PERSONNEL PSYCHOLOGY
1994, 47

PERSONALITY AND JOB PERFORMANCE: A CRITIQUE OF THE TETT, JACKSON, AND ROTHSTEIN (1991) META-ANALYSIS

DENIZ S. ONES
Department of Management
University of Houston

MICHAEL K. MOUNT, MURRAY R. BARRICK
Department of Management and Organizations
University of Iowa

JOHN E. HUNTER
Michigan State University

Tett, Jackson, and Rothstein (1991) recently presented a meta-analysis of the relationship between personality and job performance. Many of their findings, particularly those pertaining to the Big Five personality dimensions, are at odds with one other large scale meta-analytic study (Barrick & Mount, 1991) investigating the relation between personality and performance. In order to reconcile these new results with previous findings, we examined differences in the sample sizes used, the process for assigning pre-existing scales to personality dimensions, and the nature of the jobs investigated. In addition, we found four technical errors in the Tett et al. moderator meta-analyses in computing sampling error, the bias correction, sampling error for bias corrected correlations, and computing sampling error variance across studies. These errors raise serious questions about the interpretation of their results for various moderators of the personality-job performance relationship.

There has been a resurgence of interest in recent years in the validity of personality constructs as predictors of performance. Through the use of taxonomies, derived primarily from the Big Five personality dimensions, and meta-analytic methods, researchers have been able to identify predictive relationships between personality constructs and job-related criteria. At the time prior reviews were conducted, there was no widely accepted taxonomy for examining personality constructs (e.g. Guion & Gottier, 1965; Schmitt, Gooding, Noe, & Kirsch, 1984) and, consequently, meaningful relations between personality constructs and criterion measures were obscured.

The authors would like to thank Frank Schmidt and Michael Judiesch for helpful discussions during the formative stages of this article.

Correspondence and requests for reprints should be directed to Deniz Ones, College of Business Administration, Department of Management, University of Houston, Houston, TX 77204-6283.

Given the progress that has been made in recent years, it is critical that studies be carefully scrutinized in order to avoid the confused interpretations that have characterized this area in the past 25 years. Tett, Jackson, and Rothstein (1991) recently conducted a meta-analysis of the relationship between personality and job performance in order to investigate conflicting findings in previous research. They assessed the overall validity of personality measures as predictors of job performance, investigated several moderator variables, and appraised the predictability of job performance as a function of eight dimensions of personality, including the Big Five. Among the major findings were that studies using confirmatory strategies produced a corrected mean personality scale validity (.29) that was more than twice as large as those based on exploratory strategies (.12). Furthermore, validities were higher in studies where job analysis was used to select personality measures (.38). Results for analyses based on the Big Five predictor constructs revealed corrected mean validities which ranged from .16 for extraversion to .33 for agreeableness.

Some of the results of the Big Five analysis from the Tett et al. (1991) study, however, differ in important ways from a previous large scale meta-analysis (Barrick & Mount, 1991). Therefore, the first purpose of this study is to examine the potential reasons for the differences in results for the Big Five content areas obtained by Tett et al. and those of Barrick and Mount. We discuss differences in the sample sizes used for predictor constructs, how pre-existing scales were assigned to personality dimensions, and the nature of the jobs investigated. The second purpose is to discuss four sets of technical errors in other moderator analyses conducted by Tett et al. These technical errors render many of the results of the Tett et al. moderator analyses uninterpretable. Both issues are addressed in turn, in the following sections.

Comparison of Results Using the Big Five Framework

The Big Five taxonomy of personality provides a framework for summarizing validities within predictor constructs, thereby allowing knowledge to be cumulated in a meaningful way. As reported in Table 1, two recent meta-analytic studies of the validity of the Big Five dimensions of personality for predicting job proficiency (Barrick & Mount, 1991; Tett et al., 1991) arrive at different conclusions regarding the predictive validity of some of the Big Five constructs. One of the most important findings in the study by Barrick and Mount is that one of the Big Five constructs, conscientiousness, is a valid predictor of job performance in all jobs studied and for all criterion types. On the other hand, Tett et al. found that

TABLE 1

Comparisons of Meta-Analytic Validity Results for Major Dimensions of Personality in Predicting Job Proficiency

Personality dimension	Barrick & Mount (1991) meta-analysis[a]	Tett et al. (1991) meta-analysis[b]
Emotional stability[b]		
N	11,635	900
K	87	10
r_{xy}	.04	.15
p	.07	.22
Extraversion		
N	12,396	2,305
K	89	15
r_{xy}	.06	.10
p	10	.16
Openness to experience		
N	9,454	1,304
K	55	10
r_{xy}	−.02	.18
p	−.03	.27
Conscientiousness		
N	12,893	450
K	92	7
r_{xy}	.13	.12
p	.23	.18
Agreeableness		
N	11,526	280
K	80	4
r_{xy}	.04	.22
p	.06	.33

Note: Barrick and Mount (1991) results are from their Table 3, p. 15 and are only based on job proficiency. Barrick and Mount also report other analyses based on additional studies (on average, $K = 46$, $N = 8{,}500$) which assess the predictive validity of the 5-factor model with two other criterion types, training proficiency and personnel data. Tett, Jackson, and Rothstein (1991) results are from their Table 5, p. 726. N = Total sample size, K = Number of studies included in the meta-analysis, r_{xy} = uncorrected criterion-related validity coefficient, p = true criterion-related validity coefficient (meta-analytically obtained).

[a]Barrick and Mount (1991) and Tett et al. (1991) have corrected the mean validity for predictor unreliability, criterion unreliability, and range restriction.

[b]Sign reversed to make compatible with the present table label of "Emotional Stability."

conscientiousness had lower validity than three other personality constructs: agreeableness, openness to experience, and emotional stability. The highest validity in the Tett et al. study was for agreeableness. As Goldberg (1993) points out, "This inconsistency in the findings between two large-scale quantitative reviews of a similar body of literature is befuddling...." Our purpose in this section is to explain why these discrepant results may have occurred.

The first explanation of the difference in results between these two meta-analyses can be attributed to the comprehensiveness of the studies. The number of studies (K) and sample sizes (N) on which the analyses were based are shown in Table 1. It can be seen that the number of studies and sample sizes in Tett et al. (1991) are much smaller (by a factor of 10 and 20, respectively, on average) than those used in Barrick and Mount (1991) for predicting job proficiency. For example, Tett et al. included only 7 studies (N = 450) for conscientiousness, whereas Barrick and Mount included 92 studies (N = 12,893) assessing job proficiency. Given that both studies summarized the same body of literature, a relevant question is why the sample sizes reported in Tett et al. are so small.

One explanation is that Tett et al. (1991) restricted their Big Five analysis to those studies using a confirmatory strategy; that is, when there was some theoretical reason to believe there was a relationship between the personality construct and the criterion. It is unnecessary to restrict a review of the personality literature to confirmatory studies when using a meta-analysis *in combination* with a personality taxonomy. The primary purpose of such a meta-analysis is to investigate whether the correlation between a specific personality construct and performance criterion generalizes across studies or organizations within the same job category, after accounting for statistical artifacts, such as sampling error, between-study differences in test unreliability, criterion unreliability, and degree of range restriction.

Consider the following example. For theoretical reasons a researcher concludes that extraversion should be a valid predictor of managerial performance (this would be considered a confirmatory study by Tett et al., 1991). In the process of conducting a comprehensive literature review, 10 studies of similar jobs are found that investigate this relationship between extraversion and managerial performance. However, only 2 of these studies use a confirmatory strategy while 8 use an exploratory strategy. Based on Tett et al.'s rules of inclusion, only the 2 confirmatory studies would have been meta-analyzed in their Big Five analysis. In contrast, the other 8 studies, which simply "explored" the relationships between the same constructs (extraversion and job proficiency) in the same job, would be excluded. By unnecessarily restricting their analyses to confirmatory studies, and reducing the number of studies analyzed, Tett et al. reduce the likelihood of accurately assessing the magnitude of the relations between personality constructs and criteria.

Yet, even after accounting for the restriction to confirmatory studies, the relatively small sample size in Tett et al.'s (1991) Big Five meta-analyses is puzzling. The number of studies and overall sample size they used in the Big Five analyses (on average, K = 9, N = 1,050) are substantially smaller than those they report earlier for the confirmatory

analyses (K = 46, N = 9,054). However, no adequate explanation is provided for the discrepancies in sample sizes between the two analyses. Consequently, the small number of studies (and sample sizes) included in Tett et al.'s analyses mitigate the primary value of meta-analysis, which addresses the problem of inadequate statistical power. With such small samples, Tett et al. are not able to account for chance and other statistical and measurement artifacts to the same extent as Barrick and Mount (1991).

The second explanation pertains to the procedures used to assign personality scales from specific inventories to personality constructs. The procedure used to assign scales to the personality dimensions are described in detail by Barrick and Mount (1991), whereas those used by Tett et al. (1991) are not. Although Tett et al. make reference to factor analytic data by Costa and McCrae (1988), the explanation of how these results were used, for which inventories, which scales, and so forth, was incomplete. The results pertaining to the validity of the Big Five constructs are substantially influenced by the particular personality scales from the various inventories which are assigned to each construct. Without a description of the procedures used to assign scales to the Big Five constructs, it is impossible to interpret and understand the inferences about the validity of the constructs. Consequently, the differences in results between studies may be due to the fact that different personality scales were assigned to the Big Five constructs by Tett et al.

A third explanation for differences in results between the meta-analyses could be attributed to the existence of moderators. One important moderator is the nature of the jobs investigated. Barrick and Mount (1991) summarized the validity of the Big Five constructs across a number of different occupations (professionals, police, managers, sales, and skilled/semi-skilled); whereas, Tett et al. (1991) report validities *across* job categories. The moderating effect of job type is illustrated by the finding reported by Barrick and Mount that extraversion was a valid predictor of job success for sales and management occupations, but was not significantly related to job proficiency in any other occupations. Since Tett et al. did not summarize results by job category, it is quite likely that the validities obtained for the Big Five constructs may be as much a function of the types of jobs studied as the nature of the personality construct. Furthermore, the Tett et al. analysis did not indicate the extent to which the Big Five results could have been influenced by various moderators found to be significant in their own analysis (i.e., whether the subjects were incumbents or recruits, civilian or military employees, whether the studies were from articles or dissertations, or based on job analyses or not). For these reasons, it is difficult to interpret the results

of the Tett et al. Big Five meta-analysis, much less compare them to the Barrick and Mount meta-analysis.

We believe that in order to be both theoretically and practically useful, meta-analyses of personality must examine the relationship between specific personality constructs and job-related criteria in various jobs. Therefore, it is important to understand some of the reasons for differences between the Barrick and Mount (1991) and Tett et al. (1991) Big Five meta-analyses. However, it should be noted that the major thrust of the Tett et al. article is the identification of additional moderators, such as the effects of using job analysis, irrespective of the specific personality construct. In the next section we point out four technical errors in their moderator analyses that produce uninterpretable or ambiguous results.

Technical Errors in the Meta-Analysis

In conducting the moderator meta-analyses, Tett et al. (1991) averaged correlations within studies using absolute values (it should be noted that absolute values were not used in their Big Five content analyses). Although there is nothing inherently wrong with using absolute values, it is statistically unnecessary. However, if absolute values are used, it requires a relatively complicated set of mathematical transformations and, needless to say, if the inappropriate corrections are applied, the results will be uninterpretable. Unfortunately, a serious technical problem in the Tett et al. article is that statistical and mathematical errors are made in their corrections.

Four important statistical errors result from the inappropriate corrections for absolute values in the Tett et al. (1991) study. These errors apply to their findings reported in their Table 3. The first three errors were made while "correcting" individual study results before conducting the meta-analysis, while the fourth error was in the application of the meta-analytic procedures. By way of illustration, we will consider the correlation reported in Tett et al. for Bagozzi (1978)—an absolute correlation of .20 with a sample size of $N = 38$. Similar errors apply to other primary study validities in their study.

Mathematical Errors in Computing Sampling Error

In their Appendix, Tett et al. (1991) argue, correctly, that the use of absolute correlations produces a potential large upward bias in the estimated population correlation. Suppose the uncorrected population validity of the trait under consideration is 0 for a given job, with a sample size of 38, the 95% confidence interval for sample correlations as calculated using Tett et al.'s formulas would range from $-.32$ to $+.32$.

However, by using absolute correlations, the distribution would be very highly skewed with a mean of .13 and a 95% maximum of +.32. Tett et al.'s Figure 2 shows this situation, where the distribution is of absolute value correlations, with the actual $\bar{\rho} = 0$. Thus, when absolute correlations are used, the resulting estimated population correlation is .13 rather than the correct population value of 0.

Tett et al. (1991) recognize this problem and seek to solve it by subtraction; however, as we show in subsequent paragraphs, they subtract the wrong number. Since they average within study and since the within study correlations are averaged across studies in the meta-analysis, the relevant correction factor that Tett et al. should have used is the mean correlation due to sampling error, that is, .13 in this case. In other words, to overcome the effects of absolute value correlations, Tett et al. should have subtracted the mean correlation due to sampling error (.13) from the observed correlation. Thus, according to their method, the estimated correlation for the Bagozzi (1978) study should have been .20 − .13 = .07.

However, Tett et al. (1991) did not use the mean correlation based on the sampling distribution of the correlation. They used the median correlation by mistake. Whereas the mean absolute sample correlation is equal to .7979(SD_e), the median absolute sample correlation is equal to .6750(SD_e). (Both of these values can be found in Arkin and Colton (1950; p. 141), the source Tett et al. used.) Thus, using the erroneous .6750(SD_e) formula, they incorrectly computed the bias correction to be .11 instead of .13. Had there been no other errors, for the Baggozi (1978) study, this would have produced an estimated correlation of .20 − .11 = .09. An error of .09 versus .07 (i.e., the correct value) is a difference in magnitude of 29%.

Error in Computing the Bias Correction

For the Bagozzi (1978) study, Tett et al. (1991) do not produce the estimate .20 − .11 = .09 (much less the correct .20 − .13 = .07). This occurs because they made yet another error in applying the bias correction. Instead of subtracting the bias correction from the observed value, they do the following: (a) Square the observed correlation and square the (incorrect) bias correction factor, (b) subtract the smaller square from the larger square, and (c) compute the square root of the difference between the squares. To illustrate the effect of this error, consider the example. Through their procedure, they squared the observed value as well as the incorrect bias correction factor: $(.20)^2 = .0400$ and $.11^2 = .0121$. The square root of the difference between these values is $(.0400 - .0121)^{\frac{1}{2}} = (.0279)^{\frac{1}{2}} = .17$. Although the actual estimate for Tett et al. should have

been .20 − .13 = .07, due to the inclusion of both errors, they obtained an estimate of .17, which substantially differs from the correct estimate of .07. The net effect of the two errors is that they estimate the correlation to be .17; whereas the correct estimate is .07. Thus their estimate is more than two times too large.

Similarly, the cumulative net effect of the two errors on all other validities from the Tett et al. (1991) database results in a mean sample-weighted validity of .16 (their estimate) versus .12 (the correct estimate). (A table listing all the corrected validities Tett et al. should have used in their moderator analyses, after correcting for these errors, is available from the first author.) It should be noted that the differences between the Tett et al. mean correlation and the correct mean value is a lower bound estimate. This is because we were not able to correct the 20 studies Tett et al. corrected for dichotomization in our analyses. Corrections for dichotomization can be expected to increase the differences between the mean values. Nevertheless, the Tett et al. values on average are nearly one and one-third times too large.

Error in Computing Sampling Error for Continuity Corrected Correlations

In 20 of the 97 studies, Tett et al. (1991) found at least one of the variables to be artificially dichotomized. They then properly corrected the correlation (for "discontinuity" to use their terminology). However, while they corrected the correlation, they did not correct the sampling error standard deviation which they used to compute the bias correction factor.

Since they do not tell which study values were corrected and which were not, let us assume *for purposes of illustration* that the Bagozzi (1978) study was one of the corrected correlations. Let us conservatively assume that both variables were split at the median (and it could have been a more extreme split). In order to calculate what the .20 value was when artificially dichotomized, we had to apply the correction for dichotomization formula in reverse. That is, the correlation of .20 which resulted from the Bagozzi study would be attenuated to .13 due to artificial dichotomization of variables. As previously stated, given that the mean sampling error for the absolute correlation is .13 for an uncorrected correlation, the absolute value bias corrected (but attenuated) correlation would have been $r = .13 - .13 = .00$. That is, for the attenuated correlation based on dichotomized variables, the correction for biased sampling error would have produced an estimate of 0.

When the dichotomized correlation is corrected for discontinuity, the corrected correlation increases from .13 to .20. However, this correction also increases the associated sampling error standard deviation from

.1644 to .2569 (if the correction factor is based on the mean; i.e., the correct procedure) and from .1110 to .1734 (if the correction factor is based on the median, as Tett et al., 1991, do; the incorrect procedure). In this case then, the properly bias corrected correlation is .20 − (.7979)(.2569) = .20 − .21 = −.01; whereas the Tett et al. estimate would have been .20 − (.6750)(.1734) = .08.

If we assume that the Bagozzi (1978) correlation is one of those corrected for continuity, then the cumulative statistical errors in the Tett et al. (1991) procedure yield a very large error: an estimated correlation of .08 reported in Tett et al. instead of an estimated correlation of −.01.

The impact of these errors is twofold. First, the mean validity levels are uncommonly high because incorrect estimates were used at the primary study level. Second, the estimated true variance and SD_ρ is much too high because the wrong sampling error calculations were applied.

Errors in Computing Sampling Error Variance in the Meta-Analysis

Since Tett et al. (1991) used the Hunter, Schmidt, and Jackson (1982) meta-analysis formulas, they implicitly used the sampling error variance formula for an unsigned correlation. This is wrong for two reasons. First, they did not use the simple difference but rather the square root of the difference between squares. This highly nonlinear operation would have a considerable impact on the sampling error distribution (which could be estimated using Taylor's series) of the difference between squares.

Second, Tett et al. (1991) did not consider that the absolute correlation has a smaller standard deviation than the signed correlation. For a population correlation of 0 and a sample size of 38, the signed correlation has a sampling error mean of 0 and a sampling error standard deviation of .1644 (a variance of .0270). For the absolute correlation, the mean sampling error is .13 and the standard deviation is .0993 (a variance of .0099). That is, the sampling error standard deviation for the absolute correlation is 40% smaller than the sampling error standard deviation for the signed correlation. In terms of variances the comparison is .0099 versus .0270; or 63% smaller.

In sum, the Tett et al. (1991) moderator meta-analyses are mathematically incorrect. If absolute correlations are used in a meta-analysis, then Tett et al. are right to use a bias correction factor. However, they compute that factor incorrectly. First, they underestimate the bias correction because they use the median bias factor instead of the mean bias factor. Second, they miscalculate the bias correction because instead of using the simple difference they use the square root of the difference between squares. Third, they do not correct sampling error standard deviations for continuity, and thus greatly underestimate the sampling error for the

20 of 97 correlations that were corrected for continuity. The cumulative impact of these three errors is, for the Bagozzi (1978) study, a correlation of .08 instead of a correlation of −.01 (.17 vs. .07 if the Bagozzi study is not one corrected for continuity). This error is so large that it renders their numerical estimates of mean validities uninterpretable. Since the meta-analysis reported in Table 3 of Tett et al. used both distorted single study correlations and the wrong sampling error variance computations at the meta-analytic level, their estimates of variation in personality validities across studies are also uninterpretable.

In conclusion, it is apparent that great effort went into the Tett et al. (1991) study. One important contribution is that they considered a number of moderators that previous meta-analyses of the relationship between personality and job performance have not considered. Nonetheless, it is disappointing that the technical errors in their moderator meta-analyses raise serious questions about the interpretation of the findings. Furthermore, the Tett et al. conclusions from the Big Five analyses are also suspect for a number of other reasons, as previously discussed. Tett et al. state that a primary purpose of their study is to investigate conflicting findings in previous research on personality and job performance. Ironically, because of the technical errors and inadequate Big Five analyses, the study does not accomplish its intended purpose.

REFERENCES

Arkin H, Colton RR. (1950). *Tables for statisticians*. New York: Barnes & Noble.

Bagozzi RP. (1978). Salesforce performance and satisfaction as a function of individual difference, interpersonal, and situational factors. *Journal of Marketing Research, 15*, 517–531.

Barrick MR, Mount MK. (1991). The Big Five personality dimensions and job performance. PERSONNEL PSYCHOLOGY, *44*, 1–26.

Costa PT, McCrae RR. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology, 55*, 258–265.

Goldberg LR. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.

Guion RM, Gottier RF. (1965). Validity of personality measures in personnel selection. PERSONNEL PSYCHOLOGY, *18*, 135–164.

Hunter JE, Schmidt FL, Jackson GB. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publications.

Schmitt N, Gooding RZ, Noe RA, Kirsch M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. PERSONNEL PSYCHOLOGY, *37*, 407–422.

Tett RP, Jackson DN, Rothstein M. (1991). Personality measures as predictors of job performance: A meta-analytic review. PERSONNEL PSYCHOLOGY, *44*, 703–742.