

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF PSYCHOLOGY

Examining Racial Subgroup Differences in Personality Measures: A Predictive Sample
Approach

MITCHELL DOBBS
SPRING 2021

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Psychology
with honors in Psychology

Reviewed and approved* by the following:

Rustin Meyer
Assistant Professor of Psychology
Thesis Supervisor

Susan Mohammed
Professor of Psychology
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Using the largest applicant sample to date, the present study compares the personality scores of Black ($n = 17,649$), White ($n = 37,736$), and Hispanic ($n = 20,949$) police applicants. Contrary to the current literature drawing almost exclusively on incumbent samples, the present investigation found significant Black-White differences in conscientiousness favoring White applicants. Black applicants also demonstrated higher scores on Agreeableness and Neuroticism relative to White applicants, who scored remarkably similarly to Hispanic applicants on every Big 5 trait. The present results dispute the current function of personality measures as a means to reduce adverse impact in personnel selection, and carry additional implications for scholars and practitioners alike.

TABLE OF CONTENTS

LIST OF TABLES	iii
ACKNOWLEDGEMENTS	iv
Chapter 1 Introduction	1
Chapter 2 Literature Review	3
Adverse Impact	3
Subgroup Differences, Adverse Impact, & Disparate Treatment Discrimination.....	4
Adverse Impact and Cognitive Ability Measures	7
Adverse Impact and Personality Measures	10
Hypotheses	13
Chapter 3 Methodology	14
Sample.....	14
Measures	14
Procedure	15
Chapter 4 Results	16
Chapter 5 Discussion	18
Summary and Implications	18
Limitations and Directions for Future Research	19
Appendix A Tables	22

LIST OF TABLES

Table 1: Definitions and Reliability for OPQ32 Dimensions and Big Five Traits	22
Table 2: Means and Standard Deviations of Personality Scores by Subgroup	22
Table 3: Standardized Difference Scores by Subgroup	23
Table 4: Comparison of Subgroup Difference Effect Sizes Across Studies	23

ACKNOWLEDGEMENTS

While at first dreading nearly every aspect of this project, I rather quickly discovered that research can be quite an enjoyable undertaking. Such would not have been the case, however, without the individuals who guided me along the way. My utmost thanks go out to Dr. Rustin Meyer, who, in my mind, has set an incredibly high bar for what it means to be a critically-thinking scientist. His feedback and especially his patience with me throughout the past two and a half years were critical in completing the Fisher-Price “My First Research Project” that is this thesis. An absolutely **massive** “thank-you” must also be extended to Susannah Ivory, who helped guide me through this project from start to finish. I’m incredibly fortunate that she was so willing to help and easy to work with. I really do mean it when I say this project would be blank paper without her assistance; she rocks. I’d also like to thank Dr. Susan Mohammed, who kept me on track and provided invaluable feedback prior to submission. My sincere thanks must also be extended to Dr. Rick Jacobs, for his willingness to provide data.

I’m also incredibly lucky that my friends and my family showed an interest in my work on this project. The amount of phone calls and texts that began with, “yeah, and what’s really interesting is...” probably were never that interesting to them, but they listened anyway. I’m thrilled to have finished this up with their support along the way, and hope this is one of many projects I’m able to share with them. Above all else, I’m really happy that I found the research process so fun and engaging. I’m grateful to be able to assess and understand the contributions and shortcomings of the present paper, and look forward to doing more research in the future. This was a really great experience, and I have Rustin, Sus, Dr. Mohammed, Rick, and my family and friends to thank for it. Thank you.

Chapter 1

Introduction

Recommendations for reducing the adverse impact of personnel selection systems often involve the inclusion of personality measures (Avis, Kudisch, & Fortunato, 2002; Ryan, Ployhart, & Friedel, 1998; Schmitt et al., 1997). This is the case because personality traits are generally believed to have smaller subgroup differences than other commonly used predictors (especially cognitive ability; see Bobko & Roth, 2013; Bobko, Roth, & Potosky, 1999) or even scores that favor minority applicants, thereby helping to improve both validity and diversity (Hough, Oswald, & Ployhart, 2001; Foldes, Duehr, & Ones, 2008). Consistent with this perspective, various meta-analyses (e.g., Hough et al., 2001; Foldes, et al., 2008) have demonstrated small to moderate subgroup differences across Big Five traits (i.e., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism).

If personality measures are to be used to attenuate group differences, however, several unresolved matters should receive scholarly attention. Perhaps the most pressing outstanding issues concerns the fact that virtually all of the primary studies in the aforementioned meta-analyses have been disproportionately comprised of incumbents (Hough et al., 2001; Foldes et al., 2008). This leaves much to be desired in the area of practical predictions (i.e., applicant data), and might be particularly problematic in light of evidence suggesting that, relative to applicant samples, incumbent samples demonstrate a tendency to underestimate subgroup differences (Roth et al., 2001; Bobko & Roth, 2013; Stark et al., 2001).

Thus, the present investigation examines racial subgroup differences in a large ($n = 79, 339$) representative sample comprised of Black, White, and Hispanic police applicants. In order to establish a theoretical foundation for these predictions, relevant literature pertaining to subgroup differences, adverse impact, and measures of both personality and cognitive ability will be reviewed first. The sample, procedure, and measures of the current study will then be described, with empirical results presented and compared to hypotheses immediately afterwards. Finally, findings will be placed in the context of the current literature, accompanied by a discussion of methodological limitations and directions for future research.

Chapter 2

Literature Review

Adverse Impact

Selecting a workforce that is both diverse and high-performing is integral to any modern organization's success. A diverse staff increases perceived fairness, reduces economic costs (e.g., those of a lawsuit), and encourages innovation via a wide breadth of perspectives on a variety of complex tasks (Jackson, Joshi, & Erhardt, 2003). Benefits of a high performing workforce include maximizing job performance and increased economic output (Schmidt & Hunter, 1998). Yet, when subgroup differences exist on valid predictors used in selection contexts (e.g., cognitive ability, personality measures), reconciling diversity and validity of these predictors becomes an important consideration.

Thus, finding themselves mired in the “diversity-validity dilemma,” organizations are seemingly forced to choose whether the selection systems they rely upon work toward the goal of diversity or validity (Pyburn, Ployhart, & Kravitz, 2008). Cognitive ability measures are amongst the best examples of this phenomena: despite their predictive validity for various job performance criterion, they simultaneously demonstrate concerningly large subgroup differences (nearly one standard deviation between Black and White individuals) (Hunter & Hunter, 1984; Roth et al., 2001; but also see Bobko & Roth, 2013). Erring too heavily on the side of the latter (i.e., the validity of selection assessments) – whether intentionally or unintentionally – can reduce diversity, potentially leading to *adverse impact*: a substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members

of a race, sex, or ethnic group (Equal Employment Opportunity Commission, 1978, Section 60-3.16).

Subgroup Differences, Adverse Impact, & Disparate Treatment Discrimination

It is important to establish here a concrete relationship between subgroup differences and adverse impact, as the two are not one and the same. Rather, subgroup differences might be thought of as the raw materials for adverse impact: for one subgroup to be disadvantaged in matters of selection, said group must vary on a given construct (or constructs) from other subgroups. For instance, when comparing the average N.B.A. pro to the average male over twenty years old in the United States, significant group differences in height present themselves, as the former group is considerably taller than the latter.

However, adverse impact only exists so long as a protected class (i.e., a group legally protected from employment discrimination based on race, religion, sex, national origin, age, and/or disability) is shown to be measurably disadvantaged in selection practices (Equal Employment Opportunity Commission, 1978, Section 60-3.16). As an example, consider a recreational gym seeking to hire personal trainers using tests of upper-body strength. Because men typically demonstrate greater upper-body strength than women (on average), such a test would show adverse impact for female applicants. Thus, this example illustrates how the presence of subgroup differences (e.g., those in upper-body strength between the sexes) might discriminate against a protected class (i.e., female applicants), intentionality notwithstanding.

Measuring Subgroup Differences and Adverse Impact

The d Statistic. Minimizing adverse impact is thus contingent upon organizations, policymakers, and academics alike obtaining accurate estimates of group differences. The standardized difference scores (i.e., d statistic) – defined as the difference in means divided by the average of group standard deviations – is often used to provide relevant stakeholders with such estimates (Bobko & Roth, 2013). Knowledge of both the presence and the magnitude of estimated group differences via the d statistic is advantageous in determining which predictors of performance and success are vulnerable to adverse impact, making it particularly useful to those tasked with designing equitable selection systems. These standardized difference scores are also of particular use because they can be applied to applicants without their actually being selected. Thus, these values could be said to act as an index of “adverse impact potential” (Bobko & Roth, 2013).

In and of themselves, however, d statistics are insufficient for conclusively predicting adverse impact in a given selection context. Although standardized difference scores are useful in identifying specific predictors and scoring systems that might be particularly likely to lead to adverse impact, statistically significant values of d do not *guarantee* the presence of adverse impact. The opposite is also true: a non-significant d statistic does not guarantee the absence of adverse impact, either. Standardized difference scores functioning as indices of adverse impact (as opposed to guarantors) can be attributed to the fact that adverse impact analyses are based exclusively on applicants (not incumbents) and the selection ratio among those who applied. These distinctions are crucial, as they suggest that scientific investigations into group differences must account for both of the aforementioned considerations in order to best inform practice.

The Four-Fifths Rule. Thus, instead of relying on standardized d statistics calculated on the general population to determine the potential presence of adverse impact, a series of metrics has been developed to determine if adverse impact exists in an actual selection environment. The most common of these metrics is the “four-fifths rule” (see *Uniform Guidelines*, 1978), which compares the subgroup with the highest selection ratio to any subgroup with a lower selection ratio to compute a new ratio. If this new ratio falls below 80% (four-fifths), it is said to demonstrate adverse impact (Newman & Lyon, 2009). However, Roth, Bobko, & Switzer (2006) suggest that conducting a subsequent test of statistical significance (e.g., Fisher’s exact test; a chi-square test) might reduce the purportedly high false-positive rate produced by exclusive use of the four-fifths rule.

Cases where the four-fifths rule is used to measure adverse impact might be particularly problematic if the sample size is insufficient or if conclusions are used as prima facie evidence of discrimination (Roth et al., 2006). This last point is particularly useful in distinguishing adverse impact from a separate but related legal issue: disparate treatment discrimination. Returning to the previous example of tests of upper-body strength used to select personal trainers, adverse impact would likely be shown for female applicants, who, on average, possess less upper-body strength than males. Assuming a female applicant sued the recreational gym for discriminating based on sex, the gym would need to demonstrate that a certain amount of upper-body strength was necessary in order to do the job in question in order to be acquitted of adverse impact discrimination. Very different, however, is a scenario where the same gym posts a sign on its door saying “Hiring: Personal Trainers. Women need not apply!” The discrimination inferable from such a sign would make the gym immediately guilty of *disparate treatment discrimination*,

or intentional discrimination against a protected class (Civil Rights Act of 1964 § 7, 42 U.S.C. § 2000e et seq (1964)).

Adverse Impact and Cognitive Ability Measures

As a function of both class and construct, real-world cases of adverse impact or disparate treatment discrimination are often inexorably more contentious than those founded on differences in upper-body strength between the sexes. Thus, while few organizations have a vested interest in using upper-body strength to predict job performance, such is not the case for cognitive ability measures, which are amongst the most valid predictors of performance across jobs (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). However, these same measures also demonstrate the largest and most contentious mean differences between Black and White individuals. Such differences have been associated with the hiring of fewer Black employees than White (Bobko, Roth, & Potosky 1999). Despite evidence that suggests values as “low” as $d = .80$ and $d = .83$ (e.g., Murphy, Cronin, & Tam 2003; Schmitt, Clause, & Pukalos, 1996) for the difference in question, the reigning consensus in the academic literature rests at a mean difference of $d = 1.00$, which has been replicated several times (Hunter & Hunter, 1984; Hough & Oswald, 2000). Said another way, the mean scores of White individuals appear to be one standard deviation higher than the mean scores for Black individuals on cognitive ability measures.

While the accuracy of this estimate will be subsequently discussed, its far-reaching implications underscore the importance of understanding group differences in terms of their magnitude, the construct on which they exist, and such a construct’s consequences for particular

groups in selection. For instance, comparing the average height of men ($M = 70$ in., $SD = 4$ in.) and women ($M = 65$ in., $SD = 3.5$ in.) over twenty years old in the United States reveals a value of $d = 1.33$, which, for illustrative purposes, is comparable to the aforementioned Black-White standardized mean difference in cognitive ability ($d = 1.00$). The value of $d = 1.33$ indicates that on average, males over twenty years old in the U.S. are 1.33 standard deviations taller than U.S. females of the same age (translating to roughly 5 inches in this example, or about 3.75 inches if translated to a value of $d = 1.00$ using the above standard deviations). It follows that standardized difference scores of a similar magnitude are certainly non-trivial, as they might have particularly profound implications for affected groups.

However, it is also worth noting that the magnitude of such differences is only as pertinent as the predictors are valid. Comparing Black-White differences in cognitive ability measures ($d = 1.00$) to the more drastic height differences between professional basketball players and the average U.S. male ($d = 3.00$), effectively demonstrates this point. Said another way, differences in height are inconsequential compared to differences in general cognitive ability when it comes to predicting job performance, magnitude notwithstanding. Thus, it is also worth mentioning that the high criterion-oriented validity that cognitive ability measures display in predicting performance across jobs is precisely what makes them so attractive (see Hunter & Hunter, 1984, Schmidt & Hunter, 1998), a fact that further reveals the costs that Black individuals might pay in selection contexts as a result of such measures. An average difference of one standard deviation between Black and White individuals on cognitive ability measures, then, is particularly problematic when keeping in mind the magnitude, the construct, and the subsequent consequences.

However, more recent research raises much doubt as to the accuracy of the d values collected from Black and White individuals on measures of cognitive ability. Much of this doubt can be attributed to excessive reliance on range-restricted incumbent samples in a variety of studies (Bobko & Roth, 2013). Strong evidence (e.g., Schmidt, Hunter, & Urry, 1976) suggests that exclusive reliance on incumbent samples substantially underestimates values of r (the Pearson correlation coefficient for a given relationship, in this case, that between race and test performance). Considering the relationship between r and d (the two are ordinal transformations of one another), underestimations of r therefore also underestimate values of d (Bobko, Roth, & Bobko, 2001). Insofar as the current literature relies on values of d to accurately index the potential for adverse impact, underestimations of this value by means of incumbent-dominated samples present a major obstacle.

Conducting a series of studies utilizing a within-jobs design (i.e., studying applicants applying for the same position or positions that are comparably complex) Roth et al., (2001) found notably lower d values for tests of general cognitive ability. The study also accounted for moderators like employment status (e.g., applicant or incumbent) and job complexity. For Black and White applicants applying for moderately complex jobs, a d value of .72 was found on general measures of cognitive ability, while a d value of .86 was estimated for the same construct in low-complexity jobs (Roth et al., 2001). It is also worth noting that a meta-analytic within-jobs approach yielded a d value of .38 for incumbent samples, highlighting the tendency of such samples to underestimate values of d (Roth et al., 2001).

Adverse Impact and Personality Measures

Methodological improvements notwithstanding, the updated d values for general cognitive ability calculated by Roth et al. remain too high to ignore. Thus, Black-White applicant differences for other valid predictors of performance must also be examined if adverse impact is to be more optimally reduced. One such approach centers on measures of personality. Receiving support as valid predictors of overall job performance themselves (e.g., Barrick & Mount, 1991; Tett et al., 1991), personality measures typically illustrate smaller subgroup differences than those of cognitive ability (Ones, Viswesvaran, & Schmidt, 1993; Bobko, Roth, & Potosky, 1999; Hough et al., 2001; Foldes et al., 2008). Personality measures also predict job performance across work environments, and have been shown to demonstrate incremental validity above and beyond cognitive ability (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Day & Silverman, 1989; Salgado, 1998).

Current research using the Five Factor Model shows mean differences in personality scores between Black and White individuals to be seemingly negligible. At present, the only noteworthy difference concerns Black individuals scoring anywhere from 0.1 - 0.2 standard deviations below White individuals on the trait openness to experience (Hough et al., 2001; Foldes et al., 2008). Conscientiousness has consistently shown to be the personality trait most relevant to predicting overall job performance (e.g., Barrick & Mount, 1995; Salgado, 1997) and demonstrates even smaller group differences, with values of d ranging from $-.07$ – $.06$ (Hough et al., 2001; Foldes et al., 2008), thereby suggesting anything from slight group differences favoring Blacks to slight group differences favoring Whites.

Applicant vs. Incumbent Scores on Personality Measures

However, these estimates are not exempt from the methodological concerns brought up by Roth et al. concerning measures of general cognitive ability. For instance, despite utilizing a within-jobs approach (e.g., comparing applicants to applicants or incumbents to incumbents) 72% of the organizational sample from the previously cited (and most recently conducted) meta-analysis by Foldes et al., (2008) was comprised of incumbents, suggesting that any reported d values may be underestimates due to a paucity of applicant data. Reviews similar to the aforementioned (e.g., Hough et al., 2001) are privy to the same issues. Thus, there is reason to doubt the accuracy of the currently accepted estimates of Black-White differences in measures of personality, in turn warranting their remeasurement via applicant samples, ideally collected within a variety of occupations.

Predictive Contexts and Test Scores. Applicant-incumbent differences demonstrated in a variety of other constructs (e.g., response distortion, test perceptions) might also affect personality scores and thus, the resulting estimated values of d . At present, evidence exists for applicants scoring approximately one half of a standard deviation higher than incumbents on measures of personality (e.g., Weekly, Ployhart, & Harold, 2004; Stark et al., 2001; Hough et al., 1990), attributed in part to the applicant tendency to “fake good” (i.e., present oneself favorably to a potential employer) (Barrick & Mount, 1996; White & Moss, 1995). Support for this prediction comes in the form of applicant-incumbent group differences in scores being moderated by trait, so that d values for conscientiousness are generally higher than those for extraversion and agreeableness (Weekly et al., 2004). Said another way, applicants might show a greater tendency than incumbents to distort their responses on items that they perceive to be conducive to favorable self-presentation in the eyes of a potential employer, thus reducing the

construct validity of personality. However, additional evidence (e.g., Smith & Ellingson, 2002) suggests that the effects of socially desirable responding on the construct validity of personality inventories is minimal, redundant variance rather than error.

Test Perceptions and Test Scores. How applicants approach the specific situation of an employment test (i.e., test perceptions; see Ryan, 2001 for a detailed discussion pertaining to Black-White scores on cognitive ability tests) might also affect their scores on the basis of motivational, cognitive, and/or affective variables. How exactly test performance varies as a function of these variables and especially as a function of race is rather unclear, though (Ryan, 2001). For instance, evidence exists for applicants demonstrating both strong and weak effects of test perception on performance (McFarland, Lev-Arey, & Ziegert, 2003; Ryan & Ployhart, 2000). Because minority applicants are thought to perform poorly (relative to White applicants) and generally distrust employment tests, they may find themselves disproportionately disadvantaged in such a context (Ployhart, Ziegert, & McFarland, 2003; Ryan, 2001).

Additional evidence (e.g., Ployhart et al., 2003; McFarland et al., 2003) suggests that minority scores on personality measures and Situational Judgement Tests (SJT's) might be more affected by motivating contexts than majority scores by means of *stereotype threat*, or any situation where one potentially affirms a negative stereotype associated with a group that he or she identifies with (Steele, 1992; Weekly et al., 2004). However, strong evidence from Sackett, Hardison, & Cullen suggests that this is not necessarily so, finding no main effect of stereotype threat in both simulated selection contexts and large field studies (Sackett, 2003; Sackett, Hardison, & Cullen, 2004; Cullen, Hardison, & Sackett, 2004). Thus, a firm causal relationship between context and race has yet to be extensively supported.

Hypotheses

With current evidence in mind, the following predictions can be made. First, it is predicted that standardized difference scores (i.e., Cohen's d) between Black and White individuals will be greater than the current meta-analytic estimates. Higher differences are anticipated as a function of the applicant sample used in this study: previous research (e.g., Birkeland et al., 2006; Stark et al., 2001; White et al., 2001) suggests that applicants score higher on personality measures than do incumbents. While the specific magnitude and direction of the interaction between context (i.e., applicant or incumbent) and race is rather unclear in the current literature (see Ployhart et al., 2003; McFarland et al., 2003; Ryan, 2001), limited evidence (e.g., Weekly et al., 2004) suggests that minority scores on personality tests might be more affected by context than those of the majority group. Thus, it is also predicted that Black applicants will score higher on conscientiousness, extraversion, and agreeableness than White applicants. It is important to note that this prediction is offered tentatively, as causal inferences about group differences in personality represent an important direction for further research. However, this is certainly due in part to a low number of primary studies comparing racial group differences in personality in applicant samples. With this in mind, the present investigation seeks to build upon previous results by using a larger, more racially representative sample collected from a different occupation while expecting similar results. Considering the current lack of data from Hispanic individuals available in the current literature, no prediction is offered as to the magnitude or direction of their scores relative to Black and White applicants.

Chapter 3

Methodology

Sample

Participants in the present study were $n = 79,339$ police applicants collected from various law enforcement agencies across the United States. While participants spanned six racial groups, only participants reporting their race as Black, White, or Hispanic were included in the final sample due to sufficient sample sizes ($n = 17,649$, $n = 37,736$, and $n = 20,949$, respectively). To our knowledge, the final sample used in this study is amongst the largest applicant sample reported in the literature ($n = 76,334$), and is also the first of its kind to include a sufficient number of Hispanic individuals ($n = 20,949$).

Measures

SHL OPQ32 and Big Five Measures

Archival police applicant data was obtained from various law enforcement agencies across the United States. The personality measures to which applicants responded were modified from the SHL OPQ32. This assessment normally contains 104 ipsative (i.e., forced-choice) items spread out over 32 different dimensions. However, while the original dimensions were maintained, the number of items was expanded and the response format changed, so that participants rated the extent to which they agreed or disagreed with a given statement on a scale of 1 (strongly disagree) to 5 (strongly agree). Because the scales obtained were proprietary, Cronbach's alpha was incalculable. However, previous large-scale studies demonstrate strong

reliability for the available dimensions, which are displayed in Table 1. The dimensions for which data was available were then averaged and subsequently mapped onto the corresponding “Big Five” personality dimensions (i.e., Openness to Experiences, Conscientiousness, Extraversion, Agreeableness, and Neuroticism; see Bartram, 2013 for a detailed discussion). Table 1 displays a description of the dimensions used along with relevant internal consistency estimates. These low scores are likely due to the limited number of inputs and do not reflect the overall reliability of the measures.

Procedure

Composite scores for each of the Big Five traits were calculated as previously described. Means and standard deviations were then obtained for each racial group, displayed in Table 2. For the traits Openness and Agreeableness, missing data was omitted, resulting in a large reduction in sample size. However, the reduced sample for each of these traits was still very large ($n = 44,740$), seeing Black and Hispanic applicants still sufficiently represented ($n = 9,272$ and $n = 11,321$, respectively). Independent sample t-tests were calculated between each group for each of the Big Five Traits. Significance thresholds were Bonferroni corrected, resulting in a cutoff value of $p = .01$. Standardized difference scores (i.e., Cohen’s d values) were subsequently calculated between groups using the following formula to account for the variation in sample size between groups:

$$d = \frac{(\text{Mean}_2 - \text{Mean}_1)}{SD_{\text{Pooled}}}; \text{ where } SD_{\text{Pooled}} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

Chapter 4

Results

The directionality of the current findings is very different from that in the current literature. Per previous meta-analyses and primary studies, it was predicted that Black applicants would score higher than White applicants on each of the Big Five traits. However Black applicants only scored higher than White applicants on Agreeableness ($M = 3.18$ vs. $M = 3.15$) and Neuroticism ($M = 1.81$ vs. $M = 1.72$). On the other hand, Hispanic applicants' scores on each of the Big Five traits were strikingly similar to White applicants' scores, the only differences being slightly higher White scores on Openness ($M = 4.45$ vs. $M = 4.41$) and slightly higher Hispanic scores on Agreeableness ($M = 3.19$ vs. $M = 3.15$). Because of the lack of studies including Hispanic applicants available for reference, no explicit prediction regarding the direction or magnitude of their scores was offered.

Perhaps the most notable finding of the present study concerns group differences on the trait Conscientiousness. White and Hispanic applicants did not differ at all on this trait, but both scored higher than Black applicants by about .2 standard deviations (Black-White $d = 0.21$, Black-Hispanic $d = .20$). Previous meta-analyses demonstrate Black-White differences in Conscientiousness favoring Black individuals by less than one tenth of a standard deviation (i.e., $d = 0.08$). The current study, however, sees this finding almost triple in magnitude in the opposite direction, favoring White applicants ($d = .21$). This is especially surprising considering findings of a similar primary study (e.g., Weekley et al., 2004), which found that Black applicants scored about one third ($d = .33$) of a standard deviation higher than White applicants on measures of Conscientiousness. The same study found larger group differences in Agreeableness and Extraversion than the present investigation as well ($d = .50$ and $d = .24$), the

former in favor of White applicants, and the latter in favor of Black applicants. These results suggest that personality measures may serve as more of a hinderance than a help as far as reducing adverse impact is concerned.

Perhaps the difference in magnitude between said findings and those of the current study have to do with the job in which the sample was collected. The predictive sample used by Weekley et al. (2004) was comprised of individuals applying for a cashier position, where overreporting traits like Conscientiousness, Agreeableness, and Extraversion is likely to make one look particularly attractive in the eyes of an employer. Police applicants, on the other hand, may have less to lose by not overreporting on these traits, a tendency that might be further reinforced by placing a particular emphasis on honesty in the application process.

While this might aid in explaining the difference in magnitude between the findings of Weekley et al. (2004) and those of the present study, it does not do much to explain differences in directionality. In other words, the jobs from which the samples in question were constructed do not offer much to explain why *Black* cashier applicants might score higher than *Black* police applicants. This avenue for further research will be discussed later in the Discussion section.

Means and standard deviations for each racial group are displayed in Table 2. Group difference scores (i.e., Cohen's *d* values) and the corresponding *t* values are displayed in Table 3. Generally speaking, group difference values were higher than those reported in the primary literature. White and Hispanic applicant scores were strikingly similar on each of the Big Five traits, and both groups scored higher than Black applicants on Conscientiousness, Extraversion, and Openness. Black applicants scored higher than both White and Hispanic applicants on Neuroticism, and higher than Whites but not Hispanics on Agreeableness.

Chapter 5

Discussion

Summary and Implications

For decades, the inclusion of personality measures in personnel selection has been recommended as an effective means for reducing adverse impact (Bobko et al., 1999). However, the present study finds notable subgroup differences in Conscientiousness, amongst the most effective non-cognitive predictors of job performance (Barrick & Mount, 1991; Tett et al., 1991) that work *against* diversity (i.e., see higher scores from White applicants relative to Black applicants). Thus, additional subgroup differences notwithstanding, the present results suggest that including personality measures as a means of reducing adverse impact may be more harmful than it is helpful.

Insofar as performance measures that show group differences might also demonstrate adverse impact, it is important to ensure that estimates of group differences are as accurate as possible. These results demonstrate that any sample seeking to measure such differences would do well to include applicants in addition to or perhaps instead of incumbents, especially considering that any subsequent adverse impact analyses would be based exclusively on applicants (not incumbents) and the selection ratio among those who applied. The adverse impact potential of personality measures as evidenced by group difference scores is also important to consider if personality measures are to continually be coupled with cognitive ability measures as mechanisms for reducing adverse impact. These results, alongside those of Weekley et al. (2004), suggest that primary studies using applicant samples should continue to be conducted

across jobs in order to better ascertain the extent to which racial subgroup differences on personality measures exist. If the present findings persist across occupationally diverse samples, this finding would have important implications for our field's advice regarding best practices for reducing adverse impact in personnel selection.

Limitations and Directions for Future Research

Despite its contributions to the current literature, this study is certainly not without its limitations. First, the present study collected applicant data from a single job, raising questions about the generalizability of the present results to other occupations, especially those that are dissimilar. This is particularly worth considering in light of the high consequences often characteristic of policework. Because of the nature of the job, honest reporting on the part of police applicants might have been overemphasized relative to that of applicants in other lower-consequence professions (e.g., customer service). Thus, future research might be aimed at assessing the generalizability of these results to additional samples and jobs. Samples including a representative number of Hispanic applicants in particular should continue to receive attention, as their responses are disproportionately absent in current studies.

Additionally, the present investigation is not exempt from the varying amounts of response distortion that frequently accompany personality measures. This is worth considering in light of the sample used in this study, which likely reflects exaggerated levels of conscientiousness, extraversion, and openness due to applicants' desire to present themselves favorably to employers (Hough, 1998; Hough et al., 1990; Weekley et al., 2004). While not explicitly examined in this study, it is possible that an applicant's desire to "fake good" might

vary with the presented or perceived saliency of consequences in a particular job, so that when honest responding or high consequences are emphasized, applicants score lower on the aforementioned three traits. However, accurately measuring a potential reduction in such an effect hinges on whether or not “fake” responses can be differentiated from legitimate variance in personality, a rather difficult distinction to make. Thus, future work might work on detecting and reducing response distortion in personality measures.

Third, previous research (e.g., Weekley et al., 2004) points out that the effects of context on personality scores might be better understood in investigations that utilize a within-subjects design (i.e., compare the same individual’s responses when applying for and after having held a position) or research that directly compares the responses of applicants and incumbents from the same subgroup (e.g., the responses of Black applicants to Black incumbents in the same occupation). This limitation also relates to the absence of current research on exactly *why* applicants of a certain race might score higher or lower on certain personality measures. Thus, future work should focus on why Black applicant scores might be lower than Whites, or why White and Hispanic scores might be so similar. Said another way, future focus on whether or not observed personality differences across racial groups are best considered true-trait level differences or reflecting a specific group/testing-environment measurement artifact that has yet to be identified.

Finally, and perhaps most significantly, the lack of available OPQ32 dimensions available to construct Big Five personality variables certainly limits the present conclusions. Certain Big 5 traits in this study (e.g., Neuroticism, Openness) were represented by a single OPQ32 dimension, suggesting that the complexity of certain traits was not accurately represented. However, Conscientiousness, arguably the most consequential of the Big Five in

terms of predicting performance, saw the most dimensions available, and proved to be very high in validity ($\alpha = .93$). While previous studies have demonstrated strong validity of the obtained dimensions, acquiring and combining additional dimensions would greatly enhance the construct validity of the adapted Big Five measures, in turn improving the legitimacy of the obtained results.

Appendix A

Tables

Table 1: Definitions and Reliability for OPQ32 Dimensions and Big Five Traits

Dimension Definition			α_{Trait}	$\alpha_{Dimension}$
<i>Trait/Dimension</i>				
1. Openness			-	
Behavioral	tries to understand motives/behavior; enjoys analyzing people			.78
2. Conscientiousness			.93	
Conscientious	likes to get things finished on time, persists until the job is done			.79
Detail Conscious	focuses on detail, likes to be methodical and organized			.83
Forward Planning	sets longer term goals, thinks well ahead, more likely to take a strategic perspective			.84
Achieving	ambitious and career-centered; likes to set demanding goals			.75
3. Extraversion			.79	
Persuasive	enjoys selling, comfortable using negotiation and persuasion			.72
Controlling	likes to be in charge, takes the lead, tells others what to do			.81
Emotionally Controlled	rarely expresses feelings, avoids displaying emotion			.66
4. Agreeableness			.41	
Caring	feels sympathetic and considerate, particularly helpful/supportive			.72
Independent-Minded	willing to speak one's mind, support own position even if unpopular			.53
5. Neuroticism (R)			-	
Tough-Minded	not easily offended, can ignore insults, rarely takes criticisms personally			.70

Table 2: Means and Standard Deviations of Personality Scores by Subgroup

<i>Trait</i>	<i>Black</i>			<i>White</i>			<i>Hispanic</i>		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
1. Openness	9,272	4.32	.58	24,147	4.45	.49	11,321	4.41	.55
2. Conscientiousness	17,649	4.55	.45	37,736	4.64	.42	20,949	4.64	.44
3. Extraversion	17,649	3.36	.36	37,736	3.41	.34	20,949	3.40	.37
4. Agreeableness	9,272	3.18	.34	24,147	3.15	.33	11,321	3.19	.35
5. Neuroticism	17,649	1.81	.60	37,736	1.72	.62	20,949	1.71	.60

Table 3: Standardized Difference Scores by Subgroup

<i>Comparison</i>	<i>Openness</i>		<i>Conscientiousness</i>		<i>Extraversion</i>		<i>Agreeableness</i>		<i>Neuroticism</i>	
	<i>t</i>	<i>d</i>	<i>t</i>	<i>d</i>	<i>t</i>	<i>d</i>	<i>t</i>	<i>d</i>	<i>t</i>	<i>d</i>
1. B-W	20.63***	.24	23.78***	.21	18.19***	.14	-17.33***	.09	-15.84***	.15
2. W-H	6.68***	.01	1.22	0	3.02**	.03	-15.33***	.12	3.19**	.02
3. B-H	11.60***	.16	21.57***	.20	13.03***	.11	2.36*	.03	-17.18***	.17

Note. W = White, H = Hispanic, B = Black. Negative values indicate that the minority group (e.g., non-White) scored higher, or Black individuals scored higher relative to Hispanic individuals in Black-Hispanic comparison. * $p < .05$. ** $p < .005$. *** $p < .0001$.

Table 4: Comparison of Subgroup Difference Effect Sizes Across Studies

<i>Trait</i>	<i>Present Study</i>			<i>Weekley et al., 2004</i>			<i>Foldes et al., 2008</i>		
	<i>B-W</i>	<i>W-H</i>	<i>B-H</i>	<i>B-W</i>	<i>W-H</i>	<i>B-H</i>	<i>B-W</i>	<i>W-H</i>	<i>B-H</i>
Openness	.24	.01	.16	-	-	-	.10	.02	.10
Conscientiousness	.21	0	.20	-.33 ^a	-	-	-.07	-.08	.05
Extraversion	.14	.03	.11	-.24 ^a	-	-	.16	.02	.11
Agreeableness	-.09	-.12	.03	.50 ^a	-	-	.03	.05	.09
Neuroticism	-.15	.02	-.17	-	-	-	.09	-.03	.02

Note. W = White, H = Hispanic, B = Black. (a) denotes that subgroup difference values were calculated in an applicant sample. Negative values indicate that the minority group (e.g., non-White) scored higher, or Black individuals scored higher relative to Hispanic individuals in Black-Hispanic comparison.

BIBLIOGRAPHY

- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1-26.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of applied psychology*, *81*(3), 261.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and assessment*, *9*(1-2), 9-30.
- Bartram, D. (2013). Scalar equivalence of OPQ32: Big Five profiles of 31 countries. *Journal of Cross-Cultural Psychology*, *44*(1), 61-83.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317-335.
- Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of d for range restriction and unreliability. *Organizational Research Methods*, *4*(1), 46-61.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52*(3), 561-589.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black-White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology*, *66*(1), 91-126.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, *89*(2), 220.
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel psychology*, *42*(1), 25-36.

- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43, 38290–38315.
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five racial groups. *Personnel Psychology*, 61, 579–616.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of applied psychology*, 75(5), 581.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11(2-3), 209-244.
- Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future-- Remembering the past. *Annual review of psychology*, 51(1), 631-664.
- 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(1-2), 152-194.
- Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of applied Psychology*, 83(2), 179.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, 96(1), 72.
- Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of management*, 29(6), 801-830.
- McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance*, 16(3), 181-205.
- Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology*, 88(4), 660.
- Newman, D. A., & Lyon, J. S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology*, 94(2), 298.

- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of applied psychology*, 78(4), 679.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153-172.
- Ployhart, R. E., Ziegert, J. C., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance*, 16(3), 231-259.
- Pyburn Jr., K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity–validity dilemma: Overview and legal context. *Personnel Psychology*, 61(1), 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(2), 297-330.
- Roth, P. L., Bobko, P., & Switzer III, F. S. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology*, 91(3), 507.
- 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(2), 298.
- Ryan, A. M. (2001). Explaining the black-white test score gap: The role of test perceptions. *Human Performance*, 14(1), 45-75.
- Sackett, P. R. (2003). Stereotype threat in applied selection settings: A commentary. *Human Performance*, 16(3), 295-309.
- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, 59(1), 7.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied psychology*, 82(1), 30.

- Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human performance, 11*(2-3), 271-288.
- Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*(4), 473.
- 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*(2), 262.
- 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*, 115-140.
- 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*(5), 719.
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*(2), 211.
- Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied psychology, 86*(5), 943.
- Steele, C. M. (1992). Race and the schooling of Black Americans. *The Atlantic Monthly, 269*(4), 68-78.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel psychology, 44*(4), 703-742.
- Uniform guidelines on employee selection procedures. 41 C.F.R. § 60, pt. 60-3 (1978).
- 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*(4), 433-461.
- White, L. A., & Moss, M. C. (1995, May). Factors influencing concurrent versus predictive validities of personality constructs: Impact of response distortion and item job content.

Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

ACADEMIC VITA

Mitch Dobbs

The Pennsylvania State University

Email: madpsu21@gmail.com

Phone: (xxx) xxx – xxxx

Updated March 26, 2021

EDUCATION

Pennsylvania State University

Expected: May, 2021

B.A. Psychology, Schreyer Honors College

Adviser: Rustin Meyer

Examining Subgroup Differences in Personality: A Predictive Sample Approach

SUMMARY OF RESEARCH INTERESTS

- **Evolutionary Psychology:** evolutionary mismatch, error management theory, cognitive biases and decision-making errors, evolutionary consumer psychology, sexual selection
- **Race, Sex, & Class:** how race, sex, and class interact to influence the formation of social hierarchies and the basis for discriminatory behaviors and attitudes
- **Sexual Aggression/Harassment:** predictors and behavioral correlates of sexually aggressive/sexual harassment behavior in the workplace and in the formation of romantic relationships

RESEARCH EXPERIENCE

Contextual Awareness and Situational Assessment (CASA) Lab October 2018-Present

Undergraduate Research Assistant

- Completed a thorough literature review and obtained and analyzed data in preparation for potential publication of honors thesis project
- Assisted in collecting data, constructing surveys, and entering psychometric measures for use in various studies
- Constructed an undergraduate research manual for the lab detailing the honors thesis-construction process

Gender, Power, and Privilege Lab

January 2021-Present

Undergraduate Research Assistant

- Assisted in collecting and coding data for various lab projects
- Attended workshops discussing racism, sexism, power dynamics and their influence on workplace and sexual behavior

PUBLICATIONS

Manuscripts in preparation

Dobbs, M.A., Ivory, S.L., & Meyer, R.D. (2021). Not So Black & White: Analyzing Group Differences in Personality Using Police Applicants.

PRESENTATIONS

Posters in Preparation

Dobbs, M.A., Meyer, R.D. (2021) Not So Black & White: Analyzing Group Differences in Personality Using Police Applicants. Poster to be presented at the Pennsylvania State University Undergraduate Exhibition, University Park, PA.

Dobbs, M.A., Meyer, R.D. (2021) Not So Black & White: Analyzing Group Differences in Personality Using Police Applicants. Poster to be presented at the Pennsylvania State University Psi Chi Annual Undergraduate Conference, University Park, PA.

Dobbs M.A., Yarwood, M.F., (2021) Conceptualizing and Eliciting Awe: An Evolutionary Approach. Poster to be presented at the Pennsylvania State University Undergraduate Exhibition, University Park, PA.

Dobbs M.A., Yarwood, M.F., (2021) Conceptualizing and Eliciting Awe: An Evolutionary Approach. Poster to be presented at the Pennsylvania State University Psi Chi Annual Undergraduate Conference, University Park, PA.

AWARDS AND COMPETITIVE FUNDING

Department of Psychology, Pennsylvania State University
Costello Family Scholarship Award, Honorable Mention

Pennsylvania State University
J.W. Van Dyke Memorial Scholarship \$2,250

Pennsylvania State University
Baynard & Ethel Kunkle Scholarship \$1,000

Schreyer Honors College, Pennsylvania State University
Academic Excellence Scholarship, \$18,000

Pennsylvania State University
Dean's List (7/7 semesters)

TEACHING EXPERIENCE

Pennsylvania State University, Teaching Assistant
PSYCH 425, Psychology of Human Emotion

Spring 2021

REFERENCES

Dr. Rustin Meyer

Assistant Professor of Psychology
The Pennsylvania State University
Email: rxm667@psu.edu

Dr. Michelle Yarwood

Assistant Teaching Professor of Psychology
The Pennsylvania State University
Email: mxc925@psu.edu