid-1980s, and resulted in rather pessimistic conclusions regarding the validity of personality in predicting job performance. Results from the present study suggest that conclusions in the second phase covering the period from the mid-1980s to the present are more optimistic, largely due to the use of the FFM taxonomy and meta-analytic methods to cumulate results across studies after correcting for study artifacts.

At the most general level, our findings show that independently conducted meta-analytic studies are relatively consistent – perhaps more so than has been previously recognized. The apparent exception to this is the Tett *et al.* (1991) study in which the true score validities for agreeableness, emotional stability and openness to experience exceeded the 90% credibility value. Although we do not know why their results are so different, one plausible explanation is that they confined their studies only to those that included confirmatory analyses. That is, they limited their studies to only those for which the study authors formulated hypotheses or used a job analysis to choose personality measures. Another possible reason is the sample sizes in their Big Five analyses were quite small. For example, they were smaller by a factor of as much as 50 compared to the Barrick and Mount (1991) study. Consequently, the true score estimates may not be very robust. Nonetheless, considering the results overall, it would be expected that independent meta-analytic studies of the same topic would yield similar results and, for the most part, that is what we found.

Results at the FFM level differed for each trait. Beginning with conscientiousness, our results show that its validity generalizes across all criterion types and all occupations studied. Further, its validity is the highest, overall. These findings are very consistent with those reported by Barrick and Mount (1991), but the present study strengthens those findings by summarizing results across multiple, independent meta-analyses. The results for conscientiousness underscore its importance as a fundamental individual difference variable that has numerous implications for work outcomes. Conscientiousness appears to be the trait-oriented motivation variable that industrial-organizational psychologists have long searched for, and it

should occupy a central role in theories seeking to explain job performance.

Results for the remaining personality dimensions show that each predicts at least some criteria for some jobs. In the present study, emotional stability was the only FFM trait other than conscientiousness to show non-zero true score correlations with the overall work performance criterion. These findings make intuitive sense, as it would be expected that individuals who are not temperamental, not stress-prone, not anxious and not worrisome (emotional stability), and those who are hard-working, persistent, organized, efficient and achievement-oriented (conscientious), are most likely to perform well.

While the validity of emotional stability in predicting job performance appears to be distinguishable from zero, the overall relationship is smaller than the effect for conscientiousness. Thus, one might wonder why the results for emotional stability were not stronger. Barrick and Mount's (1991) meta-analysis is the largest and thus most influential meta-analysis included in our second order meta-analysis, yet it also reported the lowest validity for emotional stability. We are not sure why this is the case. However, if the Barrick and Mount (1991) meta-analytic results are replaced with results from Hough (1992) in the independent second-order meta-analysis, the true score correlation between emotional stability and work performance increases to .17 from .13. Although this is somewhat larger than that reported in this study, it is still a smaller correlation than that found for conscientiousness.

Another possible factor explaining the relatively low validity for emotional stability is the measurement and conceptualization of the construct. Research by Judge *et al.* has shown that emotional stability may be a considerably broader construct than previously considered, and that many measures of neuroticism are too narrow to capture the true breadth of the construct. In fact, research indicates that emotional stability might include self-esteem, locus of control, and other traits previously studied in isolation (Judge, Locke, Durham and Kluger 1998). When these traits are considered as part of a single nomological net, it appears that the validity of the construct in predicting job performance is considerably higher (Judge and Bono 1999). Thus, we encourage future researchers to investigate further the breadth of the emotional stability construct and whether expanding its breadth would result in higher levels of predictive validity.

The remaining three FFM traits, extraversion, agreeableness and openness to experience, predicted some aspects of performance in some occupations. However, none of them predicted

consistently across criterion types. For example, extraversion and openness to experience predicted training performance especially well. The upper bound credibility value for the two dimensions was .41, suggesting that identifying moderators may lead to higher validities. Additionally, emotional stability and agreeableness (in addition to conscientiousness) predicted teamwork moderately well. However, the small $SD\rho^{sw}$ estimates indicate that searching for moderators for these two dimensions is unlikely to be fruitful when predicting teamwork.

Results for the occupational groups revealed several meaningful relationships, in addition to those previously discussed for conscientiousness and emotional stability. Extraversion was found to be useful in predicting performance in managerial and police occupations, although the relationship is small for police. Extraversion may be also be a good predictor of performance in sales jobs, but the large $SD\rho^{sw}$ estimate indicates the presence of moderators that influence the relationship. Results for agreeableness and openness to experience indicate that they are not good predictors of performance when results are summarized at the occupational level. However, it may be beneficial to search for moderators of the relationship between agreeableness and performance in sales and skilled or semi-skilled jobs. Similarly, it may be beneficial to search for moderators of the relationship between openness to experience and performance in sales and managerial positions. However, in each of these cases the ceiling of the validities appears to be about .30.

Because this special issue of the journal focuses on global and international perspectives, it is fitting to comment specifically on the meta-analytic results obtained by Salgado (1997). Most meta-analytic studies have been conducted using samples from the United States and Canada; Salgado's study examined personality-performance relations using samples obtained in the European Community (EC). This comparison is important because it answers the question as to whether the relations observed generalize across cultural boundaries. His meta-analysis was conducted using 36 studies, none of which overlapped with previous studies included in previous meta-analyses. The results showed that both the validity of conscientiousness and emotional stability generalized across all occupations and criterion types studied. The validity of the other FFM dimensions differed by occupation or criterion type. These results are nearly identical to those we obtained in the present study for FFM dimensions for the criterion of overall work performance. (It should be noted, however, that the Salgado study is contained within the meta-analytic results reported in the present study

and therefore the comparison is somewhat biased.) Overall, however, it appears that conscientiousness and emotional stability are valid predictors of performance in the European Community as well as in the United States and Canada.

In summary, a great deal of progress has been made in the last 50 years in understanding personality and performance linkages. The use of the FFM framework coupled with meta-analysis has been especially instrumental in enabling the field to move forward. However, while much has been learned, there is still much more to be discovered. When the results of the present study, which summarizes results across independently conducted meta-analyses, are compared to the meta-analytic results of Barrick and Mount (1991), it is apparent that most of the conclusions are the same. The biggest difference is the finding in the present study that emotional stability was a non-zero predictor of overall work performance. These findings strengthen the robustness of the conclusions reached by Barrick and Mount (1991). On the one hand, the results of the present study are cause for optimism because they reveal that the validities for at least two FFM dimensions generalize for the criterion of overall work performance. At the same time they show that the other FFM dimensions are valid predictors for at least some jobs and some criteria. On the other hand, the results are somewhat disappointing, particularly with respect to the magnitude of the validities observed. It would be expected that once researchers began using the FFM typology to classify personality traits and once instruments were designed and used to measure the FFM, we would begin to see higher validities. However, despite the adoption of the FFM in the past decade by many researchers and practitioners, the magnitude of the validities of the individual FFM dimensions is modest (i.e., generally less than .30), even in the best of cases. In light of this, we believe that it is now time to embark on a new research agenda. Toward this end, we would like to call for a moratorium on meta-analytic research of the type summarized in this paper. Because the present study subsumes the results of nearly all of the previous research in the area, the incremental validity of new studies over the present one is likely to be small.

Future Research Directions

Below we highlight three areas where additional research could make important contributions to understanding personality-performance linkages. We also briefly summarize recent work surrounding the development of a global personality measure. The areas included on the list are not meant to be exhaustive, nor are the

ideas necessarily novel; rather, the list is intended to serve as examples of the types of studies that are needed to move the field forward. First, further research is needed to explore levels of analysis in personality research. This suggests that future researchers will need to use hierarchically organized taxonomies that comprehensively capture the basic lower-level categories of personality and performance, as well as the superordinate constructs. Second, additional research is needed that investigates process models of personality that seek to explain how personality affects job performance. Third, research is needed that continues to examine critical issues pertaining to the measurement of personality measures in construct valid ways.

Linking lower level predictors and criteria. Research by Hough, Ones and Viswesvaran (1998) has shown the benefits of linking specific, lower-level facets of FFM constructs to specific, lower level criteria. Their large meta-analysis of personality-performance relationships for management jobs showed that lower-level facets of conscientiousness related differently to different lower-level criteria of success for management jobs. Specifically, the facet of achievement orientation predicted lower-level criteria such as number of promotions and salary, whereas the facet of dependability did not. Thus had the broad FFM construct of conscientiousness been used along with a broad criterion measure of overall management performance, meaningful relationships may have been masked.

The aforementioned example illustrates that linking specific FFM facets and specific criterion measures can result in increased correlations and enhance understanding. However, in order to link lower level personality and criterion constructs, it is necessary to have a taxonomy of lower level personality and criterion measures. One might ask how we can expect to develop an accepted taxonomy of lower level personality constructs when there is no universal agreement regarding the higher-order, FFM constructs. This is a legitimate question for which we do not profess to have the answer. Nonetheless, we must start somewhere and the following paragraphs are intended to be a step in that direction.

Four possible frameworks based on construct valid measures of the FFM indicate the lower level personality constructs consist of 12 facets (the PCI by Mount, Barrick, Laffitte and Callans 1999), 30 facets (the NEO-PI by Costa and McCrae 1992), 44 facets (the HPI by Hogan and Hogan 1992) or 45 facets (the AB5C instrument developed by Goldberg and colleagues (Hofstee, de Raad and Goldberg 1992). The Global Personality Inventory (GPI, Schmit, Kihm and

Robie, 2000), which is described in more detail below contains 32 facets. Other recent studies have systematically examined the structure of traits subordinate to the five factor model (Ackerman and Heggstad 1997; Paunonen 1998; Saucier and Ostendorf 1999). For illustrative purposes, the findings from the Saucier and Ostendorf (1999) study are reported. In this study, they found that the five factor model can be characterized as consisting of 18 first-order traits in both an American and German data set. To summarize these sub-dimensions, extraversion was found to consist of four traits labeled sociability, unrestraint, assertiveness and activity-adventurousness; agreeableness was divided into warmth-affection, gentleness, generosity and modesty-humility; conscientiousness was composed of orderliness, decisiveness-consistency, reliability and industriousness; emotional stability could be broken into traits such as (low) irritability, insecurity, emotionality; while openness to experience was reflected by intellect, imagination-creativity, perceptiveness. Although we do not suggest this is the final lower-level personality structure, this taxonomy could serve as a useful starting point of what that structure might be. Thus, we recommend that future studies investigate the predictive validity of personality at both the superordinate level (five factor model) as well as at the subordinate level (e.g., Saucier and Ostendorf's 18-dimensional model).

Many of the same problems that hindered research in personality research before the adoption of the FFM taxonomy also appear to plague the criterion side of the equation. That is, different names have been applied to substantively similar criterion measures and substantively different criterion measures have been grouped under the same general heading. To the extent that this happens, meaningful relations between personality constructs and criterion measures are obscured.

It is not within the scope of this article to fully describe or to justify theoretically and empirically a taxonomy of criterion measures. Rather, it is our intention to discuss the potential benefits of such a framework, with the hope that this will stimulate future research. Several taxonomies provide a useful starting point (e.g. Blum and Naylor 1968; Campbell, McCloy, Oppler and Sager 1993; Hough 1992) and the reader is directed to each of these for more details. For illustrative purposes we focus on the one proposed by Viswesvaran (1993) and used subsequently by Viswesvaran, Ones and Schmidt (1996). Viswesvaran (1993) identified ten job performance dimensions that comprehensively represented the entire job performance domain. The ten performance dimensions were: overall performance, job per-

formance or productivity, quality, leadership, communication competence, administrative competence, effort, interpersonal competence, job knowledge and compliance with or acceptance of authority. These ten dimensions provide a useful framework for researchers to use when categorizing subjective ratings. The taxonomy could be expanded to include objective measures. Four major components could be: productivity data (e.g., sales, outcomes), advancement criteria (e.g., promotions, salaries), withdrawal behaviors (e.g., turnover, absenteeism), and counterproductive behaviors (e.g., theft, drug abuse).

We propose that these predictor and criterion frameworks (or something like them) could be adopted and used by researchers to formulate hypotheses and cumulate results. Linking predictors and criteria at a more specific level than that described in this article, could increase validities and enhance understanding as well. For example, our meta-analytic results presented earlier showed that while validities for emotional stability were non-zero in some cases, they were generally smaller than would be expected from a common sense perspective. Such conclusions might be more positive, however, if lower level facets of emotional stability are linked with relevant, lower-level performance constructs. For example, if the lower level facet of calmness (a facet of emotional stability from the AB5C instrument) is linked with the criterion of stress proneness (a component of interpersonal competence), it is likely that better prediction will occur. Similarly, our results revealed that agreeableness has near zero correlations with most criteria. Again, this could be because the analyses have not been conducted at the appropriate level of specificity. If the facet of cooperation (a facet of agreeableness from the PCI) is linked with the criterion of cooperation (a component of interpersonal competence) as rated by bosses, peers or customers, better prediction is likely. Other examples could easily be provided, but this is the general idea. Clearly, the taxonomies we present here are largely conceptual and need additional theoretical and empirical evidence. However, our purpose was to illustrate that when lower level personality and lower level criterion constructs are appropriately linked stronger correlations might occur.

Developing process models of personality-performance relations. Although the preponderance of evidence demonstrates that specific personality constructs are important determinants of work performance, very little is known about the mechanisms through which these distal traits affect job performance. The proximal means by which personality affects performance has long been thought to be

primarily through a person's motivation (e.g., Barrick, Mount and Strauss 1993; Kanfer 1991; Mount and Barrick 1995; Murray 1938). Again, research is hindered because an accepted framework does not exist for studying motivational constructs.

Nevertheless, three fundamental motivational constructs consistently emerge as particularly relevant mediators for personality effects in work settings. The first two are broad motivational intentions that underlie goals related to social behavior – striving for status and striving for communion. Hogan labels these two dimensions 'getting ahead' and 'getting along' (Hogan and Shelton 1998). Individuals are assumed to be motivated to seek power, achievement and status in organizational hierarchies (getting ahead or status striving) as well as striving for acceptance and intimacy in personal relationships (getting along or communion striving). The centrality of striving for status and communion as basic human motivations has been established from a variety of perspectives. For example, Wiggins and Trappnell (1996) identify these two dimensions building on concepts from evolutionary biology, anthropology and sociology. Similarly, using socio-analytic theory, Hogan and Shelton (1998) conclude these two strivings underlie interpersonal motivation.

A third motivational construct emphasizes motivation at work that is nonsocial in nature. This can be characterized as task orientation and derives from a desire for personal excellence independent of others. Thus, a third motivational force is labeled 'accomplishment striving' (getting things done) and reflects an individual's intentions to accomplish tasks. A concern for completing the task is a fundamental motivational force in research in group psychology, as Bales (1950) for example, emphasizes task inputs along with socio-emotional inputs. Similarly, leadership theory (Fiedler 1967; Fleishman 1953; Katz, Macoby and Morse 1950) frequently emphasizes concern for the task or initiating structure as a fundamental dimension of leadership. Recently, Kanfer and Heggstad (1997) have identified motivational control as a basic dimension of human motivation. We propose that these motivational constructs – communion striving, status striving and accomplishment striving – could be adopted by researchers as important mediators of personality-performance relationships. The inclusion of both proximal and distal motivation constructs into a unified motivational model will significantly advance our understanding of antecedents to job performance.

Issues regarding the measurement of personality. Another factor that may both provide an

explanation for the relatively modest validities and suggest an important avenue for future research is the measurement of the Big Five traits. Concerns with the use of self-reports of personality have been noted for as long as such measures have been around. Murray (1938), for example, was quite skeptical of the ability of individuals to provide accurate self-assessments of their personality, as revealed by his conclusion, 'Children perceive inaccurately, are very little conscious of their inner states and retain fallacious recollections of occurrences. Many adults are hardly better' (p. 15). Due to this skepticism with self-reports, Murray and his students such as McClelland advocated use of projective measures, which has generated controversies of its own (Spangler 1992). Beyond projective tests, what is the alternative to self-reports of personality?

As Johnson (1997) commented, external observers would be expected to provide more valid assessments of individual's phenotypic (i.e., externally observed or behavioral) traits than genotypic (i.e., emotional or cognitive) traits. Because the Big Five traits are relatively behavioral in nature, and since job performance is an externally observed behavior, it makes sense that measures of the Big Five supplied by external observers would correlate more strongly with job performance than self-ratings. Hogan (e.g., Hogan, Hogan and Roberts 1996) makes a complementary point in arguing that the best way of conceptualizing personality structure is by one's reputation, and the best way to measure reputation is through the ratings of knowledgeable others. As Hogan *et al.* (1996, p.469) note, 'Moreover, because reputation is built on a person's past behavior, and because past behavior is the best predictor of future behavior, this aspect of personality has important practical use.'

Indeed, there is empirical evidence to support this argument. Mount, Barrick and Strauss (1994) found that the validity of external (supervisor, co-worker and customer) ratings of personality was at least as valid as self-ratings, and those external ratings explained incremental variance in job performance beyond that accounted for by self-ratings. Mount *et al.* (1994) concluded that self-ratings may underestimate the validity of the Big Five traits. More recently, Judge, Higgins, Thoresen and Barrick (1999) showed that childhood ratings of personality supplied by trained observers achieved impressive validities in predicting occupational success and adjustment up to 50 years later. Thus, one area for future research is increased use of alternatives to self-ratings of personality. Finally, in industrial psychology, self-reports of personality are criticized due to the possibility of 'faking' (i.e.,

applicants providing socially desirable answers in order to increase their chances of being hired). Though the issue of faking is well beyond the scope of this article, and is currently being hotly debated (Ellingson, Sackett and Hough 1999; Ones, Viswesvaran and Reiss 1996; Rosse, Stecher, Miller and Levin 1998), it does provide another reason to explore the potential validity of other measurement strategies. It is ironic that some of the earliest studies of the five-factor structure were uncovered in observer, not self, ratings (e.g., Tupes and Christal 1961), yet self-ratings are the dominant means of measuring the Big Five traits. We feel the time to explore alternatives in earnest has come.

One other issue that is fundamental to this discussion pertains to the reliability of personality measures. It is possible that the reliability of measures of the FFM dimensions (and other constructs in the behavior sciences) are not as reliable as they are generally believed to be. Obviously, if reliability is low, then observed correlations between the FFM constructs and other constructs are attenuated. Schmidt and Hunter (1996 1999) have argued that only when reliability is conceptualized as a Generalizability Coefficient under Generalizability Theory (Cronbach, Gleser, Nanda and Rajaratnam 1972) and is then used in the disattenuation formula, can the true relationships among constructs be observed. They point out that the most common practice of using alpha coefficient in the disattenuation formula removes only part of the downward bias, which leads to an underestimation of the true relationship between constructs. In particular, Schmidt and Hunter (1999) believe that researchers have failed to account for transient error, which results from moods, feelings and mental states on a specific occasion, as an important source of variation in measures of constructs. Transient error is addressed through test-retest reliability and captures errors that are associated with responses that are unique to a person on a particular occasion (something coefficient alpha does not capture). According to Schmidt and Hunter (1999), the appropriate reliability is the Coefficient of Stability and Equivalence (CSE; Anastasi 1988), which accounts for transient and other sources of error, by correlating parallel forms of a scale measured on two occasions. When the CSE reliability coefficient is used in the disattenuation formula, substantially higher true-score correlations are likely to be obtained for FFM constructs.

A global measure of personality. In keeping with the focus of this special issue on global and international perspectives, one recent, ongoing research project worth highlighting is the

development of a cross-cultural measure of personality called the Global Personality Inventory (GPI; Schmit *et al.* 2000). The instrument is significant because it has the potential to further our understanding of personality-performance linkages in different countries. Typically, personality inventories have been developed using a strategy whereby an inventory developed in one country is transported to another. As pointed out by Schmit *et al.* (2000), one difficulty with this approach is that even though personality may not be different across cultures, the expression of personality is highly likely to differ. Further, exported instruments are often changed substantively when they are transported to different countries, making it difficult to compare results across cultures. To overcome this, the GPI was developed using both an emic approach whereby a culture is observed from within, and an etic approach whereby many cultures are observed from outside the culture.

The GPI was developed using ten international teams consisting of 70 members, most of whom were PhD or Master's level psychologists from the USA, the UK, France, Belgium, Sweden, Germany, Spain, Japan, Singapore, Korea, Argentina and Columbia. The five-factor model was used as the organizing structure, primarily because of the large research base surrounding it and also because evidence suggests that the five-factor model is invariant across cultures (McCrae and Costa 1997). A complex process was used to translate items from English to other languages. Sophisticated item psychometric methods such as response theory analyses, differential item functioning analyses and factor analyses were used to develop the final instrument. All told, the initial evidence of the construct validity of the GPI across cultures is impressive, although development of the instrument is on-going.

Schmit *et al.* (2000) report the results of a criterion-related, concurrent validation study using the GPI for three samples from middle management jobs in the USA. Results for the combined samples (total N = 149) showed that conscientiousness and extraversion as measured by the GPI were significantly related to overall performance ($r = .21$ and $.19$, uncorrected). These results are very similar to the results obtained in the present study for the management occupation, which revealed that both conscientiousness and extraversion were non-zero predictors of overall performance for these two occupations. Openness to experience was also related to overall performance ($r = .16$, uncorrected) in the Schmit *et al.* study, but was not in our second order meta-analysis. Overall, the GPI represents an important step forward in the measurement of FFM dimensions. Although there is only limited empirical support available

at present, the GPI has the potential to further our understanding of personality-performance linkages across cultures.

In conclusion, this study summarizes what we know about the relationships between the FFM traits and available criterion measures based on previous meta-analytic studies. While much has been learned from this body of research, we believe that little is to be gained from further meta-analytic studies of this type. Consequently, we call for a moratorium on such studies, and suggest that researchers embark on a new era of research along (but not limited to) the areas we outline above.

Note

1. The Tett, Jackson and Rothstein (1991) moderator meta-analyses have been shown to be mathematically incorrect (Ones, Mount, Barrick and Hunter, 1994). The cumulative impact of the four technical errors made in these analyses raises serious questions about the interpretation of their results for various moderators of the personality – job performance relationship. However, it is important to note that the meta-analyses pertaining to the Big Five personality dimensions did NOT use absolute correlations. Consequently, none of the technical errors distorted the meta-analytically derived Big Five estimates. For this reason, these estimates were included in this article.

References

- Ackerman, P. and Heggstad, E.D. (1997) Intelligence, Personality, and Interests: Evidence for overlapping traits. *Psychological Bulletin*, **121**, 219–245.
- Anastasi, A. (1988) *Psychological Testing*. (6th edn). New York: Macmillan Publishing Company.
- Anderson, G. and Viswesvaran, C. (1998) An update of the validity of personality scales in personnel selection: A meta-analysis of studies published after 1992. Paper presented at the 13th Annual Conference of the Society of Industrial and Organizational Psychology, Dallas.
- Barrick, M.R. and Mount, M.K. (1991) The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, **44**, 1–26.
- Barrick, M.R., Stewart, G.L., Neubert, M.J. and Mount, M.K. (1998) Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, **83**, 377–391.
- Block, J. (1995) A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, **117**, 187–215.
- Blum, M.L. (1949) *Industrial Psychology and its Social Foundations*. New York: Harper and Brothers.
- Blum, M. and Naylor, J.C. (1968) *Industrial Psychology*. New York: Harper and Row.
- Bouchard, T.J., Jr. (1997) The genetics of personality. In K. Blum and E.P. Noble (eds), *Handbook of Psychiatric Genetics*. (pp. 273–296). Boca Raton, FL: Circle Press, Inc.
- Campbell, J.P., McCloy, R.A., Oppler, S.H. and Sager, C.E. (1993) A theory of job performance. In N. Schmitt and W.C. Borman and Associates (eds), *Personnel Selection in Organizations*. San Francisco: Jossey-Bass Publishers.
- Conley, J.J. (1984) Longitudinal consistency of adult personality: Self-reported psychological characteristics across 45 years. *Journal of Personality and Social Psychology*, **47**, 1325–1333.
- Costa, P.T., Jr. and McCrae, R.R. (1988) Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, **54**, 853–863.
- Costa, P.T., Jr. and McCrae, R.R. (1992) *Revised NEO Personality Inventory and Five-factor Model Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Cronbach, L.J., Gleser, G.C., Nanda, H. and Rajaratnam, N. (1972) *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. John Wiley and Sons.
- Digman, J.M. and Shmelyov, A.G. (1996) The structure of temperament and personality in Russian children. *Journal of Personality and Social Psychology*, **71**, 341–351.
- Ellingson, J.E., Sackett, P.R. and Hough, L.M. (1999) Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, **84**, 155–166.
- Eysenck, H.J. (1992) Four ways five factors are not basic. *Personality and Individual Differences*, **13**, 667–673.
- Fiedler, F.E. (1967) *A Theory of Leadership Effectiveness*. New York: McGraw-Hill.
- Fleishman, E.A. (1953) The measurement of leadership attitudes in industry. *Journal of Applied Psychology*, **37**, 153–158.
- Goldberg, L.R. (1993) The structure of phenotypic personality traits. *American Psychologist*, **48**, 26–34.
- Guion, R.M. (1998) *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Erlbaum.
- Guion, R.M. and Gottier, R.F. (1965) Validity of personality measures in personnel selection. *Personnel Psychology*, **18**, 135–164.
- Hofstee, W.K.B., de Rand, B. and Goldberg, L.R. (1992) Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, **63**, 146–163.
- Hogan, R.T. and Hogan, J. (1992) *Hogan Personality Inventory Manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., Hogan, J. and Roberts, B.W. (1996) Personality measurement and employment decisions. *American Psychologist*, **51**, 469–477.
- Hogan, R. and Shelton, D. (1998) A socioanalytic perspective on job performance. *Human Performance*, **11**, 129–144.
- Hough, L.M. (1992) The 'Big Five' personality variables – construct confusion: Description versus prediction. *Human Performance*, **5**, 139–155.

- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. and McCloy, R.A. (1990) Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology* (Monograph), **75**, 581–595.
- Hough, L.M., Ones, D.S. and Viswesvaran, C. (1998) Personality correlates of managerial performance constructs. Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas.
- Hunter, J.E. and Schmidt, F.L. (1990) *Methods of Meta-analysis*. Newbury Park, CA: Sage Publications.
- Hurtz, G.M. and Donovan, J.J. (1998) Personality and job performance: Will the real 'Big Five' please stand up? Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas.
- Johnson, J.A. (1997) Units of analysis for the description and explanation of personality. In R. Hogan, J. Johnson and S. Briggs (eds.), *Handbook of Personality Psychology* (pp. 73–93). San Diego: Academic Press.
- Judge, T.A. and Bono, J. (1999) Core self-evaluations and construct breadth: Effects on job satisfaction and job performance. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Judge, T.A., Higgins, C., Thoresen, C.J. and Barrick, M.R. (1999) The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, **52**, 621–652.
- Judge, T.A., Locke, E.A., Durham, C.C. and Kluger, A.N. (1998) Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology*, **83**, 17–34.
- Kanfer, R. (1991) Motivation theory and industrial and organizational psychology. In M.D. Dunnette and L.M. Hough (eds.), *Handbook of Industrial and Organizational Psychology* (pp. 75–178, 2nd edn). Palo Alto, CA: Consulting Psychologists Press.
- Kanfer, R. and Heggstad, E.D. (1997) Motivational traits and skills: A person-centered approach to work motivation. *Research in Organizational Behavior*, **19**, 1–56.
- Katz, D., Macoby, N. and Morse, N.C. (1950) *Productivity, Supervision, and Morale in an Office Situation*. Ann Arbor, MI: University of Michigan Survey Research Center.
- Mount, M.K. and Barrick, M.R. (1995) The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management*, **13**, 153–200.
- Mount, M.K., Barrick, M.R., Laffitte, L.J. and Callans, M.C. (1999) *Personal Characteristics Inventory User's Manual*. Libertyville, IL: Wonderlic Consulting.
- Mount, M.K., Barrick, M.R. and Stewart, G.L. (1998) Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, **11**, 145–165.
- Mount, M.K., Barrick, M.R. and Strauss, J.P. (1994) Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, **79**, 272–280.
- Murphy, K.R. (1996) Individual differences and behavior in organizations: Much more than g. In K.R. Murphy (ed.), *Individual Differences and Behavior in Organizations* (pp. 3–30). San Francisco: Jossey-Bass.
- Murray, H.A. (1938) *Explorations in Personality*. New York: Oxford University Press.
- Ones, D.S. (1993) The construct validity of integrity tests. Unpublished doctoral dissertation, University of Iowa.
- Ones, D.S., Mount, M.K., Barrick, M.R. and Hunter, J.E. (1994) Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, **47**, 147–156.
- Ones, D.S., Viswesvaran, C. and Reiss, A.D. (1996) Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, **81**, 660–679.
- Organ, D.W. and Ryan, K. (1995) A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, **48**, 775–802.
- Paunonen, S.V. (1998) Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, **74**, 538–556.
- Robertson, I.T. and Kinder, A. (1993) Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology*, **66**, 226–244.
- Rosse, J.G., Stecher, M.D., Miller, J. and Levin, R.A. (1998) The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, **83**, 634–644.
- Salgado, J.F. (1997) The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, **82**, 30–43.
- Salgado, J.F. (1998) Criterion validity of personality measures based and non-based on the five factor model. Paper presented at the 106th Annual Convention of the American Psychological Association, San Francisco.
- Saucier, G. and Ostendorf, F. (1999) Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, **76**, 613–627.
- Schmidt, F.L. and Hunter, J.E. (1996) Measurement error in psychological research: Lessons learned from 26 research scenarios. *Psychological Methods*, **1**, 199–223.
- Schmidt, F.L. and Hunter, J.E. (1999) Theory testing and measurement error. *Intelligence*.
- Schmit, M.J., Kihm, J.A. and Robie (2000) The development of a global measure of personality. *Personnel Psychology*, **1**, 153–193.
- Spangler, W.D. (1992) Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, **112**, 140–154.
- Tett, R.P., Jackson, D. N. and Rothstein, M. (1991) Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, **44**, 703–742.
- Tupes, E.C. and Christal, R.E. (1961) *Recurrent Personality Factors Based on Trait Ratings* (ASD-TR-61-97). Lackland Air Force Base, TX: Aeronautical Systems Division, Personnel Laboratory.
- Viswesvaran, C. (1993) Modeling job performance: Is there a general factor? Unpublished doctoral dissertation, University of Iowa, Iowa City.

-
- Viswesvaran, C., Ones, D.S. and Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, **81**, 557–574.
- Weiss, H.M. and Adler, S. (1984). Personality and organizational behavior. In B.M. Staw and L.L. Cummings (eds.), *Research in Organizational Behavior* (Vol. 6, pp. 1–50). Greenwich, CT: JAI Press.
- Wiggins, J.S. and Trapnell, P.D. (1996) A dyadic-interactional perspective on the Five-Factor Model. In J.S. Wiggins (ed.), *The Five-Factor Model of Personality: Theoretical Perspectives* (pp. 88–162). New York: Guilford Press.

Appendix: Summary of prior meta-analyses of the big five model and job performance, across criteria, occupational groups, and levels of analysis

Prior Meta-analyses	Criteria						Occupational Groups				
	1	2	3	4	5	6	7	8	9	10	11
Anderson and Viswesvaran, 1998	DI(NI)	DI(NI)	NA	NA	NA	NA	NA	NA	NA	NA	NA
Barrick and Mount, 1991	DI(I)	DI(I)	DI(I)	DI(I)	NA	NA	DI(I)	DI(I)	DI(I)	DI(I)	DI(I)
Hough, 1992 ^a	DI(NI)	NA	NA	DI(NI)	DI(I)	NA	DI(NI)	NA	NA	NA	NA
Hough, Eaton, Dunnette, Kamp and McCloy, 1990 Meta-Analysis ^a Project A ^a	DA	NA	NA	DA	NA	NA	NA	NA	NA	NA	NA
Hough, Ones and Viswesvaran, 1998	DI(I)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Hurtz and Donovan, 1998	DA	NA	NA	DA	NA	NA	DI(I)	DI(I)	NA	NA	DI(I)
McHenry, Hough, Toquam, Hanson and Ashworth, 1990	DA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mount, Barrick and Stewart, 1998	NA	NA	NA	NA	DI(I)	NA	NA	NA	NA	NA	NA
Organ and Ryan, 1995	NA	NA	NA	NA	NA	DA	NA	NA	NA	NA	NA
Pulakos and Schmitt, 1996	DA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Robertson and Kinder, 1993	DA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Salgado, 1997	DI(I)	DI(I)	DI(I)	DI(I)	NA	NA	DI(I)	DI(I)	DI(I)	DI(I)	DI(I)
Salgado, 1998											
USA data (FFM only)	DI(I)	DI(I)	NA	NA	NA	NA	NA	NA	NA	NA	NA
USA data (Scale level)	DI(I)	DI(I)	NA	NA	NA	NA	NA	NA	NA	NA	NA
European data (FFM only)	DA	DA	NA	NA	NA	NA	NA	NA	NA	NA	NA
European data (Scale level)DA	DA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Tett, Jackson and Rothstein, 1991	DI(NI)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Vinchor, Schippman, Switzer and Roth, 1998 (Ratings and Sales)	NA	NA	NA	NA	NA	NA	DI(NI)	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	DI(NI)	NA	NA	NA	NA

Notes: Performance Criteria: 1 = Work Performance, 2 = Supervisor Ratings, 3 = Objective Performance, 4 = Training Proficiency, 5 = Teamwork, 6 = Organizational Citizenship Behaviors;

Occupational Groups: 7 = Sales, 8 = Managers, 9 = Professionals, 10 = Police, 11 = Skilled or Semi-skilled.

DI = Data Included in second-order meta-analysis; DI(I) = Independent Sample, DI(NI) = Non-Independent Sample. DA = Data available in prior meta-analysis, but excluded from the second-order meta-analysis. NA = Data not available in prior meta-analysis.

^aThese meta-analyses did not provide variance estimates (Obs SD or SD ρ) nor percent variance accounted for estimates.