Race Differences in Personality: An Evaluation of Moderators and Publication Bias

Brian W. Tate

Pennsylvania State University

Michael A. McDaniel

Virginia Commonwealth University

Last revised 8 1 2008

Paper to be presented at the annual conference of the Academy of Management, Anaheim, CA. August, 2008. We would like to thank Robert Davis of Matrix, Inc., IPAT, various public and private organizations and the many individuals who graciously contributed data to the study. Correspondence on this paper may be addressed to Brian Tate at bwt120@psu.edu.

Race Differences in Personality: An Evaluation of Moderators and Publication Bias

Although many have argued that mean racial group differences in scores on personality trait measures are smaller than in scores on cognitive ability tests, there has been limited quantitative evidence to confirm or deny this argument. Using 567 effect sizes and a total sample size of over one million, the present meta-analysis estimates the magnitude of White-Black differences found in scores of Big 5 personality measures as well as measures of locus of control and self-efficacy. We offer seven primary conclusions. First, we conclude that the magnitudes of the White – Black differences are very small. Our second conclusion is that the magnitude of the White – Black differences for the Big 5 is moderated by the cognitive loading of the personality scales. Our third conclusion is that most of our results are primarily consistent with the nil hypothesis which is that there are no differences between Blacks and Whites on four of the seven personality constructs. Blacks are slightly more extroverted and emotionally stable, and Whites are slightly more agreeable. Fourth, we conclude that journal data often suppress results that disfavor Blacks. Fifth, we conclude that Black college students have a slightly more favorable standing on personality variables than White college students. Our sixth conclusion is that mean racial difference in incumbent samples are not smaller than mean racial differences in applicant samples. Finally, we conclude that personnel researchers pay far too little attention to the representativeness of their samples.

Keywords:

personnel selection; selection tests; personality; race; publication bias, cognitive loading.

Race Differences in Personality: An Evaluation of Moderators and Publication Bias

The one standard deviation in scores on paper-and-pencil tests of cognitive ability that separates White from Black test takers is a ubiquitous statistic in personnel selection practice and research (Roth, Bevier, Bobko, Switzer, & Tyler, 2001; Sackett & Wilk, 1994; Schmitt, Clause, & Pulakos, 1996). In practice, this means that, despite evidence for cognitive ability as a consistently valid predictor of job performance (Schmidt & Hunter, 1998), the use of cognitive ability tests is likely to produce substantial group differences between Whites and racial minorities as well as subsequent adverse impact (Outtz, 2002).

Because group differences in cognitive ability test scores cannot be easily reduced (Murray, 2005), practitioners often use tests of constructs other than cognitive ability that might produce smaller group differences in conjunction with cognitive ability tests in applicant selection (Hogan, Hogan, & Roberts, 1996; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Studies typically find smaller group differences in tests of non-cognitive ability constructs than cognitive ability tests (Arthur, Edwards, & Barrett, 2002). For example, Ones and Viswesvaran (1998) found smaller group differences in integrity test scores and Goldstein, Yusko, Braverman, Smith, and Chung (1998) found smaller group differences in tests of managerial competence than in cognitive ability test scores. Among non-cognitive ability constructs, tests of personality constructs are often used. However, systematic, quantitative reviews of racial group differences in scores on tests of non-cognitive ability constructs, such as personality traits, are lacking despite the necessity of examining the benefits and drawbacks of non-cognitive tests. The present meta-analysis will focus on personality and will assess the magnitude of group differences in seven personality constructs.

Personality in Selection

Previous quantitative reviews have shown personality constructs, especially conscientiousness and emotional stability among Big 5 factors, to validly predict job performance criteria, in general (Barrick, Mount, & Judge, 2001; Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). In addition, several authors have argued that personality tests produce substantially smaller group differences than cognitive ability tests (Hogan, 2005; Hogan, Barrett, & Hogan, 2007). For these reasons, personality tests are widely used in selection contexts (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007; Viswesvaran, Deller, & Ones, 2007). Although there are debates as to whether the effect sizes for predicting job performance using personality are high enough to be of practical use and whether personality tests are susceptible to applicant faking (Griffth, 2006; Hogan et al., 2007; Morgeson et al., 2007; Viswesvaran, et al., 2007), the extent of use of personality tests alone warrants examinations of their mean racial differences.

Previous Research on Group Differences in Personality Constructs

Hough, Oswald, and Ployhart (2001) performed the first published review of racial group differences in scores in Big 5 traits. The authors found *d*-scores between Whites and Blacks in agreeableness, conscientiousness, emotional stability, and extraversion ranging from 0.20, for agreeableness, to 0.10, for extraversion. The largest group differences were found in openness, which was associated with a Black-White *d* of 0.21, in favor of Whites, and a Hispanic-White *d* of 0.10, also in favor of Whites. A limitation of Hough et al's. review is that their effect size estimates were based on a sample of only nine studies. Despite this limitation, Hough et al.'s results suggest that there is the potential for significant group differences in certain personality constructs, such as openness and extraversion.

Foldes, Duehr, and Ones (in press) recently completed a review of racial mean score differences in personality tests. The present study builds on Foldes et al.'s paper in three ways. First, the present study will base its effect size estimates on more data than Hough et al.'s and Foldes et al.'s estimates. For example Foldes et al. analyzed data from three samples to determine White-Black differences in global conscientiousness measure scores. We present analyses on 81 such effects. Foldes et al. also had very few samples for openness to experience (k = 9), and agreeableness (k = 9). We will analyze 78 effect sizes for openness and 73 for agreeableness.

Second, using a greater number of studies will permit us to examine more moderators. Our moderator analyses include sample type (e.g., incumbents, applicants, college students) and publication source (e.g., journals, unpublished data). We also will examine the cognitive loading of personality measures as a moderator of the magnitude of mean racial differences. We operationalize the cognitive loading of a personality scale as its correlation with a measure of cognitive ability. Whetzel, McDaniel, and Nguyen (in press) found that the cognitive loading of situational judgment tests accounted for the majority of the differences across studies in mean race differences. We anticipate that more cognitively-loaded personality tests and constructs will show the largest mean racial differences.

Third, we examine the extent to which publication bias (McDaniel, Rothstein & Whetzel, 2006) may distort the accumulated data on mean racial differences. Because the reporting of mean racial differences is a sensitive topic, it is possible that researchers who find non-trivial mean racial differences will choose not to report that data in their publications; whereas, those who find near zero differences will be more likely to report such results. The existence of

publication bias would suggest that the magnitude estimates of the mean race differences in personality measure scores are biased.

Predicting the Nature of White-Black Score Differences

Predicting the direction of group differences in personality measure scores is difficult due to existing research on the topic being limited. The two previous reviews, described above, have had few effect sizes for most distributions. The two notable exceptions are the Foldes et al.'s (in press) analyses of emotional stability (k = 128) and extraversion (k = 28), which showed Whites to score higher on the traits. Because personality tests show small to modest correlations with job performance (Hurtz & Donovan, 2000) and because most job performance measures show White-Black differences favoring Whites (Ford, Kraiger, & Schechtman, 1986; Hauenstein, Sinclair, Robson, Quintella, & Donovan, 2003; Kraiger & Ford, 1985; McKay & McDaniel, 2006; Roth, Huffcutt, & Bobko, 2003), one might anticipate White-Black differences in personality measure scores to favor Whites. Given this reasoning and Foldes et al.'s results for emotional stability and extraversion, we offer this hypothesis:

Hypothesis 1: Whites will score higher than Blacks on the Big 5, locus of control and self-efficacy.

Moderators

The present study will examine three moderators of the magnitude of group differences in scores on Big 5 traits, including sample type (e.g., applicant, incumbent, or student), publication source (e.g., journals, conference presentations), and the cognitive loading of a personality measure.

Cognitive loading. Regarding a personality measure's cognitive loading (i.e., the correlation between a personality scale and a measure of cognitive ability), we predict that, because of the

group differences associated with cognitive ability, as the relationship between a personality trait and cognitive ability increases, White-Black score differences favoring Whites will increase.

Hypothesis 2: As the correlation between personality traits and cognitive ability increases, group differences in personality trait scores favoring Whites over Blacks will increase.

Sample type. Variance on predictor variables is typically smaller in incumbent samples than in applicant samples due to the screening of the applicants. This restricted variance in incumbent samples would be expected to result in smaller mean racial difference. For example, Roth et al. (2001) showed that White-Black mean differences in incumbent samples are smaller than in the general population.

Hypothesis 3: Mean racial differences will be smaller in incumbent samples than in applicant samples.

Publication Bias

The present study will also examine whether there is evidence for publication bias in the reporting of race differences in personality traits. According to McDaniel, Rothstein, and Whetzel (2006), publication bias exists when unpublished sources of data systematically differ from published sources of data. The existence of publication bias threatens the conclusions that can be drawn from meta-analytic reviews. For example, a meta-analysis by McDaniel, Whetzel, Schmidt, and Maurer (1994) found an effect size of 0.27 for the validity of structured interviews compared to 0.19 for the validity of unstructured interviews. Duval (2005) estimated that without publication bias, structured interviews would have a lower validity of 0.21. McDaniel, et al. (2006) stated that practitioners relied on the 0.27 validity effect size as a rationale for developing structured interviews, and many researchers no longer compared structured to unstructured

interviews after McDaniel, et al.'s (1994) meta-analysis. Thus, the existence of publication bias suggests that the conclusions drawn from the 1994 meta-analysis may have been misleading for both practitioners and researchers.

Typically, according to Dickerson (2005), studies that fail to find significant results are "suppressed" in favor of studies that report significant results. The preference for the publication of statistically significant results is often due to editorial preferences and actions of authors. Editors and authors often consider statistically significant results to be more interesting than results that do not reach statistical significance. Journals have limited space and may give preference to the more interesting articles. Authors tailor their publications to editorial preferences.

Publication bias can also be a function of intentional distortion. McDaniel, Rothstein and Whetzel (2006) presented evidence suggesting that some test publishers may intentionally distort their validity effect sizes, such that small validity coefficients were suppressed. The suppression of those effect sizes likely made the test vendors' products look more useful than they were. Also, pharmaceutical companies have been accused of suppressing evidence that makes their products look ineffective (Curfman, Morrissey & Drazen, 2006). Similarly, McKay and McDaniel (2006) found published studies to yield lower magnitude differences between Whites and Blacks in job performance relative to unpublished studies, in which larger magnitude effect sizes favoring Whites were found. One interpretation of these findings is that authors have tended to report small mean racial effect sizes but not large effect sizes.

We assert that a similar distortion is likely to occur in mean differences in personality scores, as large mean racial differences, particularly those that favor Whites, will likely be suppressed in journal articles. We believe that this effect will hold for personality dimensions that have a clear

and favorable end. For the constructs in our data set, we argue that the favorable end of the personality dimension is clear for conscientiousness, agreeableness, emotional stability, locus of control and self-efficacy. For example, most everyone would agree that it is better to be conscientious than slothful. The favorable end of extraversion and openness is less clear. Authors who perceive the quieter, more reflective nature of individuals high in introversion in a positive light may perceive the ultra-gregariousness of those high in extroversion to be a negative personal characteristic, and vice-versa. Similarly, authors who value curiosity may hold high levels of openness (intellectance) as a virtue; whereas, others may consider high levels of openness to be a distraction from task work. Thus, we believe that intentional publication bias is likely for five of the seven personality constructs, excluding extroversion and openness to experience. Thus, we offer:

Hypothesis 4: Group differences in scores on personality traits, excluding extraversion and openness to experience, will tend to show publication bias in data obtained from journals and tend not to show bias from data obtained from unpublished sources.

Method

Literature Search

The literature search included several phases. In the first phase, we collected studies included in Hough et al.'s (2001) review of racial group differences in personality traits. Our literature search also benefited from the reference list of the Foldes et al. paper. All efforts were made to attain the unpublished data sources reported in Foldes et al, but at the time of publication, we had not received any data from Dr. Foldes, although we obtained some of her data by contacting her sources directly. Dr. Foldes cited confidentiality issues in not releasing her unpublished data to

us. We understand and respect her decision. In the second phase of the search, we used PsycINFO, pairing the search terms of "race" and "personality." Third, we used PsycINFO to search through all published volumes of relevant journals for relevant studies. We also performed a search using *Google Scholar* to search for relevant published and unpublished papers. To obtain data from unpublished sources, we obtained technical reports and unpublished data by contacting individuals from several organizations. We located several journal articles from 2002 and later that reported personality data but did not report racial statistics. For such articles, we requested that the authors send us summary statistics on the personality data by race.

To be included in the meta-analysis, studies had to include at least one test of a Big 5 scale, a locus of control scale, or a self-efficacy scale along with statistics that could be used to calculate a standardized mean difference effect size for Black-White score differences. Typically the standardized mean differences were calculated from means, standard deviations, and sample sizes. Authors of papers were contacted if a paper included a relevant personality test, had likely recorded race data on the respondents, and was published during or after 2002 but did not include a statistic that could be used to calculate an effect size for mean racial differences. A final set of 86 sources with 567 independent samples was obtained. Note that the number of samples is equal to sum of the samples across the seven construct analyses. If a sample contributed data to more than one construct, which was often the case, the sample was counted more than once, but the analyses of any given construct were based on independent samples. Analyses were based on 1,077,920 individuals. This sum of individuals equals the sum across samples for the seven constructs. Members of samples were counted multiple times in cases where a sample contributed data to more than one construct.

Personality measures were coded according to Hough and Ones's (2001) FFM taxonomy. Hough and Ones mapped a variety of personality measures commonly used in industrial/organizational psychology and organizational behavior research onto the FFM. The taxonomy identifies global measures of each factor as well as facets of each FFM factor. For instance, the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) contains a neuroticism scale that, if reversed coded, represents a global emotional stability scale, according to Hough and Ones. Likewise, the Jackson Personality Inventory's (Jackson, 1974) calmness subscale is an indicator of the emotional stability facet, even-tempered. Using this taxonomy allows for uniformity in selection and coding and provides greater ability to compare the current meta-analysis with findings from the Foldes et al. meta-analysis.

Despite the comprehensiveness of the Hough and Ones (2001) typology, several sources included measures not included in the typology but were believed to assess one of the Big Five traits by present study's authors. For cases in which an unlisted measure believed to assess the FFM was identified by a coder, the study was flagged and a decision was made to include or exclude the study. Decisions to include such a measure were made if it was reasoned that the measure was not included in the Hough and Ones typology due to the measure's recent development or relative obscurity. In other cases, the author specifically stated that the measure they included was a measure of the FFM. Total agreement among the authors was needed for a study to be included in the meta-analysis.

Coding Procedures

Data from multiple tests of personality were grouped into categories of Big 5 traits consistent with the taxonomy offered by Hough and Ones (2001). Only analyses of the global personality scales for the Big 5 are included in the present study. Due to paper length considerations, we

restricted the analyses to White-Black differences. For cases in which we received data on Big 5 measures not included the Hough and Ones taxonomy (such as proprietary measures used by consulting firms), we assigned the measures to the Big 5 based on information provided by the source of the data. We also coded data on locus of control and self-efficacy.

Regarding the coding of moderators, the most common sample types were applicants, incumbents, and college students. Data was most commonly obtained through personal communication with researchers and practitioners. Many organizations provided data that do not appear in any publicly available source. Data were also obtained from published articles and through personal communication with study authors, for studies that did not report codable statistics. Sources that involved personal communication with article authors in order to obtain personality statistics by race will be referred to as "journal supplemented" articles. A comparison of the journal and journal supplemented results can be useful in determining whether missing racial data are randomly missing, in which case the mean differences should be about the same, or whether data are missing as a function of the magnitude of the racial differences, in which case publication bias likely exists in the journal data.

In many cases, multiple measures of either a global factor or factor facet were administered to the same sample. This issue was especially salient with the Minnesota Multiphasic Personality Inventory (Dahlstrom & Welsh, 1960), which contains thirteen subscales identified by Hough and Ones (2001) that assess emotional stability. The inclusion of multiple effect sizes from the same sample violates the independence of data assumption. In order to maintain data independence, we first identified all instances where two or more effect sizes were present and then averaged the effect sizes to form one composite difference score. Although this does maintain independence, it does distort the sampling error estimates to some degree. We treated

each measure identified in Hough and Ones (2001) as an equivalent indicator of the construct it measured. For instance, the Hogan Personality Inventory intellectance scale (Hogan & Hogan, 1995) was treated as an equally valid measure of openness as the Goldberg Five Factor Intellect scale.

Meta-Analysis Procedure

We conducted the meta-analysis using the Comprehensive Meta-Analysis (CMA) 2.0 software (Borenstein, Hedges, Higgins, & Rothstein, 2005), which follows procedures outlined by Hedges and Olkin (1985).

Publication bias analyses

Foldes et al. (in press) argued that because each study included in their meta-analysis reported both the White mean and the Black mean that publication bias cannot exist. The rationale for this assertion is unclear. If means are reported that show no effect or favor Blacks and not reported when they favor the majority group, then only studies in the former category would be included in the meta-analysis. If the above were true, the inclusion of only studies that report effect size data such as the means would result in biased effect sizes.

One method for identifying and adjusting for publication bias is Duval and Tweedie's (2000) trim and fill technique. This is one of two methods we used in our publication bias analyses. This method plots each study based upon its difference score on the x-axis and its precision (1/sampling error variance) on the y-axis. Once plotted, asymmetry in the distribution of effect sizes indicates publication bias. Publication bias generally occurs in the lower part of the distribution to the left or right. This translates to the tendency for underpowered studies with findings not typical or not in line with the *zeitgeist* of the research community to be published less often. The tendency for these types of studies not to be published results in either an

inflation (studies underpublished to the left of the mean) or attenuation (studies underpublished to the right of the mean) of the observed population effect size. If bias exists, Duval and Tweedie's (2000) method identifies bias and then imputes the missing studies and shifts the overall observed effect size accordingly. Figure 1a shows an asymmetric funnel plot where small sample size, small effect sizes have been suppressed. Figure 1b shows the same funnel plot with imputed effect sizes (the dark circles). Note that mean has shifted to the left (.20 moved to .12). A more detailed explanation and example of publication bias is available in McDaniel et al. (2006).

The second publication bias analysis used in this study is cumulative meta-analysis. In a cumulative met-analysis, studies are sorted by a variable of interest, often time. One then conducts iterative meta-analyses adding one additional effect size for each meta-analysis. The first mean reported is the effect size from the first study. The second mean is the mean from the meta-analysis of the first and second study. The third mean is the mean of the meta-analysis of the first three studies, and so on. Historically, cumulative meta-analysis has been used to determine the time point at which a result stabilizes.

One of the most prominent examples of cumulative meta-analysis involved the streptokinase (a blood thinner) treatment of myocardial infarction (Lau, Schmid, & Chalmers, 1995). In this analysis, the studies were sorted by time of publication and the meta-analysis was iteratively conducted each time adding in one effect size. Lau et al. (1995) found that although randomized clinical trials continued until 1989, the streptokinase treatment could have been deemed an effective treatment as early as 1973. The sixteen year gap between when the drug should have been implemented as a standard therapy and when the last clinical trial was completed likely

caused unnecessary deaths. In another application, Leimu and Korichevu (2004) used cumulative meta-analysis to identify temporal trends and publication bias in the field of ecology.

When cumulative meta-analysis is used as a publication bias method, studies are sorted by standard error from low to high (or alternatively by sample size from high to low). Low standard error studies are those with the largest samples size. The means from the meta-analyses are called cumulative means and these means can be examined and plotted for evidence of drift as more studies are added to the meta-analysis. The meta-analytic means from the early studies are the estimates of the population mean from the larger samples. The meta-analytic means added in later stages of the meta-analysis are from the addition of the smaller samples to a distribution of the larger samples. If small sample size studies with small effects are being suppressed (a common publication bias scenario), the cumulative means will drift in a positive direction as the smaller sample size studies are added to the cumulative meta-analysis. This occurs because the small sample sizes available to the analyst have systematically larger magnitude effects than the largee sample studies available to the analyst. Using the same data, as in Figure 1a, Figure 1c shows an illustrative cumulative meta-analysis where effect sizes have been sorted by their standard error from low to high. Note that the small standard error (large same sizes) cumulative means at the top of the graph have a magnitude of about .10. However, as small sample studies are added, the cumulative mean shifts closer to .20. This suggests that the small samples have larger effect sizes than the large sample samples consistent with a conclusion of publication bias.

The Figures 1a, 1b, and 1c show the same pattern of publication bias. Figures 1 and 2 are based on the trim and fill approach and incorporate methods and assumptions inherent in that method. Figure 3 is based on a different method, cumulative meta-analysis. One can have greater

confidence in conclusion about publication bias when the methods agree as they do in this illustrative example.

Vector correlations

To assess Hypothesis 2, we calculated the correlation between two vectors. The first vector contains the standardized mean racial differences on a personality scale. The second vector contains the correlation between the personality scale and a measure of cognitive ability. To contribute data to this vector correlation, the sample must have data for both vectors. An example of the use of vector correlation in a similar application to this paper can be found in Whetzel, McDaniel and Nguyen (in press).Vector correlations are discussed in Jensen (1998). Hunter and Schmidt (2004, pages 294-295) use "study characteristic correlations" to describe vector correlations. Vector correlation analysis is similar to the use of a single predictor in a meta-regression (Thompson & Higgins, 2002). The vector correlations were weighted by the precision of the standardized mean difference. In the meta-analysis literature, precision is defined as the inverse of the sampling error variance and is highly correlated with sample size.

Results

White – Black Mean Differences in Conscientiousness

Based on 81 effect sizes, and a total sample size of 193,445, we obtained a mean White-Black *d* of -.02 (see Table 1). Throughout the paper, all confidence intervals are 95% confidence intervals. Based on three effect sizes and a total sample size of 21,001, Foldes, et al. (in press) obtained a mean *d* of -.17. These mean differences slightly favor Blacks. Our analyses show some variations across data source with a mean *d* of -.07 (k = 15) from journal article sources, a mean *d* of .00 (k = 7) from journal article sources supplemented with personal communication, and a mean *d* of -.01 (k = 49) from unpublished data sources.

Insert Table 1 about here

The pattern of the mean *d* by publication source suggests that journal publications, consistent with Hypothesis 4, might tend to suppress mean race data when the direction of the effect favors Whites (see Figure 2). We detail these publication bias effects in the discussion.

Regarding sample type, both incumbents and applicants have a mean *d* of .00. Among college student samples, there is a larger *d* (-.12), favoring Blacks. All three distributions showed signs of publication bias. The trim and fill analysis of the applicant data suggested that five positive direction effects are needed to bring symmetry to the distribution of applicant data which moved the mean from .00 to .04. For incumbents, the trim and fill analysis suggested that nine positive direction effects were needed to bring the distribution into symmetry and moved the mean from .00 to .25. For college respondents, the trim and fill analysis suggested that four negative direction effect sizes were need to be imputed to bring the distribution into symmetry moving the mean from -.12 to -.18. When considering the publication bias analyses, we would conclude that applicant populations show no meaningful race differences in conscientiousness (*d* = .04) but that incumbent samples show mean differences of non-trivial magnitude (*d* = .25) favoring Whites and college student samples show mean differences favoring Blacks (*d* = -.18).

We also crossed publication source by sample type and conducted analyses for those cells with five or more effect sizes. Effect sizes ranged from -.19 (journal data with college students) to .15 (journal data with incumbents). Most of these distributions showed no evidence of publication bias or were too small to conduct a credible publication bias analysis. However, publication bias was found in the incumbent data from personal communication sources. The cumulative meta-analysis showed the cumulative mean drifted from high to low as studies with

larger standard errors were added to the analysis. The trim and fill data suggested that five positive studies needed to be imputed to bring the distribution into symmetry and the mean moved from .01 to .26. This finding is consistent with the publication bias found in the distribution of all incumbents regardless of data source. The finding is inconsistent though with the lack of publication bias in the distribution of all personal communication studies regardless of sample type. The incumbent data from personal communication sources represented only 2,556 observations and 13 effect sizes of the 134,739 observations and the 49 effect sizes from the distribution of personal communication effect sizes regardless of sample type. Given that the incumbent data was such a small subset of the personal communication data, the publication bias in the incumbent data did not reveal itself in the analysis of all personal communication effect sizes.

The vector correlations between the vector of *d* effect sizes and the vector of cognitive ability correlations with conscientiousness (i.e., the cognitive loading of the measure) indicate that the magnitude of the mean White-Black differences varies with the cognitive loading of the personality measure in the sample, consistent with Hypothesis 2. For the 81 samples with effect sizes, 38 had cognitive loadings resulting in a vector correlation of .43. This vector correlation indicates that larger magnitude mean racial differences favoring Whites are more likely to be found to the extent that the personality measure is positively correlated with cognitive ability.

White – Black Mean Differences in Agreeableness

Based on 73 effect sizes, and a total sample size of 202,211, we obtained a mean *d* of .09 (see Table 2). Based on nine effect sizes and a total sample size of 3,297, Foldes et al. (in press) obtained a mean *d* of .03 although it not clear from their paper whether that distribution included only global measures of agreeableness (it might include effect sizes for facets of agreeableness).

These mean differences slightly favor Whites. Our analyses show some variations across data source with the mean *d* from journal articles of .10 (*k* = 14). The mean *d* from journal articles where the race data was not in the article but was obtained from the author was -.17 (*k* = 6) and the mean *d* from unpublished data sources was .12 (*k* = 42).

Insert Table 2 about here

The pattern of the mean *d* by publication source is different from that for conscientiousness in that for conscientiousness effect sizes obtained from journals were smaller than effect sizes from journals where the race data were obtained from the authors (journal supplemented). For the agreeableness data, the mean effect size from journals was positive (.10) but for the journal supplemented data the mean effects size was negative (-.17). However, we did find evidence of publication bias for both the journals and journal supplemented data, both indicating that positive direction effects (i.e., effect sizes disfavoring Blacks), may be suppressed (see Figure 3). For both categories of journal data, the cumulative mean drifted from right to left with the addition of larger standard error (smaller sample size) studies suggesting publication bias that makes the mean effect favoring Whites smaller. The trim and fill analyses also indicated that higher magnitude positive effects may be suppressed. For the journal effect sizes, five positive effects needed to be imputed to bring the effect sizes into symmetry and moved the mean from .10 to .27. For the journal supplemented effect sizes, three positive effects needed to be imputed to bring the distribution into symmetry and moved the mean from -.17 to .08. There was no evidence of publication bias in the effect sizes from personal communication.

By type of respondent, incumbents and applicants have similar means (.14 and .12) favoring Whites while results based on college students is smaller (-.02) favoring Blacks. These

three sample type distributions were examined for publication bias. The application data showed some drift from right to left as larger standard error studies were added but trim and fill identified no evidence of suppression. The trim and fill analysis of the college student data indicated that six negative effect sizes would need to be imputed to bring the distribution into symmetry and the mean moved from -.02 to -.16. There was little evidence of publication bias for the incumbent data.

We also crossed publication source by sample type. Effect sizes ranged from .00 (journal data with college students and personal communication with college students) to .23 (personal communication with incumbents). Most of these distributions showed no evidence of publication bias or the distributions were too small to conduct a credible publication bias analysis. Although publication bias was found in the incumbent data from personal communication sources, these incumbent data represented a small subset of the personal communication data. Therefore, the publication bias in the incumbent data did not reveal itself in the analysis of all personal communication effect sizes.

The vector correlations between the vector of *d* effect sizes and the vector of cognitive ability correlations with agreeableness (i.e., the cognitive loading of the measure) indicate that the magnitude of the mean White-Black differences in agreeableness vary with the cognitive loading of the personality measure in the sample, consistent with Hypothesis 2. For the 73 samples with effect sizes, 32 had cognitive loadings resulting in a vector correlation of .11.

White – Black Mean Differences in Emotional Stability

Based on 140 effect sizes, and a total sample size of 168,898, we obtained a mean *d* of -.06 (see Table 3). Based on 143 effect sizes and a total sample size of 151,523, Foldes et al. (in press) obtained a mean *d* of .09. Though the Foldes et al. mean difference slightly favors Whites, our

mean difference slightly favors Blacks. Our analyses show some variations across data sources. There was a mean *d* from journal articles of -.06 (*k* = 25). The mean *d* from journal articles whose authors had to be contacted for race data was .00 (*k* = 6), and the mean *d* from unpublished data sources was -.05 (*k* = 100).

Insert Table 3 about here

The pattern of the mean *d*'s by publication source suggests that, consistent with Hypothesis 4, journal publications might tend to suppress test score statistics that favors Whites and disfavors Blacks. Specifically, journal articles that include race data have mean differences favoring Blacks (-.06); whereas, journal articles that do not report race analyses (the journal supplemented distribution) have an effect size of zero.

Figure 4 provides the cumulative meta-analyses and trim and fill analyses shed further light on this apparent publication bias effect. For journal articles that report race data, the cumulative meta-analysis indicates that the cumulative mean shifts from positive to negative with the addition of effect sizes with larger standard errors (i.e., effect sizes with smaller sampler sizes). The trim and fill analysis suggests that eight additional studies of positive direction are needed to bring symmetry to the distribution and changes the mean *d* from -.06 to .10. For journal articles that did not report race data, the cumulative meta-analysis indicates that the cumulative mean shifts from negative to near zero with the addition of larger standard error studies. However, the trim and fill analysis suggests that the distribution is symmetrical. We suggest that the trim and fill algorithm was not robust to the odd shape of the distribution and the limited number of effect sizes reduced the power of the trim and fill statistical tests. Our results are consistent with the conclusion that statistics favoring Whites and disfavoring Blacks are less

likely to be reported in published articles than statistics favoring Blacks and disfavoring Whites. Unexpectedly, we also found publication bias in the distribution 100 effect sizes obtained from personal communication sources. The cumulative meta-analysis (not displayed because of its ungainly size) shows a drift in the cumulative mean toward negative effect sizes with the addition larger standard error (smaller sampler sizes). The trim and fill analysis imputed 33 effect sizes to bring the distribution into symmetry although the mean only moved from -.05 to -.01.

Analyses were also conduct by type of sample. The mean *d* for applicants (-.04, slightly favoring Blacks) was near the mean *d* for incumbents (.01, slightly favoring Whites). The college student samples had a mean *d* of -.16, favoring Blacks. We also had six effect sizes for samples described as general population. General population samples include those that were not entirely student, applicant, or incumbent and did not fall into any other potential sample type category. The general population samples had a mean *d* of -.09, favoring Blacks. There was evidence of publication bias in the applicant data. The trim and fill analyses identified 30 studies that would need to be imputed on the right of the distribution and this moved the mean from -.04 to .01. There was also evidence of publication bias in the general population sample three effect sizes were imputed and the mean shifted from -.09 to .04.

For our moderator analyses, we also crossed publication source by sample type and conducted analyses for those cells with five or more effect sizes. Effect sizes ranged from -.17 (journal data with college students) to .05 (journal data with incumbents). The distribution of personal communication data using applicants accounts for 80 of the 85 effect sizes for personal communication data. Not surprisingly, there was similar publication bias found in data obtained

through personal communication and involving applications with 30 effect sizes being imputed and the mean moving from -.05 to .01.

The vector correlation analyses were consistent with the analyses for conscientiousness and agreeableness by indicating the cognitive loading of the personality test covaries with the magnitude of the *d*. As the cognitive loading of emotional stability increases, so does the magnitude of the effect size favoring Whites.

White – Black Mean Differences in Extroversion

Based on 117 effect sizes, and a total sample size of 57,820, we obtained a mean *d* of -.18 (see Table 4). Based on 55 effect sizes and a total sample size of 109,922, Foldes et al. (in press) obtained a mean *d* of -.16. Both mean racial differences are similar and favor Blacks. Our analyses show some variations across data source with the mean *d* from journal articles of -.05 (*k* = 20). The mean *d* from journal articles where race data was not in the article but was obtained from the author was .02 (*k* = 5) and the mean *d* from unpublished data sources was -.30 (*k* = 84).

Insert Table 4 about here

The pattern of the mean *d* by publication source suggests that journal publications, consistent with Hypothesis 4, might tend to suppress mean race data when the direction of the extroversion effect disfavors Blacks. Specifically, the journal articles that do include race data have mean differences favoring Blacks (-.05) while the journal articles that do not report race analyses (the journal supplemented distribution) have an effect size of .02. Figure 5 provides the cumulative meta-analyses and trim and fill analyses shed further light on this apparent publication bias effect. For journal articles that report race data, the cumulative meta-analysis indicates that the cumulative mean shifts from negative to positive with the addition of effect sizes with larger

standard errors (i.e., effect sizes with smaller sampler sizes). The trim and fill analysis suggests that three additional studies of negative direction are needed to bring symmetry to the distribution and changes the mean *d* from -.05 to -.10. For journal articles that did not report race data, the cumulative meta-analysis indicates that the cumulative mean shifts from positive to negative with the addition of larger standard error studies with the mean moving from .02 to .07 with the addition of one imputed study. Our results are consistent with the conclusion that when mean differences show Blacks to be more introverted relative to Whites, the results tend to be reported in journal articles but when they show Blacks to be more extroverted the results tend not to be reported. Unexpectedly, we also found publication bias in the distribution of 84 effect sizes obtained from personal communication sources. The cumulative meta-analysis (not displayed because of its ungainly size) shows a drift in the cumulative mean toward from a more negative mean to a less negative mean with the addition larger standard error (smaller sampler sizes). The trim and fill analysis imputed 30 effect sizes to bring the distribution into symmetry and the moved substantially from -.30 to -.02.

Analyses by sample type yielded a mean estimate of -.38 for applicants (Blacks are more extroverted), -.03 for incumbents (Blacks are slightly more extroverted), and .17 for college students (Blacks are more introverted). The interpretation of these analyses is complicated by publication bias in all three distributions. For applicants, the trim and fill analysis added 28 effect sizes to the right of the distribution moving the mean from -.38 to -.10. Trim and fill also imputed seven effect sizes to the right of the mean for the incumbents moving the mean from -.03 to .07. Finally, for the college students, one effect size to the left of the mean was imputed moving the mean from .14 to .09.

Analyses that crossed publication source with sample types, yielded mean racial differences ranging from -0.41 (personal communication data with applicant samples) to .25 (personal communication data with college students). The distributions with more than 10 effects sizes were examined for publication bias. Of the 84 personal communication samples, 65 were for applicants. Because the personal communication samples showed substantial publication bias, it is not surprising that the personal communication data with applicants also showed substantial publication bias. Trim and fill imputed 29 effects to the right of the mean to bring to move the mean from -.41 to -.06. The personal communication data with incumbents also showed publication bias with four effect sizes imputed to the right of the mean moving the mean from .02 to .19.

White – Black Mean Differences in Openness to Experience

Based on 78 effect sizes, and a total sample size of 148,097, we obtained a mean *d* of .02 (see Table 5). Based on 9 effect sizes and a total sample size of 24,957, Foldes et al. (in press) obtained a mean *d* of .10. Both mean racial differences are small and favor Whites. Our analyses show very little variation in effect sizes across data sources. There was a mean *d* from journal articles of .07 (k = 12). The mean *d* from journal articles whose authors had to be contacted for codable data was -.02 (k = 5), and the mean *d* from unpublished data sources was .02 (k = 53).

Data obtained from journal articles suffered from publication bias (see Figure 6). Note that the cumulative means shift from form the right to the left with the addition of four large standard error effects. Trim and fill imputed four effects to the left of the mean and the mean moved from .07 to -.18. There was no apparent publication bias in the journal supplemented data. The personal communication data cumulative analysis (graphic not provided due to its size)

showed a drift from right to left with the additional of larger standard error samples and the mean moved from .02 to .07.

Insert Table 5 about here

Analyses by sample type showed little difference by sample types with applicants having a mean *d* of .01, incumbents having a mean *d* of .00 and college students having a mean *d* of .10. Each of these distributions showed evidence of publication bias, with each moving the mean in a more positive direction, favoring Whites. For incumbents, trim and fill imputed four studies and moved the mean from .06 to .00.

Analyses that crossed publication source with sample type, yielded mean racial differences ranging from -.06 (personal communication data with incumbents) to .10 (journal data with college students). Two distributions showed small amounts of publication bias. Applicants accounted for 34 of the 53 personal communication samples and also showed a small publication effect, with four imputed studies and the mean moving from .02 to .08. Effect sizes for incumbents from personal communication sources required one sample to be imputed to bring symmetry to the distribution and moved the mean from -.06 to -.03.

White – Black Mean Differences in Locus of Control

The locus of control data are primarily from one source, a consulting firm. The data from that source were based on applicants. We could not compare our results with the Foldes et al. results because that study did not report findings for locus of control. We could not examine the cognitive loading of the locus of control measure as a moderator of the mean difference because only one sample reported a correlation between locus of control and a cognitive measure.

The locus of control data distribution comparing White and Black differences consisted of 16 effect sizes (see Table 6). One was from a conference paper and was based on incumbent data. Another was from a journal and based on a community sample of older adults. The remaining 14 were unpublished data from one consulting firm and were based on applicant data. For the full distribution of data, the mean *d* was .05, indicating that Whites had slightly greater internal locus of control, on average. One of the 14 effect sizes from the consulting firm was an outlier (*d* = -.82). With that effect size was removed, the mean *d* was .10. There was no evidence of publication bias. We conclude that the magnitude of mean racial differences in locus of control is very small and to the extent there is a difference, it favors Whites. Because the distribution is based on only 16 effect sizes and because 14 of the 16 are from one source, we encourage additional analyses to be completed.

White – Black Mean Differences in Self-Efficacy

All of the self-efficacy effect sizes came from a single human resource consulting firm and all of the respondents were applicants. Based on 37 samples and 236,906 individuals, the mean *d* was .02 (Table 7). One sample was very large (*N* = 42,364) and the *d* was an extreme outlier (*d* = -1.52). When that outlier was removed the mean *d* moved to -.03 and the confidence interval shrunk dramatically. We suspect but cannot verify that the outlier is an erroneous data point and assert that the -.03 is the best estimate. This distribution shows no publication bias.

Discussion

This meta-analysis has sought to summarize the White – Black mean racial differences in the Big 5, locus of control and self-efficacy. We draw seven primary conclusions. The first and primary conclusion drawn from our results are that the magnitude of the White – Black differences are very small. For applied purposes, such as whether a personality test will likely

result in adverse impact in employment hiring decisions, this primary conclusion may be sufficient. A second primary conclusion is that the magnitude of the White – Black differences for the Big 5 is moderated by the cognitive loading of the personality scales. As the cognitive loading of the personality scale increases (i.e., as the correlation between the scale and cognitive ability increases), the mean racial difference moves in the direction favoring Whites. These results support Hypothesis 2 and we will discuss the reasons why this tends to occur.

The third primary conclusion is that our results are primarily consistent with the nil hypothesis which is that there are no differences between Blacks and Whites on four of the seven personality constructs. We will discuss why the nil hypothesis is reasonable.

Our fourth primary conclusion is that data on mean racial differences show publication bias effects for data from journal sources but not from unpublished data sources. We will discuss how these biases are primarily in the direction of suppressing personality differences that are favorable to Whites (or alternatively suppressing personality differences that are disfavorable to Blacks). We will detail these differences and offer that the journal results reflect norms against discussing race differences.

Our fifth and sixth conclusions relate to sample type moderators. Our fifth primary conclusion is that Black college students generally score more favorably on personality scales than Whites. We will offer several post-hoc explanations for these effects. Our sixth primary conclusion is that incumbents do not show smaller mean racial differences than applicants. This is an expected finding that is inconsistent with much personnel literature and we will discuss various post-hoc explanations.

Our seventh primary conclusion is that researchers in employment selection research need to pay much more attention to the representativeness of their samples and how departures from

representativeness affect their conclusions. The authors find it very odd that our discipline devotes substantial effort to precision in measurement and devotes very little attention to sampling.

Conclusion 1: The magnitude of the White – Black differences are very small.

Tables 1 through 7 summarize means differences for seven personality constructs. The results are presented for all available data and are analyzed separately by publication source and sample type. When evidence suggests publication bias, we offer additional analyses that estimate what the mean effect size might be in the absence of publication bias. Very few of these effect sizes are above .2 standard deviations. To bring perspective to this magnitude, a standardized mean difference of .2 converts to a correlation coefficient of about .1. If one considers a correlation of .1 to be quite small, as we suspect most readers will, a *d* of .2 should also be considered quite small. Of the 93 mean effect sizes reported in Tables 1 through 7, only nine have absolute magnitudes greater than .2. Of these nine effect sizes, four are estimates from trim and fill imputed distributions and there is not a consensus that the trim and fill adjusted means are accurate estimates of the population effect.

Conclusion 2: The magnitude of the White – Black differences in the Big 5 is moderated by the cognitive loading of the personality scales.

The majority of the vector correlations between mean racial differences and the cognitive loading of the personality scales are zero. It is important to note that these correlations are calculated separately by construct. Thus, the .43 vector correlation for conscientiousness indicates that across samples the larger the correlation between the conscientiousness scale and cognitive ability the larger the mean racial differences favoring Whites.

Whetzel, McDaniel and Nguyen (in press) found a positive vector correlation between cognitive loading and the magnitude of the mean racial differences in situational judgment tests. Situational judgment tests tend to be construct heterogeneous and a sizable portion of their variance is associated with cognitive ability. Thus, Whetzel et al. argued compelling that larger mean racial differences in SJTs were due to the SJTs having greater cognitive variance. This argument could be made for the vector correlations for personality variables but it is less compelling for two reasons. First, unlike situational judgment tests, personality scales tend to be homogeneous, and second, little of their variance is associated with cognitive ability.

Although it is likely that some measures of a personality construct are more correlated with cognitive ability than others, this might not be the best explanation for the vector correlations in this study. An additional explanation for the vector correlations concerns differences in range variation due to variations in the efficiency of the screening procedures. If an employer or a university has an inefficient hiring system for screening applicants on cognitive ability and conscientiousness, those who are accepted will have substantial variance in both cognitive ability and conscientiousness. This variance would permit the vector correlation to be larger to the extent that there is a positive vector correlation in the population. Also, if the employer or university seeks to advance racial diversity by using lower standards for Blacks than for Whites, the magnitude of the vector correlations should increase.

Conclusion 3: Most results are primarily consistent with the nil hypothesis, that there are no differences between Blacks and Whites on the personality constructs.

Estimating the direction of near zero effects is very difficult. There are three primary reasons for this difficulty. First, when mean differences are near zero, their confidence intervals usually range from negative values to positive values. For example, for all 81 conscientiousness effect

sizes, the mean *d* for conscientiousness is -.02, and the confidence interval ranges from -.08 to .03. With such results, it is difficult to make a compelling conclusion that one group is more conscientious than the other. Second, our data show some differences by publication source and type of respondents. For example, in Table 1, the conscientiousness mean differences vary from positive to negative values depending on which moderator subgroup one examines. Third, the effect size distributions often show evidence of publication bias, and while the publication effects are typically small, they sometimes change the direction of effect from favoring one group to favoring another. For example, the journal data for conscientiousness yield a mean of -.07 favoring Blacks. The estimate of the mean in the absence of the suppression (i.e., the trim and fill adjusted mean) is .04, favoring Whites. With these results, one could argue that the mean racial difference in conscientiousness is small, but it is difficult to make a compelling statement concerning which group has an advantage in conscientiousness. Our Hypothesis 1 argues that Whites will score higher than Blacks on all personality scales. Although one could cite specific sub-distributions favoring Whites or favoring Blacks, the results supporting one group scoring more favorably than another will usually not be compelling. Thus, our results do not provide compelling support for Hypothesis 1.

The nil hypothesis is that there is a zero difference between groups. We acknowledge that a zero relationship is impossible. If one reported many decimal points summarizing the difference between two groups, one group will eventually be higher. Still our results appear to be close enough to nil for us to assert a zero difference.

Exceptions to our conclusion of nil effects could be made for agreeableness, extroversion, and emotional stability. From the perspective of statistical significance testing, one could argue that a difference exists if the confidence interval does not include zero. Using this criteria for the

distribution of all effect sizes for each construct, one would conclude that Whites are slightly more agreeable than Blacks (*d* = .09, confidence interval = .05 to .14). One would also conclude that Blacks are slightly more extroverted than Whites (*d* = -.18, confidence interval = -.24 to -.12) and slightly more emotionally stable (*d* = -.06; confidence interval -.10 to -.03).

Conclusion 4: Publication bias distorts the mean racial difference estimates in journals but not in unpublished data.

Hypothesis 4 states that group differences in scores on personality traits will tend to show publication bias in data obtained from journals and tend not to show bias from data obtained from unpublished sources. To understand publication bias, it is useful to review the distinctions between random and systematic sampling error. Random sampling error is a function of sample size and the magnitude of the effect in the population. Random sampling error is unbiased in that the mean of a set of representative studies will be an unbiased estimate of the mean in the population. That is, the mean across studies will be correct on average and its departures from the population mean, if any, will grow smaller as more data are added to the analysis. Publication bias concerns the representativeness of the data to be analyzed. Expressed another way, publication bias exists to the extent that there is systematic sampling error in the data.

If systematic error is present in the sampling of the studies (if the studies being analyzed are not representative of the population), the mean effect from the set of studies will be a biased estimate of the population. The estimate might be too high or too low and the estimate will not become more accurate as more unrepresentative data are added to the distribution. Because non-representativeness is primarily a matter of lack of availability of some studies, publication bias may be described more accurately as “availability bias.” However, because decisions made by

authors and editors in the publication process often create the availability bias, "publication bias" has become the most frequently used term to describe the phenomenon.

When presenting Hypothesis 4, we argued that journal data are more likely to show publication bias than unpublished data. In the United States, discussions of race differences make many uncomfortable. There are social and legal norms that assume that all groups are equal and when departures from this equality occur they are often attributed to discrimination or unequal opportunity. The norms against the discussion of race differences present a problem for personnel researchers who routinely encounter mean racial differences disfavoring non-Asian minorities in selection measures (Roth et al., 2001) and job performance (Ford, Kraiger, & Schechtman, 1986; Hauenstein, Sinclair, Robson, Quintella, & Donovan, 2003; Kraiger & Ford, 1985; McKay & McDaniel, 2006; Roth, Huffcutt, & Bobko, 2003).

Among personnel researchers, there is frequent discussion about the need for employers to choose between hiring the best employees or enhancing demographic diversity (Pyburn, Ployhart, & Kravitz, 2008). Journals are a public forum and the presentation of data on mean racial differences, particularly evidence that disfavors Blacks, is a norm violation. Thus, when researchers find mean racial differences in their data, we believe that they are more likely to mention the data in their journal articles when the differences are small or favor Blacks then when the differences favor Whites. This causes the data reported in journals to be unrepresentative of the population of all data. We believe that this scenario for publication bias makes the most sense for conscientiousness, agreeableness, emotional stability, locus of control, and self-efficacy because there is substantial consensus on which end of these personality dimensions is most favorable. However, it is unclear if there is a consensus on the favorable ends of extraversion and openness.

Publication Bias in Journal Data. Mean racial effect sizes drawn from journal articles showed evidence of publication bias in the distributions for conscientiousness, agreeableness, and emotional stability suggesting that larger magnitude effect sizes, favoring Whites, had been suppressed. This can be seen graphically in Figures 1 through 3. The cumulative mean graphs show that the cumulative means shifted from the right to the left with the addition of smaller sample, larger standard error effect sizes. The trim and fill graph shows that the imputed effect sizes were added to the right of the distribution. For conscientiousness, there was an 11 point difference (the mean moved from -.07 to .04). For agreeableness, there was a 17 point difference (the mean moved from .10 to .27). For emotional stability, there was a 16 point difference (the mean moved from -.06 to .10). Still, for both the observed and trim and fill adjusted means, one would conclude that the mean racial differences are relatively small compared to cognitive ability tests. However, if one sought to answer which racial group was favored on each scale, the observed data would cause one to conclude that Blacks were more conscientious and emotionally stable than Whites, but lower on agreeableness. In contrast, the trim and fill adjusted distributions would cause one to conclude that Whites were favored on all three personality traits. Thus, we conclude that the mean racial differences between Blacks and Whites on conscientiousness and emotional stability are very small but that the data do not permit a firm conclusion on whether one group scored more favorably. For agreeableness, we conclude that Whites, on average, are more agreeable and that the magnitude of the effect is small (.10 or .27). We also conclude that, consistent with Hypothesis 4, mean racial differences in conscientiousness, agreeableness, and emotional stability reported in journal data show evidence of publication bias and this bias is in the direction of suppressing data favorable to Whites.

If journal articles that report mean racial data tend to suppress mean racial effect sizes that disfavor Blacks, journal articles that do not report race data are more likely to be based on data that favor Whites. Thus, we argue that both sets of data are not representative and the mean estimates will be biased in opposite directions. This was the case for conscientiousness in that the mean racial difference for journals was -.07 (favoring Blacks), the mean racial difference for journal supplemented data was .00. Consistent with our reasoning, both distributions showed publication bias and the bias was in the opposite directions. In the journal data, the trim and fill analysis moved the mean in the direction of being more favorable to Whites (-.07 to .04) and in the journal supplemented data the mean was moved in the direction of favoring Blacks (.00 to -.14).

The results for agreeableness showed partial support for our reasoning. There was support for the journal data. The trim and fill analyses moved the mean to be more favorable to Whites (.10 to .27). And while the journal supplemented data also showed publication bias, the trim and fill adjusted mean favored Whites and not Blacks as expected (-.17 to .08).

Emotional stability data was mostly supportive of our reasoning. The journal data showed publication bias with the trim and fill adjustment moving the mean from favoring Blacks to favoring Whites. The journal supplemented mean *d* was also supportive of our reasoning because the mean *d* was more favorable to Whites (.00) than the mean *d* for the journal data which favored Blacks (-.06). However, there was little evidence of publication bias in the supplemented journal data when we would expect it show suppression of effects favorable to Blacks.

Publication bias in unpublished data. We would not expect publication bias in most sources of unpublished data. Much of our unpublished data were drawn from operational applicant

screening data sets. To the extent that these data were reported in their entirety to the authors, we know of no mechanism that would cause publication bias. Some of our unpublished data came from a consulting firm and was restricted to those data sets from projects in which the research staff was actively involved. These data may or not be representative of the population. Some unpublished data involved incumbents. Some incumbent data were likely part of concurrent validation studies. We do not know if respondents in concurrent validation studies are representative. Some concurrent studies use pre-selected groups. For example, some concurrent validity studies select participants only from the exceptionally high performers and the exceptionally low performers. Such a sample would not be representative. However, on the whole, we would not expect unpublished data to exhibit publication bias.

For the most part, we did not find publication bias in our unpublished data sets. For the distribution of 49 personal communication effect sizes for conscientiousness, there was no publication bias. For agreeableness, the distribution of 42 personal communication effect sizes showed no publication bias. For the 100 personal communication effect sizes for emotional stability, there were only very small effects attributable to publication bias (the trim and fill analysis shifted the mean from -.01 to -.05). The distribution of 53 personal communication effect sizes for openness, showed small publication bias effects (the trim and fill analysis shifted the effect from .02 to .07). Nearly the entire locus of control data was unpublished and there was no evidence of publication bias. All of the self-efficacy data was unpublished and offered no evidence of publication bias. The exception to this trend was for extroversion which showed evidence of publication bias such that effects favoring Whites were suppressed (the trim and fill analysis shifted the mean from -.38 to -.10). Thus, there was little to no evidence of publication bias in unpublished data for six of the seven personality traits examined.

Conclusion 5: Among college students, personality test scorers favor Blacks over Whites.

Blacks show small advantages over Whites in conscientiousness (d = -.12 or -.18; see Table 1), in agreeableness (d = -.02 or -.16; see Table 2), and emotional stability (d = -.16; see Table 3 3). These differences are small but consistent. We offer a post hoc explanation for our finding that Black college students score higher on conscientiousness than White college students. Given mean differences in academic preparation (e.g., SAT scores and perhaps quality of pre-college education) Black college students may be more conscientiousness than White college students to compensate for lower, on average, academic preparedness. We argue that Whites low in conscientiousness are more likely to enter and survive college than Blacks low in conscientiousness because Whites on average, more so than Blacks, can make up for their slothful habits through greater academic preparation. There are also likely to be motivational or financial explanations for the mean difference in conscientiousness. For example, because Black families, on average, have lower income than White families (Webster & Bishaw, 2007), Black college students may be more likely to have to pay for their own tuition and living expenses than White college students. Also, if Blacks are more likely than Whites to require financial aid to attend college, they have greater need to maintain an acceptable grade point average and meet other aid-related standards, such as a minimum number of courses to be taken. Thus, greater conscientiousness may be required of Blacks, relative to Whites, in order to enter and complete college. We have no compelling explanation for the slight Black advantage in agreeableness and emotional stability.

Conclusion 6: Incumbents do not have smaller mean differences than applicants.

Our results when comparing applicants and incumbents is curious because typically one would expect range restriction in incumbent samples to cause mean differences to be smaller.

The range restriction typically occurs because the incumbents are a subset of the applicants and have been selected as employees because they are more qualified subset of all the applicants. However, in our data, the mean racial differences in the incumbents are not smaller than applicants and are often larger. For conscientiousness, the observed means for applicants and incumbents are both zero and for the trim and fill adjusted means, the incumbent differences (*d* = .25) are larger than the applicant differences (*d* = .04). For agreeableness, the incumbent and applicant mean differences are about the same (.14 vs .12), with the incumbent data showing a slightly larger difference. For emotional stability, the incumbent and applicant data are about the same (-.04 vs .01). When one uses the trim and fill adjusted mean for the applicants, the mean differences for applicants are identical (.01). For openness, the mean differences are about the same for both incumbents (.00) and applicants (.01). This comparison cannot be made for locus of control and self-efficacy, because the data are almost entirely from applicants. The only exception to this trend is for extroversion where effect sizes favoring Blacks are meaningful larger in applicant samples than incumbent samples. We do not have a data-based explanation for these results and encourage addition research attention to this issue.

Conclusion 7: Researchers in employment selection research need to pay much more attention to the representativeness of their samples.

Personnel researchers rely very heavily on convenience samples. Perhaps because they are so pervasive in our discipline, we seldom consider the impact that our non-representative samples may have on our results. The Foldes et al. paper relied heavily on unpublished police screening data. For the emotional stability scales, our data set includes a very large police applicant data set. We have a large amount of data from a consulting firm that has large private sector clients. We doubt that we have much data from small business clients. We have a very large sample of air

traffic controllers. We do not know the extent to which these characteristics of our data cause our data to be non-representative of the population. Indeed, given the preponderance of some occupations in our data set, and the complete absence of others in our data set, it would be difficult for us to define the population to which our results generalize. This problem is not unique to this meta-analysis but is common to meta-analyses throughout the management literature.

We have presented evidence that our data from journals in non-representative. This is disheartening given that most researchers and practitioners in personnel areas are socialized to believe that the best research evidence is presented in journals. For the topic of race differences in personality, journal policy can help reduce the publication bias by requiring mean differences to always be reported. However, this is unlikely to solve the problem. If the reporting of race differences is required, we could expect many organizations to refuse to contribute data to research projects.

Comparison to Foldes et al. (in press)

Foldes et al.'s results are similar to those of the present study in that all mean racial differences are very small. However, our statistics do not exactly replicate those of Foldes et al. and at times the direction of the effects do not agree. This is likely due in part to a substantial non-overlap of data. Dr. Foldes declined to release her unpublished data to us citing confidentiality concerns. We understand and respect her position because we too received data on the condition that we respect its confidentiality. Although it is important to keep promises made to data sources, the situation is unfortunate for the advancement of science because it prevents discovery for reasons for differences across meta-analyses that have different sources of data. Although our paper is based on a substantial amount of data some moderator sub-

distributions have relatively few effect sizes, the sharing of data would have increased the *k* for some distributions and permitted more accurate mean estimates.

Journal supplemented data as a publication bias tool

To our knowledge, this paper is the first to compare journal data to journal supplemented data. The rationale of the analysis is compelling because it permits tests of publication bias that involve hypothesized bias in opposing directions. Specifically, if journals suppress data in one direction, the journal data collected from authors who chose not to publish it should show bias in the opposite direction. These directional bias hypotheses were largely supported in our analysis. Future research should explore this analysis strategy in examining publication bias.

Limitations of data and analyses

Although we have a large data set, it was not without problems. We know that we are missing much of the unpublished data and a few unpublished papers found in the Foldes et al. analysis. Many of our requests to organizations for unpublished data were denied because of the sensitivity of the data. Many of our requests for journal supplemented data were unanswered. There are sections of the population who do not attend college, have seldom applied for work and have never been job incumbents. We suspect that this subset of the population scores rather unfavorably on desirable personality traits such as conscientiousness and emotional stability, and our samples likely did not include such individuals. We cannot claim that our data are representative of a population and have difficulty identifying the population from which we sampled. Given that race differences in personnel research is a much debated issue, we were surprised by how little data are available in journals.

Finally, there is a need for methodological advancements in publication bias methods. Trim and fill, for example, relies on the assumption that the only variance in the distribution is

sampling error and that departures from sampling error are potential evidence of publication bias. There is need for publication bias methods that can work well with heterogeneous data (i.e., data containing moderators or other forms of systematic variance).

Conclusion

This paper makes several important contributions to the research literature. First, we offer the largest set of data addressing White-Black mean racial differences in seven personality constructs. Second, we show that the White-Black differences are very small and for the most part it is difficult to conclude that one group is favored over the other. Third, we show that the White-Black mean differences are moderated by the cognitive loading of the personality measures. Fourth, we show that incumbent means are not smaller than applicant means. Fifth, we offer advances in publication bias detection. Specifically, we show that one can compare journals data to journal supplemented data to test directional publication bias hypotheses. In addition, we also introduce cumulative meta-analysis into the personnel research literature as a useful publication bias approach that complements other publication bias detection methods. Sixth, our publication findings show that journal data can yield distorted estimates of population values. This should worry those who rely on journal data for state-of-the art knowledge. Finally, we issue a call for researchers in personnel to pay more attention to the representativeness of their samples and to consider the impact on their conclusions of departures from representativeness.

References

Arthur Jr., W., Edwards, B. D., & Barrett, G. V. (2002). Multiple-choice and constructed response tests of ability: Race-based subgroup performance differences on alternative paper-and-pencil formats. *Personnel Psychology, 55,* 985-1008.

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

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9,* 9-30.

*Bartram, D., Brown, A., Fleck, S., Inceoglu, I., & Ward, K. (2006). *OPQ32 Technical Manual.* UK: SHL Group, Plc.

*Biderman, M. D., Sebren, J., & Nguyen, N. T. (2007). *Time on task mediates the conscientiousness-performance relationship*. Paper presented at the 22nd annual conference of the Society for Industrial and Organizational Psychology. New York, NY.

Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2005). *Comprehensive meta-analysis. Version 2*. Englewood, NJ: Biostat.

*Brannick, M. (2007). Personal communication from Michael Brannick to M. McDaniel on Aug. 26, 2007.

*Brown, N. W. (1997). Description of personality similarities and differences of a sample of Black and White female engineering students. *Psychological Reports, 81*, 603-610.

*Brown, N. W. & Cross, E. J. Jr. (1991). Capitalizing on personality differences of Black and White engineering students. *Journal of Instructional Psychology, 18,* 43-50.

*Bruce, T. A., Leslie, L. M., Gelfand, M. J., Ryan, A. M., Hui, C. H., & Radford, M. (2005, April). A cultural perspective on frame-of-reference effects. In M.J. Gelfand & A.M. Ryan (Chairs), Is personality research culture bound? Examining personality effectsacross cultures. Symposium conducted at the Society for Industrial and Organizational Psychology Conference. Los Angeles, CA.

*Butcher, J., Ball, B., & Ray, E. (1964). Effects of socio-economic level on MMPI differences in Negro-White college students. *Journal of Counseling Psychology, 11,* 83-87.

*Chan, D., Schmitt, N., Sacco, J. M., & DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology, 83,* 471-485.

*Cucina, J. M. (2008). Personal communication from Jeffrey M. Cucina to Michael A. McDaniel on Feb. 26, 2008.

*Cucina, J. M., Vasilopoulos, N. L., & Sehgal, K. (2005). Personality-based job analysis and the self-serving bias. *Journal of Business and Psychology, 20,* 275-290.

Curfman, G. D., Morrissey, S., & Drazen, J.M. (2006). Expression of concern reaffirmed. *New England Journal of Medicine, 354,* 1193.

Dahlstrom, W. G. & Walsh, G. S. (1960). *An MMPI handbook: A guide to use in clinical practice and research.* Minneapolis, MN: University of Minnesota Press.

*Davis, R. (2007). Personal communication from Robert D. Davis to Michael A. McDaniel on on February 27, 2008. We believe that these data are the same as or substantially overlap data from the Foldes et al. (in press) reference: Davis, R. & Rostow, C. D. (2004). [Personality scale scores of various demographic groups for multiple personality inventories]. Unpublished raw data.

Dickerson, K. (2005). Publication bias: Recognizing the problem, understandings its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments* (pp. 11-34). Chichester, UK: Wiley.

*Donnellan, B. M., Conger, R. D., & Burzette, B. G. (2005). Criterion-related validity, self other agreement, and longitudinal analyses for the Iowa personality questionnaire: A short alternative to the MPQ. *Journal of Research in Personality, 39,* 458-485.

*Dudley, N. M., McFarland, L. A., Goodman, S. A., Hunt, S. T., & Sydell, E. J. (2005). Racial differences in socially desirable responding in selection contexts: Magnitude and consequences. *Journal of Personality Assessment, 85,* 50-64.

Duval, S. J. (2005). The "trim and fill" method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments* (pp. 127-144). Chichester, UK: Wiley.

Duval, S. J. & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 276-284.

*Edwards, J. R. (2007). Personal communication from Jeffery R. Edwards to Michael A. McDaniel on August 9, 2007.

Eysenck, H. J. & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire* (adult and junior). London: Hodder & Stoughton.

*FAA Aerospace Human Factors Research Division (1997). [16PF and Air Traffic Control Specialist (ATCS) written aptitude test scores, 1982-1992.] Unpublished data.

*Federal Government Agency A. (2008). Personal communication from Federal Government Agency A to Michael A. McDaniel on January 13, 2008.

Foldes, H., Duehr, E., & Ones, D.S. (in press). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology*.

Ford, J. K., Kraiger, K., & Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. *Psychological Bulletin, 99,* 330–337.

*Goldberg, L. R., Sweeney, D., Merenda, P. F., & Hughes, J. E. Jr. (1998). Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. *Journal of Individual Differences, 24,* 393- 403.

Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51,* 357-374.

*Harris, P. R. & Lightsey, O. R. (2005). Constructive thinking as a mediator of the relationship between extraversion, neuroticism, and subjective well-being. *European Journal of Personality, 19,* 409-426. Supplemented by data from a personal communication from Richard Lightsey to Michael A. McDaniel on February 26, 2008.

*Harrison, R. H. & Kass, E. H. (1967). Differences between Negro and White pregnant women on the MMPI. *Journal of Consulting Psychology, 31,* 454-463.

Hauenstein, N. M. A., Sinclair, A. L., Robson, V., Quintella, Y., & Donovan, J. J. (2003, April). *Performance dimensionality and the occurrence of ratee race effects.* Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Academic Press.

*Heuchert, J. W. P., Parker, W. D., Stumpf, H., & Myburgh, C. P. H. (2000). The five factor model of personality in South African college students. *American Behavioral Scientist, 44,* 112-125.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92,* 1270-1285.

Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18,* 331-341.

*Hogan, R. & Hogan, J. (1995). *Hogan Personality Inventory Manual (2nd Ed.).* Tulsa, OK: Hogan Assessment Systems.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions. *American Psychologist, 51,* 469-477.

*Hough, L. M. (1997). Issues and evidence: Use of personality variables for predicting job performance. Symposium conducted at the 12th annual conference of the Society for Industrial and Organizational Psychology. St. Louis, MO.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology—Volume 1: Personnel Psychology* (Vol. 1). London: Sage.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

*Human Resources Firm A. (2007). Personal communication from Human Resources Firm A to Michael A. McDaniel.

Hunter, J.E., & Schmidt, F. L. (2004). *Methods of Meta-analysis: Correcting error and bias in research findings*. Second edition. Newbury Park, CA: Sage Publications.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869–879.

*Huth (2007). Personal communication from Huth to Michael A. McDaniel.

*IPAT (2008). Personal communication of unpublished data from IPAT to Michael A. McDaniel on March 13, 2008.

Jackson, D. N. (1974). *Jackson personality inventory manual*. Goshen, NY: Psychologists.

*Jacobs, R. R. (2005). Balancing validity and adverse impact in police selection: Two steps forward and one step back. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology. Los Angeles, CA.

Jensen, A.R. (1998). *The g factor*. Westport, CT: Praeger.

*Jensen, J. (2007). Personal communication from Jaclyn Jensen to Michael McDaniel on Aug, 13, 2007.

*Jones, E. E. (1978). Black-White personality differences: Another look. *Journal of Personality Assessment, 42,* 244-252.

*King, E. B., George, J. M., & Hebl, M. R. (2005). Linking personality to helping behaviors at work: An interactional perspective. *Journal of Personality, 73,* 585-607. Supplemented by data from a personal communication from Eden B. King to Michael A. McDaniel on Feb. 26, 2008.

*King, H. F., Carroll, J. L., & Fuller, G. B. (1977). Comparison of nonpsychiatric Blacks and Whites on the MMPI. *Journal of Clinical Psychology, 33,* 725-728.

*Klein, K. J., Lim, B. C., Saltz, J. L., & Mayer, D. M. (2004). How do they get there? An examination of the antecedents of centrality in team networks. *Academy of Management Journal, 47,* 952-963. Supplemented by data from a personal communication from Katherine J. Klein to Michael A. McDaniel on March 6, 2008.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70,* 56–65.

Lau, J., Schmid, C. H., & Chalmers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology, 48,* 45-57.

Leimu, R. & Koricheva, J. (2004). Cumulative meta-analysis: A new tool for detection of temporal trends and publication bias in ecology. *Proc. R. Soc. Lond. Ser. B-Biol. Sci., 271,* 1961-1966.

*Maraist, C. C. & Russell, M. T. (2002). 16PF Fifth Edition Norm Supplement, Release 2002. Institute for Personality and Ability Testing, Inc. Champaign, IL.

*McAdams, D. P., Anyidoho, N. A., Brown, C. Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and states: Links between dispositional and narrative features of personality. *Journal of Personality, 72,* 761-784. Supplemented by data from Dan P. McAdams to Michael A. McDaniel on April 2, 2008.

McDaniel, M.A., Rothstein, H.R. & Whetzel, D.L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology, 59,* 927-953.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599–616.

McKay, P. & McDaniel, M.A. (2006). A re-examination of Black-White mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91,* 531-554.

*Morey, L. C. (1991). *Personality Assessment Inventory Test Manual.* Port Huron, MI: Sigma Assessment Systems, Inc.

*Mount, M. K. & Barrick, M. R. (2002). *Wonderlic Personal Characteristics Inventory.* Libertyville, IL, Author.

Murray, C. (2005, September). The inequality taboo. *Commentary, 120,* 13-22.

* Nguyen, N. T. (2004) A further study of personality and academic performance. Unpublished manuscript.

*Nguyen, N.T., McNary, L., & Biderman, M.D. (2008). A validation study of the Cross-Cultural Adaptability Inventory. Paper accepted for presentation at the 2008 annual conference of the Southern Management Association. St. Pete Beach, FL.

*O'Connell, M. S., McDaniel, M. A., Grubb, W. L. III, Hartman, N. S., & Lawrence, A. (2002). Incremental validity of situational judgment tests for task and contextual job performance. Paper presented at the 17th annual conference of the Society for Industrial and Organizational Psychology. Toronto, ON.

*Ones, D. S. & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology, 75,* 255-276.

Ones, D. S. & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83,* 35-42.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15,* 161-171.

*Pace, V. (2005). *Creative performance on the job: Does openness to experience matter*. Unpublished thesis, University of South Florida, Florida.

Pyburn, K., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61,* 143-151.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer III, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54,* 297-330.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88*, 694–706.

*Russell, T. L. & Peterson, N. G. (2001). The experimental battery: Basic attribute scores for predicting performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum.

*Ryan, A. M., Ployhart, R. E., & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83,* 298-307.

Sackett, P. R. & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*, 929-954.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-273.

*Schmitt, N. (2007). Personal communication from Neal Schmitt to Michael McDaniel on August 8, 2007.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. *International Review of Industrial and Organizational Psychology, 11,* 115-139.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82,* 719-730.

*Sitzmann, T., Brown, K. G., Ely, K., & Kraiger, K. (2008). *A longitudinal evaluation of a dynamic model of motivation to learn.* Unpublished manuscript.

*Spence, J. T., Helmreich, R. L., & Pred, R. S. (1987). Impatience versus achievement strivings in the Type A pattern: Differential effects on students' health and academic achievement. *Journal of Applied Psychology, 72,* 522-528. Supplemented by data from a personal communication from Janet T. Spence to Michael A. McDaniel.

*Strauss, J. P. & Connerly, M. L. (2003). Demographics, personality, contact, and universal diverse orientation: An exploratory examination. *Human Resource Management, 42,* 159-174.

Tett, R. P., Jackson, D. N, & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Thompson, S.G. & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine, 21, 1559 - 1573.

*Timbrook, R. E. & Graham, J. R. (1994). Ethnic differences on the MMPI-2. *Psychological Assessment, 6,* 212-217.

*Tyler, F. B., Gease, E. I., Moran, J. A., & Gatz, M. J. (1982). Individual psychosocial competence and aging. *Academic Psychology Bulletin, 4,* 503-513.

*Van Rooy, D. L., Alonso, A., & Viswesvaran, C. (2005). Group differences in emotional intelligence test scores: Theoretical and practical implications. *Personality and Individual Differences, 38,* 689-700.

*Vasilopoulos, N. L., Cucina, J. M., & Hunter, A. E. (2007). Personality and training proficiency: Issues of bandwidth-fidelity and curvilinearity. *Journal of Occupational and Organizational Psychology, 80,* 109-131. Supplemented by data from a personal communication from Nicholas L. Vasilopoulos to Michael A. McDaniel on Feb. 28, 2008.

Viswesvaran, C., Deller, J., & Ones, D. S. (2007). Personality measures in personnel selection: Some new contributions. *International Journal of Selection and Assessment, 15,* 354-358.

Webster, B. H. Jr. & Bishaw, A. (2007). *Income, earnings, and poverty data from the 2006 American Community Survey* (American Community Survey Reports, ACS-08). Washington, DC: U.S. Government Printing Office.

*Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance, 17,* 433-461.

Whetzel, D.L., McDaniel, M.A., & Nguyen, N. T. (in press). Subgroup differences in situational judgment test performance: A meta-analysis. Human Performance.

Table 1. White – Black mean differences in conscientiousness

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	81	193,455	-.02	-.08 to .03	.43	38
Source: Journal	15	11,142	-.07 (.04)	-.34 to .21	.30	8
Source: Journal (supplemented)	7	7,262	.00 (-.14)	-.18 to .17	-	-
Source: Personal communication	49	134,739	-.01	-.06 to .03	.44	25
Respondents: Applicants	30	150,160	.00 (.04)	-.07 to .06	.79	7
Respondents: College Students	22	17,376	-.12 (-.18)	-.21 to -.02	.13	12
Respondents: Incumbents	23	13,394	.00 (.25)	-.21 to .21 (.06 to .44)	.41	16
Source: Journal Respondents: College Students	8	1,901	-.19	-.41 to .04	.32	5
Source: Journal Respondents: Incumbents	5	5,698	.15	-.39 to .69	-	-
Source: Personal communication Respondents: Applicants	25	128,071	-.01	-.06 to .04	.95	4
Source: Personal communication Respondents: College Students	8	3,445	-.07	-.15 to .02	-.39	6
Source: Personal communication Respondents: Incumbents	13	2,566	.01 (.26)	-.23 to .25	.41	12

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that publication bias analyses were not conducted due to too few studies or the analysis was conducted and there was no publication bias or a trivial change in the mean *d*.

Table 2. White – Black mean differences in agreeableness

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	73	202,211	.09	.05 to .14	.11	32
Source: Journal	14	10,945	.10 (.27)	-.04 to .11	-.33	7
Source: Journal (supplemented)	6	1,834	-.17 (.08)	-.04 to .11	-	-
Source: Personal communication	42	169,405	.12	.07 to .17	.37	21
Respondents: Applicants	27	164,607	.12	.07 to .18	-.43	6
Respondents: College Students	19	11,699	-.02 (-.16)	-.14 to .10	.28	10
Respondents: Incumbents	22	13,394	.14	.04 to .25	.02	14
Source: Journal Respondents: College Students	7	1,704	.00 (.07)	-.21 to .20	-.26	4
Source: Journal Respondents: Incumbents	5	5,698	.19	-.08 to .46	-	-
Source: Personal communication Respondents: Applicants	22	154,066	.12	.06 to .17	.72	4
Source: Personal communication Respondents: College Students	7	3,395	.00	-.08 to .08	.56	5
Source: Personal communication Respondents: Incumbents	11	2,301	.23 (.12)	.07 to .40	-.11	10

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that the publication bias analyses were not conducted due to too few studies or the analysis was conducted and there was no publication bias or a trivial change in the mean *d*.

Table 3. White – Black mean differences in emotional stability

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	140	168,898	-.06	-.10 to -.03	.26	31
Source: Journal	25	12,085	-.06 (.10)	-.16 to .04	.93	7
Source: Journal (supplemented)	6	1,834	.00	-.23 to .23	-	-
Source: Personal communication	100	118,723	-.05 (-.01)	-.09 to -.02	.13	21
Respondents: Applicants	85	123,346	-.04 (.01)	-.08 to .00	.48	7
Respondents: College Students	26	12,325	-.16	-.25 to -.07	.01	10
Respondents: Incumbents	20	10,253	.01	-.10 to .12	.14	12
Respondentss: General population	6	21,559	-.09 (-.04)	-.24 to .07	-	-
Source: Journal Respondents: College Students	14	2,329	-.17 (-.26)	-.32 to -.01	1.00	4
Source: Journal Respondents: Incumbents	5	3,780	.05	-.06 to .15	-	-
Source: Personal communication Respondents: Applicants	80	112,387	-.05 (.01)	-.08 to -.01	.55	4
Source: Personal communication Respondents: College Students	7	3,394	-.04	-.12 to .04	-.37	5
Source: Personal communication Respondents: Incumbents	11	2,299	-.11	-.31 to .08	.07	10

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that the publication bias analyses were not conducted due to too few studies or the analysis was conducted and there was no publication bias or a trivial change in the mean *d*.

Table 4. White – Black mean differences in extroversion

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	117	57,820	-.18	-.24 to -.12	.06	30
Source: Journal	20	12,614	-.05 (-.10)	-.16 to .07	-.61	7
Source: Journal (supplemented)	5	1,637	.02 (.07)	-.22 to .25	-	-
Source: Personal communication	84	29,329	-.30 (-.02)	-.39 to -.20	.26	21
Respondents: Applicants	68	27,368	-.38 (-.10)	-.48 to -.28	.33	5
Respondents: College Students	19	5,805	.17 (-.02)	.01 to .34	-.57	10
Respondents: Incumbents	22	12,523	-.03	-.13 to .07	.64	13
Source: Journal Respondents: College Students	9	1,905	.14 (.09)	-.16 to .45	.02	4
Source: Journal Respondents: Incumbents	6	5,810	-.16	-.35 to .03	-	-
Source: Personal communication Respondents: Applicants	65	23,166	-.41 (-.06)	-.52 to -.30	.69	4
Source: Personal communication Respondents: College Students	6	3,246	.25	-.04 to .53	-.54	5
Source: Personal communication Respondents: Incumbents	11	2,274	.02 (.19)	-.18 to .21	.65	10

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that the publication bias analyses were not conducted due to too few studies or the analysis was conducted and there was no publication bias or a trivial change in the mean *d*. For locus of control, a positive *d* indicates that Whites have more internal locus of control than Blacks, on average.

Table 5. White – Black mean differences in openness

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	78	148,097	.02	-.03 to .06	.19	28
Source: Journal	12	2,771	.07 (-.18)	-.21 to .35	-.61	4
Source: Journal (supplemented)	5	1,637	-.02	-.39 to .35	-	-
Source: Personal communication	53	130,349	.02 (.07)	-.04 to .07	.38	21
Respondents: Applicants	36	125,712	.01 (.08)	-.04 to .07	-.41	4
Respondents: College Students	19	5,402	.05 (.15)	-.14 to .23	-.19	19
Respondents: Incumbents	19	7,183	.00 (.06)	-.11 to .12	.55	11
Source: Journal Respondents: College Students	8	1,434	.10	-.36 to .55	-.61	4
Source: Personal communication Respondents: Applicants	34	124,163	.02 (.08)	-.03 to .08	-.41	4
Source: Personal communication Respondents: College Students	6	3,246	.08	.00 to .17	.53	5
Source: Personal communication Respondents: Incumbents	11	2,297	-.06	-.28 to .15	.54	10

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that the publication bias analyses were not conducted due to too few studies or the analysis was conducted and there was no publication bias or a trivial change in the mean *d*. For locus of control, a positive *d* indicates that Whites have more internal locus of control than Blacks, on average.

Table 6. White – Black mean differences in locus of control

Distribution description	k	Total *N*	Mean *d*[1]	Confidence Interval	Vector Correlation	k for vector correlation
All data	16	68,095	.05	-.01 to .12	-	-
All data except d = -.82 outlier removed	15	67,314	.10	.05 to .14	-	-

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis. A lack of a mean d in parentheses indicates that the analysis was not conducted due to too few studies or the analysis was conducted and there were no publication bias analyses conducted or a trivial change in the mean *d*. For locus of control, a positive *d* indicates that Whites have more internal locus of control than Blacks, on average.

Table 7. White – Black mean differences in self-efficacy

Distribution description	k	Total *N*	Mean d^1	Confidence Interval	Vector Correlation	k for vector correlation
All data	37	236,906	.02	-.20 to .25	-	-
All data except *d* = 1.52 outlier removed	36	184,542	-.03	-.08 to .03	-	-

Note: A positive *d* indicates that the White mean is higher than the Black mean. A mean *d* in parentheses is the mean d estimated from a trim and fill analysis.

Figure 1. Illustrative Meta-Analysis Graphics

Figure 1a. An asymmetric funnel plot

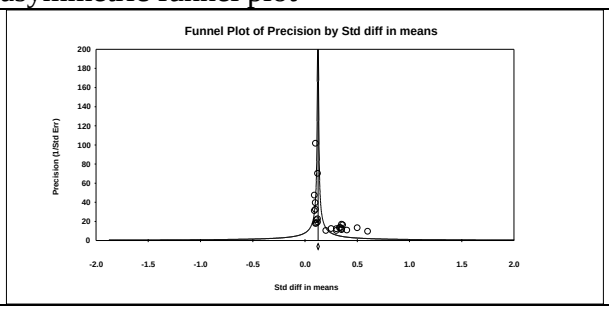
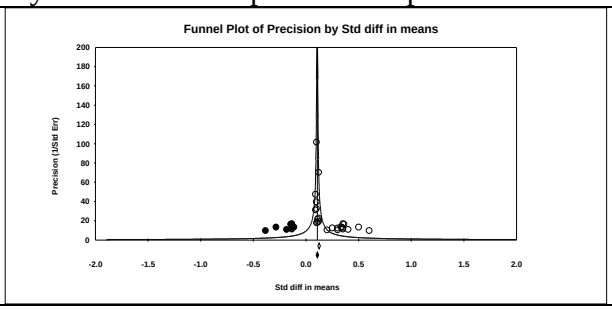


Figure 1b. An asymmetric funnel plot with imputed studies



Funnel 1c. A cumulative means graph

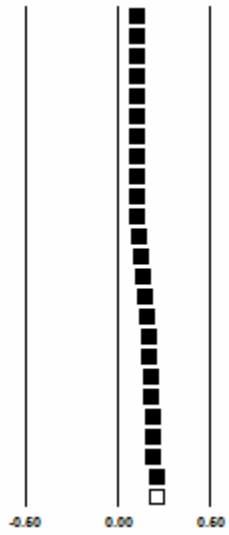


Figure 2. Publication bias graphics for global conscientiousness effect sizes from journals

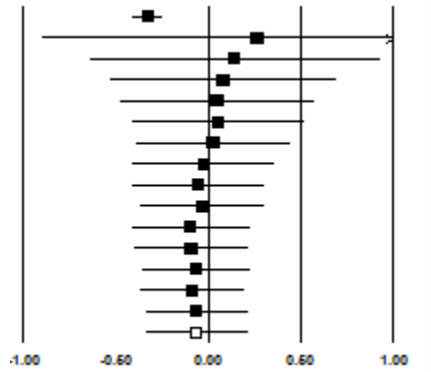
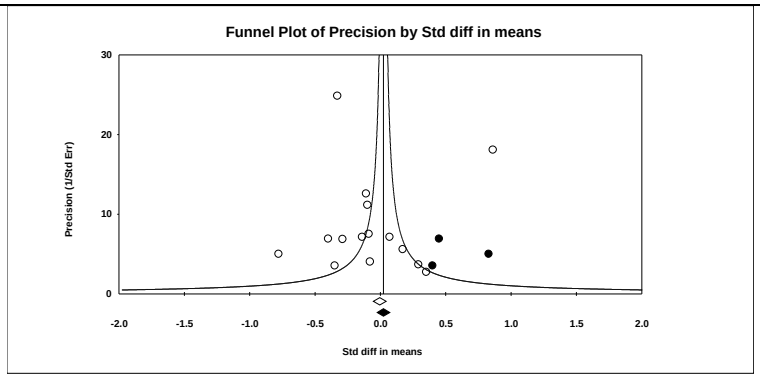
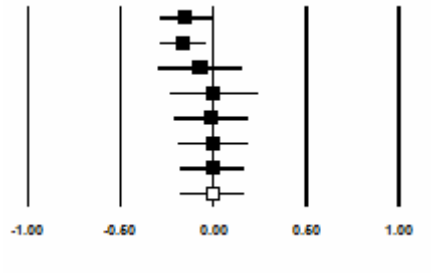
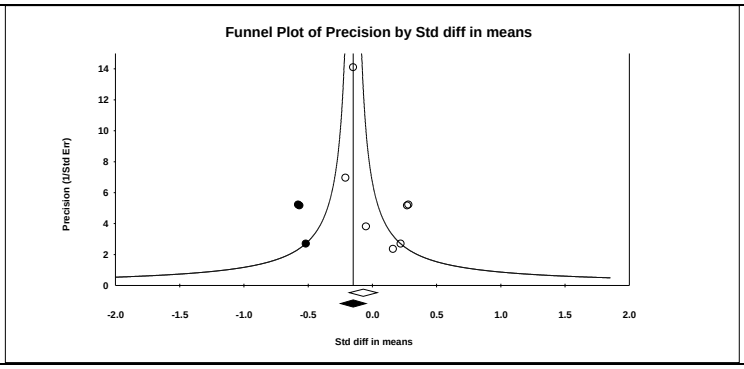
Journal data	
Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
	
Journal Data Where Race Data Was Obtained Through Personal Communication	
Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
	

Figure 3. Publication bias graphics for agreeableness effect sizes from journals

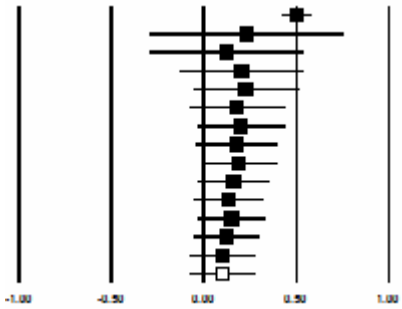
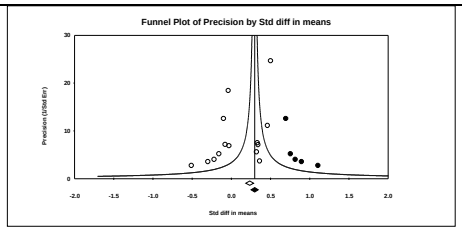
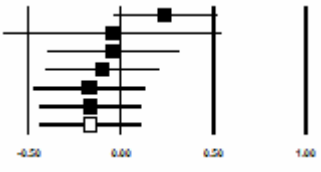
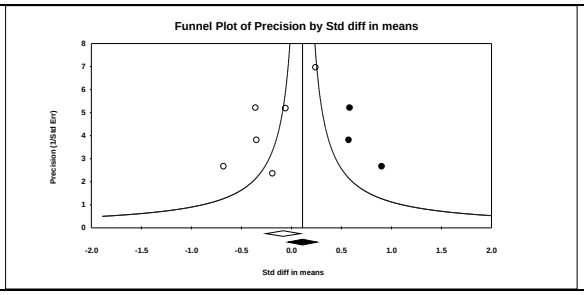
<table>
<tr><td colspan="2">Journal data</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td>  </td><td> 
</td></tr>
<tr><td colspan="2">Journal Data Where Race Data Was Obtained Through Personal Communication</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td>  </td><td> 
</td></tr>
</table>

Figure 4. Publication bias graphics for emotional stability effect sizes from journals

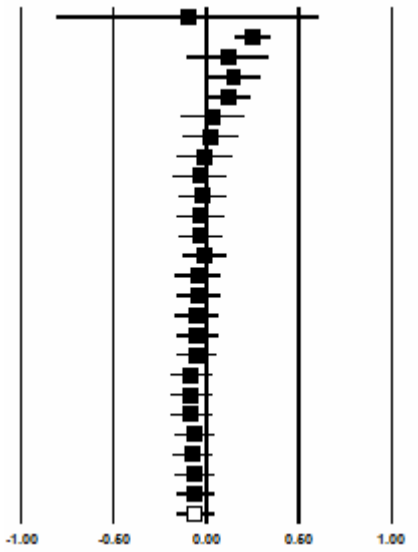
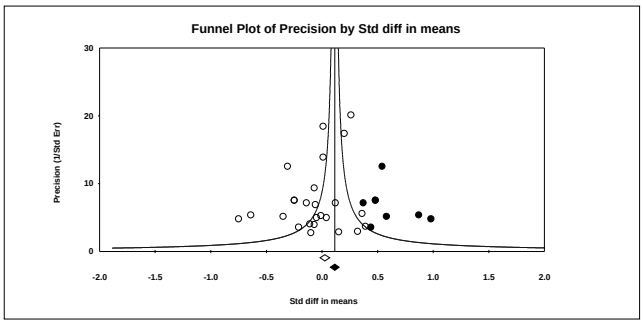
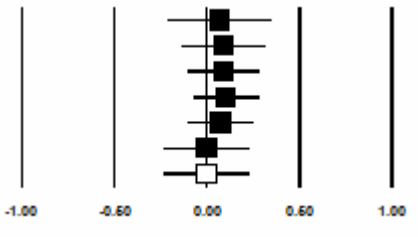
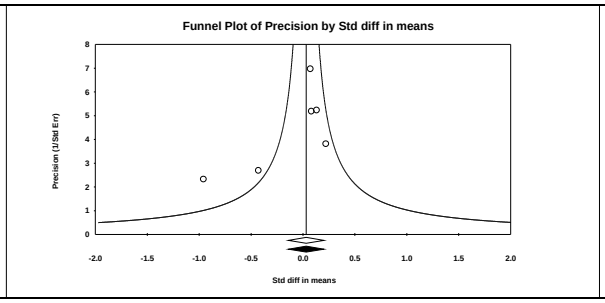
<table>
<tr><td colspan="2">Journal data</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td colspan="2">Journal Data Where Race Data Was Obtained Through Personal Communication</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>

Figure 4. Publication bias graphics for extraversion effect sizes from journals

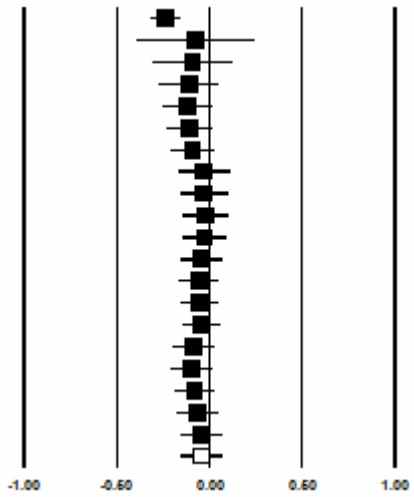
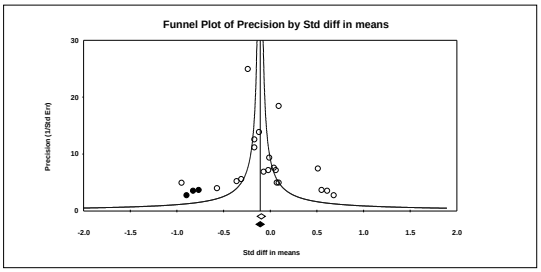
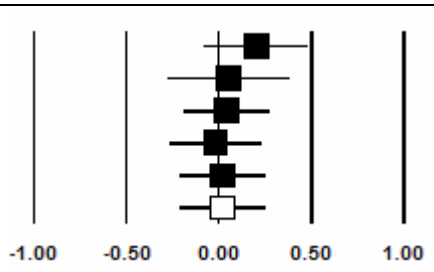
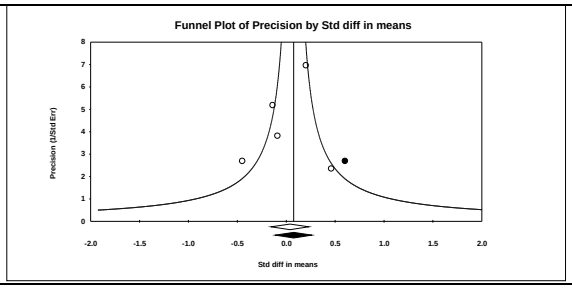
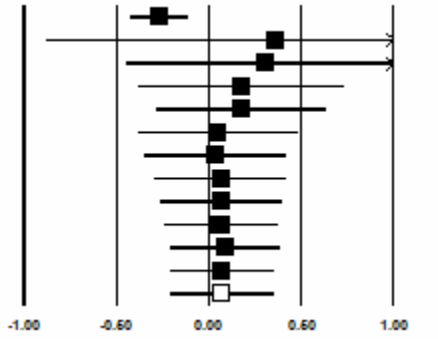
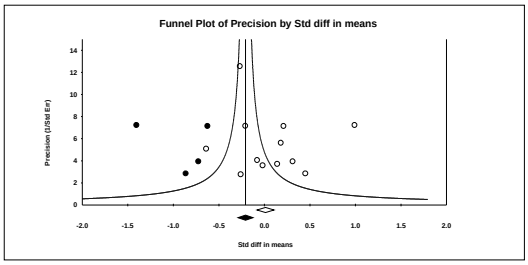
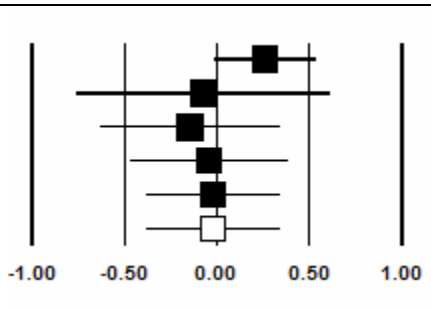
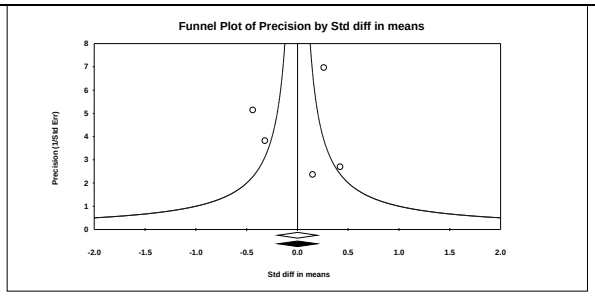
Journal data	
Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
	
Journal Data Where Race Data Was Obtained Through Personal Communication	
Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
	

Figure 5. Publication bias graphics for openness effect sizes from journals

<table>
<tr><td colspan="2">Journal data</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td colspan="2">Journal Data Where Race Data Was Obtained Through Personal Communication</td></tr>
<tr><td>Cumulative Meta-Analysis.
Studies sorted low to high by standard error.</td><td>Trim and Fill Plot</td></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>