Impact of Meta-Analysis Methods on Understanding Personality-Performance Relationships

Murray R. Barrick
Department of Management and Organizations
Tippie College of Business
University of Iowa

Michael K. Mount
Department of Management and Organizations
Tippie College of Business
University of Iowa

Running Head: Impact of Meta-Analyses

Murray and Mick—I have put comments and suggestions in boldface throughout
Impact of Meta-Analysis Methods on Understanding Personality-Performance Relationships

Understanding the relations among personality dimensions and job performance is a fundamental concern of industrial-organizational psychologists. In this chapter we discuss the impact that meta-analysis has had on furthering our knowledge in this area. We first discuss the results and conclusions from research conducted prior to 1990, which was based largely on narrative reviews of personality performance relations and did not utilize a taxonomy for classifying personality traits. We then examine the results and conclusions derived from the literature based on studies conducted between 1990 and 2000 that used both meta-analysis and the Five Factor Model (FFM) to classify personality traits. Next, we discuss the overall impact that meta-analysis has had on understanding personality performance relations. Finally, we provide suggestions for the role that meta-analysis can play in the future in furthering our understanding of personality-performance relations.

In order to understand the impact that meta-analysis has had on understanding personality – performance relations, it is useful to review the methods and conclusions of studies conducted prior to the meta-analytic studies, as well as the reasons for the conclusions. Literally hundreds of studies have investigated this topic over the past 25 years, which attests to its importance to the I-O field. One of the first major reviews of personality tests was conducted by Ellis and Conrad (1948). The purpose of their study was to examine the validity of personality inventories used in the military compared to those used in civilian settings. Although the validities were higher overall for inventories used in the military, this conclusion must be qualified by several factors. First, the review did not provide precise quantitative estimates of validity, making the comparisons tentative at best. Further, many of the favorable conclusions arrived at in the
military setting were based on predicting psychiatric criteria such as neuropsychiatric unfitness for military duty. Also, according to the authors there were spurious factors that may have accounted for the higher correlations in the military setting. Examples of these included criterion contamination, whereby those who had knowledge of the criterion also had some knowledge of predictor scores. And, in some cases there was criterion overlap, in which there was duplication of questions in the predictor and criterion measures.

A few years later, Ghiselli and Bartol (1953) conducted the first major review of the validity of personality measures for selection purposes. Their research examined only those studies for which the personality traits had relevance to the job in question. In general their findings were positive, but they argued for caution in the use of personality measures as selection instruments, largely because results varied widely across job category and by study. Locke and Hulin (1962) examined 18 studies that had used The Activity Vector Analysis, and reported generally low validities, and also pointed out that there were numerous methodological problems in some of the validation studies. Shortly after that Guion and Gottier (1965) published their influential study, which was a qualitative review of 12 years of research in the Journal of Applied Psychology and Personnel Psychology pertaining to personality – performance relations. They concluded that personality measures and job performance were not related in any meaningful way across traits and across situations.

Schmitt, Gooding, Noe, and Kirsch (1984) conducted the first meta-analytic review of personality – performance relations. They examined criterion-related validity studies published in Personnel Psychology and the Journal of Applied Psychology between 1964 and 1982. They did not examine the validity of specific personality constructs, nor did they attempt to analyze the data according to conditions of the studies (e.g. whether job analysis was used or whether
researchers tested specific hypotheses). They reported the average uncorrected correlation between personality measures and measures of job success (across criteria and occupations) was $r = .149$. They also report uncorrected mean correlations between personality and five different criteria: performance ratings, $r = .206$; turnover, $r = .121$; achievement/grades, $r = .152$; status change, $r = .126$; and wages, $r = .268$. Their conclusions were similar to those of previous researchers: “Results concerning different types of predictors are consistent with previous literature reviews (Ghiselli, 1973; Guion and Gottier, 1965) which conclude that personality tests have low validity” (p. 420). Interestingly, the average observed correlation of .149 reported by Schmitt et al. over all criteria is larger than that obtained by Barrick and Mount (1991) for conscientiousness ($r = .13$ across occupations). Yet, Schmitt et al.’s conclusion about the usefulness of personality measures for selection purposes are more pessimistic than Barrick and Mount’s. One reason for Barrick and Mount’s more optimistic view pertains to the purpose of their study and the methodology used, which focused on understanding personality-performance linkages. Thus, the finding that the validity of one personality dimension, conscientiousness, generalized across occupations was more important than the magnitude of the validities obtained.

Needless to say the pessimistic conclusions by Guion and Gottier (1965) and Schmitt et al. (1984) led to a decline in the use of personality measures for selection purposes. Subsequently, however, researchers began to explore several reasons for these pessimistic conclusions, and began to examine alternative ways to examine personality-performance relations. For example, there are thousands of personality traits that pertain to specific aspects of human behavior, yet prior to the late 1980’s, no accepted classification system was available. Prior reviews drew broad conclusions about the predictive validity of personality as if personality were a “unitary” construct. Thus, one explanation for these findings was that personality is not a
good predictor of job performance when dealt with as one concept. In contrast, we believed that understanding could only be achieved if we examined the relationship of specific personality traits with performance.

During the 1980s and 1990s researchers intensely searched for the dimensionality of personality in order to identify a parsimonious, yet comprehensive taxonomy of human behavior. Without the capability to reduce the number of personality traits into a smaller, more manageable number, the process of studying personality-performance linkages was unwieldy, and resulted in ambiguity about which personality traits were being measured. 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 the 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 higher-order construct. Furthermore, many researchers simply correlated all of the scales on the personality inventories with the criteria without regard to the relevance of the relations. This approach was problematic because in the absence of theory driven hypotheses, it is difficult to interpret the results meaningfully. Finally, the reviews of the literature at this time were largely narrative (with the exception of Schmitt et al., 1984), rather than quantitative meta-analytic reviews. Because study artifacts were not corrected across studies, the validity estimates were downwardly biased. Given these problems, it was difficult if not impossible to identify consistent relationships among personality traits and criteria. It is understandable that little advancement was made in understanding personality-performance relationships.
In the past decade significant progress has been made due largely to the convergence of two developments in the psychology literature. The first is the emergence of a widely accepted taxonomy, the Five-Factor Model of personality (FFM), which could be used to classify the thousands of personality traits into a manageable number of personality dimensions. Each personality dimension from the FFM pertains to a specific aspect of human behavior, one that is relatively independent of others. These five traits 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 framework to systematically examine the relationship between specific personality traits and job performance.

The second development was the formulation and refinement of meta-analytic methods (e.g., Hunter & Schmidt, 1990). As discussed above, numerous problems characterized personality research that hindered the ability to draw meaningful conclusions. Some reviews were largely qualitative in nature, whereas others computed mean validities for personality traits or sets of traits, but did not correct for statistical artifacts across studies such as sampling error, range restriction, and differences in reliabilities for predictors and criteria. The use of meta-analysis to investigate these relations has led to more positive conclusions. In fact, Guion (1998) recently noted that, “Meta-analyses have provided grounds for optimism” (p. 145).
We agree that meta-analysis has had a positive impact on our knowledge in this area. Nonetheless, we would like to point out that it was not meta-analysis alone that led to these advancements. For example, the Schmitt et al. (1984) study was a meta-analysis and was helpful in understanding the magnitude of the validity of personality measures. However, because they did not use a taxonomy like the FFM to categorize personality traits, their findings did not enhance understanding of how well specific, meaningful personality constructs predicted performance. Subsequently, when the literature was examined using the FFM taxonomy coupled with meta-analytic methods, the resulting conclusions were more positive than before. The results were useful not so much because they shed light on the magnitude of the validities, but because they increased understanding at the construct level. Prediction that does not enhance understanding is not theoretically meaningful.

Armed with an understanding about the predictive validity of different personality constructs from meta-analytic research, researchers can develop, test and refine theories of job performance. One example comes from the study by Barrick, Mount and Strauss (1993), who used structural equations modeling to test the theory that distal motivational processes associated with conscientiousness were related to proximal motivational variables such as goal setting behaviors, which in turn, were linked with higher performance. Their results confirmed the hypothesis that proximal goal-setting behaviors mediated the relations between distal personality traits and performance. Thus meta-analysis provided the initial, broad level of understanding about the predictive validity of conscientiousness, which was then used as the basis for developing and testing a more specific set of hypotheses about the way Conscientiousness relates to performance.
Thus we believe that the recent optimism about the usefulness of personality measures for personnel selection purposes is due equally to the use of the FFM to classify personality traits and to methods of meta-analysis to cumulate results across studies. It is difficult to imagine how understanding in this area could have advanced by using only one and not the other. In our opinion, this underscores the need for any meta-analytic study to use a well-defined taxonomy for classifying both predictors and criteria. Having said this, we should point out that very few meta-analyses have actually met this standard. That is, while most recent meta-analyses of personality-performance relations have used the FFM taxonomy, there has been little consensus regarding the classification of criterion measures. Further, considering the issue more broadly, most selection-oriented meta-analyses have focused on the validity of various methods of selection such as interviews, integrity tests, and ability tests, rather than on the validity of the constructs measured by the methods.

**Meta-analytic Studies of Personality-Performance Relations**

Since 1990 at least 16 meta-analytic studies investigated relations between personality traits using the Five-Factor Model (FFM) of personality (or some variant of it) and job performance. This statement is both striking and provocative in its implications. On one hand it is striking because of the sheer number of studies: 16 meta-analytic studies investigating the same topic in the short time span of a decade is remarkable. (Prior to 1990 only one meta-analytic study had investigated these relations (Schmitt et al., 1984) and it did not use the FFM taxonomy.) To our knowledge, there is no other topic in the I/O Psychology literature that has been the focus of so many meta-analyses in such a short period of time.

The above statement is also provocative in several ways. The purpose of meta-analysis is to cumulate the results across studies to establish facts (Hunter & Schmidt, 1990). Because a
meta-analysis necessarily summarizes results across large numbers of studies while correcting for study artifacts, it would be expected that the conclusions would be very robust. Hence, if a meta-analytic study were comprehensive and well executed, it would seem that additional meta-analyses on the same topic would be unnecessary. In light of this, it begs the question: Why have so many researchers conducted meta-analyses on essentially the same topic over the past decade? And, why have editors of major journals continued to publish them?

There are several possible explanations. One reason is that emergence of meta-analysis and the FFM occurred at about the same time, and were viewed as natural partners in the study of personality – performance linkages, which stimulated further research using both of them. A second reason is that relations between personality and performance are complex, which means that they can be difficult to discover and that they can be investigated in several ways. Further, meta-analysis involves formulating numerous decision rules about which studies to include, how to classify measures, how to code data, and so on. Although there are numerous judgment calls that must be made in any meta-analysis, this is especially true in personality research given the countless number of personality traits that exist and the ambiguity regarding their labels and definitions. Similarly, judgment calls also are made on the criterion side, where little consensus exists regarding a taxonomy of performance dimensions. Because these decision rules obviously have a bearing on the outcome of the meta-analysis, subsequent researchers may believe that different conclusions would be reached if different decision rules were used. Another reason that it may be fruitful to conduct a new meta-analysis on the same topic is that over time, new primary studies are conducted and when a sufficient number have become available, researchers recognize that re-analyses could alter the conclusions of the original meta-analysis.
But what have the additional meta-analyses of personality-performance relations shown? Have they resulted in different conclusions? Are additional meta-analyses needed? By coincidence there were three meta-analyses of personality and performance relations that appeared in published journals in the I-O Psychology field at about the same time (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Hough, 1992). Each of these studies examined personality-performance linkages using the FFM framework (or some variant of it) and covered much of the same literature; yet, conclusions regarding the validity of the personality constructs differed somewhat across these studies. 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 et al. (1991) found that only Emotional Stability displayed nonzero correlations with performance, and two other Big Five traits--Agreeableness and Openness to Experience--displayed higher correlations with performance than Conscientiousness. Goldberg (1993) described these differences in findings based on a similar body of knowledge as “befuddling” (p. 31).

The methodological differences between the Barrick and Mount (1991) and Tett et al. (1991) studies that might have accounted for the disparities in the FFM results have been discussed at length elsewhere (Ones, Mount, Barrick, & Hunter, 1994; Tett, Jackson, Rothstein & Reddon, 1994), so we will not reiterate them here. The important point is that while both the Barrick and Mount (1991) and the Tett et al. (1991) meta-analyses were designed to enhance understanding of personality and performance relations, the studies had quite different purposes and examined fundamentally different questions. The primary purpose of the Barrick and Mount (1991) meta-analysis was to enhance understanding at the construct level. That is, we used a construct-oriented approach to examine whether there were generalizable relationships between
FFM dimensions and job performance. Based on both theory and common sense, we reasoned that there should be generalizable relations between two personality dimensions, Conscientiousness and Emotional Stability, with job performance. That is people who are hard working, organized, prudent, persistent, dependable, and achievement oriented (Conscientiousness) should be better performers in about any job. Similarly, those who are insecure, lack confidence, are stress-prone, and are moody (low in Emotional Stability or Neuroticism), are likely to be poor performers in about any job. The availability of the FFM enabled us to classify personality traits into the appropriate categories to examine whether these hypotheses were true.

On the other hand, the Tett et al. meta-analysis was designed to enhance understanding of the conditions under which personality best predicts performance. Although their study summarized results at the FFM level, these analyses were actually secondary in importance. Tett et al. (1991) were primarily interested in the magnitude of the validities that could be obtained when the authors of primary studies had formulated hypotheses or had used a job analysis to choose personality measures, irrespective of which construct the personality measure assessed. In the 1994 re-analysis of this data (Tett et al., 1994), they showed that in those situations when researchers had formulated hypotheses about personality-performance relationships (a confirmatory strategy), the estimated mean true score correlation was higher (.24) than when an exploratory strategy was used (.04). And when job analysis was used explicitly to choose predictors, the corrected mean validity was slightly higher (.25).

The point of the preceding discussion is not that one of these meta-analyses is superior to the other (methodological differences notwithstanding). Rather, the point is that two meta-analyses that examined the same body of literature, and that were conducted independently and at
about the same time, can appropriately yield different results because they investigated quite different questions. In fact, our view is that these two meta-analyses should be viewed as complimentary, in that each provides information that the other does not. Barrick and Mount (1991) did not try to identify the optimal circumstances or conditions that might yield higher validities for Conscientiousness (or other personality constructs). And in their analyses, Tett et al. did not examine which personality constructs accounted for the higher validities in the confirmatory or job analysis based studies.

But the preceding discussion does not explain why at least 13 other meta-analyses were conducted on this same topic after 1992. Most of these meta-analyses seemed to follow the construct-oriented approach taken by Barrick and Mount (1991), which focused on increasing understanding of which personality constructs were related to performance constructs. One possible reason why the additional meta-analytic studies were conducted was to clarify some apparent conflicting findings in the Barrick and Mount (1991) and Tett et al. (1991) studies with respect to the FFM dimensions. For example, some of these discrepancies may have occurred because of different decision rules used in the studies (e.g. which scales from personality inventories were assigned to FFM dimensions). Other meta-analyses were conducted because they used different populations (international rather than U.S. participants) or because they used only inventories that measured personality at the FFM level rather than individual scales from personality inventories.

We think that it is especially useful to examine three of these meta-analytic studies (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1997), because they had a similar purpose (i.e. they examined personality and performance relations using the FFM framework), but the meta-analyses were operationalized differently (i.e. different decision rules and non-
overlapping samples). Comparison of the results of these studies provides a unique opportunity to determine the convergent validity of three meta-analyses where samples and decision rules differ in important ways. For example, Salgado’s (1997) study used a similar methodology as Barrick and Mount (1991), but used only participants in the European community. Hurtz and Donovan (2000) also used a similar methodology as Barrick and Mount but considered only studies where personality was measured at the FFM level.

Table 1 presents the results from these three meta-analyses as well as the sample-weighted mean estimates of these prior meta-analytic results (i.e., average estimates). For each study, there are five columns, which contain respectively, the total sample size for each trait (N), the mean observed correlation ($r$ and $r^{sw}$), the estimated true score correlation ($\rho$ or $\rho^{sw}$), the estimated true residual standard deviation (SD$\rho$ or SD$\rho^{sw}$), and the percentage of observed variance that was accounted for by statistical artifacts ($(\%V$ and $\%V^{sw}$; where the four artifacts are sampling error and between-study differences in predictor and criterion unreliability, and degree of range restriction).

The best estimate of the true population parameters is provided by the average estimates reported in the last five columns of Table 1. Inspection of these results reveals that among the Five Factor Model constructs, Conscientiousness and to a lesser extent, Emotional Stability have produced consistent, acceptable validities against various job performance criteria across numerous jobs. Consequently, these two personality traits appear to be “universal” predictors across many, if not all jobs. Although the results are not reported here, a meta-analysis of meta-analyses (Barrick, Mount, & Judge, 2001) has shown that the other three personality traits, Agreeableness, Extraversion, and Openness to Experience, are also valuable predictors of
performance, but only in some jobs and in some situations. Thus, these three traits are “contingent” predictors, as they will be relevant only in some situations.

These findings are important and would not have been discovered without meta-analysis and the FFM. However, meta-analysis yields important information other than the magnitude of the estimated true score correlation, namely the variability in the estimates of the true score correlation. We believe that this is a critical area that needs to be scrutinized in more detail. As shown in Table 1, the estimated true score correlations across these three meta-analyses are strikingly similar. For example, the estimated true score correlations for Conscientiousness are .20, .22, and .25 (Hurtz & Donovan, 2000; Barrick & Mount, 1991; Salgado, 1997; respectively). Thus, the meta-analytic estimates were remarkably stable across these three studies, despite the inclusion of different primary studies, different decision rules, and even different operationalizations of personality traits (one study only included FFM level measures (Hurtz & Donovan, 2000), while the other two studies averaged the validities across components of the constructs). These results reveal the robustness of conclusions drawn from large-scale meta-analyses.

The obvious conclusion from Table 1 is that a comprehensive meta-analysis, in combination with a meaningful taxonomy to classify personality traits, has allowed researchers to determine the magnitude of the relationship between personality and performance that is more accurate, credible, and stable than can be derived in any one primary study. Another important conclusion that can be drawn from these results is that the point now has been reached where there is no need for future meta-analyses of this type, as they are likely to result in quite similar findings and conclusions.

Future Role of Meta-Analysis Examining Personality – Performance Relations
Before we discuss the future role of meta-analysis in this area, we would like to offer several recommendations for conducting meta-analyses of personality-performance relations. These recommendations are based on our review of 16 previous meta-analyses in this area and are aimed specifically at meta-analyses of personality traits, rather than at meta-analyses of other topics (e.g. whether unpublished studies are included, how to treat multiple correlations from the same sample). The first recommendation pertains to the classification schemes used. Researchers should clearly define their predictor and criterion taxonomies. If the FFM taxonomy was used to assign personality traits, the definitions for the FFM dimensions should be provided (or made available), as there is not universal agreement on their meaning. If a different personality taxonomy is adopted, there should also be some discussion regarding the theoretical support and comprehensiveness of the taxonomy, in addition to providing definitions of the various dimensions in that taxonomy. If multiple criterion categories (e.g. contextual vs. task performance; objective vs. subjective criteria; organizational citizenship and counterproductive behaviors) are used, the rationale for using them and their definitions should be provided. In addition, researchers should provide (or make available) which specific personality traits or criteria were assigned to each category. Researchers should also clearly describe the jobs that are studied, and if they are assigned to occupational categories, the classification scheme should be explicit.

The second recommendation pertains to the conditions under which the correlations were obtained in the primary studies. We recommend that the following study conditions be coded so that potential moderator analyses can be conducted. Was the correlation hypothesized? Was the personality dimension selected by primary researchers based on a job analysis? Were correlations obtained from a study that reported only significant correlations or from a study that reported all
correlations? Was the personality trait a lower-level facet from a scale on a personality inventory, or was it a higher-level personality construct measured at the FFM level? Was the study obtained from a refereed journal article or was it unpublished?

The third recommendation pertains to the way researchers code and analyze signed correlations. This is a particularly relevant issue for personality traits because of the potential ambiguity regarding the meaning of certain traits. We recommend that the sign of the correlation always should be coded, provided it is done in a consistent way. For example, if in a sample Conscientiousness is found to have a negative correlation with the criterion of number of accidents, this relation is actually in the expected direction and should be coded as a positive correlation. On the other hand, if in a particular sample Conscientiousness has a negative correlation with supervisor ratings of amount of effort exerted, this is in the opposite direction expected and should be coded negatively. These are a few of the practices that if implemented, might lead to greater consistency in conducting and interpreting meta-analytic studies.

We have argued that meta-analysis is a very powerful technique that researchers (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 2000; Tett et al., 1991) have applied to reverse the longstanding belief that personality measures do not contribute to individual differences in work performance. Given that conclusions about the usefulness of personality in work situations are now more optimistic than they were 25-30 years ago, we believe research attention should shift to other questions. Two critical questions are how personality traits are converted into behaviors, and what factors may moderate these relationships. While these issues have been examined in some previous primary studies (e.g., Barrick et al., 1993; Barrick, Stewart, & Piotrowski, in press; Borman, White, Pulakos, & Oppler, 1991; Gellatly, 1996), there is very little research that has systematically examined potential mediator and moderator
variables. This is not surprising since we only recently have been able to conclude “personality matters.” Thus, we expect a large number of primary studies will be conducted over the next decade to examine the effect of mediators and moderators of the relationship between personality and performance.

Does this mean there is no future role for meta-analysis in this area of research? Although there may be a reduction in the number of meta-analyses conducted in the next few years, we do believe meta-analysis will play as pivotal a role during the next decade as was true of the past decade. However, we expect the purposes of those meta-analyses to change dramatically. In the following section, we discuss the role meta-analysis should play in future research.

Clearly, meta-analyses have played a crucial role in advancing our theoretical understanding of the nature of the relationship between specific personality constructs and success at work. However, relatively little attention has been paid to the magnitude of these relationships, which are modest at best. The estimated true score correlations for specific traits rarely exceed .30, which is relatively modest compared with the estimated true score correlation for cognitive ability. We believe meta-analysis can play a key role in trying to understand the conditions that produce higher predictive validities, especially in the search for moderator variables. A critical outcome in any meta-analysis of selection studies is the amount of variation in the predictive validities that is attributed to different work settings. Inspection of the results reported in Table 1 (%V and $\%\bar{V}^{sw}$ columns) suggests that differences in correlations may exist across subpopulations. The sample-weighted mean estimate of the percentage of variance explained by the four artifactual errors ($\%\bar{V}^{sw}$) fails to exceed the 75% rule (Hunter & Schmidt, 1990) for all five of the FFM personality traits. Stated another way, the estimated true residual
standard deviation ($SD_\rho$) values are large in comparison to the estimated true scores ($\rho$). These findings point to the urgent need to further search for moderators of these relationships.

A number of meta-analyses have examined the nature of these relationships across different job types and criteria. For example, the three meta-analyses reviewed in Table 1 examine estimated true score correlations across different jobs (managers, sales, skilled/semi-skilled, etc.) as well as across different criteria (job proficiency, training proficiency, and personnel data, etc.). Differences in job content and criteria serve as important potential moderators and warrant further examination. Tett et al. (1991) examined the effects of these same moderators, including those found across types of criteria (objective vs. subjective); across different jobs (professional vs. non-professional; managerial vs. non-managerial; civilian vs. military); and across applicant samples vs. incumbent samples. They also examined differences across studies where researchers developed a priori hypotheses (confirmatory approach) or did not (exploratory approach); and by source of data (researchers vs. company; articles vs. dissertations). These are also important types of moderators that future research should examine.

We are particularly intrigued by Tett et al.’s finding (1994) that when researchers follow a confirmatory approach rather than just an exploratory approach, the magnitude of the effects for personality are substantially larger ($\rho = .24$ vs. .04). Specifically, what attributes are researchers focusing on when deriving a priori hypotheses? Robertson and Kinder (1993) found that when asked to generate hypotheses linking a specific personality test or scale to job performance on the basis of written descriptions, practitioner’s predictions were supported by empirical data less than 45 percent of the time. Thus, the usefulness of a priori judgments about the traits relevance to a particular criteria will have to be examined carefully. Nevertheless, the findings by Tett et al. (1994) suggest there is merit in such an approach, particularly when researchers examine
Impact of Meta-Analyses

theoretically meaningful factors that influence personality-performance relationships. Below, we discuss a few of the moderating factors researchers should consider.

Researchers have long been interested in personality as a means to explain a person’s behavior. This is one of the primary reasons personality is of such interest to researchers in the field of industrial/organizational psychology. Yet, there is very little research that accounts for the effects of personality on behavior at work. Instead, much of this research relies on outcomes (sales, turnover, promotions, etc.) and supervisory ratings of performance. These measures are certainly influenced by the actions people take (behavior), but they also include, to an unknown extent, factors outside the individual’s control. To the extent these performance measures include variance that is due to factors the individual cannot control, the correlates are likely to underestimate how well personality predicts behavior.

Thus, we suggest that future research should focus on behavior rather than outcomes. Although job analysts have devoted considerable effort to describing behavior at work (Harvey, 1991), there has been inadequate theoretical and empirical work to link the domain of individual differences (e.g., cognitive abilities, personality traits, and interests) with the domain of job behavior constructs (e.g., delegating/coordinating, exchanging information, operating machines, etc.). We believe that our understanding of the relationship between personality and behavior will be enhanced if we develop and apply theories of the structure and content of work behavior.

There is also considerable recognition that situations are important determinants of the nature of the relationship between personality and performance. To date, there have only been a few papers (e.g., Barrick, Mitchell, & Stewart, in press; Murtha, Kanfer, & Heggested, 1996; Stewart & Barrick, in press) that attempt to conceptualize the basic kinds of situations and links those situations to personality traits. We believe one fundamental situational variable that must
be accounted for in personality research is the level of autonomy. Autonomy captures the extent to which the external environment constrains a person’s freedom to choose how to behave (Barrick & Mount, 1993). In *strong situations* (low in autonomy), the organization exerts considerable pressure to induce behavioral conformity. These controlling forces “press” the individual to behave in a specific way or exhibit a very narrow range of behaviors. In such circumstances, personality has at best, a weak effect on behavior. In contrast, *weak situations* (high in autonomy) present few demands or “presses” to conform. In such settings, the individual’s “free will” determines which behaviors, if any, to undertake. The magnitude of the relationship between personality traits and behavior is likely to be greater in weak situations, where people have greater discretion or choice in how they will perform their job. Thus, an important moderator researchers must consider involves the influence autonomy has on the relationships between personality and behavior.

In research on cognitive ability, a theory has been developed and tested that focuses on the complexity of the job to explain the relationship between ability and performance (Hunter, 1986; Schmidt, Hunter, Outerbridge, & Trattner, 1986). This research shows that cognitive ability relates to job performance indirectly through its relationship to job knowledge, which in turn relates to performance. These results also demonstrate that cognitive ability relates directly to performance, although the effects are smaller than the indirect effects. This theory shows that more complex jobs require even more job knowledge and learning, which in turn leads to higher performance. Thus, job complexity is an important situational variable. Is job complexity likely to be an important moderator for personality? We do not know. However, if it is, it may be because complexity is associated with greater discretion or autonomy, in addition to a need for more knowledge.
There are also a number of methodological issues researchers should consider as potential moderators. Mount, Barrick, and Strauss (1994) found that job-relevant personality constructs were valid predictors of performance when based on ratings from observers (e.g., supervisor, co-workers, and customers). Further, they showed observer ratings of sales representatives on Conscientiousness (corrected $r$ ranged from .32 to .42) and Extraversion (corrected $r$ ranged from .28 to .38) were correlated with ratings of performance, and that these correlations accounted for additional variance beyond that due to self-ratings alone. We believe this distinction is important because observer ratings capture the social reputation of the individual (Hogan, 1991). Further, these reputational effects are primarily derived from work settings for these raters (supervisors, co-workers, customers). In contrast, self-ratings tend to capture what people are like “way down deep”. Ratings from the individual’s perspective are influenced by an individual’s perceptions of himself or herself in numerous situations, incorporate less observable information about motives, intentions, feelings, and past behavior, and also includes the effects of one’s attempt at impression management, that is to consciously manage the impression one conveys to others. Taken together, it may well be that observer ratings capture more valid variance from work behaviors, particularly if those ratings are obtained almost exclusively in the work environment. Thus, one methodological moderator researchers should consider is rating source (self vs. observer).

There is considerable debate in the field about the appropriate level of analysis for assessing relationships between personality and performance. Schneider, Hough, and Dunnette (1996) argue that FFM traits are too broad and that prediction will improve if you rely on more precise facets of these traits, while Ones and Viswesvaran (1996) argue for greater breadth in the use of personality traits. We believe the appropriate level of analysis depends on the purpose. If
the purpose is to make a selection decision, we argue that a broader, more comprehensive measure is appropriate for predicting an equally broad measure of overall success at work. Support for this position can be found in two recent meta-analyses, one pertaining to integrity (Ones, Viswesvaran, & Schmidt, 1993) and the other to Customer Service (Frei & McDaniel, 1998). The estimated true score validity for integrity measures and customer service measures was .41 and .50 when predicting supervisory ratings of overall job performance, respectively. In both cases, construct validity evidence suggests these measures were positively and strongly related to Conscientiousness, Emotional Stability, and Agreeableness. Thus, broad measures of personality have very high predictive validities of overall performance, higher even than those typically reported for specific FFM personality traits. More importantly, these broad measures of success at work are the appropriate criteria to consider when making selection decisions.

In contrast, if the purpose is to enhance understanding, linking specific, lower level facets of FFM constructs to specific, lower level criteria may result in stronger correlations. For example, Mount and Barrick (1995) found that lower level personality traits predicted specific performance criteria better when they were conceptually related to the criterion. In this meta-analysis, they examined the strength of relations for Conscientiousness and two lower-level facets, dependability and achievement. As hypothesized, they found dependability was a better predictor of employee reliability (\( \rho = .47 \)) and quality (\( \rho = .48 \)) than was conscientiousness (\( \rho = .41 \) and .44, respectively) or achievement (\( \rho = .33 \) and .22, respectively). In contrast, achievement was a better predictor of effort (\( \rho = .58 \)) and creativity (\( \rho = .19 \)) than was Conscientiousness (\( \rho = .51 \) and .13, respectively) or dependability (\( \rho = .43 \) and -.04, respectively). Although the estimates for conscientiousness were not larger than the “relevant” lower-level component, they were always larger than the “nonrelevant” lower-level component.
Researchers can examine the influence of this moderator by using meta-analysis to generate a more accurate and more credible estimate of the predictive validity of higher-level or lower-level personality traits across these various purposes.

Past research has focused on differences among individuals on single traits (e.g., Conscientiousness, Emotional Stability, etc). While this was the appropriate step when initially trying to establish the effectiveness of personality in selection settings, it is time to examine the multivariate relationship among personality traits when predicting performance. Can meta-analysis play a role in examining indexes of multivariate predictability? Meta-analysis used in combination with structural equation modeling could be used to establish the estimated true score correlation between multiple personality dimensions and performance measures (Viswesvaran & Ones, 1995). We propose researchers apply structural equation modeling to the matrix of estimated true score correlations derived from two separate meta-analyses; one assessing the relationship between personality traits, the other assessing the relationship between personality and performance. Such research will enable researchers and practitioners to begin talking about the multivariate validity of multiple personality traits.

A second “multivariate” role for meta-analysis, in combination with structural equations modeling would be to find whether FFM personality traits accounted for incremental variance over that accounted for by cognitive ability, the interview, biodata, and other selection predictors. Some recent evidence (Goffin, Rothstein, & Johnston, 1996; McManus & Kelly, 1999; Mount, Witt, & Barrick, 2000; McHenry et al., 1990) demonstrates that increments for Conscientiousness and Emotional Stability over these predictors typically range between 0.05 and 0.15, which may seem small when contrasted with cognitive ability constructs.

Nevertheless, it should be noted that such gains in incremental validity yield large economic and
social gains. For example, personality measures typically result in minimal disparate impact compared to cognitive ability measures (Mount & Barrick, 1995). Moreover, a practical issue is that it often is quite inexpensive to obtain information from personality measures. More importantly, personality traits also predict many other outcomes organizations value, including whether the person shows up for work on time (Judge, Martocchio, & Thoresen, 1997), remains with the firm (voluntary turnover; Barrick & Mount, 1996), contributes to the firm beyond that which is required by the job (organizational citizenship behavior; Organ & Ryan, 1995), and avoids harmful behaviors, including theft, sabotage, and breaking the rules (counterproductive behaviors; Hough et al., 1990). Meta-analysis along with structural equations modeling could play a substantial role in determining the likely effects firms would realize by using personality traits along with other predictors during selection.

Two recent examples illustrate the integration and application of these two methods when exploring incremental validity. Cortina and associates (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000) formed a “meta-correlation matrix” representing the relationships between cognitive ability, Conscientiousness, interviews, and job performance to examine the incremental validity of 3 levels of structured interviews on job performance. Although this study did not report whether Conscientiousness predicts job performance above and beyond cognitive ability and interviews, it could have. Schmidt and Hunter (1998) did use this methodology to illustrate that for most jobs, Conscientiousness combined with an intelligence test is 18 percent more valid than an intelligence test alone. Combining meta-analysis with structural equation modeling will enable researchers and practitioners to begin talking about the incremental validity of personality, across multiple dimensions of personality.
Another “multivariate” role for meta-analysis is to account for the effect of impression management on these validities. There has been concern raised that the factor structure of personality scales differs for applicant responses, due to the demands to manage one’s impressions in an applicant setting (Ellingson, Sackett, & Hough, 1999; Montag & Comrey, 1990; Schmit & Ryan, 1993). Recent evidence suggests that social desirability in responses can erode the dimensionality of personality (Ellingson et al., 1999) or results in an additional “halo/good impression” factor (Montag & Comrey, 1990; Schmit & Ryan, 1993). In either case, one implication is the FFM factors become more highly intercorrelated, due to the influence of impression management.

To test this, one could apply meta-analysis to explore the relationships found between personality traits from applicant studies and compare that matrix to one found from incumbent samples. The resulting estimated true score correlation matrixes for the two sets of analyses should differ substantially, with much higher intercorrelations found in the applicant sample, if factor structures do indeed differ across applicant and incumbent settings. If this is so, the predictive validity of a structural equations model including all FFM traits should be significantly lower when using applicant samples instead of incumbent samples. Ironically, Tett et al. (1994) found just the opposite in their meta-analysis of individual personality scales (not a multivariate analysis though), as the estimated true score validity was significantly higher for recruits ($\rho = .27$) than for incumbents ($\rho = .12$).

Finally, another “multivariate” approach to personality assessment focuses on the patterning and organization of traits within a person. This approach recognizes we are each characterized by a unique constellation of personality traits that defines who we are and how we behave. This “profile configural” approach aims to discover the basic categories of human
nature and emphasizes how various personality dimensions may interact to predict behavior. For example, Simonton’s (1999; 2000) work suggests that eminence will only emerge when a number of traits all are present in high levels. This theory specifies a complex interaction among multiple personality variables and stands in sharp contrast to past attempts to predict job performance with specific personality traits. Given “configural” data would not be available in studies focusing on the “main effects” of personality, researchers must first conduct a number of primary studies. Consequently, in contrast to the other moderators explored here, the role for meta-analysis is likely to be a number of years away, however, as there is a dearth of primary studies assessing this effect in the literature.

Taken together, we argue that meta-analysis should play a meaningful role in future theory development, as it provides the promise of clarifying the magnitude and establishing the importance of a number of theoretically oriented moderator variables. The second way meta-analysis can contribute to theory development is to facilitate the discovery of the effects of various mediators on this body of research. Prior meta-analyses have not examined the mediating effect of other variables on the personality – performance relationship. We believe this type of application of meta-analysis will become particularly important to researchers during the next decade.

Although there is clear evidence 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 primary way personality is thought to affect performance is through a person’s motivation (Barrick et al., in press; Kanfer, 1991; Mount, & Barrick, 1995; Murray, 1938). To date, research has been significantly hindered because an accepted framework does not exist for studying motivational constructs. Nevertheless, with the
development of a theoretically relevant motivational taxonomy, we believe meta-analysis can effectively be used to explore the nature and magnitude of the relationships between specific motivational variables and personality traits. We expect the next decade will witness an explosion of interest in the structure of motivation. Through this research, we believe a meaningful taxonomy will emerge that will enable researchers to examine the process through which personality affects motivation, and motivation in turn, affects job performance.

To illustrate, one could study how Conscientiousness affects performance. Conscientious individuals are achievement oriented, hard working, persistent, decisive, responsible, dependable, reliable, careful, organized and planful. These traits are fundamentally linked to motivation at work. Motivation is defined by Campbell (1991) as: “[it is] a combined effect from three choice behaviors – choice to expend effort, choice of level of effort to expend, and choice to persist in that level of effort” (p. 706). A researcher could use these three choice behaviors to develop a motivational taxonomy, and than apply meta-analytic procedures to examine how Conscientiousness affects these three motivational variables, and in turn, how these motivational variables influence performance.

In conclusion, meta-analysis combined with the FFM has enabled researchers to further our understanding of the true nature of the relationship between specific personality traits and success at work in the past decade. Meta-analysis will continue to play a critical role in the development of our theoretical understanding of these relationships. However, the focus of future meta-analyses should shift from explaining the relationship between personality traits and performance towards explaining the motivational process through which personality affects performance (mediation) as well as clarifying the influence of situational and methodological factors on those relationships (moderators). Such work will be crucial both theoretically in the
continued development of models of job performance, and practically, as it will clarify the value of using personality-based tests during the selection or promotion of employees.
References


Footnotes

1 The results reported here are from Tett, Jackson, Rothstein, and Reddon (1994). This meta-analysis is a re-analysis of Tett et al’s (1991) non-FFM data, and is based on a number of revisions to address the statistical concerns raised by Ones, Mount, Barrick, and Hunter (1994).

2 The re-analysis from Tett et al. (1994) resulted in different magnitudes than those reported in Tett et al. (1991). For example, across the two analyses, the mean correlation when a confirmatory strategy was used was .29 in 1991 and .24 in 1994, when an exploratory strategy was used the estimate was .12 in 1991 and .04 in 1994. Finally when job analysis was used, the estimate was .38 in 1991 and .25 in 1994. The latter estimate was not significantly higher than the “no job analysis” estimate (.24) in the 1994 re-analysis. In 1991, this distinction had been found to significantly differ. We report the 1994 estimate here because they are likely to be more accurate.
Table 1. Meta-Analytic Results for Personality Factors Pooled Across Criteria and Occupations.

<table>
<thead>
<tr>
<th>Personality Trait</th>
<th>Barrick &amp; Mount</th>
<th>Salgado</th>
<th>Hurtz &amp; Donovan</th>
<th>Average Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>(\bar{r})</td>
<td>(\rho)</td>
<td>SD(\rho)</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>18,719</td>
<td>.05</td>
<td>.08</td>
<td>.10</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>17,520</td>
<td>.04</td>
<td>.07</td>
<td>.09</td>
</tr>
<tr>
<td>Extraversion</td>
<td>19,511</td>
<td>.08</td>
<td>.13</td>
<td>.11</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>14,326</td>
<td>.03</td>
<td>.04</td>
<td>.13</td>
</tr>
</tbody>
</table>

Note: N = sample size; \(\bar{r}\) = Mean Observed Correlation; \(\rho\) = Mean True Score Correlation; SD\(\rho\) = True Score Residual Standard Deviation; %V = Percentage of Variance Explained By Artifactual Errors; \(\bar{r}_{sw}\) = Estimated Sample Weighted Mean Observed Correlation; \(\bar{\rho}_{sw}\) = Estimated Sample Weighted Mean True Score Correlation; SD\(\bar{\rho}_{sw}\) = Estimated Sample Weighted True Score Standard Deviation; And %\(\bar{\rho}_{sw}\) = Estimated Sample Weighted Percentage Of Variance Explained By Artifactual Errors.

*The 90% Credibility Value around the True Score excludes zero.