8-2004

Reducing Adverse Impact While Maintaining Validity: Finding the Balance Between Competing Employee Selection Goals

John Ashley Henderson
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by John Ashley Henderson entitled "Reducing Adverse Impact While Maintaining Validity: Finding the Balance Between Competing Employee Selection Goals." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

Robert T. Ladd, Major Professor

We have read this dissertation and recommend its acceptance:

Lawrence James, Michael Lane Morris, Chanaka Edirisinghe

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by John Ashley Henderson entitled "Reducing Adverse Impact While Maintaining Validity: Finding the Balance Between Competing Employee Selection Goals." I have examined the electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

Robert T. Ladd, Major Professor

We have read this dissertation
and recommend its acceptance:

Lawrence James

Michael Lane Morris

Chanaka Edirisinghe

Accepted for the Council:

Anne Mayhew
Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

REDUCING ADVERSE IMPACT WHILE MAINTAINING VALIDITY: FINDING THE BALANCE BETWEEN COMPETING EMPLOYEE SELECTION GOALS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John Ashley Henderson

August 2004

DEDICATION

This dissertation, as well as the entirety of my education, is dedicated to my loving wife, Greta Ann Henderson. She has remained steadfast in her support of my pursuits, and without her love, comfort, and tolerance, I would have lost my way long ago. My thankfulness for her sharing her life is immeasurable.

ACKNOWLEDGMENTS

As the culmination of my educational experience at the University of Tennessee, this dissertation would have not been possible if not for the assistance, teaching, and support from a number of individuals. The faculty and staff of the Industrial and Organizational Psychology program have been absolutely wonderful and have benefited me in too many ways to note. However, I would like to acknowledge two individuals in particular. I would like to thank Dr. Mike Rush for constantly challenging me to become a better writer and a more complete researcher. Your critical analyses have been invaluable. I would also like to thank Ms. June Trbovich not only for her commitment to the students but also for her consistent support and sympathy when the road became a little rough. June, there will never be another like you, and that is a shame.

My wife and I would like to express gratitude toward our families as well. You have both represented beacons of support – financially, emotionally, and spiritually. I am forever grateful.

While I would also like to thank my peers and fellow students in both Chattanooga and Knoxville, there are simply too many to note. Jen, Stephen, Kate, Laura, James, Mark, Dave, and everyone else, you know who you are, and hopefully, you know what you've meant to me.

A special thanks is reserved for Dr. Ken Carson. Without your teachings, encouragement, and confidence in my ability, I would never have embarked on this journey.

Dr. Lane Morris and Dr. Chanaka Edirisinghe, from my dissertation committee, I appreciate your insights, comments, and suggestions. Both of you have contributed to this project in unique ways, and this has resulted in a higher quality and more comprehensible product. Thank you for your time and effort, and I hope to repay your generosity someday.

Dr. Larry James also deserves special mention. Along with your incredible insights on almost all of the subjects within our field, you have taught me to think creatively while always enjoining the basics tenets of good science. In fact, I consider myself a scientist because of your contributions, and for that, I am in your debt.

Finally, I would like to acknowledge my dissertation chair and mentor, Dr. Robert Ladd. I will never be able to sufficiently relate the personal benefits I have received from our relationship as professor and student, nor will I even begin to describe the long-lasting impact of your advice and friendship. Whether it involved improving my critical thinking skills, advancing this research, or simply teaching me to make better decisions, your efforts have resulted in me becoming a better researcher, scientist, consultant, statistician, and person. Unfortunately, I fear that this will have to remain a debt never fully paid, but I will always looking for ways to make a dent. Thank you.

ABSTRACT

Adverse (or disparate) impact has probably represented one of the most persistent and pervasive problems in employee selection. Innumerable approaches to eliminating its presence have been attempted, but most have been met with limited success. To date, this success has been measured in only slight reductions in adverse impact unless substantial losses in validity are accepted. While a number of reasons for these results have been advanced, this research asserted that part of the problem originated in the narrow perspective with which employee selection is often defined. This narrow perspective has resulted in a singular focus on validity with insufficient attention allocated to multiple criteria. The purpose of the present research was to expand upon an earlier study (Henderson & Ladd, 2001) that introduced a methodology (constrained estimation) that incorporated multiple objectives into the decision-making process associated with employee selection. Specifically, the goals of the methodology included reducing adverse impact while maintaining validity. In order to test the efficacy of this methodology, constrained estimation was applied to both Monte Carlo data as well as archival data obtained from an assessment project conducted from 1992 to 1993. It was also compared to two commonly used predictor weighting methodologies – Ordinary Least Squares regression and Unit Weighting. Results suggested that constrained estimation was moderately successful in reducing, but not eliminating, adverse impact while maintaining validity. Implications, limitations, and suggestions for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Organizations live, work, and function within boundaries, boundaries set forth economically, socially, and legally. From an economic perspective, despite hardships such as downturns in the market or difficulty recruiting new employees, organizational leaders, individual workers, and, when applicable, shareholders expect a degree of profit from their initiatives. Socially, organizations are often expected to make contributions to their communities, promote a general sense of citizenship throughout their culture, and to remain directly responsible not only for their actions but also for the impact of their actions on society as a whole. Finally, organizations are expected to achieve their economic and social goals within the confines of legally accepted behavior. Frequently, however, difficult issues arise that traverse these three parameters in a conflicting manner. Sometimes they simply pressure the boundaries to stretch; but occasionally they operate well within the confines of one boundary while simultaneously puncturing the walls of one or both of the remaining boundaries. One such issue is that of adverse impact in personnel decision-making.

Adverse impact basically refers to a substantial difference in employment-centered selection rates of individual subgroups (e.g., Caucasians and African Americans). It can result from a number of organizational procedures such as promotions, training, layoffs, and even performance appraisals, but it is probably most often associated with initial hiring decisions. These hiring decisions are typically based, at least in part, on applicant performance on one or more predictors that have been weighted according to some pre-specified guidelines. In general, predictors can be weighted in any number of ways, but ordinary least squares (OLS) regression and unit weighting are probably the most common methodologies. OLS regression focuses on minimizing errors of estimation and results in an equation that defines the relationship between the predictors and the criterion of interest (e.g., job performance) in terms of a weighted composite of variables. For example:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k$$

Y refers to the predicted scores on the criterion (e.g., job performance), $b_1, \ldots, b_k$ refer to the weights (i.e., regression coefficients) attributed to each predictor variable, and $X_1, \ldots, X_k$ refer to individual scores on each predictor. When working with Z scores, the regression coefficients (now called beta weights), as well as the criterion and predictor scores, are transformed to standardized values:

$$Z_Y = \beta_1 Z_{X1} + \beta_2 Z_{X2} + \ldots + \beta_k X_{Xk}$$

In contrast, unit weighting uses a reduced variance model where each of the predictors are weighted 1.0. That is, the various predictors are simply added together. Regardless of the weighting scheme, the ultimate goal is to achieve a predictor composite (the result of the above noted equations) that can be used to make distinctions between those individuals predicted to be more or less successful on the job. Subgroup differences on this predictor composite play a key role in determining how much adverse impact a selection scheme produces.

Within this context, adverse impact can be generally defined as some substantial difference in a predictor or predictor composite that results in the disproportionate selection of one subgroup's members (e.g., the majority group) over another (e.g., the minority group) with the basis for subgroup membership defined by factors such as race, sex, and national origin. As a rule of thumb, the traditional test of this difference is the "four-fifths" rule (see Appendix B for an example of how this test is applied), which states that adverse impact exists if the selection rate of the minority group is less than four-fifths of the majority group (*Uniform Guidelines*; Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978). While the presence of adverse impact does not always reflect evidence of illegal discrimination (Guion, 1998), it is typically viewed as an undesirable characteristic in selection and other employment decisions, and its existence may leave an organization vulnerable to legal challenge. Valid selection systems significantly reduce this vulnerability, but potential court costs and social perception can carry difficult burdens in and of themselves. Thus, considerable time and energy have been devoted to understanding and exploring why adverse impact occurs and what should be done to combat its effects.

For many, the use of cognitive ability testing represents one of the primary forces driving adverse impact. While the empirical literature would suggest that for many or most job categories cognitive ability generally predicts job performance as well as, if not better than, other predictors (Hunter & Hunter, 1984; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Nathan & Alexander, 1988; Ree & Earles, 1991; Schmidt & Hunter, 1998; Schmidt, Hunter, McKenzie, & Muldrow, 1979; Schmitt, Gooding, Noe, &

Kirsch, 1984), a substantial mean test score difference (about one standard deviation) is found between African-American and Caucasian subgroups (Gottfredson, 1988) for all commonly used cognitive ability tests. Whereas this substantial difference routinely results in significantly fewer minority group members being chosen for jobs where cognitive ability is utilized as a predictor, cognitive ability testing has been found to be equally valid for both Blacks and Whites (Steffy & Ledvinka, 1989) and is generally viewed as practical, moderately inexpensive, and highly reliable (Wagner, 1997). It has also been shown to possess a high degree of utility with resultant savings in employee training as well as improved productivity from the workforce (Schmidt & Hunter, 1981; Schmidt et al., 1979). Nevertheless, while cognitive ability predicts job performance validly and proficiently, and maybe even more universally than any other predictor available (Huffcutt, Roth, & McDaniel, 1996), we can also not escape the social and potential legal ramifications of it's use. A one standard deviation difference between majority and minority subgroups on a predictor or predictor composite almost guarantees significantly (and substantially) different hiring rates. Therefore, the reduction of adverse impact in selection poses a significant quandary for human resource professionals and researchers alike. On one hand, employers are motivated to select individuals who will maximize workforce productivity; however, on the other hand, societal and legal pressures stipulate a need to employ a diverse workforce where the opportunities to succeed are not limited by, for example, the color of one's skin. These seemingly conflicting goals have caused some researchers to question, "How can [employers] use valid procedures in a manner that optimizes the expected performance of their workforce and at the same time employ a demographically diverse workforce?" (Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997, p. 719).

While numerous approaches to answering this question have been addressed, few have met with more than limited success, and most have demonstrated an inability to rectify this issue. For example, administrative options such as subgroup norming and score adjustments (Sackett & Wilk, 1994) appeared to be gaining momentum in the 1980s until the Civil Rights Act of 1991 made these practices illegal. Because of its success in increasing minority representation, the use of banding with minority preference (Cascio, Outtz, Zedeck, & Goldstein, 1991) obtained a degree of prominence in the late 1980s and early 1990s, but it has since lost some of its luster because it is generally viewed as inconsistent with the goals and verbiage outlined by Civil Rights laws (Guttman, 2000). Moreover, the use of those banding strategies that reduce adverse impact consistently result in less effective workforces when compared with traditional top-down selection (Schmidt, 1991).

In recent years, the field of Industrial and Organizational Psychology has shifted much of its focus toward noncognitive predictors (Murphy, 1996), and many researchers have hoped to serendipitously benefit from the fact that these predictors often exhibit less adverse impact than cognitive ability. However, like banding, the use of predictors other than cognitive ability has often been associated with a drop-off in expected job performance (Campbell, 1996) and thus, detracts from the goal of optimizing a workforce's productivity. One approach to reducing adverse impact in selection systems that has garnered considerable attention is the idea of combining predictors that exhibit smaller group differences with cognitive ability. In fact, up until the late 1990s, there were sufficient studies and papers espousing the benefits of this approach (see Campbell, 1996; Hunter & Hunter, 1984; Reilly, 1996) to suggest that this procedure might be the panacea that closes the lid of Pandora's box. The rationale for this methodology was quite simple and very appealing. Use a predictor composite of cognitive ability along with one or more predictors displaying smaller group differences, and in return, dilute the mean group differences associated with cognitive ability and gain incremental validity (Sackett & Ellingson, 1997). However, although this procedure has reduced adverse impact to a degree and generally increased the observed validity, it has failed to sufficiently compensate for the population differences often attributed to cognitive ability testing (Pulakos & Schmitt, 1996; Ryan, Ployhart, & Friedel, 1998).

Finally, some researchers have suggested differentially weighting criterion domains within a criterion composite as a potential resolution. Basically, if different predictors are associated with different facets of performance, and some of these predictors manifest minimal subgroup differences, then placing greater emphasis on alternative criteria could reduce adverse impact when multiple predictors are used (Campbell, 1996). However, as with previous approaches, although some modest reductions in adverse impact occur as a result of differential criterion weighting, these reductions have yet to become sufficient in eliminating adverse impact, and they are often associated with reductions in the anticipated job performance of the selected workforce (Hattrup, Rock, & Scalia, 1997).

In short, though we have been able to reduce adverse impact, even substantially at times, we have been unable to remove its presence while sustaining the validity and overall utility accrued from the use of cognitive ability. From a predictor standpoint, to assuage the conflict of adverse impact associated with our selection schemes, removing cognitive ability testing from the equation would appear to be our only answer. However, it is important to recognize that cognitive ability is not necessarily the most troubling issue. In fact, the actual problem lies in the real world differences observed within the labor market. Cognitive ability is not the only predictor that results in large group differences. We also often find adverse impact with other predictors as well as with combinations of predictors that are generally purported to represent alternatives to cognitive ability (Bobko, Potosky, & Roth, 1999). Moreover, several reviews have shown that there are substantial differences in the job performance of majority versus minority hires (see Bernardin, 1984; Ford, Kraiger, & Schechtman, 1986; Hunter, Schmidt, & Rauschenberger, 1977; Kraiger & Ford, 1985; Roth, Huffcutt, & Bobko, 2003), which would necessarily increase the occurrence of adverse impact.

In the end, these differences, while controversial, are real, and researchers and practitioners in the fields of selection and employment law are faced with the challenge of generating some resolution. Because previous attempts have been less than successful, this study argues that the problem should be viewed from a new perspective and possibly from the lens of different disciplines.

Reviewing the situation, there are two conflicting goals – maximizing productivity and minimizing adverse impact. Hence, there should be at least some focus on dealing with multiple criteria. The problem with this perspective becomes three-fold. First, adverse impact is not generally considered as a direct criterion. Selection systems are frequently designed with the goal of maximizing validity with some ancillary thought to other issues such as increasing tenure or diversity. However, if the objective is to solve a specific problem (e.g., adverse impact), then that problem should be explicitly addressed. Secondly, there would also be some question about the legality of this approach. To protect against legal challenge, any selection system resulting from the use of adverse impact as a criterion would have to demonstrate substantially similar validity to any system that excludes adverse impact from consideration. Per the *Uniform Guidelines on Employee Selection Procedures* (EEOC et al., 1978):

> where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact (p. 38297)

Finally, the field most associated with this issue, Industrial and Organizational Psychology, has not demonstrated an acute or masterful command for dealing with multiple criteria. Much of this stems from traditional methodologies and procedures. When designing selection systems, some predictor-weighting scheme is required to make final decisions on who is eventually hired. These schemes are sometimes generated from unit weights but probably more often from ordinary least squares (OLS) regression (Schmidt, 1971). However, OLS regression and unit weighting are univariate procedures when considering criteria; multiple predictors can be utilized, but only a single criterion can be predicted. Thus, researchers have been forced to either rely on these single criterion predictions or generate some criterion composite that many would argue could not be accurately interpreted (Schmidt & Kaplan, 1971).

As a result, these weighting schemes, although sufficient in many situations, simply cannot provide the necessary information when utilizing two or more criteria. However, the single index (in this case the weights attributed to each predictor), typically provided by the use of regression or unit weighting and a composite criterion, is necessary when making administrative decisions (Landy, 1989). Therefore, whether a researcher's interests lie in decreasing employee theft or maximizing job performance while minimizing adverse impact, a procedure or methodology is needed that can accurately reflect the importance of multiple criteria in administrative decisions while still providing a specific decision-making model.

The purpose of this research was to expand upon an earlier study (Henderson & Ladd, 2001) that introduced just such a methodology in the form of constrained estimation. Basically an optimization technique drawn from the field of Management Science, constrained estimation provides researchers with the ability to optimize multiple criteria in order to accomplish complementary, or sometimes conflicting, objectives. Comparatively, while the sole objective of OLS regression is the minimization of errors in estimation (i.e., maximizing prediction), constrained estimation allows for the minimization or maximization of multiple objectives. In the present instance, the objectives were to minimize errors in estimation while simultaneously reducing the expected adverse impact of a selection program by minimizing subgroup differences on the predictor composite.

A secondary purpose of this paper involved explicating the origins of adverse impact while also discussing both previous attempts at amelioration and why these attempts have generally failed. This was done by reviewing the literature surrounding adverse impact and exploring the various topics related to its existence. Emerging from this review, this study hoped to support the argument that rather than focusing most of our efforts on validity, we should be looking at utility while searching for some optimal balance between the economic, social, and legal issues constraining our decisions.

CHAPTER II

REVIEW OF THE LITERATURE

Beginning in the mid 1960's and continuing through both numerous court decisions and congressional legislation, the issue of unfair or illegal discrimination has been defined, codified, and further adapted to cement the role of Industrial Psychology within the legal realm of employment testing. While a decision in *Myart v. Motorola* (1964) set the precedent for hearing these types of claims (Arvey & Faley, 1992; Cohen, 1974), the first major step intended to address concerns of this nature occurred with the implementation of Title VII of the Civil Rights Act of 1964. The intention of this legislation was to eradicate discriminatory barriers in any and all employment decisions. Illegal employment discrimination at this time was referred to as disparate treatment and defined as "evil intent" (Bolick, 1988). However, over the course of the next several years, influential members of congress and various communities were successful in helping to redefine illegal discrimination in terms of both intent and effects (Sharf, 1988). Consequently, in *Griggs v. Duke Power Co.* (1971), a unanimous Supreme Court ruled that discriminatory motive was not required and that "Congress directed the thrust of the Act to the consequences of employment practices, not simply the motivation". With this decision came a second definition of discrimination and the official birth of adverse impact as an employment concept.

With the *Griggs* decision in hand, industrial psychologists and other human resource specialists were confronted with new challenges revolving around the defense of their testing procedures. Previously, the focus of illegal employment practices was on intentional discrimination with responsibility falling on human resource departments and decision makers that make individual hiring decisions rather than the field of Industrial Psychology that generates group-based tests and selection systems. Moreover, traditional testing methodologies were seemingly insulated by section 703h of Title VII (also known as the Tower Amendment) which states that it shall not "…be an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided such test is not designed, intended, or used to discriminate because of race, color, religion, sex, or national origin". However, the *Griggs* decision altered the landscape so dramatically as to insulate all protected classifications from the onerous "effects" of these same traditional practices. Moreover, shortly following the resolution of this case, the courts rendered decisions in a number of disputes associated with adverse impact challenging inappropriate validation practices, arbitrary cutoffs, and inadequate job analyses (see *Fowler v. Schwarzwalden*, 1972; *United States v. Detroit Edison Co*., 1973; *Boston Chapter, NAACP v. Beecher*, 1974; *Albermarle Paper Company v. Moody*, 1975). Thus, the rise of adverse impact forced industrial psychologists to remain more cognizant in their efforts to eliminate discriminatory barriers and to specifically attend to the use of those attributes that might make adverse impact more likely. Although the basic practices of validation work, job analyses, and setting cutoffs have improved, employers are still faced with the problem of determining what applicant characteristics will be used to select candidates for employment that will not result in adverse impact. While scrutiny of some applicant characteristics can be more readily resolved due to their seemingly closer, more recognizable relationships with the demands of the job in question (e.g., height, weight, age requirements, and physical ability), the most complicated and contentious issue appears to be with the use of employment tests (Cascio, 1991) – specifically, cognitive ability tests.

The Role of Cognitive Ability

Some researchers have argued for the near universal acceptance and use of cognitive ability testing in selection (Wagner, 1997), and their arguments show some validity. Study after study has demonstrated that cognitive ability is not only a very strong predictor of training results (Ree & Earles, 1991) but also of both educational achievement (Campbell, 1996) as well as job performance across a wide variety of jobs (Hunter, 1986; Schmidt & Hunter, 1998). Observed validity coefficients between cognitive ability and job performance regularly fall at least in the range of .20 to .30 (Ghiselli, 1966, 1973; Hartigan & Wigdor, 1989; Wigdor & Garner, 1982) and oftentimes reach as high as .60 or more when corrected for artifacts such as restriction of range and unreliability in the criterion (Hunter & Hunter, 1984; Schmidt, Hunter, & Caplan, 1981; Schmitt, Gooding, Noe, & Kirsch, 1984). Supplementing this evidence, cognitive ability testing is also highly reliable (Wagner, 1997) and oftentimes less expensive and more feasible than other predictors showing similar validity (Gatewood & Feild, 1998). Moreover, substantial utility may be realized with the use of cognitive ability measures. For example, Schmidt, Hunter, McKenzie, & Muldrow (1979) found a savings of about $376 million if the federal government were to adopt cognitive ability testing in the selection of computer programmers. Hunter and Schmidt (1982) estimated a national economic bonus in terms of productivity of about $80 billion per year if this type of testing was universally utilized in selection. Thus, with the above criteria in mind, one could easily understand why cognitive ability might gain universal acceptance.

However, on the other side of this equation, one could argue that cognitive ability testing is a major roadblock in the resolution of adverse impact issues. While it's use often results in a workforce that greatly increases productivity, cognitive ability testing commonly results in about a one standard deviation difference between African American and Caucasian subgroups (Gordon, 1986; Hernstein & Murray, 1994; Jensen, 1980, 1985; Loehlin, Lindzey, & Spuhler, 1975; Schmidt, Greenthal, Hunter, Berner, & Seaton, 1977). Given that a standardized group difference of about 1.0 would result in minority hiring rates of .013, .159, and .610 for selection ratios of .10, .50, and .90, respectively (Sackett & Wilk, 1994), adverse impact is likely a resulting consequence when cognitive ability is used in selection.

At the same time, there is little to no evidence that this adverse impact is the result of any bias toward African Americans. In fact, considering the proliferation of studies surrounding this issue over the last 30 some odd years, research would specifically suggest otherwise. Initially, investigations focused on possible differences in subgroup validity coefficients – namely instances of single-group validity (i.e., tests may be valid for the majority but invalid for the minority) and differential validity (i.e., significant differences between the validity coefficients obtained for two subgroups). However, the accumulated evidence revealed that findings of single-group validity (e.g., Boehm, 1977; Katzell & Dyer, 1977; O'Connor, Wexley, & Alexander, 1975; Schmidt, Berner, & Hunter, 1973) and differential validity (e.g., Bartlett, Bobko, Mosier, & Hannan, 1978; Hunter, Schmidt, & Hunter, 1979; Ruch, 1972) were rare and generally occurred no more frequently than what might be expected by chance alone.

Because the lack of single-group validity or differential validity does not preclude a lack of predictive bias (Cascio, 1991), researchers moved on to investigations of differences in the slopes and intercepts between subgroup regression lines. This arena was called differential prediction, and the results of this research followed a similar pattern as with previous research with only chance level differences found in slope comparisons but occasionally significant intercept differences. For instance, in 1,190 racial group comparisons, Bartlett et al. (1978) observed significant slope differences about 5% of the time whereas significant intercept differences maintained a rate of about 18%. Moreover, in a review of 72 General Aptitude Test Battery (GATB) validity studies, Hartigan and Wigdor (1989) noted that significant slope and intercept differences occurred at rates of about 3% and 37%, respectively. So, while significant slope differences rarely appear, significant intercept differences (typically favoring majority group members) are a more common phenomenon. At one point, Jensen (1980) argued that these intercept differences could have resulted from less that perfect reliability in the predictors; but regardless of the origin, it is important to remember that significant intercept differences do not equate to a bias that would lead to adverse impact. On the contrary, these particular differences, along with the use of a common regression line, would result in the overprediction of the minority group's performance.

All in all, it would appear that the prediction of job performance with cognitive ability pursues basically the same path when comparing majority and minority applicants (Schmidt, 1988; Wigdor & Garner, 1982). In general, low test scores are associated with lower job performance for both subgroups just as high test scores are associated with higher job performance (Bartlett et al., 1978; Grant & Bray, 1970; Jensen, 1980; Schmidt, Pearlman, & Hunter, 1980). As a whole, the evidence would suggest that cognitive tests are equally valid for both African Americans and Caucasians, and that adverse impact is not a result of any bias associated with these tests.

Present day, the real issue surrounding the use of cognitive ability testing is not subgroup validity or necessarily bias but rather some perception of fairness. While bias is interpreted as a psychometric issue, fairness carries a more social, judgment-laden connotation (Steffy & Ledvinka, 1989). Unfortunately, this is a much more difficult subject to tackle because not only are there numerous perspectives on what constitutes fair test use, there have also been several definitions of fairness proposed in the literature with each often contradicting the findings of others. Some examples of the various definitions of fairness include the subjective regression model (Darlington, 1971), the equal risk model (Einhorn & Bass, 1971; Guion, 1966), the constant ratio model (Thorndike, 1971), and the conditional probability model (Cole, 1973). However, the most commonly accepted definition, as well as the one approved by the *Uniform Guideline on Employment Selection Procedures* (EEOC et al., 1978), the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 2002), and the *Standards for Educational and Psychological Tests* (American Psychological Association, 1985) is Cleary's (1968) regression model. This model was originally suggested by Humphreys (1952) and holds that selection is fair only if the prediction errors sum to zero for all groups considered. Stated another way, fairness is investigated by comparing the slopes and intercepts associated with each subgroup for significant differences and determining if a specific predictor score results in the same predicted criterion score regardless of which subgroup the score comes from.

While the Cleary model is considered to be the least biased because it maximizes prediction and minimizes error, there has been some argument that it maximizes utility while minimizing minority hiring (Cronbach, Yalow, & Schaeffer, 1980; Hunter, Schmidt, & Rauschenberger, 1977). For some, the central problem revolving around the use of this model concerns its institutional perspective. By focusing its efforts on maximizing the number of individuals who would be successful on the job and minimizing the proportion of those individuals who would not (i.e., false positive error), the regression model ignores that population who would not be selected for the job but would otherwise be successful (false negatives). This culminates in some inequity for minorities because a predictor mean score difference between two groups results in a greater proportion of false negatives for the lower scoring group regardless of whether or not differential prediction is detected (Campbell, 1996; Cascio, 1991). Given this fact, the use of cognitive ability in selection is obviously further complicated by its typical one standard deviation mean score difference between African Americans and Caucasians.

In the end, cognitive ability testing presents a quandary to those who wish to employ a productive workforce while eliminating adverse impact. Cognitive ability testing is technically fair, inexpensive, reliable, and highly feasible. It also shows tremendous utility and predicts job performance with at least about the same validity as, if not more than, other predictors. However, its use will almost definitely result in adverse impact and most assuredly lead to a greater number of false negatives for the lower scoring minority group. Though these two opposing viewpoints suggest a mathematically inescapable tradeoff between minority employment and maximum productivity (Steffy & Ledvinka, 1989), researchers have endeavored to find methods and additional predictors that would concurrently satisfy two of our most important goals.

Reducing Adverse Impact

The basic premise behind selection is fairly simple – select the best, most qualified individuals for employment within an organization and, subsequently, reap the benefits of productivity and organizational success that comes with those individuals. To this end, researchers have utilized a number of different predictors with the goal of achieving the highest level of validity possible. Cognitive ability testing has achieved that goal both effectively and efficiently. However, racial diversity became an important, if not legislatively mandated, goal with the passage of Title VII of the Civil Rights Act of 1964 (Murphy & Shiarella, 1997). Thus, the pinnacle of selection research has now been cast in the mold of high validity with small to nonexistent subgroup differences in the workforce. In scaling this peak of prediction, numerous approaches have been attempted in order to secure or maintain this degree of validity while eliminating or reducing adverse impact. For the most part, these approaches fall into four main categories: (1) Procedural/Administrative methods, (2) Alternative predictors, (3) Expanding the criterion domain, and (4) Predictor combinations.

Procedural/Administrative Methods

One of the first procedural methods used to combat group differences and adverse impact came in the form of score adjustments (e.g., within-group norming, bonus points, and separate cutoffs) for minority test scores. The use of quotas and other race norming tactics are also usually discussed within this arena. One example of the use of score adjustments occurred in the early 1980's when the United States Employment Services (USES) converted GATB test scores to percentile scores within racial groups for the purposes of job referral. This was done because of the mean group differences found in the test scores of African Americans, Hispanics, and Caucasians. As a result, when an employer requested candidates for a particular position, they received a list of those individuals that scored within a specific range for their racial group rather than the raw scores across all groups. Advocates argued that this type of minority preference was warranted because of: (1) the value of increased minority representation, (2) bias in test measurement, and (3) disagreements over what constituted fair test use (Sackett & Wilk, 1994). However, though supported by the National Academy of Sciences (NAS; Hartigan & Wigdor, 1989), public outcry and opposition by the U.S. Department of Justice resulted in Congress declaring these practices illegal in section 106 of the Civil Rights Act of 1991. Specifically, this provision made the following unlawful:

> in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of employment-related tests on the basis of race, color, religion, sex, or national origin

Another procedural option emerging in the late 1980's was banding (Cascio, Outtz, Zedek, & Goldstein, 1991). Banding refers to defining a range of scores based on the standard error of measurement and then considering all of the scores found within a band to be equivalent. The idea behind this technique is that the differences between scores within a band may reflect measurement error or imperfection rather than differences in ability. Though researchers have shown that banding can reduce adverse impact, this decrease depends upon the specific banding strategy chosen (Hauenstein, Bess, Swartz, & Byrd, 2001). Campbell (1996) suggested four different strategies: (1) selecting everyone within a band, (2) selecting at random within a band, (3) basing selection on additional predictor information for those located within a specific band, and (4) using minority preference as additional information within a band. One can also distinguish between fixed bands (i.e., score intervals that do not change) and sliding bands (i.e., bands slide down to create new bands as soon as every individual with the highest score within a band is chosen). Obviously, the first of Campbell's strategies is similar to top-down selection and has little value in reducing adverse impact. Random selection and the use of additional predictors result in decreases, but these decreases are rarely sufficient. In general, the most successful strategy comes with using sliding bands and minority preference (Cascio, et al., 1991; Sackett & Wilk, 1994). However, aside from psychometric challenges to the methodology itself (see Schmidt, 1991; Schmidt & Hunter, 1995), the use of minority preference should be viewed as a red flag that raises significant legal questions. Additionally, while the use of banding has withstood legal scrutiny by the courts (see *Bridgeport Guardians v. City of Bridgeport*, 1991, *Officers for Justice v. CSC*, 1992, and *United States v. City and County of San Francisco*, 1992), this scrutiny involved cases decided before enactment of the Civil Rights Act of 1991. Moreover, the support from the courts in these cases was based on using minority preference to break ties within bands rather than for the use sliding bands with strict minority preference. In fact, both the 2nd and 9th Circuits specifically frowned upon this practice (Guttman, 2000). This is important because utilizing minority preference only to break ties is unlikely to reduce adverse impact to any significant degree (see Cascio et al., 1995; Murphy, Osten, & Myors, 1995). In sum, while banding is capable of playing an important role when sliding bands are available due to some type of affirmative action consent decree, its use in day-to-day selection activities appears limited.

A final administrative option was suggested and researched by Sackett and Roth (1996). Realizing that cost and logistical constraints can sometimes hamper the selection of employees, these researchers looked at how the use of hurdles influenced group differences in final selection decisions. Hurdles are basically used by screening on one or more predictors and then making hiring decisions on either other predictors or some composite of predictors. Sackett and Roth generated a number of different scenarios based on two predictors (modeled after cognitive ability and integrity) that were uncorrelated with each other and that displayed drastically different degrees of adverse impact. They also varied the scenarios by making the two predictors either equal or unequal in validity and by observing every possible combination of the two predictors within various hurdling strategies. As a result, they found that the use of hurdles in selection could lead to increased rates of minority hiring. However, these increases were consistently associated with decreases in the overall anticipated performance of those being hired and only occurred when the predictor displaying large group differences (and in this case the greatest validity) played a minor role in the selection system. As an additional complication, the use of hurdles in this manner would just as likely suffer adverse impact claims as using cognitive ability alone. Under the “bottom line” rule (see *Connecticut v. Teal*, 1982), the Supreme Court noted that an overall finding of no disparate impact from a selection system means little if at any stage in that system some group of individuals were impacted adversely. Therefore, organizations would be forced to choose between either eliminating a predictor with large group differences regardless of its validity (and suffer with lower overall performance by their workforce) or use the predictor within a hurdling strategy with the very strong possibility that adverse impact would occur at some stage.

Alternative Predictors

As an initial result of Title VII and the *Griggs* decision, the use of cognitive ability testing dropped off from fear of legal challenge (Arvey & Faley, 1992). In its place, researchers began searching for "alternative" predictors that would both demonstrate similar levels of validity and satisfy our social conscience by exhibiting small group differences. Thus began a shift toward more noncognitive (or maybe less cognitive) predictors that rarely display the same degree of adverse impact as cognitive ability (Schmitt, Clause, & Pulakos, 1996). For example, in a meta-analytic review of subgroup differences on job sample tests between Caucasians and African Americans, Bernardin (1984) found an average difference of .54 standard deviations. Similarly, Verive and McDaniel (1996) discovered a .42 standard deviation difference using short-term memory tests. While Hoffman and Thornton (1997) report that observations of significant Black-White differences in assessment centers are fairly evenly split between studies (e.g., Byham, 1983; Friedman, 1980; Huck & Bray, 1976; Jaffee, Cohen, & Cherry, 1972), when differences are found, they are typically smaller than with cognitive ability (Goldstein, Yusko, & Braverman, 1996) with exercises like leaderless group discussions and role plays falling under .25 standard deviations. Reduced Black-White differences can also be seen with structured interviews (.24 standard deviations; Huffcutt & Roth, 1998), biodata instruments (.33 standard deviations; Bobko, Roth, & Potosky, 1999), and both integrity tests as well as personality factors such as conscientiousness (below .10 standard deviations; Bobko et al., 1999; Ones & Viswesvaran, 1998). Additionally, studies have shown that the observed validities for these predictors have traditionally ranged anywhere from .16 for the factor conscientiousness (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990) to .35 for the structured interview (Pulakos & Schmitt, 1995).

With low-fidelity simulations remerging in the 1990's, Motowidlo, Dunnette, and Carter (1990) created a type of situational judgment test where applicants are presented with various scenarios in written form and asked to endorse one of several responses provided. Research in this area suggests that the validity for this type of test falls in the range of .13 to .37 (see Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996), but this is usually associated with less adverse impact. This study in particular found group differences (favoring Caucasians) of .21 standard deviations with job incumbents and .41 standard deviations with job applicants. More recently, Chan and Schmitt (1997) argued that different methods of testing might alleviate adverse impact concerns and supported this assumption by demonstrating substantially smaller Black-White differences on a video-based situational judgment test (.21 standard deviations) versus the traditional paper-and-pencil variety (.95).

These observations have given rise to the hope that predictors such as these could possibly replace cognitive ability in our various schemes. However, regardless of the benefits to minority representation, it is important to note that many of the reviewed predictors rarely surpassed cognitive ability in terms of validity on a consistent basis, and thus, there use would often result in an explicit drop-off in expected job performance. Furthermore, Sackett and Wilk (1994) have shown that a standardized difference of only .25 is needed to produce adverse impact when the selection ratio is .50 and only .14 is needed when the selection ratio is .10. Given the fact that predictive accuracy is decreased along with the idea that, out of the above predictors, only personality and integrity testing might give reasonable adverse impact rates at low selection ratios, it is difficult to find reason to recommend these predictors over cognitive ability when a specific situation does not necessarily warrant a specific alternative predictor.

On one final note, the term noncognitive in this situation is sometimes a misnomer because many of the alternative predictors can, and frequently do, reveal a significant cognitive element (e.g., biodata, structured interviews, and assessment centers). This common cognitive element might help to explain why the elimination of adverse impact has been so difficult.

Expanding the Criterion Domain

Because different elements of performance might be best predicted by an array of different predictor variables, some researchers have contended that the differential weighting of multiple criterion variables along with the use of multiple predictor variables might facilitate the reduction of adverse impact. This would be especially true if some of the predictor variables exhibited small group differences. Hattrup, Rock, and Scalia (1997) tested this hypothesis by varying the importance of two different types of criteria (contextual and task performance) in a criterion composite. However, using cognitive ability and a measure of work orientation as predictors, Hattrup et al. found that adverse impact, though reduced in most scenarios, was only eliminated when they used a selection ratio of at least .80 and when either cognitive ability was dropped from the equation or contextual performance was given five times as much weight as task performance. Moreover, this situation resulted in a lower explanation of variance (.12) than when equally weighted criteria were used (.19) as well as an uncertain explanation as to how much actual task performance could be expected from the workforce because of the substantially different weight it received in the composite. Thus, while adverse impact was reduced, this approach was unable to fully eliminate group differences while preserving validity.

Predictor Combinations

Over the last 10 years, one of the most common approaches to the adverse impact-productivity dilemma has been combining predictors that display small group differences with those that exhibit large group differences. The hope was that the resulting predictor composite would demonstrate sufficiently reduced group differences as to remove adverse impact from the scenario. However, this has not worked quite as expected. For example, Pulakos and Schmitt (1996) used a composite consisting of a structured interview, a biodata instrument, and a situational judgment test both with and without a verbal ability test and observed adverse impact in every scenario that included verbal ability. Similarly, Ryan, Ployhart, and Friedel (1998) generated differentially weighted composites of a cognitive ability test along with three personality scales for two different populations (police officers and firefighters). Though the use of differential weighting resulted in sometimes substantially reduced subgroup differences, adverse impact was still observed in almost every situation explored. The lone exception occurred in the firefighters sample, and adverse impact in this case was only acceptable if cognitive ability was removed or if the selection ratio was .60 or higher and cognitive ability was given a 25% weighting.

Taking this approach a step further, Schmitt, Rogers, Chan, Sheppard, and Jennings (1997) investigated the influence of number of predictors, predictor intercorrelations, predictor validity, and level of subgroup difference on composite validity and adverse impact in a Monte Carlo simulation. The predictors consisted of meta-analyzed estimates of the structured interview, a biodata instrument, and a personality measure (conscientiousness) with cognitive ability either specifically included with or excluded from the predictor composite. The subgroup difference on the criterion was assumed to be .45 and taken from the Ford, Kraiger, and Schechtman (1986) meta-analysis. Similar to previous research, the results showed that adverse impact would still occur at all but the very highest of selection rates (.90 or above) if all four predictors were used but only at the very lowest of selection rates (less than .30) if the three alternative predictors were used alone. This study also showed that although the number of predictors had little effect on subgroup differences, the smallest differences were found when the simulated composite included predictors that exhibited strong validity (.30 or more), small to no subgroup differences individually, and high levels of intercorrelation (.50 or more). Interestingly, from the results of this study, the authors surmised that while uncorrelated predictors are best for optimal prediction, correlated predictors have a greater effect on reducing group differences. This lends some insight into how the type of variance within each predictor as well as the type of shared variance across predictors has significant impact on meaningful selection outcomes (e.g., diversity and productivity).

In a similar context, Bobko, Potosky, and Roth (1999) recalculated the meta-analytic estimates used in Schmitt et al. (1997) and included additional studies to form a more extensive matrix for use in future studies. Like the research it followed, Bobko et al. found that simply combining the alternative predictors with cognitive ability did little in the way of eliminating adverse impact. However, unlike the previous study, they also found that adverse impact occurred even when only the three alternative predictors were used at selection ratios less than .70. Thus, use of the alternative predictors exclusively would probably lead to adverse impact at typical selection ratios.

Finally, Sackett and Ellingson (1997) challenged the perceived benefits of differential predictor weighting as well as the perception that simply adding predictors with small group differences to equations involving cognitive ability would result in less adverse impact. These researchers demonstrated how the standardized group difference for a predictor composite was a function of the sum of every predictor's standardized group difference, the number of predictors in the composite, and the average intercorrelation across all predictors. Not surprisingly, they found that the addition of predictors with zero group differences as well as increases in the average intercorrelation across all predictors generally reduced the overall standardized group difference. However, they also showed how combining one predictor with a large associated group difference with another predictor exhibiting a small group difference can actually produce a composite demonstrating a larger overall group difference than with using the first predictor alone. Most surprising though was their demonstration of the pervasiveness of adverse impact despite changes in average intercorrelations, increases in number of predictors, and varying levels of group differences. For example, their research showed that if five predictors each with zero group differences were combined with cognitive ability with its typical one standard deviation difference, the composite group difference would be .41 if all of the predictors were completely uncorrelated; but the same scenario with an average intercorrelation of .50 would still result in a .22 standard deviation difference. Referring to their table on p.710, both of these situations would end up violating the four-fifths rule, the first at selection ratios under .75 and the second at selection rates under .50.

The Pervasiveness of Adverse Impact

All in all, it would appear we are not really that much closer to solving this dilemma than we were some 30 years ago. In fact, a review of our attempts firmly suggests that the trade-off between maximizing workforce productivity and reducing adverse impact is probably as complex an issue as we have faced (Maxwell & Arvey, 1993). Schmidt's (1993) claim that these conflicting goals have "hobbled" personnel selection may be more accurate than we might wish because adverse impact has proven to be a very resilient adversary. Some researchers have gone so far as to state that to remove the possibility of adverse impact, we would have to discontinue the use of cognitive ability altogether (Schmitt et al., 1997); but this approach to solving the dilemma raises three strong arguments. First, cognitive ability has proven to be an efficient and valid predictor that shows a great deal of utility. To discontinue its use would equate to ignoring the degree to which cognitive ability is associated with general job performance as well as overlooking the practical and utilitarian benefits it provides. Secondly, simply removing cognitive ability from our selection equations does not guarantee the elimination of adverse impact. Bobko et al. (1999) showed where the use of alternative predictors often results in adverse impact despite the exclusion of cognitive ability. Moreover, Sackett and Ellingson (1997) demonstrated that even the small, standardized group differences associated with some of the commonly used alternative predictors would culminate in adverse impact at typical selection ratios. Finally, this type of thinking disregards the possibility that true population differences exist in specific qualifications for different jobs (Ironson, Guion, & Ostrander, 1982). As Jensen (1980) notes:

> Test scores themselves are merely correlates, predictors, and indicators of other socially important variables, which would not be altered in the least if tests did not exist. The problem of individual differences and group differences would not be made to disappear by abolishing tests. **One cannot treat a fever by throwing away the thermometer** [emphasis added] (p. xi).

The bottom line in this argument is that there might be a real "glass ceiling" when it comes to judging the performance of different groups. This view is further bolstered by the fact that the performance ratings of minority hires have been traditionally found to be lower, on average, than the performance ratings of majority hires (Bernardin, 1984; Campbell, Crooks, Mahoney, & Rock, 1973; Gael & Grant, 1972; Gael, Grant, & Ritchie, 1975). These differences have generally ranged upwards of one-half a standard deviation (Bobko et al., 1999; Hunter, Schmidt, & Rauschenberger, 1977; Kraiger & Ford, 1985). Moreover, these findings are consistent across both subjective and objective measures (Ford, Kraiger, & Schechtman, 1986; Sackett & Wilk, 1994). Therefore, instead of a testing problem, we have more of a social problem represented by real differences in the job-related abilities that tests capture (Gottfredson, 1988). To put it another way, "…adverse impact is a property of labor markets, not employment tests" (Wollack, 1994, p.218). This is not to say that these differences are genetic or inborn in nature, the answer to that question is for an entirely different venue; but these differences are real, and our work in this field will continue to be complicated by the economic, social, and legal realities that these differences generate. Organizations still expect to prosper, society will continue to expect the promotion and perception of fairness and equity across individuals, and the courts will ensure that legal standards are upheld.

Balancing these three scales has clearly been a momentous problem. Given the fact that our previous balancing attempts have been less than successful, this study argues that if we intend to "weigh in" on this issue, then we must shift the focus of our attention.

New Perspective

Personnel selection is often conceptualized as determining the relationship between a predictor or a set of predictors and job performance (Sekiguchi, 2001). However, because organizations function through their members (Guion, 1998), selected individuals show value and impact organizational success in more ways than simply through their rote productivity on the job. Borman and Motowidlo's (1993; 1997) contextual behavior would be one example of this, but other examples might include promoting the organization's image or influencing organizational creativity through diversity. The point here is that personnel selection plays a major role in the attainment of various organizational objectives. As such, the benefits of an appropriate selection system should go beyond the typical prediction of job performance. However, the bulk of our organizational research in selection has centered on maximizing job performance and improving individual effectiveness and not in specifically considering organizational effectiveness (Dunnette, 1963; Dunnette, Goldstein, Hough, Jones, Outtz, Prien, Schmitt, Siskin, & Zedeck, 1997). The traditional view of selection research represents a narrow perspective on how we can facilitate organizational success. Moreover, our almost singular (Wallace, 1965), sometimes blinded (Hoffman & Thornton, 1997), focus on the validity of our selection procedures has hindered our ability to resolve ancillary issues that come up throughout the course of research and practice. Although we do not necessarily ignore these ancillary issues, they often become secondary in importance to validity. However, the utility that organizations search for goes well beyond maximizing performance and will ultimately come from more than the slope from a regression. True utility reflects the attainment of a variety of organizational goals. Thus, our chosen focus begs the question, “Can we see the forest for the trees?” Obviously, organizations are concerned with selecting the best employees for their workforce; but they are also quite concerned with issues such as cost, turnover, tardiness, contextual behavior, and of course adverse impact. While we primarily focus on the “trees” (i.e., validity), organizations are rightfully concerned about the “forest” (i.e., the big picture). Therefore, if the goal of our selection procedures is to identify those employees that will facilitate organizational success, our interests should better coincide with those of the organizations we wish to benefit. This requires a much greater, and to some degree different, emphasis on multiple criteria.

Multiple Criteria

Few would argue with the multifaceted or multidimensional nature of either organizational success or job performance in general. However, much of what we know about our selection devices is based on analyses of univariate relationships between tests and criterion measures (Murphy, 1996; Murphy & Shiarella, 1997). In fact, personnel research in this century has been dominated by the single criterion measure (Campbell, McCloy, Oppler, & Sager, 1993). But as Guion (1987) notes, when you have more than one specific problem or issue, more than one specific criterion measure is called for. Additionally, because personnel selection most often involves, at least implicitly, a multivariate process with multiple independent and dependent variables (Murphy & Shiarella, 1997), a fully multivariate perspective is required when examining those variables believed to best describe, explain, and predict the criteria of interest. The problem with this perspective is that our traditional procedures are geared toward more of a univariate framework. While working with multiple predictors is easily resolved with techniques such as multiple regression, the default method of dealing with multiple criteria simultaneously (when it is attempted at all) has focused on using a method such as ordinary least squares (OLS) regression to generate regression weights associated with a composite that supposedly represents the “ultimate criterion” of an employee’s value to the organization. However, the concept of an ultimate criterion is fraught with a number of problems, ranging from compensatory issues (Schmidt & Kaplan, 1971) to questions about ambiguous interpretations (Cattell, 1957; Ghiselli, 1956), and from the words of the man who coined the term, probably does not, and never will, exist (Thorndike, 1949).

Therefore, the resolution of this adverse impact/productivity controversy is confronted with a specific set of impasses or constraints that are made more difficult by the customary treatment of multiple criteria. First, there is some recognition that neither removing cognitive ability nor simply focusing on alternative predictors necessarily solves the issue at hand. Second, a realization presents itself, not without reluctance, that job performance and validity are not the ultimate goals but rather utility and organizational success in general. Third, an honest assessment of organizational success reveals many influential factors with at least two of those coming in the form of diversity and avoiding legally questionable situations. Finally, though it would be preferable to incorporate multiple goals such as increasing productivity and decreasing adverse impact into our selection models, a determination is made that, short of using some potentially ambiguous composite criterion, traditional methodologies are insufficient for including multiple objectives in establishing a single index with which we can make administrative decisions.

This would suggest that a new or novel approach to addressing this issue is needed. One plausible alternative would be to observe how different disciplines resolve similarly patterned problems. From these observations, it is discovered that the Management Science community has been dealing with these types of problems for decades, initially concentrating on linear programming but later evolving their methodologies to encompass more complex problems such as solving for multiple criteria. These evolved methodologies fall into the field of optimization.

Optimization & Constrained Estimation

Optimization is a form of mathematical programming intended to facilitate decision-making by generating optimal solutions to mathematically modeled situations. With these techniques, the objective is to find the best possible solution to a particular problem (Beale, 1988). This is accomplished by applying specific algorithms (sometimes referred to as methods of solution) through an iterative sequence to minimize or maximize a function of n variables, $f(x_1,\ldots,x_n)$, subject to any constraints placed on the model. The specific algorithm depends on the type of technique used, the linearity of the model, and whether the model is constrained or unconstrained. Optimal solutions result in values for structural or decision variables over which the user has some control (e.g., how many units of a particular product should be produced daily; how to weight a particular predictor variable), and constraints represent boundary conditions that the model cannot exceed (e.g., limits on the amount of raw materials utilized per day; the selection ratio).

Initial efforts in this arena concentrated on Dantzig's work in 1947 on the general linear programming model and a general method of solution designated the simplex method (Dantzig, 1982). For reference, Ignizio (1982, pp. 247-249) provides an example problem that illustrates the general optimization process in linear programming terms and compares it to the use of OLS regression. However, because linear programming is limited in its ability to effectively address problems involving multiple objectives, a number of researchers proposed additional methods and algorithms to satisfy this concern (Ignizio, 1985). Examples of this include Kuhn and Tucker's (1951) vector-maximum model, Charnes and Cooper's (1961) use of constrained regressions and goal programming, and the present day extrapolations of the Kuhn-Tucker equations and sequential quadratic programming (see Optimization Toolbox, Coleman, Branch, & Grace, 1999).

These types of techniques have long been used in a variety of disciplines such as economics, business, and engineering (Feiring, 1986) but are most commonly associated with operational research (Beale, 1988). Though certainly not an exhaustive list, real world applications include the establishment of a proper diet, transportation of goods, personnel planning, portfolio selection, logistics, and oil refinery scheduling. An example with multiple objectives might include production planning where important goals are represented by maximizing total net revenue and minimum net revenue in any period while at the same time minimizing backorders and overtime.

Constrained Estimation

For the purposes of the present study, constrained estimation is a form of mathematical programming and a variation of constrained optimization customized to fit personnel selection decisions and modeled through the computer programming package MATLAB to optimize multiple criteria. More specifically, constrained estimation, in the context of the present study, seeks a solution that minimizes adverse impact while simultaneously maximizing selection validity. The procedure involves determining which predictors to utilize, deciding on a selection ratio, and generating initial β-weights (i.e., predictor weights) through OLS regression to create a starting point for the constrained estimation program. The weights ($\beta_1,\ldots, \beta_k$) from the final solution represent the unknowns as well as the ultimate goal of the optimization process, and the selection ratio reflects a constraint. This routine also allows for a determination of how much importance is placed on minimizing subgroup differences as opposed to maintaining the validity achieved with OLS regression. However, an important distinction is required when considering this weighting of criteria versus the weighting of criteria in a composite criterion as described by Thorndike (1949) or Schmidt and Kaplan (1971). Constrained estimation does not combine criterion elements into one ultimate criterion; instead, the weighting system used with this technique refers to how much of a balance is to be achieved when optimizing each criterion concurrently. Thus, there are no compensatory issues to resolve because the final solution encapsulates the prediction of each criterion independently subject to the requested balance as reflected by the degree of importance placed on one versus the other criterion.

Overall, the objectives of this routine are to minimize errors in estimation while simultaneously minimizing the expected adverse impact of a selection program by reducing subgroup differences on the predictor composite. This is accomplished by manipulating β-weights from an initial OLS solution through an iterative process to minimize a function of both objectives. In this case, and from a different perspective, constrained estimation adjusts a set of predictor weights that have already maximized validity in order to minimize mean subgroup differences while sustaining some degree of validity in accordance with the amount of importance placed on each goal. This function can be mathematically represented by:

$$f\left(\beta_i, AIW, Z_{cutpoint}, Y, X, Gp\right) = AIW\left(1 - \frac{p\left(z_{cutpoint} + abs\left(z_{gp1} - z_{gp2}\right)\right)}{p\left(z_{cutpoint}\right)}\right) + \left(1 - AIW\right)\left(1 - \frac{r_i^2}{r_{ols}^2}\right)$$

where β_i refers to the β-weight function to be optimized, *AIW* refers to the adverse impact weight (ranging from 0 to 1.0 and reflecting the ratio or degree of importance placed on minimizing group mean differences as opposed to minimizing errors of estimation), $z_{cutpoint}$ refers to the standard score where the area of the associated normal curve (density function) above the cutpoint divided by the total area of the curve reflects the selection ratio, $abs(z_{gp1} - z_{gp2})$ refers to the absolute mean difference in expected job performance between the subgroups in standardized form, *Y* refers to the criterion scores, *X* refers to scores associated with a predictor composite, *Gp* refers to group membership, *p* reflects a probability function (relating to the normal curve), r_i^2 refers to R-squared of the constrained solution, and r_{ols}^2 refers to the R-squared of the initial solution.

Additionally, the first half of the right side of the equation represents minimizing subgroup differences whereas the second half represents the maximization of validity. This minimization and maximization is accomplished by setting each portion of the equation equal to zero (completely minimized) when there are no subgroup differences (i.e., $abs(z_{gp1} - z_{gp2}) = 0$) and when the constrained solution achieves the same validity as that with OLS regression (i.e., $r_i^2 = r_{ols}^2$).

As an example, consider a situation where three predictors (cognitive ability, a biographical inventory, and a structured interview) are used to predict general job performance, and the selection ratio is set at .25. The first step in this program would include using OLS regression to determine the initial (or starting) predictor weights along with the standardized subgroup difference associated with the predictor composite formed by these weights. Furthermore, suppose that the initial estimates resulted in an R-squared of .21, an adverse impact ratio of .52, and predictor weights of .39, .21, and .13, for cognitive ability, biographical data, and the structured interview, respectively. At this point, the researcher or practitioner would be required to decide upon an adverse impact weight (i.e., *AIW*) representing how important eliminating adverse impact is as compared to maintaining initial validity. It is important to note at this junction, that if *AIW* is set to zero, the program should revert to OLS regression because validity has already been maximized, and the program is instructed to ignore any importance associated with reducing subgroup differences on the predictor composite. Thus, for the sake of illustration, imagine that *AIW* is set at .50 (equal importance between objectives). The program would then begin an iterative process of solving for a set of predictor weights by minimizing a function of the specified objectives that would achieve an optimal balance between the two competing goals. However, this process would also be constrained by the selection ratio set at .25, which determines the number of individuals to be selected. Thus, the ultimate objective of the program is to generate a new predictor composite by manipulating predictor weights to a point where subgroup differences associated with the new predictor composite and a selection ratio of .25 are reduced sufficiently to eliminate adverse impact while the prediction of job performance remains as similar as possible to that achieved with OLS regression.

In reviewing the literature within our discipline, only one study was found to use this type of programming. Interestingly enough, this study was also concerned with the same basic problem. De Corte (1999) used constrained nonlinear programming to address the reduction of adverse impact while preserving quality in the workforce. Using the base data provided by Hattrup et al. (1997), De Corte explored the usefulness of optimization techniques by generating a Monte Carlo study using two predictors (cognitive ability and work orientation). Similar to the Hattrup et al. study, he further manipulated the weights of two criterion dimensions (contextual and task performance) in a criterion composite and varied different scenarios based on disparate selection rates. In order to determine the weights assigned to each predictor, De Corte constrained selection (using a Fortran computer program) to maximize average quality (i.e., expected job performance) and limit the adverse impact ratio to some acceptable level (between .80 and 1.25) while attaining a predetermined selection rate. This resulted in a consistent elimination of adverse impact but also culminated in consistent and substantial decreases in the average job performance of selected individuals as compared with a multiple regression approach except when the selection rate was unusually high (i.e., .80 or greater).

In the present study, it is suggested that rather than limiting the adverse impact ratio to fall within some acceptable boundaries, the focus should be on obtaining an optimal balance between the two competing goals: minimizing errors in estimation and eliminating adverse impact. Thus, this reflects the major difference between De Corte's (1999) constrained nonlinear program and the constrained program presented in this research. While De Corte's method constrains the adverse impact ratio to fall between .80 and 1.25 and then maximizes validity within these set limits, constrained estimation actually minimizes the errors in estimation while simultaneously minimizing subgroup differences with no direct controls or limits placed on the adverse impact ratio. In other words, both goals are optimized to the extent possible, and the data is allowed to guide the final solution characteristics. The distinction also reflects a difference in the use of optimization techniques. De Corte optimizes one criterion while constraining adverse impact; constrained estimation optimizes two criterion variables concurrently.

Taken in total, the goal of constrained estimation is to find the "best" minimized function that strikes a complete as possible balance between the competing objectives. This is accomplished by minimizing a function of both objectives of interest with the resulting solution providing regression-type weights and an explanation of variance that should mirror, as best as possible, OLS regression. While there may be many instances when eliminating adverse impact is just not possible, this study argues that the present procedure will still provide a narrowly tailored selection model that achieves an optimal balance between the goals of maximizing workforce productivity and increasing the diversity of the workplace.

Constrained Estimation Pilot Study

As an initial step in gauging the efficacy of this approach, constrained estimation was applied to a subset of the data used in this study (Henderson & Ladd, 2001). This data originated with a study within the confines of an assessment center performed at a large Southeastern utility company (Gniatczyk, 2000; Gniatczyk & Ladd, 2001). For this study, four predictors were extracted from the data: a cognitive ability test (CA), a biographical inventory (Bio), consensed ratings from the assessment center (Rate), and the overall performance on a managerial video simulation (Vscore). The criteria included the current salary of each employee and, from a constrained programming perspective, the reduction of group mean differences on the predictor composite. The sample size was 188 with 12 African Americans and 176 Caucasians.

For the purposes of illustration, three types of solutions were utilized (OLS regression, unit weighting, and constrained estimation) with an overall selection ratio of .25. Additionally, the ensuing constrained solution was generated by varying the amount of importance (i.e., *AIW*) placed on minimizing group mean differences as opposed to maximizing validity. Specifically, adverse impact weights of .50, .25, and .15 (all ranging from zero to one with .50 reflecting an equal weighting between the two criterion objectives) were chosen to show how the methodology operated under disparate conditions. Results included the predictor weightings, overall R-squared values, and the adverse impact associated with each solution. The adverse impact ratio was determined using the four-fifths rule outlined in the *Uniform Guidelines* (EEOC et al., 1978) where impact occurs when the minority selection rate is less than four-fifths (or .80) of the majority selection rate.

As can be seen in Table 1, the initial OLS regression resulted in an R-squared of .26 with an associated adverse impact ratio of .67. Comparatively, unit weighting resulted in a decrease in R-squared (.20) and an exacerbated adverse impact (.35). However, whereas the constrained solution with equal weightings (i.e., *AIW* = .50) accomplished the objective of eliminating adverse impact (adverse impact ratio = 1.0), it also culminated in only a small reduction in the amount of variance explained (.23 compared

Table 1: R-squared, Adverse Impact Ratio, and Predictor Weights

	R^2	Adverse Impact Ratio	Weights			
			Cognitive Ability	**Biodata**	**Assessment Center Ratings**	**Video Simulation**
OLS	.2610	.6709	.0119	.3740	.2420	.1382
Unit	.2010	.3470	1.0	1.0	1.0	1.0
Constrained						
AIW = .50	.2348	1.0	-.1460	.3542	.2815	.1151
AIW = .25	.2348	1.0	-.1462	.3562	.2795	.1143
AIW = .15	.2529	.8131	-.0788	.3728	.2713	.1286

to .26). In fact, the proportionate loss in R-squared ($1 - r^2_{ols}/r^2_{constrained}$) was only .11. When slightly less emphasis was placed on minimizing adverse impact (i.e., lowering *AIW* from .50 to .25), no change was found in either the level of diversity or explanation of variance. Finally, when *AIW* was lowered to .15, R-squared increased to .25 while the associated adverse impact ratio decreased to a level (.81) just acceptable under the 4/5ths rule.

In order to assess the stability of these estimates, cross validation procedures were performed. Ideally, it would have been preferable to split the sample, use constrained estimation to fit the predictor weightings on one half, and then apply those same weightings to the second half. However, as is often the case with adverse impact research, there were so few individuals found in the minority sample, this was not feasible. Therefore, bootstrap models (Efron, 1982) were generated to evaluate the standard errors associated with each estimate. In this capacity, resampling with replacement was used over 100 iterations to derive bootstrap estimates of R-squared values, adverse impact ratios, and predictor weightings. The mean values of each

Table 2: Bootstrapped Estimates Based on Various Adverse Impact Weights

	R^2	Adverse Impact Ratio	Weights			
			Cognitive Ability	**Biodata**	**Assessment Center Ratings**	**Video Simulation**
OLS	.2656 (.062)	.7033 (.134)	.0104 (.070)	.3752 (.068)	.2345 (.060)	.1341 (.065)
Unit	.2010 (.061)	.3415 (.069)	1.0 (N/A)	1.0 (N/A)	1.0 (N/A)	1.0 (N/A)
Constrained						
AIW = .50	.2348 (.059)	1.0 (.050)	-.1460 (.077)	.3542 (.068)	.2815 (.064)	.1151 (.066)
AIW = .25	.2519 (.058)	.9305 (.159)	-.0854 (.091)	.3637 (.060)	.2608 (.072)	.1063 (.072)
AIW = .15	.2598 (.059)	.8496 (.164)	-.0578 (.075)	.3726 (.064)	.2565 (.068)	.1251 (.066)

Note: Standard errors included in parentheses below estimates.

estimate as well as each estimate's associated standard error across all of the 100 repetitions can be found in Table 2. Bootstrapping results showed consistently small standard errors associated with almost every estimate (ranging from .05 to .08) except those relating to adverse impact ratios. While the unit-weighted approach resulted in a generally small standard error for adverse impact (.06), the standard error for OLS regression and for constrained estimation where emphasis was placed on minimizing errors in estimation resulted in considerably larger standard errors (OLS = .13; Constrained = .16). When equal emphasis was placed on minimizing adverse impact as well as maximizing validity (*AIW* = .50), the standard error for adverse impact was significantly reduced to .05.

Overall, constrained estimation significantly reduced or eliminated adverse impact while still preserving nearly all of the explained variance found with OLS regression. However, there is still a question about how constrained estimation would hold up in disparate situations involving different predictors and additional predictor combinations. Moreover, although the bootstrap procedure has been generally accepted as an appropriate cross validation methodology (Cooil, Winer, & Rados, 1987; Efron, 1982; Mooney & Duval, 1993), further evaluation of the stability of these estimates with different procedures would lend support to the efficacy of this new approach. The present research was constructed to address these concerns.

Research Questions & Predictions

The overarching purpose of the present research was to address and answer three major research questions. First, "How can [employers] use valid procedures in a manner that optimizes the expected performance of their workforce and at the same time employ a demographically diverse workforce?" (Schmitt et al., 1997, p. 719). Second, does the use of constrained estimation provide a viable alternative to other selection strategies? Finally, how does constrained estimation compare to other selection schemes? Based on these general questions, a previous study of constrained estimation, and a review of the literature, two studies were designed to explicate the role of constrained estimation in selection.

In the first study, a Monte Carlo simulation was used to test the effectiveness and sensitivity of the methodology. The major objective revolved around showing the viability of constrained estimation under a number of different study characteristics. The second study consisted of data from a previous assessment experience and research effort

(Gniatczyk, 2000; Gniatczyk & Ladd, 2001). The rationale for this study was to look at the implications of using constrained estimation in a very specific case. While the simulation work was intended more for exploratory purposes, both studies adhered to the same set of general questions and predictions.

Importantly, this research focused on predictions rather than strict hypotheses. While the study of constrained estimation should be concerned with a number of relevant hypotheses, many of the salient questions surrounding this level of exploratory work with this type of programming present themselves as almost mathematical certainties. However, this does not diminish the importance of these questions, and because they have yet to be demonstrated, it is imperative to solidify the foundation of this research, as well as future work, with supported answers. Therefore, subsequent questions, predictions, and hypotheses are denoted simply as predictions.

<u>Predictions</u>

The first characteristic most people look for when deciding on a set of predictors or an overall selection system is validity. In general, OLS regression provides the pinnacle of validity partly because of its focus on minimizing the presence of residuals but also because it capitalizes on sample specific features of a particular dataset (Dawes, 1979). However, OLS regression is subject to an exaggerated influence from outliers (Ignizio, 1982), and thus, frequently demonstrates shrinkage when weighting schemes are applied to new samples (Einhorn & Hogarth, 1975). Unit weights are more stable and result in less shrinkage but only outperform regression weights when the ratio of observations to predictors is less than 25 to 1 unless the presence of suppression is discovered. In this instance, the ratio drops somewhere in the range of 15 to 1 (Schmidt, 1971).

Constrained estimation should show similar validity to OLS regression because it shares a common purpose (i.e., optimizing performance) and works with the same set of predictors. However, part of this similarity is dependent upon the amount of importance placed on eliminating adverse impact as opposed to maximizing performance. When the adverse impact weight is zero, constrained estimation and OLS regression should provide exactly equivalent results. As the adverse impact weight increases, additional deviations should be found between the two methods. This is due to constrained estimation's (1) additional focus on reducing subgroup differences and (2) less optimal predictor weightings and use of predictor variance. At the same time, the predictor weights associated with constrained estimation should perform better than those of a completely reduced variance model (i.e., unit weighting). Therefore, based on this information and consistent with previous work, it was predicted that:

> *Prediction 1a*: Constrained estimation will reduce to OLS regression when no importance is placed on adverse impact in the criteria.
>
> *Prediction 1b*: Constrained estimation and OLS regression will provide similar explanations of criterion variance.
>
> *Prediction 1c*: Constrained estimation will provide a greater explanation of criterion variance than that of unit weighting.

Constrained estimation deviates from OLS regression weights in the minimization of adverse impact ($\underline{AIW} > 0$); as a result, it also deviates, at least to some degree, from the overcapitalization on sample specific features that plague OLS regression in cross validation. Moreover, just as ridge regression attempts to improve stability in new samples by adding common variance to the predictor weights (Darlington, 1978), the consistent application of an additional goal (i.e., optimizing adverse impact) should add common variance as well. Therefore, one would expect constrained estimation to demonstrate less shrinkage under cross validation. However, the bootstrap estimates provided in the pilot study exhibited no real differences in the standard errors associated with either constrained estimation or OLS regression across either validity or specific predictor weights. On the other hand, the inclusion of different predictors and additional predictor combinations along with a larger sample could provide a more solid platform for constrained estimation to display stronger results. Additionally, while the standard errors demonstrated in the pilot study were very small, there is no reason to expect that the weights generated with constrained estimation should be as stable as unit weights because of the empirical nature of the procedure. Thus:

> *Prediction 2a*: Across disparate situations, constrained estimation will more frequently result in less shrinkage than OLS regression.
>
> *Prediction 2b*: Unit weighting will show less shrinkage than OLS regression.
>
> *Prediction 2c*: Unit weighting will display less shrinkage than constrained estimation.

Along with maximizing workforce productivity (i.e., validity), the goal of constrained estimation is to reduce adverse impact. This reduction should be directly related to the amount of importance (*AIW*; a ratio ranging from 0 to 1.0) placed on minimizing subgroup differences as compared to maximizing validity. Obviously, an adverse impact weight of zero should result in the same adverse impact as that shown with OLS regression; but as this weight increases, so should the level of adverse impact decrease.

However, the ability of constrained estimation to reduce adverse impact is also heavily dependent on the types of predictors used and the type of variance displayed. Constrained estimation basically repartitions the variance associated with each predictor in a selection scheme in such a manner as to reduce the subgroup differences related to the new predictor composite. Thus, along with a sufficient adverse impact weight, constrained estimation requires more than one predictor to work with. Whereas the use of a single predictor obviously eliminates any available weighting options, additional predictors provide a larger, more encompassing palette from which to devise the predictor composite. Moreover, for constrained estimation to have any influence on decreasing subgroup differences while maintaining validity, there must be available variance that relates to the first criterion of interest (e.g., job performance) as well as variance showing little or no relationship with group membership. This additional variance can come from single predictors that exhibit both characteristics (e.g., biographical data) or from some combination of predictors. A variety of predictors with each possessing unique variance with either one or both criteria would be ideal; however, simply possessing somewhat different variance should enable the constrained estimation program to manipulate the use of that variance in an effort to achieve both objectives.

Therefore, given a situation where more than one predictor is available and where at least some differences are present in the types of variance each predictor explains, the use of constrained estimation should benefit a selection system intended to reduce adverse impact. In addition, by specifically and simultaneously concentrating efforts on optimizing two objectives of interest, constrained estimation should be better able to meet each goal concurrent with the other with greater success than either OLS regression or unit weighting. Subsequently, it was predicted that:

> *Prediction 3*: When sufficient importance is placed on reducing subgroup differences as well as maximizing validity, and when multiple predictors possessing unique variance associated with both adverse impact and other criteria of interest are available, the constrained estimation routine will partition variance such that adverse impact is eliminated and validity is sustained at acceptable levels when compared to that of OLS regression and unit weighting.

CHAPTER III

METHODOLOGY

Overview

Two studies were utilized in illustrating the value of constrained estimation when working with multiple, conflicting criteria. The data consisted of selection scenarios where maximizing performance and decreasing or eliminating adverse impact were desirable goals. In the first study, a Monte Carlo simulation was performed to explore how the proposed methodology reacted under varying conditions. In the second, constrained estimation was applied to a dataset from a consulting project utilizing several predictors. Because constrained estimation ultimately results in a set of regression type predictor weightings, comparisons were made between these results and those found with two common predictor weighting strategies: OLS regression and Unit weighting. Moreover, cross validation was planned in each situation presented to examine the stability of each approach. While specific procedures are discussed within the framework of each study's description, general procedures common to both datasets as well as all analyses are described in the last section of this chapter.

Study One

Simulation Design

A mathematical programming package (MATLAB) was used to generate Monte Carlo data with four predictor variates and one criterion for 200,000 subjects. Minority representation was initially set at 20%. However, because the percentage of minorities in the workforce might impact the degree of adverse impact found within a particular selection system, this study incorporated minority representation as a specific variable of interest. Moreover, while previous research in the arena of adverse impact has incorporated several different minority base rates into their simulations (e.g., Schmitt et al. (1997) used .20; Hattrup et al. (1997) used .25 based on projections of minority representation in the U.S. workforce for the year 2000), this simulation followed the lead of Sackett and Roth (1996) by establishing multiple levels of minority membership (e.g., they used .20 and .40) in order to provide an array of sample characteristics. The levels used within this data were 5%, 20%, and 40%. Subsequently, three samples of 10,000 subjects each were randomly selected from the overall dataset of 200,000 reflecting these specific levels of minority representation.

Based on generalized estimates reported in Bobko et al. (1999), the simulated predictor and criterion set included construct variates representative of cognitive ability, the structured interview, conscientiousness, biographical data, and general job performance. As can be seen in Table 3, these estimates consisted of both correlations among all of the variables involved as well as subgroup differences regarding race. However, because of some initial discrepancies in the newly created correlation matrix,

Table 3. Bobko et al.'s (1999) Matrix of Correlations and *d* Values

	d	**Subgroup**	**CA**	**SI**	**Con**	**Bio**
CA	1.00	.37				
SI	.23	.09	.24			
Con	.09	.04	.00	.12		
Bio	.33	.13	.19	.16	.51	
Job Perf	.45	.18	.30	.30	.18	.28

Note. Cognitive ability, structured interviews, conscientiousness, biographical data, and job performance are referred to as CA, SI, Con, Bio, and Job Perf, respectively.

the correlation between the cognitive ability variate and race was changed from .37 to .40 to provide an accurate match between the variable relationships reported by Bobko et al. (1999) and the Monte Carlo generated matrix used in this study.

Subgroup differences are described in the form of *d* values (standardized difference scores). These values are computed by subtracting the minority group mean from the majority group mean and then dividing by the pooled within-group standard deviation. This is often useful when comparing across predictors that exhibit disparate standard deviations (Sackett & Ellingson, 1997). Larger *d* values equate to larger mean differences favoring majority group members. So, for example, this simulated dataset shows a one standard deviation difference in the performance of minority and majority members on the cognitive ability variate, whereas there is about a half a standard deviation difference reported for the job performance variate.

Importantly, the matrix of relationships between all of the predictors and job performance should be considered an update from a matrix generated by Schmitt et al. (1997). Both studies used meta-analytically derived estimates to construct their matrices, and both focused only on the “operational use” of uncorrected, as opposed to corrected, correlation estimates. However, Bobko et al. (1999) included additional research and weighted studies by sample size. As such, the Bobko et al. matrix was chosen for use because it represented the most accurate and up-to-date correlational and subgroup difference information about these specific types of predictors and job performance.

Study Two

Participants

The total sample consisted of 535 individuals employed by a large Southeastern utility company in 1992 through 1993. A review of various demographic characteristics reveals that the participants were largely white (86%) males (81%) with an average age of 45. Additionally, 7% reported having obtained only a High School degree and 41% a Bachelor's degree, whereas 48% reported holding a graduate degree.

Because the client organization was experiencing a restructuring effort, an external consulting group was employed to facilitate a number of re-organization related appointments. This facilitation included the administration of a three-phase managerial assessment process in which all of the individuals participated. The initial data for the pilot study originated from archival records maintained by the external consulting group and primarily focused on those participants who had completed all three stages of the assessment process and for whom criterion data was available. However, because only a limited number of individuals were allowed to proceed to the third stage, this sample consisted of a disproportionate number of Caucasians (176) as opposed to African Americans (12). While the small number of African Americans provided a test of constrained estimation's ability to reduce or eliminate adverse impact, this also made the results from cross-validation both more difficult to examine and possibly tenuous (due to a lack of power) in interpretation. Therefore, an alternate sample was extracted from the data focusing on those predictors utilized in the second stage after additional criterion data had been obtained from the client organization. This resulted in only a small increase in the number of African Americans. Missing data across all of the variables further reduced the benefit of this expanded sample.

Sample. The alternate sample consisted of 340 individuals of whom 310 were Caucasian with 26 African Americans and 292 were male with 48 females. The average age was 46 with a range from 28 to 61 and a standard deviation of 5.716. Additionally, 27 individuals reported holding only a High School degree, 135 an undergraduate degree, and 178 a graduate degree.

Outline of the Assessment Process

Collaboration between the consulting group and the organization resulted in a multi-stage assessment process where only those individuals deemed to have successfully completed each stage moved on for further review. The first stage consisted of a pre-screening process administered by the organization to ensure that all individuals met the minimum qualifications for the job(s) in question. The second stage involved a battery of assessment devices designed to assess managerial skills. The specific assessments utilized at this point included the Watson-Glaser Critical Thinking Appraisal (CTA), a managerial video simulation, the Manager Profile Record (MPR), a reasoning by inference test (RBI), and a strategic in-basket. The third and final stage of this assessment process included a daylong managerial assessment center and the completion of a personality test: the California Psychological Inventory (CPI). During the assessment center, selected individuals were subjected to a simulation exercise, an in-basket exercise, a case analysis, and a leaderless group discussion.

Predictor Measures

Watson-Glaser Critical Thinking Appraisal. The CTA is an 80-item paper and pencil instrument measuring an individual’s critical thinking skills. It is comprised of five subtests: (1) inference – discriminating among degrees of truth or falsity of inferences drawn from given data; (2) recognition of assumptions – recognizing unstated assumptions or presuppositions in given statements or assertions; (3) deduction – determining whether certain conclusions necessarily follow from information in given statements or premises; (4) interpretation – weighing evidence and deciding if generalizations or conclusions based on the given data are warranted; (5) evaluation of arguments – distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular question at issue. Scores on this instrument range from 0 to 80 with higher scores reflecting greater critical thinking skills.

Managerial Video Simulation. This exercise was designed to assess how individuals exert influence, show initiative, and manage subordinates in a low-fidelity simulation (e.g., Motowidlo, Dunnette, & Carter, 1990). It is comprised of fourteen, two-part video vignettes demonstrating situations where supervisory personnel are interacting with others in a work environment. Participants are asked to view the vignettes and, at predetermined points, choose one of four behavioral options to indicate how they would respond in a similar situation. Scores on this instrument range from 0 to 100 with higher scores reflecting better performance. Although overall performance was the main focus in this study, this exercise also provides sub-scores in three different areas: customer relations, judgment, and attracting new business.

Manager Profile Record. The MPR is a traditional biographical inventory (Owens & Schoenfeldt, 1979; Stokes, Mumford, & Owens, 1994) designed to identify managerial potential. It is divided into two sections with the first part containing 196 multiple choice items pertaining to personal, educational, and employment histories and the second part containing 46 multiple choice items assessing management philosophies and/or styles. The second part is presented as a low-fidelity simulation where participants are asked to choose the best and second best responses to different management scenarios. Results from the MPR are broken down into background (biographical) and judgment (simulation). Each portion of the MPR, as well as overall performance, result in score ranges of eight to 32.

Reasoning by Inference Test. The RBI is a 25-item paper and pencil instrument designed to assess the relative motive strength of an individual's achievement motivation in relation to that same individual's fear of failure. The basis of this testing format comes from the concept of conditional reasoning which argues that individuals choose behaviors that appear to be sensible, logical, and consistent with their own reasoning process (James, 1998). However, conditional reasoning also argues that an individual's reasoning process is influenced by a number of justification mechanisms such that different individuals may view a single situation in completely different ways and thus choose disparate behaviors in response to that situation. The relative motive strength proffered by the RBI reflects this reasoning process and gives insight into how an individual is likely to approach differing situations. Written instructions inform participants that the instrument is designed to measuring reasoning ability, and respondents are asked to select the most reasonable alternative to each item. Scoring is based on the number of responses indicating either achievement motivation or fear of failure, and ultimately, three performance scores are generated: (1) achievement motivation, (2) fear of failure, and (3) the difference between the two scores.

Strategic In-Basket. This exercise provides a partial simulation of the administrative tasks that executives and supervisory personnel might face throughout the course of a typical day. Participants are presented with a number of memoranda and other written documents that require quick analysis and action. It is similar in form to the in-basket used in the assessment center but presents itself with more difficult situations that focus on more strategic issues. Scores on this instrument come in the form of an overall rating as well as five dimension ratings (e.g., analysis, judgment, initiative, team building, and planning and organizing). All of the ratings follow a scoring range of one to five.

Criteria

Multiple criteria were used to determine the efficacy of constrained estimation. Specifically, the criteria were represented by measures or variates of general job performance and adverse impact (or diversity). The first study simulated job performance with Monte Carlo work in accordance with Bobko et al.'s (1999) estimations of the relationships between general job performance and four types of predictors (see Table 3). The second study used a surrogate measure of job performance in the form of the then (1993) current salary of each participant during the assessment process. Initially, this information was gathered from organizational records in 1993, but additional information was obtained in 2003 (but still originating in 1993) to enlarge the study. The rationale behind using this criterion was based on the assumption that better workers eventually achieve higher salaries on average than those who do not perform adequately. The salaries from this data ranged from 33754 to 100000 with a mean of 69285 and a standard deviation of 12340. The second criterion used in both datasets was adverse impact operationalized as the reduction of group mean differences on the predictor composite such that the eventual adverse impact ratio of minority selection rates to majority selection rates reached at least .80 or four-fifths with an optimal value of 1.0.

General Procedures & Analyses

In order to test the efficacy of constrained estimation, an optimization routine (a copy of this program is available upon request) was generated with the mathematical program MATLAB (Version 5) along with information and code provided by an associated secondary manual, Optimization Toolbox (Coleman, Branch, & Grace, 1999). In that all of the hypotheses involved a comparison of weighting schemes, this routine incorporated three predictor weighting strategies in the generation of regression-type weights: (1) constrained estimation, (2) OLS regression, and (3) unit weighting. The resulting output from this program included a listing of the weights assigned to each predictor as well as both R-squared values (i.e., explanation of variance) and adverse impact ratios associated with each weighting strategy. The routine used a selection ratio of .25, and all variables were examined for both kurtosis and skewness, and subsequently treated as continuous, normally distributed variables.

Additional dataset manipulations occurred to further explore comparisons between the three weighting strategies. First, the Monte Carlo study included three datasets varied by minority representation (5%, 20%, and 40%). Moreover, 15 different combinations were presented within each dataset representing every possible composite of the four predictor variates (i.e., cognitive ability, structured interview, biographical data, and conscientiousness). This assortment of predictors can be seen in Table 4.

Table 4: Various Monte Carlo Predictor Combinations

Cognitive Ability only	Cognitive Ability, Structured Interview, & Biographical Data
Structured Interview only	Cognitive Ability, Structured Interview, & Conscientiousness
Biographical Data only	Cognitive Ability, Biographical Data, & Conscientiousness
Conscientiousness only	Structured Interview, Biographical Data, & Conscientiousness
Cognitive Ability & Structured Interview	Cognitive Ability, Structured Interview, Biographical Data, & Conscientiousness
Cognitive Ability & Biographical Data	
Cognitive Ability & Conscientiousness	
Structured Interview & Biographical Data	
Structured Interview & Conscientiousness	
Biographical Data & Conscientiousness	

Evaluation of the second study followed a similar but somewhat more controlled appraisal without differing minority populations. Specifically, all of the predictors utilized in the second phase of the assessment process (i.e., the Critical Thinking Appraisal, the background portion of the Manager Profile Record, the situational judgment test associated with the Manager Profile Record, the Reasoning by Inference Test, a managerial video simulation, and a strategic in-basket) were utilized both independently as well as with the cognitive ability test (CTA). Additional predictor schemes were also devised based on the traditional compatibility of different predictors. Biographical data and personality are often used in conjunction with cognitive ability to battle adverse impact, so the Reasoning by Inference Test (RBI) and the background portion of the Manager Profile Record (Bio) were used both with and without the CTA. Finally, the reemergence of low-fidelity simulations has sparked researcher interest in recent years, so the situational judgment test (SJT; low fidelity), the managerial video simulation (Vscore; medium fidelity), and the strategic in-basket (SIB; high fidelity) were used in combination as well as in conjunction with the CTA.

Further comparisons were made in both studies within the constrained estimation scheme based on the degree of importance (i.e., *AIW*) placed on minimizing adverse impact as opposed to optimizing validity. Specifically, adverse impact weights of .50, .35, .25, and .15 (all ranging from zero to one with .50 reflecting equal importance between the two criterion objectives) were chosen to show how the methodology operated under disparate conditions.

Moreover, all analyses were subjected to both initial examinations of relevant output as well as cross-validation. This was accomplished traditionally by first splitting the data randomly, allowing the three weighting schemes to fit predictor weights on the first part of the data, and then applying those weights to the second part of the data. In the Monte Carlo study, the samples were simply split approximately in half. In the second study, 227 individuals were used to generate initial weights whereas 113 individuals were used to cross-validate the results with some variation occurring due to missing values.

Additionally, when working with iterative techniques, some concern must be displayed for achieving a globally optimal solution. By manipulating a number of values to minimize a function, researchers can find numerous solutions that fit the objectives of the program. However, a solution is only optimal when it is both efficient and nondominated (Steuer, 1986). In other words, a solution is considered optimal when it is not feasibly possible to improve the performance of that solution through further manipulation of before said values. While we are often satisfied with "near-optimal" solutions in practice (Steuer, 1986), global optima, rather than local optima, remain the goal. One method of testing whether a global optima has been achieved is to determine if the same initial constrained solution is reached from different starting points (Using MATLAB, 1998). In this research, this test was carried out by initiating the constrained routine from two starting points: OLS regression and unit weighting.

Finally, while these procedures represented the core of what this research hoped to accomplish, additional analyses were expected based on anomalous but relevant findings from both the Monte Carlo simulation and the second study.

CHAPTER IV

RESULTS

The results are divided into two major sections. The first section deals with the Monte Carlo study, whereas the second section details the results from a dataset extracted from an assessment project completed in 1993. Within each section, base analyses discussing descriptive statistics come first followed by primary analyses involving all of the predictions. Each section concludes with an examination of any relevant supplementary analyses.

Before delving into the various analyses associated with each study, an important note is warranted. During the initial phase of data analysis, it became apparent that the constrained estimation procedure was not performing optimally. When the adverse impact weights were set to lower levels (e.g., .15, .25) the routine provided estimates along the lines of what was expected. However, as the weights increased from around .30 to .50, the routine began to sometimes work against itself when no positive predictor weight could be assigned to cognitive ability (or a variate thereof). At times, this allowance of a negative weight for cognitive ability increased the potential for greater R-squares and less overall adverse impact. However, this increased potential was not consistent and occasionally resulted in a disproportionately large importance being placed on minimizing subgroup differences. For example, the R-squared and adverse impact ratio in one analysis changed from .17 and .31 to .08 and 1.03, respectively, with only a minimal change in the adverse impact weight. In essence, the optimization routine ignored a range of important predictor weights that would have resulted in some middle ground between the results found with differing adverse impact weights.

To rectify this issue, two steps were taken. First, the subgroup differences minimization portion of the program was simplified by transforming the minimization penalties into functions of the OLS solution. Within this process, the R-squared penalty became zero for OLS regression while the adverse impact penalty became one. In other words, the constrained estimation routine considered the validity from OLS regression as the maximum validity possible whereas the subgroup difference associated with the OLS predictor composite was considered to be the minimum. These penalties were weighted, as in the initial equation, using the adverse impact weight (i.e., AIW) and its complement. The mathematical representation of this adjusted function can be seen as follows:

$$f\left(\beta_{\mathrm{i}}, AIW, Z_{cutpo\,\mathrm{int}}, Y, X, Gp\right) = AIW\left(\frac{abs\left(z_{gp1} - z_{gp2}\right)}{abs\left(z_{gp1_{ols}} - z_{gp2_{ols}}\right)}\right) + \left(1 - AIW\right)\left(1 - \frac{r_i^2}{r_{ols}^2}\right)$$

Secondly, a constraint was placed on the estimation to limit predictor weights to positive values. In other words, negative predictor weights were specifically excluded from the estimation. While this approach based its rationale in stabilizing the estimates, it also found support from an additional standpoint. Allowing negative predictor weights basically translates as a non-traditional selection strategy where, in this case, suppression becomes the primary vehicle of the selection system's validity. For example, if cognitive ability were to be chosen for inclusion in a selection system based on a high correlation with job performance, it would be assumed that greater cognitive ability generally resulted in greater job performance. However, placing a negative weight on cognitive ability within a selection system containing multiple predictors might provide more optimal prediction if (1) one of the other predictors shared variance with cognitive ability that was specifically related to job performance, (2) this second predictor was given a positive weight, and (3) the variance in cognitive ability that was not shared with the second predictor was also unrelated to job performance (thus representing the negatively weighted variance). This is the essence of suppression, and its effects can appropriately boost the explained variance of a criterion. However, this becomes particularly troublesome when considering the fact that most predictors are at least initially chosen based on their individual relationships with the criterion of interest. The use of suppression would probably be interpreted by many as penalizing someone for scoring too high on a predictor that has shown a strong relationship with the criterion. While these types of strategies have been upheld in court (see *Jordan v. City of New London,* 1999; Demonte & Arnold, 2000), the cases generally dealt with situations substantially different from predicting performance (*Jordan* involved cognitive ability and the prediction of employment longevity for a police officer position) and will probably always present a difficult proposition for defense in a court of law. Therefore, for the purposes of this study, predictor weights were constrained to retain a positive weight.

Study One

Descriptive Statistics

Before proceeding with the analyses, an effort was undertaken to ensure that the routine responsible for generating the Monte Carlo data (N = 200,000) accurately replicated Bobko et al.'s (1999) model matrix. The initial examination progressed using the model matrix in its exact form. However, during this process, the resulting subgroup difference associated with cognitive ability in the Monte Carlo data was found to be somewhat smaller than expected (around .90 rather than 1.0). Further review found that the program was working accurately but that the subgroup correlation for cognitive ability (i.e., the correlation between cognitive ability and race) was still slightly lower than what was needed to replicate the original matrix. To adjust for this, the subgroup correlation for cognitive ability (initially established by Bobko et al.) was increased from .37 to .40. After this change was made, the program was used to generate 20 samples of 20,000 subjects each. The criterion and predictor variates were designed to have a mean of zero and a standard deviation of one. A minority proportion of 20% was used in this process because that was the majority to minority ratio used by Bobko et al. The 20 samples were then individually compared to the model matrix using root mean squared residuals (RMS). These RMSs can be found in Table 5 along with overall deviations in *d* for each criterion and predictor variate. The descriptive statistics associated with these comparisons can be found in Table 6.

These tables show that there was very little difference between the model matrix and that of the Monte Carlo generated matrices. In fact, the average RMS across all of the 20 comparisons was .006 with a standard deviation of .001. Moreover, the largest average subgroup difference discrepancy within these comparisons came with the conscientiousness variate at -.014. Given that Bobko et al. (1999) reported these coefficients with only two decimals of accuracy, this evidence led to the conclusion that the Monte Carlo generated matrices were almost exact replicas of the model matrix upon which they were based, and the primary Monte Carlo matrix of 200,000 subjects was created. The *d* values and correlation matrix from this dataset were then contrasted with the originating base matrix from Bobko et al. in Table 7. This table reveals only two

Table 5: Overall Deviations in *d* Across all Predictor Variates and General Performance Between the Model Matrix and Twenty Monte Carlo Generated Matrices Using Populations of 20,000 Individuals

Sample	RMS	Perf	CA	SI	Con	Bio
1	.0097	.0138	.0226	.0558	-.0390	-.0225
2	.0063	-.0176	.0053	-.0084	-.0240	-.0183
3	.0064	-.0064	-.0012	-.0225	-.0234	-.0157
4	.0057	-.0148	-.0181	-.0116	-.0011	-.0059
5	.0076	.0046	-.0361	-.0085	-.0351	-.0038
6	.0068	-.0303	-.0028	.0292	.0214	-.0020
7	.0061	-.0127	.0010	.0065	-.0077	-.0209
8	.0071	-.0294	-.0192	.0050	.0170	.0294
9	.0070	-.0052	.0118	-.0058	-.0477	-.0118
10	.0073	-.0065	-.0202	-.0169	.0109	.0224
11	.0044	-.0047	-.0090	-.0008	-.0104	.0276
12	.0044	.0226	-.0210	-.0068	-.0140	.0120
13	.0061	.0153	.0010	-.0095	-.0113	.0001
14	.0046	.0016	.0037	.0061	-.0281	-.0021
15	.0075	-.0030	-.0207	.0324	-.0438	-.0206
16	.0052	-.0042	.0059	-.0042	-.0132	-.0067
17	.0076	-.0138	.0006	.0314	-.0131	.0181
18	.0051	-.0219	-.0264	-.0035	.0019	.0150
19	.0066	-.0060	-.0285	-.0167	.0041	.0116
20	.0059	-.0105	.0134	-.0364	-.0141	.0070

Note: RMS refers to the root mean squared residual comparing the model matrix to the generated matrix from each line or sample. The other values refer to the differences between the expected race mean differences and the observed differences. Cognitive ability, structured interviews, conscientiousness, biographical data, and job performance are referred to as CA, SI, Con, Bio, and Job Perf, respectively.

Table 6: Descriptive Statistics Associated with Model to Monte Carlo Data Comparisons

	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
RMS	.006	.001	.412	.021	.006	.004	.010	20000
Perf	-.006	.014	.255	-.444	-.006	-.030	.023	20000
CA	-.007	.016	-.077	-1.179	-.002	-.036	.023	20000
SI	.001	.022	.833	.181	-.005	-.036	.056	20000
Con	-.014	.019	-.016	-.934	-.013	-.048	.021	20000
Bio	.001	.017	.211	-1.300	-.002	-.023	.029	20000

Note: RMS refers to the root mean squared residual. Cognitive ability, structured interviews, conscientiousness, biographical data, and job performance are referred to as CA, SI, Con, Bio, and Perf, respectively.

Table 7: Matrix of Deviations Between the Model Matrix's Correlations and *d* Values to that of the Monte Carlo Generated Dataset

			Intercorrelations			
	d	**Subgroup**	**1**	**2**	**3**	**4**
1. Cognitive Ability	.00	-.03				
2. Structured Interview	.00	.00	.00			
3. Conscientiousness	-.01	.00	.00	.00		
4. Biographical Data	.00	.00	.00	.00	.00	
5. General Performance	.00	.00	.00	.00	.00	.00

Note: Negative values indicate that the Model Matrix provided smaller estimates than the Monte Carlo dataset.

minor differences between the two matrices (both of these matrices as well as the deviation matrix can be found in Appendix C). First, as mentioned previously, the subgroup correlation for cognitive ability was increased to .40 (a change of .03) to account for cognitive ability's one standard deviation difference in performance (*d*) between the groups. Second, the subgroup difference for conscientiousness was .10 in the Monte Carlo dataset as opposed to .09 from Bobko et al. (a change of only .01). As such, it was determined that any error occurring between the two matrices was compatible within an expected range.

Once this overall dataset was created, three samples of 10,000 subjects were randomly drawn conforming to minority proportions of 5%, 20%, and 40%. While the samples closely resembled the overall dataset of 200,000, there were some small differences. The intercorrelations and subgroup correlations appeared to get larger as the percentage of minorities expanded. This should be expected when increasing the number of individuals who belong to a generally lower scoring group (all of the subgroup differences favored the majority group). In effect, the increase equates simply to an increase in variance. The differences in *d* values did not demonstrate the same trend. However, at the same time, none of these values were more than about .05 different from each other. For reference, the correlation and subgroup difference matrices associated with each of these samples can be found in Appendix D.

The next step involved randomly splitting each of these three samples approximately in half to allow for testing under cross-validation. This resulted in a total of six samples (three for validation and three for cross-validation) of about 5000 subjects each. The descriptive statistics associated with each sample can be viewed in Appendix E.

Primary Analyses

The primary analyses involved generating R-squares, adverse impact ratios, shrinkage, and predictor weights for each of the predictor combinations noted in Table 4. This process included data from six disparate samples that varied by purpose (validation and cross-validation) and minority proportion (5%, 20%, and 40%). Additionally, three types of predictor weighting methodologies were incorporated: (1) OLS regression, (2) Unit weighting, and (3) Constrained estimation. Once predictor weights were established in each of the validation samples, these weights were then used with each of the corresponding cross-validation samples to derive the same estimates. Table 8 presents the overall results from this effort (complete results including predictor weights can be found in Appendices F, G, and H). For each specific estimate (i.e., R-squared, adverse impact ratio, and shrinkage) and across minority samples and sample purpose, 33 comparisons were made between OLS regression and unit weights, whereas 132 comparisons were

Table 8: Monte Carlo R-squares, Adverse Impact Ratios, Cross-Validation Results, and Shrinkage estimates for Various Predictor Combinations in Three Minority Samples

	5% Minority						20% Minority						40% Minority					
	Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage	
Predictors	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR
CA	.08	.12	.07	.11	-.01	-.01	.10	.15	.08	.13	-.02	-.02	.09	.17	.10	.16	.01	-.01
SI	.09	.62	.08	.93	-.01	.31	.09	.84	.10	.84	.01	*	.10	.71	.10	.76	*	.05
Con	.03	.88	.03	1.02	*	.14	.03	.92	.03	.87	*	-.05	.04	.84	.03	.96	-.01	.12
Bio	.07	.53	.07	.60	*	.07	.08	.67	.08	.62	*	-.05	.08	.63	.08	.65	*	.02
CA+SI																		
OLS	.14	.23	.13	.22	-.01	-.01	.15	.29	.15	.30	*	.01	.15	.29	.16	.29	.01	*
Unit	.14	.23	.13	.22	-.01	-.01	.15	.30	.15	.30	*	*	.15	.29	.16	.28	.01	-.01
Constrained																		
AIW=0.15	.14	.31	.13	.26	-.01	-.05	.15	.34	.15	.33	*	-.01	.15	.32	.16	.34	.01	.02
AIW=0.25	.14	.31	.12	.30	-.02	-.01	.14	.39	.14	.38	*	-.01	.14	.36	.16	.37	.02	.01
AIW=0.35	.13	.35	.12	.47	-.01	.12	.13	.51	.13	.49	*	-.02	.13	.45	.15	.46	.02	.01
AIW=0.50	.09	.62	.08	.93	-.01	.31	.09	.84	.10	.84	.01	*	.10	.71	.10	.76	*	.05
CA+Con																		
OLS	.11	.14	.10	.17	-.01	.03	.13	.21	.12	.23	-.01	.02	.13	.21	.13	.24	*	.03
Unit	.11	.24	.10	.25	-.01	.01	.12	.30	.11	.30	-.01	*	.12	.28	.12	.30	*	.02
Constrained																		
AIW=0.15	.11	.15	.10	.17	-.01	.02	.13	.22	.12	.23	-.01	.01	.13	.23	.13	.25	*	.02
AIW=0.25	.11	.15	.10	.17	-.01	.02	.13	.24	.11	.24	-.02	*	.13	.24	.13	.27	*	.03
AIW=0.35	.11	.21	.10	.30	-.01	.09	.13	.26	.11	.28	-.02	.02	.12	.27	.13	.28	.01	.01
AIW=0.50	.08	.42	.08	.44	*	.02	.03	.92	.03	.87	*	-.05	.04	.84	.03	.96	-.01	.12
CA+Bio																		
OLS	.13	.12	.12	.15	-.01	.03	.15	.19	.14	.24	-.01	.05	.14	.23	.15	.24	.01	.01
Unit	.13	.12	.12	.19	-.01	.07	.15	.21	.14	.26	-.01	.05	.14	.23	.15	.24	.01	.01
Constrained																		
AIW=0.15	.13	.12	.12	.19	-.01	.07	.15	.21	.14	.27	-.01	.06	.14	.24	.15	.25	.01	.01
AIW=0.25	.13	.14	.12	.19	-.01	.05	.15	.23	.13	.28	-.02	.05	.14	.26	.15	.26	.01	*
AIW=0.35	.13	.12	.11	.33	-.02	.21	.14	.25	.13	.32	-.01	.07	.14	.31	.14	.32	*	.01
AIW=0.50	.07	.53	.07	.60	*	.07	.08	.67	.08	.62	*	-.05	.08	.63	.08	.65	*	.02

Table 8: Continued

	5% Minority						20% Minority						40% Minority					
	Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage	
Predictors	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR
SI+Con																		
OLS	.11	.64	.11	.90	*	.26	.11	.82	.12	.76	.01	-.06	.12	.71	.12	.78	*	.07
Unit	.10	.74	.10	.78	*	.04	.11	.84	.11	.77	*	-.07	.11	.71	.11	.82	*	.11
Constrained																		
AIW=0.15	.11	.65	.11	.90	*	.25	.11	.81	.12	.76	.01	-.05	.12	.71	.12	.78	*	.07
AIW=0.25	.11	.65	.11	.90	*	.25	.11	.82	.12	.76	.01	-.06	.12	.71	.12	.78	*	.07
AIW=0.35	.11	.64	.11	.90	*	.26	.11	.82	.12	.76	.01	-.06	.12	.71	.12	.78	*	.07
AIW=0.50	.11	.70	.11	.88	*	.18	.11	.84	.12	.75	.01	-.09	.12	.71	.12	.79	*	.08
SI+Bio																		
OLS	.14	.53	.14	.70	*	.17	.14	.67	.15	.66	.01	-.01	.15	.61	.16	.63	.01	.02
Unit	.14	.51	.14	.70	*	.19	.14	.66	.15	.65	.01	-.01	.15	.60	.15	.62	*	.02
Constrained																		
AIW=0.15	.14	.54	.14	.70	*	.16	.14	.66	.15	.67	.01	.01	.15	.61	.16	.63	.01	.02
AIW=0.25	.14	.54	.14	.72	*	.18	.14	.66	.15	.67	.01	.01	.15	.62	.16	.64	.01	.02
AIW=0.35	.14	.57	.14	.70	*	.13	.14	.66	.15	.68	.01	.02	.15	.63	.16	.64	.01	.01
AIW=0.50	.13	.61	.13	.73	*	.12	.13	.74	.14	.69	.01	-.05	.14	.63	.15	.67	.01	.04
Con+Bio																		
OLS	.07	.59	.08	.65	.01	.06	.08	.74	.08	.66	*	-.08	.08	.66	.08	.69	*	.03
Unit	.06	.67	.07	.75	.01	.08	.07	.80	.07	.76	*	-.04	.08	.73	.07	.76	-.01	.03
Constrained																		
AIW=0.15	.07	.59	.08	.65	.01	.06	.08	.75	.08	.66	*	-.09	.08	.67	.08	.69	*	.02
AIW=0.25	.07	.54	.08	.63	.01	.09	.08	.78	.08	.68	*	-.10	.08	.68	.08	.68	*	*
AIW=0.35	.07	.56	.08	.65	.01	.09	.08	.79	.08	.69	*	-.10	.08	.68	.08	.70	*	.02
AIW=0.50	.07	.66	.07	.73	*	.07	.06	.86	.06	.80	*	-.06	.07	.73	.06	.80	-.01	.07

Table 8: Continued

	5% Minority						20% Minority						40% Minority					
	Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage	
Predictors	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR
CA+SI+Con																		
OLS	.16	.31	.15	.22	-.01	-.09	.17	.31	.17	.31	*	*	.17	.32	.18	.30	.01	-.02
Unit	.16	.37	.15	.34	-.01	-.03	.17	.39	.16	.38	-.01	-.01	.17	.35	.18	.35	.01	*
Constrained																		
AIW=0.15	.16	.32	.15	.30	-.01	-.02	.17	.37	.17	.36	*	-.01	.17	.35	.18	.35	.01	*
AIW=0.25	.16	.37	.15	.34	-.01	-.03	.17	.43	.17	.43	*	*	.17	.39	.18	.39	.01	*
AIW=0.35	.15	.42	.14	.54	-.01	.12	.15	.58	.15	.58	*	*	.15	.49	.16	.50	.01	.01
AIW=0.50	.11	.64	.11	.90	*	.26	.11	.82	.12	.76	.01	-.06	.12	.71	.12	.78	*	.07
CA+SI+Bio																		
OLS	.18	.20	.17	.28	-.01	.08	.19	.27	.18	.33	-.01	.06	.19	.32	.20	.32	.01	*
Unit	.18	.20	.17	.28	-.01	.08	.19	.28	.18	.35	-.01	.07	.19	.31	.20	.31	.01	*
Constrained																		
AIW=0.15	.18	.23	.17	.31	-.01	.08	.19	.33	.18	.40	-.01	.07	.19	.36	.20	.37	.01	.01
AIW=0.25	.17	.28	.16	.39	-.01	.11	.18	.38	.18	.45	*	.07	.18	.41	.19	.42	.01	.01
AIW=0.35	.16	.40	.15	.52	-.01	.12	.16	.54	.13	.69	-.03	.15	.17	.50	.17	.54	*	.04
AIW=0.50	.14	.59	.14	.70	*	.11	.14	.65	.14	.68	*	.03	.15	.62	.16	.65	.01	.03
CA+Con+Bio																		
OLS	.14	.14	.13	.14	-.01	*	.15	.20	.14	.24	-.01	.04	.15	.22	.16	.24	.01	.02
Unit	.12	.21	.12	.30	*	.09	.14	.38	.13	.37	-.01	-.01	.14	.34	.14	.35	*	.01
Constrained																		
AIW=0.15	.14	.14	.13	.15	-.01	.01	.15	.21	.14	.26	-.01	.05	.15	.24	.15	.25	*	.01
AIW=0.25	.14	.14	.13	.19	-.01	.05	.15	.24	.14	.28	-.01	.04	.15	.27	.15	.28	*	.01
AIW=0.35	.13	.14	.12	.26	-.01	.12	.14	.30	.14	.33	*	.03	.14	.32	.14	.32	*	*
AIW=0.50	.07	.61	.07	.67	*	.06	.08	.80	.08	.70	*	-.10	.08	.71	.07	.71	-.01	*

Table 8: Continued

	5% Minority						20% Minority						40% Minority					
	Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage		Validation		Cross-Val		Shrinkage	
Predictors	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR	R^2	AIR
SI+Con+Bio																		
OLS	.14	.57	.14	.70	*	.13	.15	.69	.15	.68	*	-.01	.15	.63	.16	.63	.01	*
Unit	.12	.64	.12	.70	*	.06	.13	.73	.13	.66	*	-.07	.13	.63	.13	.64	*	.01
Constrained																		
AIW=0.15	.14	.57	.14	.70	*	.14	.14	.70	.15	.67	.01	-.03	.15	.63	.16	.63	.01	*
AIW=0.25	.14	.61	.14	.65	*	.04	.14	.71	.15	.69	.01	-.02	.15	.63	.15	.64	*	.01
AIW=0.35	.14	.62	.14	.67	*	.05	.14	.71	.15	.69	.01	-.02	.15	.63	.15	.66	*	.03
AIW=0.50	.12	.64	.12	.77	*	.13	.11	.84	.12	.76	.01	-.08	.13	.69	.13	.75	*	.06
CA+SI+Con+Bio																		
OLS	.18	.21	.17	.28	-.01	.07	.19	.29	.19	.36	*	.07	.19	.32	.20	.33	.01	.01
Unit	.17	.31	.16	.34	-.01	.03	.18	.38	.17	.42	-.01	.04	.18	.36	.18	.38	*	.02
Constrained																		
AIW=0.15	.18	.26	.17	.33	-.01	.07	.19	.32	.19	.38	*	.06	.19	.36	.20	.37	.01	.01
AIW=0.25	.17	.31	.17	.37	*	.06	.18	.39	.18	.47	*	.08	.19	.39	.19	.41	*	.02
AIW=0.35	.16	.45	.16	.52	*	.07	.16	.58	.17	.60	.01	.02	.17	.51	.18	.53	.01	.02
AIW=0.50	.14	.62	.14	.68	*	.06	.14	.71	.15	.69	.01	-.02	.14	.63	.15	.66	.01	.03

Note: R^2 refers to the amount of variance explained. AIR refers to the adverse impact ratio. AIW refers to the adverse impact weight. Cognitive ability, structured interviews, conscientiousness, and biographical data are referred to as CA, SI, Con, and Bio, respectively. * indicates those instances where no shrinkage occurred.

made between either OLS or unit weights and those from constrained estimation.

Predictions 1a, 1b, and 1c. Prediction 1a served as somewhat of a manipulation check of whether or not the results from the optimization routine replicated the results from OLS regression when no importance was placed on minimizing subgroup differences. Because the constrained estimation procedure actually begins the iterative process of manipulating predictor weights with the weights provided by OLS regression, the results from these two methods should only deviate when the adverse impact weight is greater than zero. In other words, when no importance is placed on minimizing subgroup differences (*AIW* = 0), constrained estimation should produce the exact same results as that of OLS regression. This prediction was confirmed across every sample and predictor combination.

Prediction 1b examined the explained variance (i.e., R-squared) provided by constrained estimation, compared it to that of OLS regression, and predicted similar estimates. Referring to Table 8, this prediction held true in every situation save for those where *AIW* = .50. In fact, Table 9 shows that across each of the minority samples and throughout all of the predictor combination comparisons where the adverse impact weight was set to .15, .25, and .35 (99 in total), constrained estimation most frequently resulted in the exact same (72), or at most .01 lower (17), R-squared values as that of OLS regression. Moreover, there were only eight times that these results differed by as much as .02 and only twice did they reach a difference of .03 (both CA+SI+Bio and CA+SI+Con+Bio at an *AIW* = .35 in the 20% minority sample). When these results were broken down by specific adverse impact weight, an ensuing trend was found where more differences occurred as the weight was increased. For example, when this weight was set

Table 9: Frequency and Degree of R-square Change Between both OLS Regression and Unit Weighting and that of Constrained Estimation Broken Down by Adverse Impact Weight and Minority Proportion

	OLS Regression					Unit Weighting			
	Minority Proportion					Minority Proportion			
Change	**5%**	**20%**	**40%**	**Total**	**Change**	**5%**	**20%**	**40%**	**Total**
AIW = .15					**AIW = .15**				
+.02	*	*	*	0	**+.02**	2	*	1	3
+.01	*	*	*	0	**+.01**	3	5	4	12
.00	11	10	11	32	**.00**	6	6	6	18
-.01	*	1	*	1	**-.01**	*	*	*	0
-.02	*	*	*	0	**-.02**	*	*	*	0
-.03	*	*	*	0	**-.03**	*	*	*	0
AIW = .25					**AIW = .25**				
+.02	*	*	*	0	**+.02**	2	*	1	3
+.01	*	*	*	0	**+.01**	2	4	4	10
.00	9	7	9	25	**.00**	6	5	4	15
-.01	2	4	2	8	**-.01**	1	2	2	5
-.02	*	*	*	0	**-.02**	*	*	*	0
-.03	*	*	*	0	**-.03**	*	*	*	0
AIW = .35					**AIW = .35**				
+.02	*	*	*	0	**+.02**	1	*	1	2
+.01	*	*	*	0	**+.01**	3	3	1	7
.00	6	4	5	15	**.00**	3	3	5	11
-.01	3	3	2	8	**-.01**	3	1	1	5
-.02	2	2	4	8	**-.02**	1	4	3	8
-.03	*	2	*	2	**-.03**	*	*	*	0
Total	33	33	33	99	**Total**	33	33	33	99

Note: AIW refers to the adverse impact weight. The "Total" columns reflect the total number of R^2 deviations across the three samples at each level of change. The row totals reflect the total number of counts or comparisons found within each sample as well as the overall total. * refers to zero but is used to make the table easier to read. Positive changes reflect larger constrained estimation estimates.

to .15, there was only one time when constrained estimation and OLS regression differed at all, and that was only at .01. Furthermore, eight comparisons were found to differ by .01 when the weight was set to .25 whereas both the number and size of the differences increased when the adverse impact weight was .35.

When the adverse impact weight was set to .50, relatively larger decreases in R-squared were more common. These reductions ranged from 0 to .10 with an average of .04. Interestingly, the average decrease from those scenarios including the cognitive ability variate as a predictor was much larger (.06) than the average from those excluding the cognitive ability variate from use (.01).

Prediction 1c was concerned with comparisons between R-squared results from unit weighting and those from constrained estimation. Specifically, it was expected that constrained estimation would produce greater validity. As with the comparisons to OLS regression, the estimates from these two weighting schemes were more similar than dissimilar with larger differences occurring when the adverse impact weight was set to .50. Specifically, across all of the scenarios when the weight was set to .15, .25, and .35 (see Table 9), 37 comparisons were found to favor constrained estimation (29 by .01 and 8 by .02), 44 displayed no differences, and 18 were found to favor unit weighting (10 by .01 and 8 by .02). There was also some variation, however slight, found when the results were separated by specific adverse impact weight. When the weight was set to .15, constrained estimation produced larger or equal R-squares in every instance, but the average difference was only .014. Unit weighting resulted in a greater explanation of variance in five cases when the weight was set to .25 (as opposed to 13 comparisons favoring constrained estimation) and 13 cases when the weight was set to .35 (as opposed to nine comparisons favoring constrained estimation). However, the average difference between constrained estimation and unit weighting when the adverse impact weights were set to .25 and .35 remained small (.006 and -.005, respectively).

When the adverse impact weight was set to .50, the results ran somewhat parallel to that of the comparisons with OLS regression. Differences between constrained estimation and unit weighting ranged from +1 to -.09 (with negative values indicating higher unit weighted validity) with an average of .04. Of some importance, there was again large variation in the size of differences based on whether the cognitive ability variate was included (.05) and when it was excluded (.003).

In summary, the predictions involving validity and explanations of variance received mixed support. Prediction 1a was fully confirmed, and although larger and more frequent differences occurred when the adverse impact weight was set to .50, prediction 1b was supported by virtue of the generally small discrepancies found between OLS regression and constrained estimation. Prediction 1c received partial support from the number of those comparisons favoring constrained estimation over unit weighting (37) when the adverse impact weight was set to .15, .25, and .35. However, within this range of adverse impact weights, there were still 18 comparisons favoring unit weighting, and the average difference at each of these weights was very small (.014, .006, and -.0047, respectively). Given the additional differences that occurred when the weight was set to .50 (three with constrained estimation improving prediction and 27 where unit weighting resulted in larger estimates) as well as the overall difference of .04 favoring unit weighting in this circumstance, prediction 1c was not supported.

Predictions 2a, 2b, and 2c. These predictions examined the cross-validation estimates between the methodologies of OLS regression, unit weighting, and constrained estimation. It was predicted that constrained estimation would result in less shrinkage than OLS regression and that unit weighting would reveal less shrinkage than both OLS regression and constrained estimation. In general, there was very little difference between the three methodologies across all of the predictor combinations and minority proportions, and the overall shrinkage was minimal.

As can be seen in Table 10, of the 198 scenarios presented, most of the cross-validation work resulted in either no shrinkage (85) or a change in validity (from initial validation work) of .01 (54 at .01 and 51 at -.01). In fact, there were only eight instances that found differences of .02 or .03 (two at .02, five at -.02, and one at -.03). From a comparative standpoint, OLS regression and unit weighting resulted in the exact same shrinkage under cross-validation 24 times (out of 33) and differed by only .01 for the other nine scenarios. Moreover, regardless of the adverse impact weight used, constrained estimation most often reflected the same level of shrinkage as OLS regression (86 of 132 comparisons) with 45 differences of.01 and only a single shrinkage estimate reaching a difference of .02. A similar trend ensued when comparing unit weighting to constrained estimation. Of the 132 comparisons, 70 resulted in no shrinkage differences, and 58 found a contrast of .01. There were only four scenarios under cross-validation that found differences of .02, and these all showed unit weighting with slightly less shrinkage. Therefore, because none of the methodologies appeared to outperform another under cross-validation, no support was found for these three predictions.

Table 10: Actual and Average Shrinkage by Weighting Method and Minority Proportion

5% Minority Proportion							
	Weighting Method						
Predictors	**OLS**	**Unit**	**AIW=.15**	**AIW=.25**	**AIW=.35**	**AIW=.50**	**Average**
CA+SI	-.01	-.01	-.01	-.02	-.01	-.01	-.0117
CA+Con	-.01	-.01	-.01	-.01	-.01	*	-.0083
CA+Bio	-.01	-.01	-.01	-.01	-.02	*	-.01
SI+Con	*	*	*	*	*	*	0
SI+Bio	*	*	*	*	*	*	0
Con+Bio	.01	.01	.01	.01	.01	*	.0083
CA+SI+Con	-.01	-.01	-.01	-.01	-.01	*	-.0083
CA+SI+Bio	-.01	-.01	-.01	-.01	-.01	*	-.0083
CA+Con+Bio	-.01	*	-.01	-.01	-.01	*	-.0067
SI+Con+Bio	*	*	*	*	*	*	0
CA+SI+Con+Bio	-.01	-.01	-.01	*	*	*	-.005
Average	-.0055	-.0045	-.0055	-.0055	-.0055	-.0009	

20% Minority Proportion							
	Weighting Method						
Predictors	**OLS**	**Unit**	**AIW=.15**	**AIW=.25**	**AIW=.35**	**AIW=.50**	**Average**
CA+SI	*	*	*	*	*	.01	0017
CA+Con	-.01	-.01	-.01	-.02	-.02	*	-.0117
CA+Bio	-.01	-.01	-.01	-.02	-.01	*	-.01
SI+Con	.01	*	.01	.01	.01	.01	.0083
SI+Bio	.01	.01	.01	.01	.01	.01	.01
Con+Bio	*	*	*	*	*	*	0
CA+SI+Con	*	-.01	*	*	*	.01	0
CA+SI+Bio	-.01	-.01	-.01	*	-.03	*	-.01
CA+Con+Bio	-.01	-.01	-.01	-.01	*	*	-.0067
SI+Con+Bio	*	*	.01	.01	.01	.01	.0067
CA+SI+Con+Bio	*	-.01	*	*	.01	.01	.0017
	-.0018	.0045	-.0019	-.0018	-.0018	.0055	

40% Minority Proportion							
	Weighting Method						
Predictors	**OLS**	**Unit**	**AIW=.15**	**AIW=.25**	**AIW=.35**	**AIW=.50**	**Average**
CA+SI	.01	.01	.01	.02	.02	*	.0117
CA+Con	*	*	*	*	.01	-.01	0
CA+Bio	.01	.01	.01	.01	*	*	.0067
SI+Con	*	*	*	*	*	*	0
SI+Bio	.01	*	.01	.01	.01	.01	.0083
Con+Bio	*	-.01	*	*	*	-.01	-.0033
CA+SI+Con	.01	.01	.01	.01	.01	*	.0083
CA+SI+Bio	.01	.01	.01	.01	*	.01	.0083
CA+Con+Bio	.01	*	*	*	*	-.01	0
SI+Con+Bio	.01	*	.01	*	*	*	.0033
CA+SI+Con+Bio	.01	*	.01	*	.01	.01	.0067
	.0073	.0027	.0064	.0055	.0055	0	

Note: AIW refers to the adverse impact weight. Cognitive ability, structured interviews, conscientiousness, and biographical data are referred to as CA, SI, Con, and Bio, respectively. * indicates no shrinkage.

Prediction 3. This prediction explored the effectiveness of constrained estimation in reducing adverse impact while preserving validity. Specifically, it was predicted that when sufficient importance was placed on minimizing subgroup differences and if sufficient predictor variance was available, the use of constrained estimation would simultaneously eliminate adverse impact and sustain validity at comparable levels to that of OLS regression and unit weighting. Overall, this prediction received mixed results.

A quick glimpse at Table 8 shows that neither OLS regression nor unit weighting fared very well with regards to adverse impact. In fact, across all of the minority samples in the validation datasets, these two methodologies only resulted in an acceptable adverse impact ratio (greater than or equal to .80) in three instances (both OLS regression and unit weighting with the predictor combination of the structured interview and conscientiousness and unit weighting with conscientiousness and biographical data). However, a closer look at these results reveals a couple of facets about how predictor variance influences selection-based estimates.

Table 11 breaks down both adverse impact and validity by number of predictors and predictor combination content (with and without cognitive ability included). This table shows that while OLS regression demonstrated higher validity regardless of the scenario examined, unit weighting appeared to result in less adverse impact. Presumably, this was derived by the manner with which each methodology attends to weighting predictors. OLS regression gives the most weight to those predictors that display the strongest relationship to the criterion of interest (e.g., job performance), whereas unit weighting assigns the same weight across all predictors. In this situation, the predictors with the strongest relationships to the criterion also culminated in the largest subgroup

Table 11: Average R-squared and Adverse Impact Ratio by Number of Predictors and Predictor Combination Content using OLS Regression and Unit Weighting

	R-squared						Adverse Impact Ratio					
	OLS			Unit			OLS			Unit		
	5%	20%	40%	5%	20%	40%	5%	20%	40%	5%	20%	40%
2 Predictors	.12	.13	.13	.11	.12	.13	.38	.49	.45	.42	.52	.47
3 Predictors	.16	.17	.17	.15	.16	.16	.31	.37	.37	.36	.45	.41
4 Predictors	.18	.19	.19	.17	.18	.18	.21	.29	.32	.31	.38	.36
All Predictor Combinations												
CA included	.15	.16	.16	.14	.16	.16	.19	.25	.27	.24	.32	.31
CA excluded	.12	.12	.13	.11	.11	.12	.58	.73	.65	.64	.76	.67
2 Predictors												
CA included	.13	.14	.14	.13	.14	.14	.16	.23	.24	.20	.27	.27
CA excluded	.11	.11	.12	.10	.11	.11	.59	.74	.66	.64	.77	.68
3 Predictors												
CA included	.16	.17	.17	.15	.17	.17	.22	.26	.29	.26	.35	.33
CA excluded	.14	.15	.15	.12	.13	.13	.57	.69	.63	.64	.73	.63

Note: CA refers to the cognitive ability variate.

differences on this criterion. Therefore, when these predictors were used in OLS regression, increased validity ensued. However, these same predictors were given a smaller relative weight in unit weighting, and thus the predictors with weaker relationships to the criterion (but, in this case, also displaying smaller subgroup differences) were allowed to have more relative influence over the resulting adverse impact ratios.

Another interesting aspect revealed by this table is the fact that while greater validity was attained when cognitive ability was included as a predictor and as the number of predictors was increased, the exact opposite occurred when viewing the adverse impact ratios. Using additional predictors actually exacerbated the amount of adverse impact as did including cognitive ability in the selection procedure. However, notice that when the first variation (i.e., increasing predictors) was further broken down by whether or not cognitive ability was entered into the equation, a slightly different picture developed. Including cognitive ability and moving from a two- to three-predictor combination lessened the amount of adverse impact; but when cognitive ability was excluded, this same movement caused the average level of adverse impact to increase. This increase was probably due to the average addition of predictors with different but additive variance associated with subgroup differences. The decrease when cognitive ability was used was likely caused by a dilution of subgroup differences when predictors with less adverse impact were added. Still, it is important to note that the overall level of adverse impact was substantially lower when cognitive ability was excluded. However, these results give rise to the notion that individual predictor variance (whether associated with validity or adverse impact) can play an enormous role in the resulting estimates obtained from any particular predictor composite. The role of individual predictor variance is even more prominent within constrained estimation.

Observing the results of constrained estimation in Table 8, it is clear that the use of this optimization technique influenced both validity and adverse impact. As expected, this influence varied based on three major factors: (1) the number of predictors, (2) predictor combination content, and (3) the specific adverse impact weight utilized. In general, the benefits of constrained estimation increased as the number of predictors increased and when cognitive ability was included as a predictor. When two predictors were used, there was very little change in either validity or adverse impact. The lone exception came from the use of cognitive ability and the structured interview. While this combination reverted to the estimates offered by the structured interview alone when the adverse impact weight was set to .50, there was a meaningful reduction in adverse impact as compared to OLS regression when the weight was set to .35 (e.g., .29 versus .51 in the 20% minority sample). Moreover, the loss of validity from this scenario was minimal (from an R-squared of .15 to .13).

As can be seen in Table 12, larger changes in adverse impact were found with three-predictor combinations. Using an adverse impact weight of .15 typically replicated what was found with OLS regression. At a weight of .25, the reduction to adverse impact ranged from .01 to .12 with an average of .06. This average increased slightly to .07 when only those predictor combinations including cognitive ability were examined (compared to .02 when cognitive ability was excluded). When a weight of .35 was utilized, a mean reduction of .11 was found with an even larger discrepancy between the average from

Table 12: Average Reductions in Adverse Impact Ratios using Constrained Estimation in Three-Predictor Combinations

	Adverse impact weight			
	AIW=.15	**AIW=.25**	**AIW=.35**	**AIW=.50**
Overall Average	.023	.057	.114	.345
CA included	.029	.069	.144	.428
CA excluded	.003	.020	.023	.093
Range	0.0 - .06	0.0 - .12	0.0 - .20	.06 - .60

Note: AIW refers to the adverse impact weight. CA refers to cognitive ability.

those combinations including cognitive ability (.14) as opposed to those that did not (.02). Yet, the most dramatic changes occurred with an adverse impact weight of .50. Using this weight, an overall reduction in adverse impact of .35 was found with substantial variation based on predictor combination content. This difference highlights the role of predictor variance when using constrained estimation. This process resulted in small to sometimes almost insignificant changes in both validity and adverse impact when cognitive ability was excluded from the predictor combinations.

However, even when cognitive ability was included in any of these scenarios, it generally required an adverse impact weight of .35, and most often.50, for a substantial reduction in adverse impact. More importantly, this reduction was usually the result of greatly diminishing the weight of whichever predictor generated the largest subgroup difference (e.g., cognitive ability). Thus, validity was sometimes reduced to that provided by the predictor or predictors with the smallest subgroup differences. On the other hand, the use of constrained estimation often resulted in a more optimal balance between maximizing validity and minimizing adverse impact. For instance, when cognitive ability, the structured interview, and conscientiousness were used with an adverse impact weight of .35 in the 20% minority sample, the adverse impact ratio was improved from .31 (with OLS regression) to .58 with only a .02 drop in the amount of variance explained by the predictor composite (.17 to .15). Similarly, when all of the predictors were used in the same sample with the same weight, adverse impact was reduced from a ratio of .29 to .58 with a concurrent drop in R-squared of .19 to .16. Moreover, when the adverse impact weight was set to .50, there was a .05 reduction in validity (.19 to .14) but a very strong improvement in the amount of adverse impact exhibited (from .29 to .71).

Overall, because validity was often substantially reduced when the adverse impact weight was set to .50, and because adverse impact was rarely eliminated, very little support was found for prediction three.

Supplementary Analyses

In the end, Table 8 shows that there is a distinct trade-off between validity and adverse impact when using this optimization technique. As the adverse impact weight increased, adverse impact decreased with associated reductions in validity. In fact, adverse impact can often be eliminated completely (given sufficient predictor variance), but the costs to validity can be significant. Table 13 and Figure 1 demonstrate much of this "give and take" relationship by plotting R-squared and associated adverse impact ratios at 20 different adverse impact weights (.05 to 1.0 at .05 intervals) for two separate predictor combinations. The data for these analyses came from the 10,000 subjects found in the entire 20% minority sample. Table 13 shows that adverse impact was indeed eliminated at very high adverse impact weights (somewhere between .70 and .80 for a composite containing all of the predictors and between .55 and .70 for a composite of the

Table 13: Monte Carlo Plot Estimates for R-squared and Adverse Impact Ratios by Adverse Impact Weight and Minority Population using Two Predictor Combinations

	CA+SI+Con+Bio						SI+Con+Bio					
	MinProp 5%		MinProp 20%		MinProp 40%		MinProp 5%		MinProp 20%		MinProp 40%	
AIW	**R^2**	**AIR**	**R^2**	**AIR**	**R^2**	**AIR**	**R^2**	**AIR**	**R^2**	**AIR**	**R^2**	**AIR**
0.05	.17	.27	.19	.34	.20	.34	.14	.63	.15	.67	.15	.63
0.10	.17	.29	.19	.35	.20	.35	.14	.63	.15	.68	.15	.63
0.15	.17	.30	.19	.37	.20	.36	.14	.64	.15	.69	.15	.63
0.20	.17	.32	.19	.41	.19	.38	.14	.64	.15	.69	.15	.64
0.25	.17	.36	.18	.45	.19	.40	.14	.64	.15	.70	.15	.64
0.30	.16	.47	.17	.51	.19	.45	.14	.66	.14	.70	.15	.64
0.35	.15	.55	.16	.62	.17	.51	.13	.66	.14	.70	.15	.65
0.40	.14	.64	.15	.70	.15	.64	.13	.69	.13	.73	.15	.67
0.45	.14	.66	.14	.70	.15	.64	.12	.69	.12	.78	.14	.70
0.50	.13	.66	.14	.70	.15	.66	.11	.73	.11	.79	.12	.74
0.55	.13	.69	.14	.72	.14	.68	.11	.75	.11	.80	.12	.75
0.60	.12	.70	.12	.77	.13	.71	.10	.77	.10	.82	.11	.76
0.65	.11	.73	.11	.79	.12	.74	.10	.78	.03	.89	.03	.90
0.70	.11	.75	.11	.80	.11	.75	.03	.93	.03	.89	.03	.90
0.75	.10	.75	.10	.84	.03	.90	.03	.93	.03	.89	.03	.90
0.80	.03	.93	.03	.89	.03	.90	.03	.93	.03	.89	.03	.90
0.85	.03	.93	.03	.89	.03	.90	.03	.93	.03	.89	.03	.90
0.90	.03	.93	.03	.89	.03	.90	.03	.93	.03	.89	.03	.90
0.95	.03	.93	.03	.89	.03	.90	.03	.93	.03	.89	.03	.90
1.00	.03	.93	.03	.89	.03	.90	.03	.93	.03	.89	.03	.90

Note: R^2 refers to the amount of variance explained. AIR refers to the adverse impact ratio. AIW refers to the adverse impact weight. Cognitive ability, structured interviews, conscientiousness, and biographical data are referred to as CA, SI, Con, and Bio, respectively.

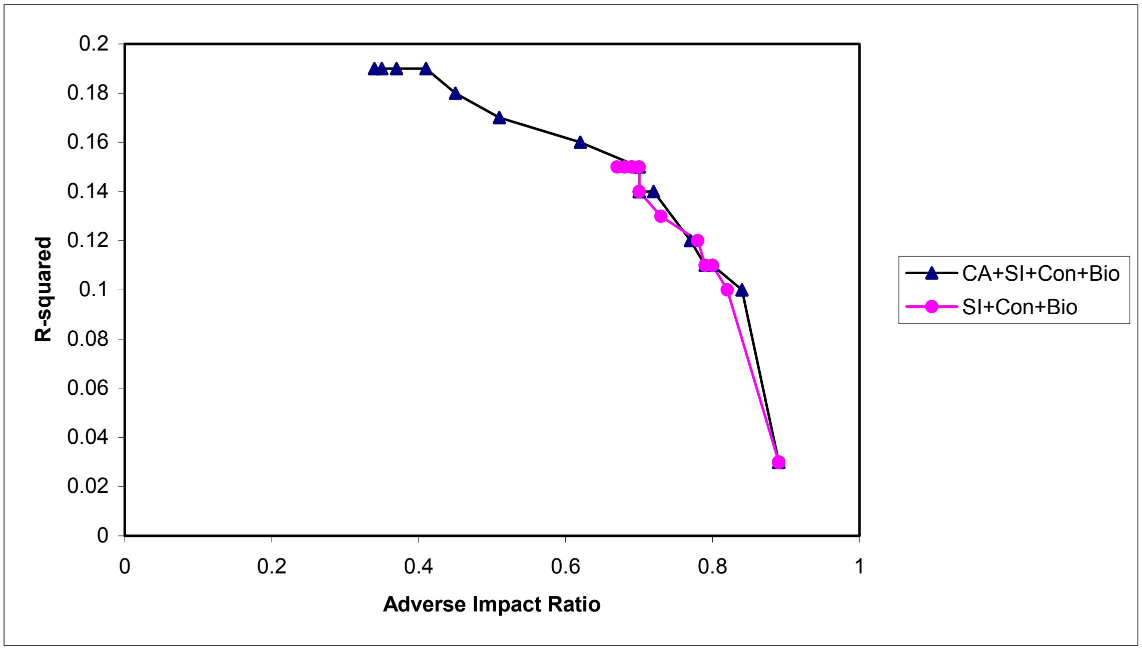

Figure 1. R-squared by Adverse Impact Ratio at Various Adverse Impact Weights (from left to right, .05 to 1.0) in a 20% Minority Population

structured interview, conscientiousness, and biographical data) with constrained estimation continuing to reduce subgroup differences beyond the point required (i.e., an adverse impact ratio of .80) as the weight was further increased. At the same time, the graph illustrates that to improve upon the adverse impact found with either OLS regression or unit weighting, validity must be sacrificed. For instance, in the 5% minority sample using all of the predictors, 82% of the variance explained was lost at the point where adverse impact became technically acceptable under the 4/5ths rule. However, although adverse impact was rarely eliminated without substantial losses in validity, there were also several instances where constrained estimation resulted in meaningful reductions while maintaining much of the original validity. Using all of the predictors in the 20% minority sample with an adverse impact weight of .40 resulted in an R-squared of .15 and an adverse impact ratio of .70 (just short of the .80 goal). This represented an improvement in the adverse impact ratio of .36 (from .34 to .70) with a .04 (.19 to .15) loss of validity. Interestingly enough, one of the common and often suggested methods of eliminating or reducing adverse impact is to remove cognitive ability from the selection process. Doing so in this scenario while using OLS regression resulted in the exact same validity (.15) as with all of the predictors in constrained estimation using an adverse impact weight of .40. However, it also culminated in an increase of adverse impact from an adverse impact ratio of .70 to .66. Admittedly, a .04 difference in the adverse impact ratio is not substantial, but the difference helps to highlight the value of constrained estimation in providing additional options.

Another way of viewing the options that constrained estimation provides is by looking at how it performs under a number of different selection ratios. For this Monte Carlo work, a selection ratio of .25 was used across the board. However, Table 14 presents both R-squared and adverse impact estimates for the three minority samples (entire datasets; 10,000 subjects each) varied by weighting method and selection ratio. The predictor sets included a composite of all the predictors used in this research as well as a composite of alternate predictors (terminology from Bobko, Roth, & Potosky, 1999) consisting of the structured interview, biographical information, and conscientiousness (excluding cognitive ability). This table confirmed much of what has been previously presented in this research as well as in previous studies of the subject. First, constrained estimation, like any selection procedure, showed more value as the selection ratio became smaller. Second, the composite including cognitive ability outperformed the composite of

Table 14: Adverse Impact Ratios by Predictor Combination, Weighting Method, Selection Ratio, and Minority Sample

5% Minority Sample

SR	Composite of Four Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50	Composite of Alternative Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50
.90	.80	.84	.83	.86	.92	.94	.95	.95	.95	.94	.94	.94
.70	.60	.65	.65	.70	.77	.81	.81	.86	.80	.81	.82	.87
.50	.43	.53	.50	.54	.69	.76	.73	.72	.73	.74	.76	.81
.30	.30	.34	.35	.43	.59	.70	.65	.68	.65	.67	.70	.76
.10	.06	.11	.15	.23	.47	.69	.53	.55	.53	.61	.69	.87
	(R^2 =.17)	*(R^2 =.16)*	*(R^2 =.17)*	*(R^2 =.17)*	*(R^2 =.15)*	*(R^2 =.13)*	*(R^2 =.14)*	*(R^2 =.12)*	*(R^2 =.14)*	*(R^2 =.14)*	*(R^2 =.13)*	*(R^2 =.11)*

20% Minority Sample

SR	Composite of Four Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50	Composite of Alternative Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50
.90	.83	.86	.85	.88	.92	.95	.94	.94	.94	.94	.95	.97
.70	.65	.70	.69	.74	.83	.88	.85	.88	.86	.86	.88	.92
.50	.49	.56	.54	.59	.72	.81	.77	.83	.78	.77	.81	.87
.30	.36	.42	.42	.49	.61	.73	.69	.72	.70	.71	.73	.80
.10	.21	.28	.26	.32	.49	.61	.59	.61	.59	.59	.63	.73
	(R^2 =.19)	*(R^2 =.18)*	*(R^2 =.19)*	*(R^2 =.18)*	*(R^2 =.16)*	*(R^2 =.14)*	*(R^2 =.15)*	*(R^2 =.13)*	*(R^2 =.15)*	*(R^2 =.15)*	*(R^2 =.14)*	*(R^2 =.11)*

40% Minority Sample

SR	Composite of Four Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50	Composite of Alternative Predictors: OLS	Unit	AIW=0.15	AIW=0.25	AIW=0.35	AIW=0.50
.90	.86	.89	.88	.89	.92	.95	.94	.95	.94	.94	.95	.96
.70	.66	.70	.69	.71	.78	.85	.83	.85	.84	.84	.85	.88
.50	.49	.54	.53	.57	.65	.74	.72	.76	.72	.72	.73	.80
.30	.35	.39	.38	.45	.56	.67	.66	.68	.66	.66	.67	.75
.10	.19	.28	.23	.29	.41	.60	.52	.60	.54	.55	.58	.68
	(R^2 =.20)	*(R^2 =.18)*	*(R^2 =.20)*	*(R^2 =.19)*	*(R^2 =.17)*	*(R^2 =.15)*	*(R^2 =.15)*	*(R^2 =.13)*	*(R^2 =.15)*	*(R^2 =.15)*	*(R^2 =.15)*	*(R^2 =.12)*

Note: SR refers to the selection ratio. AIW refers to the adverse impact weight. R^2 refers to the amount of variance explained. Shaded areas represent those scenarios with no adverse impact.

alternate predictors in validity, but the opposite occurred when reviewing adverse impact estimates with a similar trend occurring when comparing OLS regression to unit weighting. Third, for the most part, the 4/5ths rule was violated at all but the highest of selection ratios regardless of the weighting methodology used. In fact, when the selection ratios were set to .30 and .50, there were only five instances where acceptable adverse impact ratios were found, and that was when the adverse impact weight was set to .50. Fourth, there were consistent reductions in both validity and adverse impact as the adverse impact weight was increased with larger differences occurring with the composite including all of the predictors.

However, what also mimicked earlier presentations of this research was the fact that constrained estimation typically found a balance between the goals of minimizing adverse impact and maximizing validity. At the lower selection ratios, constrained estimation often resulted in substantial reductions in adverse impact. There were losses in validity, but there was never a loss of more than .05 to the R-squared value when compared to OLS regression. While this would be considered a substantial drop in the amount of variance explained by the predictor composite, it was offset, at least to some degree, by increases in diversity. Additionally, given that the adverse impact weight can be set to any value between zero and one, the losses in validity were mitigated further when less importance was placed on minimizing subgroup differences.

Moreover, although much of the focus has been on comparing the results of constrained estimation with what is generally considered optimal prediction (OLS regression), the specific comparisons to unit weighting were very intriguing. As noted earlier, unit weighting regularly produced less adverse impact (as well as less validity) than OLS regression. However, starting with an adverse impact weight of .15, constrained estimation frequently generated similar adverse impact ratios as unit weighting while sustaining the validity accorded by OLS regression. Typically, it required an adverse impact weight of .50 for unit weighting to consistently outperform constrained estimation in validity, but this was also the point at which constrained estimation usually demonstrated a substantial advantage in reducing adverse impact.

Check for Global Minima

Optimization techniques often fall prey to producing local minima as opposed to global minima. This basically means that while different techniques may minimize the function of interest to some degree, researchers must guard against a set of results that produces a generalized minimum versus the absolute minimum that can be achieved. These same researchers are also frequently pleased with results that approach that absolute minimum, but the absolute minimum remains the goal. Therefore, optimization methodology should be examined in this vein. One way of performing this test is to start the iterative sequence indicative of optimization techniques from different starting points. This research complied by beginning the constrained estimation program from weights provided by OLS regression and then again from weights provided by unit weighting. A predictor composite including all of the predictors was used. Table 15 reveals that while the estimates provided by each starting point were very similar, there were several small differences that suggested that this procedure only approached rather than attained global minima. However, this was not perceived as a major impediment to the interpretation of these results.

Table 15: Comparison of Estimates when Constrained Estimation Starts with Both OLS and Unit Predictor Weights in a 20% Minority Population (N = 10,000)

	Estimates		**Predictor Weights**			
	R^2	**AIR**	**CA**	**SI**	**Con**	**Bio**
Weight						
AIW=0.0	.19	.33	.2157	.2155	.0735	.1647
	(.19)	(.33)	(.2157)	(.2155)	(.0735)	(.1647)
AIW=.15	.19	.37	.1739	.2402	.0790	.1720
	(.19)	(.39)	(.1656)	(.2443)	(.0798)	(.1730)
AIW=.25	.18	.45	.1254	.2608	.0830	.1758
	(.18)	(.48)	(.1047)	(.2673)	(.0840)	(.1761)
AIW=.35	.16	.62	.0377	.2788	.0846	.1713
	(.15)	(.71)	(.0000)	(.2786)	(.0846)	(.1619)
AIW=.50	.14	.70	.0000	.2836	.1145	.1203
	(.14)	(.71)	(.0000)	(.2841)	(.1302)	(.0960)

Note: The first numbers in each column refer to those estimates that were generated by starting the program with ordinary least squares regression weights. The numbers in parentheses are associated with the program beginning with equal or unit weights. R^2 refers to the amount of variance explained. AIR refers to the adverse impact ratio. AIW refers to the adverse impact weight. Cognitive ability, structured interviews, conscientiousness, and biographical data are referred to as CA, SI, Con, and Bio, respectively.

Study Two

Descriptive Statistics

The initial plan for Study Two was to replicate the analyses performed in the Monte Carlo work from Study One and to expand upon the results of an earlier Pilot Study. This expansion was warranted because the original study contained such a small proportion of African Americans (6.8%) that any generalizations from cross-validation work would have been questionable. To this end, additional criterion data was requested and received from the same large Southeastern utility company used in the Pilot Study. As a result, the sample size increased from 188 to 535. However, because a number of individuals participated in some phases of the assessment process and not others, this sample size was reduced to 340. While additional missing data further attenuated the sample sizes associated with different predictor combinations, it was believed that a sample of around 300 individuals would be easily sufficient for these analyses. However, after reviewing the data, it was found that the proportion of African Americans was only slightly increased (from 6.8% to 7.7%). Moreover, as Table 16 demonstrates, when the total sample was randomly split into a validation sample and a cross-validation sample, the resulting differences in correlations and subgroup differences (*d*) were surprisingly large. In effect, these two attributes negated much of the benefit that the additional data provided. With this in mind, a judgment was made to perform all of the analyses on the total sample and to omit those analyses associated with cross-validation.

In the end, a total sample size of 340 was used for this study. The descriptive statistics for this sample can be found in Table 17 (the descriptive statistics for the

Table 16: Matrix of Deviations Between the Validation and Cross-Validation Samples

		Intercorrelations					
	d	1	2	3	4	5	6
1. Critical Thinking Appraisal	-.10						
2. MPR – Background	.22	.03					
3. MPR – Judgment	.07	.07	0				
4. Reasoning by Inference	.29	.24	-.05	.16			
5. Mgr. Video Simulation	-.24	-.09	.12	-.14	-.15		
6. Strategic In-Basket	-.07	-.08	-.18	.03	.04	-.07	
7. Salary 1993	0	-.05	.01	.06	0	.04	-.15

Note: Negative values indicate that the validation sample provided smaller estimates than the cross-validation sample.

Table 17: Study Two Descriptive Statistics

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Critical Thinking Appraisal	322	36	78	65.63	7.293	-.917	.661
Manager Profile Record							
Background (Bio)	334	13	32	24.72	3.612	-.518	.303
Judgment	334	14	32	23.09	2.912	-.043	.264
Reasoning by Inference	308	7	27	17.80	3.859	-.272	-.270
Managerial Video Simulation	313	10	99	66.79	19.514	-.293	-.950
Strategic In-Basket	298	2	4	3.04	.607	-.019	-.273
Salary 1993	340	33754	100000	69284.79	12339.63	-.390	.220

omitted validation and cross-validation samples can be found in Appendix I). As noted previously, the sample sizes for each predictor varied from 298 to 340 as a result of some individuals not participating in all of the exercises. All of the variables were treated as continuous and normally distributed, and, save for potentially two variables, an examination of skewness and kurtosis supported this treatment. Scores on the Managerial Video Simulation appeared to be a little flat, and the scores on the Critical Thinking Appraisal showed a slight skew. However, these results were not considered to be overly abnormal. In fact, the skewness associated with the Critical Thinking Appraisal should be expected. There appeared to be a ceiling effect with many of the scores bunched toward the high end. This is not unusual given that this was a managerial and executive sample where about 40% of the individuals reported having earned a Bachelor's degree and another 48% reported having earned a graduate degree.

The subgroup differences and intercorrelations associated with all of these variables can be found in Table 18 (the subgroup differences and intercorrelations associated with the omitted validation and cross-validation samples can be found in Appendix J). This matrix shows that while almost all of the variables appeared to result in fairly substantial subgroup differences, the background portion of the Manager Profile Record (MPR) and the salary criterion demonstrated the smallest *d* values. Additionally, all of the variables were significantly correlated with the exception of two instances involving the Managerial Video Simulation with the background portion of the MPR and the Strategic In-Basket with the judgment portion of the MPR.

Table 18: Study Two Matrix of Correlations and *d* Values for the Total Sample

		Intercorrelations					
	d	**1**	**2**	**3**	**4**	**5**	**6**
1. Critical Thinking Appraisal	1.0						
2. MPR – Background	.27	.25** (318)					
3. MPR – Judgment	.65	.29** (318)	.27** (334)				
4. Reasoning by Inference	.79	.39** (291)	.17** (305)	.31** (305)			
5. Mgr. Video Simulation	.83	.21** (296)	.01 (310)	.11** (310)	.13* (308)		
6. Strategic In-Basket	.57	.27** (281)	.17** (296)	.08 (296)	.12* (293)	.15** (298)	
7. Salary 1993	.24	.32** (322)	.46** (334)	.26** (334)	.18** (308)	.18** (313)	.21** (298)

Note: The numbers within parentheses represent the sample sizes associated with each correlation.

* $p < 0.05$. ** $p < 0.01$.

Primary Analyses

As with the Monte Carlo work in Study One, the primary analyses involved generating R-squares, adverse impact ratios, and predictor weights for a number of different predictor combinations. The predictor sets included combinations of each predictor along with the Critical Thinking Appraisal (CTA) in two-predictor combinations, the background portion of the Manager Profile Record (Bio) and the Reasoning by Inference test (RBI) both with and without the CTA, and the Managerial Video Simulation (Vscore), the Strategic In-Basket (SIB), and the situational judgment portion of the Manager Profile Record (SJT) both with and without the CTA. The analyses associated with the planned cross-validation work were omitted due to sampling issues. For each specific estimate (i.e., R-squared and adverse impact ratios), nine comparisons were made between OLS regression and unit weights, whereas 36 comparisons were available between either OLS regression or unit weights and those from constrained estimation. Table 19 presents the results of this work (complete results including predictor weights both with and without cross-validation efforts can be found in Appendices K and L, respectively).

Predictions 1a, 1b, and 1c. Prediction 1a stated that constrained estimation would reduce to OLS regression, and thus provide the same results, when no importance was placed on minimizing subgroup differences. This prediction was completely supported across all of the predictor combinations.

Prediction 1b contrasted the results from OLS regression with those from constrained estimation and predicted that the two methodologies would generate similar estimates of validity. Across all of the predictor sets and adverse impact weights, constrained estimation and OLS regression provided the exact same explanations of criterion variance 24 times (out of 36). There were only eight instances when the two methods differed by as much as .01 with four additional comparisons where OLS regression resulted in an improvement of .04 or more (two at .04 and two at .05). Of these larger discrepancies, and similar to the Monte Carlo work, one was found when the adverse impact weight was set to .35 while the other three were found at a weight of .50. Of note, the largest differences occurred in the two predictor sets where the biographical portion of the MPR was combined with the CTA. Overall, this prediction was mostly supported.

Table 19: Study Two R-squares and Adverse Impact Ratios for Various Predictor Combinations in the Total Sample

Predictors	R^2	AIR
Bio+CTA		
OLS	.26	.35
Unit	.25	.35
Constrained		
AIW=0.15	.26	.35
AIW=0.25	.25	.53
AIW=0.35	.22	.53
AIW=0.50	.21	.53
SJT+CTA		
OLS	.15	.35
Unit	.15	.35
Constrained		
AIW=0.15	.15	.35
AIW=0.25	.15	.35
AIW=0.35	.15	.35
AIW=0.50	.15	.35
RBI+CTA		
OLS	.10	.19
Unit	.09	.38
Constrained		
AIW=0.15	.10	.19
AIW=0.25	.10	.19
AIW=0.35	.10	.19
AIW=0.50	.09	.19
Vscore+CTA		
OLS	.11	.18
Unit	.10	.18
Constrained		
AIW=0.15	.11	.18
AIW=0.25	.11	.18
AIW=0.35	.11	.18
AIW=0.50	.11	.18
SIB+CTA		
OLS	.14	.20
Unit	.13	.20
Constrained		
AIW=0.15	.14	.40
AIW=0.25	.14	.40
AIW=0.35	.14	.40
AIW=0.50	.13	.40

Predictors	R^2	AIR
Bio+RBI		
OLS	.19	.56
Unit	.15	.56
Constrained		
AIW=0.15	.19	.56
AIW=0.25	.18	.56
AIW=0.35	.18	.56
AIW=0.50	.18	.56
Bio+RBI+CTA		
OLS	.23	.40
Unit	.19	.40
Constrained		
AIW=0.15	.23	.40
AIW=0.25	.22	.40
AIW=0.35	.19	.40
AIW=0.50	.18	.40

Predictors	R^2	AIR
SJT+Vscore+SIB		
OLS	.12	.00
Unit	.12	.00
Constrained		
AIW=0.15	.12	.00
AIW=0.25	.12	.00
AIW=0.35	.12	.00
AIW=0.50	.11	.19
SJT+Vscore+SIB+CTA		
OLS	.19	.21
Unit	.18	.21
Constrained		
AIW=0.15	.19	.21
AIW=0.25	.19	.21
AIW=0.35	.19	.21
AIW=0.50	.19	.21

Note: R^2 refers to the amount of variance explained. AIR refers to the adverse impact ratio. AIW refers to the adverse impact weight. CTA, Bio, SJT, RBI, Vscore, and SIB refer to the Critical Thinking Appraisal, the biographical portion of the Manager Profile Record (MPR), the situational judgment portion of the MPR, the Reasoning by Inference Test, and the Strategic In-Basket, respectively.

Prediction 1c compared the results from constrained estimation with those from unit weighting and predicted that constrained estimation would culminate in greater explanations of variance. Of the 36 possible comparisons, unit weighting produced more validity in four cases (two at .02 and one at both .03 and .04) with 11 displaying no differences. In contrast, constrained estimation outperformed unit weighting a total of 21 times (15 at .01, four at .03, and two at .04). Moreover, the two larger differences favoring unit weighting came at adverse impact weights of .35 and .50 when the biographical portion of the MPR was combined with the CTA. Other than these two instances, constrained estimation generated more validity with much more frequency than unit weighting. While this suggested support for this prediction, it is important to recognize that 28 of the 36 comparisons revealed either no difference or at most a difference of .01. Thus, although this prediction received some support, it was not very strong.

Predictions 2a, 2b, and 2c. These predictions were omitted from analysis because the minority proportion was too small to warrant cross-validation work.

Prediction 3. This prediction stated that constrained estimation would eliminate adverse impact while maintaining similar levels of validity as compared to OLS regression and unit weighting when sufficient importance was placed on minimizing subgroup differences and when sufficient predictor variance was available. In general, while the constrained estimation program sustained comparable degrees of validity, adverse impact was never eliminated using a selection ratio of .25 along with the primary adverse impact weights presented in this study. In fact, six of the nine predictor combinations resulted in no changes to adverse impact even at an adverse impact weight as high as .50.

There was some success when the biographical portion of the MPR (Bio) and the Strategic In-Basket were combined with the CTA in the two-predictor combinations at adverse impact weights of .25 and .35, respectively. In each instance, the adverse impact ratio was reduced by about .20 while losing at most .01 from R-squared. Additionally, the three-predictor combination of the situational judgment portion of the MPR (SJT), the Managerial Video Simulation (Vscore), and the Strategic In-Basket (SIB) along with a weight of .50 culminated in a .19 decrease in the adverse impact ratio with a similar reduction in R-squared of .01. Yet, the only demonstrable trend surrounding this success was found when combining those predictors with the largest subgroup differences (the CTA, SJT, and Vscore) with predictors that resulted in smaller individual subgroup differences (Bio and the SIB). However, this finding was diminished by the fact that there was no change in adverse impact in any of the combinations including the Reasoning by Inference Test (with a d of .79). Moreover, whereas the Monte Carlo study showed that increasing the adverse impact weight had a continued effect on reducing adverse impact, the same results were not evident with this data. Increasing this weight generated a more static change in three of the nine scenarios with reductions occurring at adverse impact weights of .15, .25, and .50 but also with no continued change to adverse impact as the weight was increased. At the same time, increasing the adverse impact weight in each of these scenarios did result in smaller validities. Therefore, based on this information and the estimates provided in Table 19, this prediction received no support.

Supplementary Analyses

Given the estimates obtained from the Monte Carlo study, these findings were surprising. Because of this, a more detailed examination of the descriptive statistics and the actual data was performed. After a careful review, there were a number of factors that complicated the interpretation of these results and that appeared to be likely culprits for the difficulties faced in this study.

The simulation work from Study One presented a number of situations where adverse impact was difficult to eliminate; but if sufficient weight was placed on minimizing subgroup differences and if one was willing to allow large reductions in validity, it could often be accomplished. However, this success was based on the ability to actually reduce subgroup differences. Thus, there was some question as to the size of the subgroup difference reductions found in this study. In many of the scenarios presented with this dataset, the decreases in *d* were often smaller than expected even as the adverse impact weight approached 1.0. Interestingly, a common result of increasing this weight was for the constrained estimation procedure to assign a zero or near-zero weight to the predictors with the largest subgroup differences, and thus allow only the variance associated with the predictor with the smallest *d* value to have any influence. This suggested that there might not have been sufficient predictor variance for the constrained estimation program to operate effectively, and there is some evidence of this fact. A quick look at Table 18 shows that almost all of the predictors were associated with somewhat sizeable subgroup differences. Combine this finding with the fact that most of the intercorrelations were not extraordinarily large, and a likely explanation is that there was not enough variance associated with either subgroup differences or validity for the program to find an appropriate balance between the competing goals.

However, the reduction of subgroup differences was not always the problem. Figure 2 provides a graphical demonstration of this with plots of *d* values by adverse impact weights (ranging from .05 to 1.0 in .05 increments) for two predictor combinations (Bio+RBI+CTA and Bio+CTA). Note that while there were some sizeable reductions in *d* as the adverse impact weight was increased, Table 19 revealed very little movement regarding the adverse impact ratios. In fact, increasing this weight with the predictor combination of the biographical portion of the MPR, the Reasoning by Inference Test, and the CTA showed no change whatsoever in the adverse impact ratio despite the subgroup difference reductions found in Figure 2. Moreover, when the

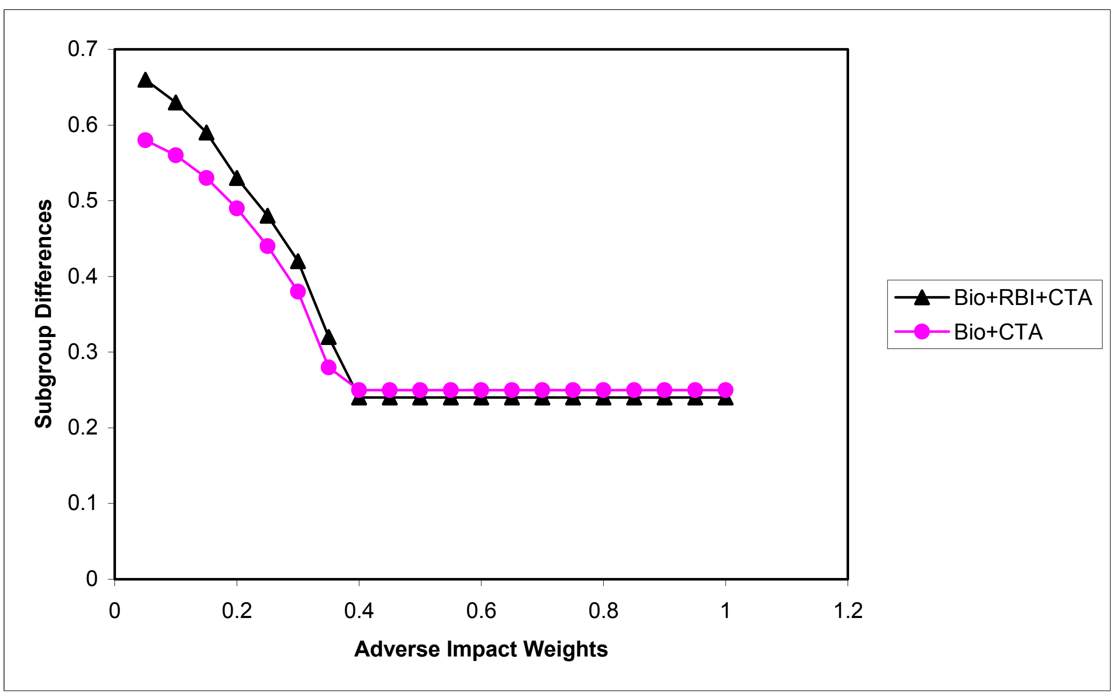

Figure 2: Subgroup Differences by Various Adverse Impact Weights (from left to right, .05 to 1.0) for Two Predictor Combinations

biographical portion of the MPR and CTA predictor set was examined, the adverse impact ratio changed once from .35 to .53 at an adverse impact weight of .20 and remained at that level all the way through a weight of 1.0.

Of additional concern was how the selection ratio sometimes influenced adverse impact ratios. Typically, it would be assumed that greater adverse impact would be associated with smaller selection rates if the predictor composite resulted in moderate to even small subgroup differences. However, Table 20 demonstrates that this was not necessarily the case in this dataset. Using the predictor combination of the biographical portion of the MPR, the Reasoning by Inference Test, and the CTA, adverse impact increased expectedly as the selection ratio was decreased (and thus became more inclusive) from .90 to .70. In contrast, adverse impact actually decreased when the selection ratio was dropped from .70 to .50 using unit weighting and then again with constrained estimation at adverse impact weights of .25 and .35 when the selection ratio

Table 20:Adverse Impact Ratios and R-squared Estimates by Weighting Method and Selection Ratio for One Predictor Combination

	Bio+RBI+CTA					
SR	**OLS**	**Unit**	**AIW=.15**	**AIW=.25**	**AIW=.35**	**AIW=.50**
.90	.93	.75	.93	.93	.93	.93
.70	.51	.29	.51	.74	.82	.82
.50	.40	.40	.40	.40	.40	.51
.30	.34	.34	.34	.51	.51	.51
.10	.50	.50	.50	.50	.50	.50
	(R^2 = .23)	*(R^2 = .19)*	*(R^2 = .23)*	*(R^2 = .22)*	*(R^2 = .19)*	*(R^2 = .18)*

Note: SR refers to the selection ratio. R^2 refers to the amount of variance explained. AIW refers to the adverse impact weight. Bio, RBI, and CTA refer to the biographical portion of the Manager Profile Record, the Reasoning by Inference Test, and the Critical Thinking Appraisal, respectively.

was lowered from .50 to .30. Moreover, OLS regression, unit weighting, and constrained estimation (at an adverse impact weight of .15) resulted in less adverse impact when the selection ratio was reduced from .30 to .10. This suggested a possible challenging issue in the form of exactly how individuals from different subgroups performed and were ranked on each of the predictors. Further complicating this issue was the small proportion of minorities in the sample.

All of the predictors used in this study would have resulted in adverse impact if utilized alone. This stems, in part, from the various subgroup differences; however, the degree of adverse impact was not as related to the size of individual predictor subgroup differences as one might expect. For example, the Reasoning by Inference Test and the Strategic In-Basket showed *d*'s of .79 and .57, respectively; but both of these predictors also displayed less adverse impact (ratios of .71 and .74, respectively) than the biographical portion of the MPR, which generated an adverse impact ratio of .65 with a *d* of .27. This occurred despite the fact that at a selection ratio of .25, each of the predictors would lead to the selection of four minority candidates. Obviously, differing overall sample sizes associated with each predictor were the cause of the disparate proportions, and the size of the minority sample for each predictor and predictor combination limited the benefits of any strategy aimed at reducing adverse impact while preserving validity. It would have taken just one more African American selection (from four to five individuals) for the adverse impact ratio for the biographical portion of the MPR to rise from .65 to .82 (technically acceptable). Interestingly, the fifth highest scoring minority member scored a 26 on this test as opposed to a low score of 27 for the last six individuals that would be selected by this predictor. In actuality, if this had occurred within a predictor composite of multiple predictors, constrained estimation might have been able to alleviate much if not all of the adverse impact. However, constrained estimation was often prevented from working optimally because of the small overall minority proportion as well as surprisingly little overlap in top scores between African Americans across the predictors. Only two to three minority individuals appeared to perform well on most or all of the predictors. This is likely the reason for such consistency in the number of minorities selected as well as the consistency of adverse impact ratios obtained across a number of adverse impact weights. When multiple predictors are used to form a composite, there must be some set of weights that effectively combines those predictors to provide something of a "selection" profile. If there are very few minority individuals that match the profile provided by all of the predictors or a routine like constrained estimation exhibits difficulty in even producing a profile that mirrors the validity of OLS regression or unit weighting because of a lack of matching characteristics, then the battle over adverse impact is probably lost before any soldiers take the field.

Therefore, because of various issues associated with this study, it is questionable as to whether or not this was an accurate reflection of constrained estimation's ability to reduce adverse impact while preserving validity. However, it should be understood that few datasets are without problems and that the use of constrained estimation in its present form might be limited to more robust samples.

Table 21 provides a summary of the research findings broken down by study and specific prediction.

Table 21. Summary of Research Findings

Predictions		Findings: Study One	Findings: Study Two
1a	Constrained estimation will reduce to OLS regression when no importance is placed on adverse impact in the criterion	Full support	Full support
1b	Constrained estimation and OLS regression will provide similar explanations of criterion variance.	Supported	Supported
1c	Constrained estimation will provide a greater explanation of criterion variance than that of unit weighting.	Not supported; Comparisons were more similar than dissimilar.	Partially supported; Minor advantage favoring constrained estimation.
2a	Across disparate situations, constrained estimation will more frequently result in less shrinkage than OLS regression.	Not supported; Shrinkage was similar across methodologies.	Prediction eliminated
2b	Unit weighting will show less shrinkage than OLS regression.	Not supported	Prediction eliminated
2c	Unit weighting will display less shrinkage than constrained estimation.	Not supported	Prediction eliminated
3	When sufficient importance is placed on reducing subgroup differences as well as maximizing validity, and when multiple predictors possessing unique variance associated with both adverse impact and other criteria of interest are available, the constrained estimation routine will partition variance such that adverse impact is eliminated and validity is sustained at acceptable levels when compared to that of OLS regression and unit weighting.	Mixed support; Constrained estimation worked in reducing (but not eliminating) adverse impact while maintaining validity.	Not supported

CHAPTER V

DISCUSSION

An underlying purpose of this research was to introduce a somewhat novel paradigm concerning both the potential benefits and proposed future tactics associated with employee selection. Within this paradigm, a number of perspectives were advanced. First, the value of employee selection goes beyond the singular prediction of job performance or productivity. For true utility, practitioners in this field should focus on optimizing multiple objectives related to organizational success. Second, the treatment of multiple criteria should be revisited. Traditional methodologies have proven incapable of predicting, describing, and explaining multiple criteria of interest with both accuracy and efficiency, and thus, new directions should be explored when faced with criteria that are not easily, and understandably, combined into a composite. Third, adverse impact presents itself as a difficult criterion because of the conflicting objectives related to its resolution. Validity must often be sacrificed in order to assuage the social and legal ramifications associated with its existence. Finally, because optimization techniques have demonstrated success when dealing with multiple criteria, constrained estimation is proffered as a potential solution to adverse impact and as a means for understanding.

Constrained estimation was designed to optimize two conflicting objectives that have frustrated the I/O community for some 40 years – eliminating adverse impact while sustaining the validity accorded from OLS regression. It was expected that while the elimination of adverse impact might not always be possible, constrained estimation would at least offer the most optimal balance between these competing goals.

Specifically, and in this study, the procedure attempted to accomplish this balance by manipulating predictor weights (provided by OLS regression) such that the resulting predictor composite revealed reduced subgroup differences. The underlying assumption behind this estimation routine was that if subgroup differences were reduced sufficiently, then adverse impact would be eliminated.

Questions Answered

In testing the efficacy of this approach, three questions were proposed. First, "How can [employers] use valid procedures in a manner than optimizes the expected performance of their workforce and at the same time employ a demographically diverse workforce?" (Schmitt et al., 1997, p. 719). Second, does the use of constrained estimation provide a viable alternative to other selection strategies? Finally, how does constrained estimation compare to other selection schemes? The answers to these questions revolve around constrained estimation's success within this research.

For the most part, the Monte Carlo work showed that constrained estimation succeeded in reducing adverse impact while maintaining validity. Unfortunately, these reductions rarely resulted in the elimination of adverse impact unless an unusually large amount of importance was placed on minimizing subgroup differences within the routine. Part of the problem was that as more focus was placed on minimizing subgroup differences, less focus was placed on maintaining validity. Thus, constrained estimation met with much the same fate as that of other solutions to the adverse impact dilemma. There was typically a tradeoff of validity for increased diversity.

The size of this tradeoff as well as the ultimate benefit of using constrained estimation was determined by three major factors. First, there was generally very little change in either validity or adverse impact when only two predictors were used to create a predictor composite. It typically required at least three predictors for the benefits of constrained estimation to arise with the use of four predictors revealing the greatest changes. This was to be expected because limiting the number of predictors also limits the number of predictor weights for the estimation routine to manipulate. Additional predictors simply provide more avenues for change.

Second, the value assigned to the adverse impact weight played a key role. In general, as the adverse impact weight was increased, both validity and adverse impact decreased. Up to an adverse impact weight of about .35, constrained estimation typically revealed slightly more validity than unit weighting and almost the same validity as OLS regression with most discrepancies falling at about .01. At the same time, it is important to remember that this occurred with a rather large dataset. With a much smaller dataset one would probably expect unit weighting to perform on par with that of both constrained estimation and OLS regression (Schmidt, 1971). As the adverse impact weight was increased to .50 or more, these discrepancies became more notable with OLS regression and unit weighting often displaying greater prediction; but at this point the losses to R-squared were generally no more than about .05 with many falling somewhere below this level. Moreover, as the adverse impact weight was increased to .15 or more, there were also meaningful reductions in adverse impact with most of these occurring at adverse impact weights of .25 or more. For example, using an adverse impact weight of .35 would, on average, result in about a .11 improvement to the adverse impact ratio with little loss of validity. Additionally, although constrained estimation with an adverse impact weight of .50 would usually reveal the moderately large decrease in R-squared noted previously, the average benefit to the adverse impact ratio was about .35.

The third determinant of constrained estimation's value to employee selection was predictor combination content. It was initially predicted that given sufficient predictor variance associated with both validity and adverse impact (or really subgroup differences), constrained estimation would be able to eliminate adverse impact while maintaining validity. While this is probably true, neither of the studies demonstrated strong evidence of this statement. The lack of evidence could be a testament to one of the general perceptions that frequently surrounds this issue – the subgroup differences associated with cognitive ability, as well as other predictors and predictor composites, are simply too much to overcome. It could also relate to a need to look at other predictors that display different kinds of variance and distributions. However, both of these reasons would fall under the likely cause – there was not enough predictor variance available. Given the variety of predictors included in these two studies, it is just as likely that no set of predictors presently available would have fared any better. On the other hand, there were several indicators that predictor variance played an enormous role in this process. The very fact that constrained estimation demonstrated more success when larger predictor sets were used implies that the additional variance associated with those larger sets facilitated the overall objectives. Moreover, the Monte Carlo work revealed substantial differences in the changes to validity and adverse impact based upon whether or not cognitive ability was included in the predictor composite. Finally, the results of Study Two point to a lack of predictor variance as one of the reasons constrained estimation led to such small changes. Almost all of the predictors used in this study displayed rather large subgroup differences. Combine this with the fact that most of the intercorrelations were not very large and variance definitely becomes an issue. If, for example, the intercorrelations had been larger, constrained estimation might have been able to partition certain portions of shared variance (associated with the subgroup differences) and manipulated the predictor weights to also focus on other aspects of prediction. In this case, smaller intercorrelations probably resulted in additive variance associated with adverse impact.

In the end, a lack of predictor variance often limited the ability of constrained estimation to manipulate predictor weights in an effort to optimize the objectives of interest. With this in mind, there is some question as to how much predictor variance constrained estimation might need as opposed to what might be reasonably expected.

However, given the comparisons associated with this research, constrained estimation appears to be a viable selection strategy that compares favorably to two commonly used methodologies for weighting predictors – OLS regression and unit weighting. All three methodologies provided similar estimates of both validity and shrinkage with constrained estimation showing an additional benefit of reducing adverse impact. While it is by far the most complicated of the three methods, it should be considered when reductions to adverse impact are at a premium.

Pilot Study Comparisons

It was interesting to note the lack of convergence between the results from the Pilot Study and those of Study Two. Given that adverse impact was completely eliminated with a loss in R-squared of at most .03 in the Pilot Study, similar results were expected in Study Two. However, there were a number of differences between the two studies that probably led to the discrepancies. First, the samples were quite different. The

Pilot Study used individuals that had progressed through three stages of an assessment process whereas Study Two focused on those individuals that had only progressed beyond the first stage and through the second. Thus, these individuals included the entire pool of applicants used for selection into the third stage. Moreover, there were a number of individuals assigned by the organization to complete only the third stage. These individuals were included in the Pilot Study but not in Study Two. Second, because different individuals had completed different exercises and stages, the set of assessments included in the Pilot Study were very different. One of these assessments (overall ratings of Assessment Center performance) was shown to have very strong validity while also demonstrating almost no subgroup differences. Finally, some note must be made about the differences between the two constrained estimation procedures. It was found that the present research required some modifications to the estimation routine. Specifically, the routine was simplified and predictor weights were constrained to maintain positive values. This constraint was obviously not present in the earlier version of this program, and thus, a negative weight was placed on the Critical Thinking Appraisal (CTA) in the Pilot Study. Because there were two other predictors (overall ratings of Assessment Center performance and the Manager Profile Record (MPR)) that shared similar variance with the CTA, the negative weight served as a suppressor. In effect, validity was bolstered while the subgroup differences associated with the CTA that were not shared by the other predictors were removed.

Overall Implications

It is clear from both the literature review and the results of this research that the adverse impact problem is far from resolved. Constrained estimation was moderately successful in reducing adverse impact while maintaining validity, but adverse impact was frequently observed regardless of the situation presented. However, from a technical standpoint, constrained estimation achieved what it was initially designed to accomplish. When sufficient predictor variance was available, it consistently reduced subgroup differences associated with a new predictor composite. Unfortunately, it was revealed that reduced subgroup differences were not always adequate for the elimination of adverse impact. Even standardized group differences in the range of .15 to .20 can be nearly impossible to overcome at lower selection ratios, but as Study Two showed, applicant rankings can also play a significant role regardless of the subgroup difference exhibited. This relates to a base rate problem that presents itself as an additional labor market issue. The base rate refers to the proportion of those individuals judged to be successful using a particular selection procedure (Cascio, 1991). The difficulty found in Study Two was that there were typically only two or three minority candidates deemed successful within each predictor. In addition, there was really only one or two that performed well across all of the predictors. Therefore, each predictor composite resulted in the selection of at most two or three minority candidates when five or six were needed for a technically acceptable adverse impact ratio (i.e., .80 or above). Moreover, the small pool of minority candidates exacerbated this issue. Changes could be made to the predictor weights that resulted in reduced subgroup differences, but sampling error prevented acceptable adverse impact ratios. This becomes especially troublesome when the realization sets in that many selection scenarios and most adverse impact research face this issue on a regular basis.

Thus, even though the estimation routine succeeded in reducing subgroup differences, additional constraints limited the overall value of the procedure in this situation. This might suggest that constrained estimation, in its current form, is best suited for those projects where the overall sample size and, in particular, the minority sample size is not unusually small. However, constrained estimation can still add value in these situations by providing alternatives and by increasing our understanding. By specifically delineating the potential tradeoff of validity for increased diversity through predictor weight manipulation, constrained estimation can facilitate the management of a selection process by presenting the options available. Additionally, the actual weights assigned by constrained estimation to each predictor within a predictor composite when two or more criteria are used can lead to a better understanding of how predictor variance can be utilized in the description, explanation, and future prediction of various objectives.

One potential caveat should be noted about the value of constrained estimation. The perceptions about the legality of this approach might be mixed because the minimizing function incorporates the objective of reducing group mean differences instead of allowing the predictor weights to be exclusively determined by predictor relationships with some more objective criterion. However, because the variance in predictor weightings is always subject to issues such as multicollinearity and validity concentration, they are rarely determined exclusively by their predictiveness of a particular criterion (Bobko, 1990; Budescu, 1993; Cohen & Cohen, 1983; Darlington, 1968). Moreover, probably the most common method of reducing adverse impact in selection systems, removing cognitive components and tests, is rarely questioned on the basis of legality but would result in a dramatic change to the remaining predictor weights within a composite because the process essentially removes a predictor that generally displays a very strong relationship with the criterion. Additionally, a review of Section 106 of the Civil Rights Act of 1991 regarding the prohibition against group norming suggests that this approach is technically legal. Because no scores are adjusted and no differential cutoffs are used regarding subgroups, constrained estimation with a focus on reducing group mean differences follows the guidelines put forth by this section. It can also be argued that this type of approach closely follows the "letter" of how the Uniform Guidelines on Employee Selection Procedures (EEOC, 1978) explicitly instructs employers to consider adverse impact when choosing between alternative selection devices:

> where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact (p. 38297).

Furthermore, when one considers the plaintiff's burden of proof (see *Griggs v. Duke Power*, 1971) in the third step of determining adverse impact in a court of law (presenting some predictor with equal or substantially similar validity that shows no adverse impact), it can be suggested that this approach not only follows the guidelines outlined by the EEOC, but also that of congressional intent. This view is further bolstered by the recent Supreme Court decision in *Grutter v. Bollinger* (2003) stating that diversity is a compelling government interest and that narrowly tailored selection plans are acceptable means to an end. Although this decision was made within an educational setting, it is not a giant leap of logic to expect a similar framework in employee selection. Thus, while the novelty of constrained estimation might raise questions as to its use, these questions should not necessarily pertain to its legality if similar validity can be attained. Additionally, though the current form of constrained estimation might not be appropriate in a number of different situations, it provides an immediate benefit for those organizations forced to comply with difficult diversity requirements ordered through a consent decree.

Research Limitations

There were a number of limitations within this study that require mentioning. By far and away, the largest limitation to this study involved sampling error. The initial design of this research included a balance of work with both large-scale Monte Carlo data (Study One) as well as archival data obtained from a field setting (Study Two). However, the small minority sample in Study Two prevented a confident analysis of constrained estimation's ability to balance the goals of maintaining validity while eliminating adverse impact. With a larger minority sample, constrained estimation might have resulted in more consistent estimates (as in Study One). In addition, the small minority sample precluded an analysis of cross-validation statistics. It was hoped that by obtaining a larger sampling of criterion data from the organization, the sampling issues related to the Pilot Study could be resolved. However, the increase in the minority sample was only marginal, and this remains an issue to be observed in future research.

These difficulties were further intensified by the minority applicant rankings on each of the individual predictors. Although the data was judged to be normal, it is clear that the minority distributions on each of the individual predictors in Study Two resulted in a base rate problem that almost guaranteed adverse impact regardless of the method or approach used in selection. Simply reducing subgroup differences often had little effect on adverse impact ratios. While this is probably a function of the labor market and indicative of the problems faced by researchers in this area, the effects of minority distributions should certainly be considered and looked at more closely in research dealing with adverse impact.

Another limitation of this research came with the criterion data used in Study Two. It would have been preferable to utilize performance data rather than salary reports, and in fact, this data was acquired from the organization. However, it was found that the obtained performance appraisals were based on a three-point scale and had almost no variance whatsoever. Almost all of the ratings were clustered at the high end of the spectrum, which might be typical of a sample of individuals who have been specifically chosen to participate in an assessment process where the goal is to select those individuals most qualified for promotion. This implies an additional limitation in the type of sample used. Most of the individuals from this assessment project were mid- to upper-level managers who self-reported more education than the typical blue-collar worker. It would be interesting to judge the effectiveness of constrained estimation with lower level jobs.

Finally, there has been a significant change in the literature since Bobko, Potosky, and Roth (1999) first reported their model matrix. Specifically, a recent article by Roth, Huffcutt, and Bobko (2003) reported that the meta-analyzed subgroup difference associated with job performance is closer to .30 than the .45 noted in the previous study. Although the reduction of subgroup differences in constrained estimation primarily depends on those differences related to the predictor composite, a smaller *d* value associated with the job performance criterion might have lasting effects on the routine's ability to optimize the two objectives of interest.

Future Research

In addition to the research directions suggested in the Limitations section, this research opens up a host of opportunities for future research. The first obvious directions would include the utilization of different populations, other forms of criteria, and samples with increased minority representation. Constrained estimation is still in the infant stage, and the routine would benefit from a variety of work testing it from a number of different angles. Along these same lines, the use of different predictor variables would appear to be necessary. While both of the studies used in this research displayed a multitude of disparate predictor variables, additional variables as well as modes of testing could be looked at for a better understanding of how predictor variance influences the results of constrained estimation. In particular, short-term memory tests (Verive & McDaniel, 1996), video-based testing (see Chan & Schmitt, 1997), and well-developed situational judgment tests (Motowidlo, Dunnette, & Carter, 1990; Pulakos & Schmitt, 1995) present fruitful avenues. All of these options have been shown to exhibit a great deal of validity with smaller subgroup differences than that of cognitive ability.

Additionally, because constrained estimation partitions variance to achieve multiple objectives, the actual use of this variance should be studied. It would be informative to observe how the different variances of multiple predictors combine to achieve the desired result. The partial and semi-partial correlations associated with newly generated composites could provide insight into how predictor composites could and should be created. Furthermore, more work should be focused on studying the influence of suppression. While this would certainly be a contentious issue to discuss in a court of law, the value of this statistical artifact could be substantial. The Pilot Study demonstrated how placing a negative weight on cognitive ability could boost validity while eliminating adverse impact (given the presence of other predictor that shared much of the same criterion-related variance). Because cognitive ability usually exhibits a greater subgroup difference than what is found with job performance, this negative weight could become somewhat necessary when attempting to alleviate the adverse impact found with many predictor composites. At the very least, the study of variance partitioning and suppression would improve our understanding of how multiple predictors combine to describe, explain, and predict multiple criteria.

It would also be enlightening to recreate the studies performed by Hattrup, Rock, and Scalia (1997) and DeCorte (1999). Both of these studies combined two criteria (task and contextual performance) into a criterion composite in their research on adverse impact. DeCorte went one step further and used nonlinear constrained programming to constrain the adverse impact ratio to acceptable levels. The use of constrained estimation with these parameters would require some modification to the estimation routine to include two performance criteria (whereas only one was used before) along with the reduction of subgroup differences. However, this is exactly one of the directions future research on constrained estimation should take. In fact, other conceptualization of criteria as well as multiple criteria excluding adverse impact should be considered. Schmidt and Kaplan (1971) used an example of combining speed and accuracy in bemoaning the use of criterion composites. Their major complaint was that a criterion composite would allow for high ratings on one of the individual criterion components to compensate for low ratings on the other. A modified version of constrained estimation would instead allow for the optimization of each criterion component individually. While this is not to suggest that constrained estimation represents a solution to the "composite versus multiple criteria" dilemma that has raged through the I/O literature for some time, it could certainly represent a step in the right direction.

Finally, while not necessarily related directly to constrained estimation, this research serves as an additional call for the creation of new predictors that exhibit validity with small subgroup differences. Much of the literature has come to understand that even those composites composed of "alternate" predictors possess sufficient subgroup differences to frequently result in adverse impact. Thus, more effort should be focused on the types of constructs used as well as the manner in which new predictors are created. One potential avenue would be to mimic the work performed with biographical information. Biographical information has a long history of showing strong validity with very little adverse impact. This is likely because items included in biographical inventories are initially created with both validity and adverse impact in mind. There is no reason that other predictors cannot follow the lead of a predictor that has displayed more than a moderate level of success. Constrained estimation can help in this regard by demonstrating how predictor variance influences important outcomes.

Conclusions

The findings of this research show how optimization techniques, specifically constrained linear programming, can be used to accomplish multiple objectives by optimizing two or more criteria. Overall, constrained estimation was successful at reducing adverse impact while maintaining validity, but adverse impact was rarely eliminated without substantial losses to prediction. However, it is believed that constrained estimation represents a solid first step in searching for some optimal balance between the economic, social, and legal issues that often constrain employee selection decisions.

REFERENCES

REFERENCES

Albemarle v. Moody Paper Company, 422 U. S. 405 (1975).

American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC: APA.

Arvey, R. D., & Faley, R. H. (1992). Fairness in selecting employees (2nd ed.). Reading, MA: Addison-Wexley.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression: An alternative to differential analysis. Personnel Psychology, 31, 233-241.

Beale, E. M. L. (1988). Introduction to optimization. New York: John Wiley & Sons.

Bernardin, H. J. (1984). An analysis of black-white differences in job performance. Paper presented at the 44th Annual Meeting of the Academy of Management, Boston, MA.

Bobko, P. (1990). Multivariate correlational analysis. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 2, 2nd ed., pp. 637-686). Palo Alto, CA: Consulting Psychologists Press, Inc.

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, 561-589.

Boehm, V. R. (1977). Differential prediction: A methodological artifact? Journal of Applied Psychology, 62, 146-154.

Bolick, C. (1988). Legal and policy aspects of testing. Journal of Vocational

Behavior, 33, 320-330.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp.71-98). San Francisco: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10, 99-100.

Boston Chapter, NAACP v. Beecher, 8 FEP 855 (1974).

Bridgeport Guardians v. City of Bridgeport, 933 F.2d 1140 (CA2 1991).

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multivariate regression. Psychological Bulletin, 114, 542-551.

Byham, W. (1983). Impact of sex and race variables on a middle management assessment center. Pittsburgh, PA: Development Dimensions, Inc.

Campbell, J. P. (1996). Group differences and personnel decisions: Validity, fairness, and affirmative action. Journal of Vocational Behavior, 49, 122-158.

Campbell, J., Crooks, L. A., Mahoney, M. H., & Rock, D. A. (1973). An investigation of sources of bias in the prediction of job performance – A six year study. (Tech. Rep. PF73-37). Princeton, N. J.: Educational Testing Service.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W. C. Borman, & Associates (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco, CA: Jossey-Bass.

Cascio, W. F. (1991). Applied psychology and personnel management.

Englewood Cliffs, NJ: Prentice Hall.

Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. Human Performance, 4, 233-264.

Cattell, R. B. (1957). Personality and motivation: Structure and measurement. New York: Harcourt, Brace and World.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. Journal of Applied Psychology, 82, 143-159.

Charnes, A., & Cooper, W. W. (1961). Management models and industrial applications of linear programming, Vols. 1 and 2. New York: Wiley.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Cohen, J. L. (1974). Issues in the selection of minority group employees. Human Resource Management, Spring, 17-18.

Cole, N. S. (1973). Bias in selection. Journal of Educational Measurement, 10, 237-255.

Coleman, T., Branch, M. A., & Grace, A. (1999). Optimization toolbox (Version 2). Natick, MA: The Math Works, Inc.

Cooil, B., Winer, R. S., & Rados, D. L. (1987). Cross-validation for prediction. Journal of Marketing Research, 24, 271-279.

Connecticut v. Teal, 457 U. S. 440 (1982).

Cronbach, L. J., Yalow, E., & Schaeffer, G. A. (1980). A mathematical structure for analyzing fairness in selection. Personnel Psychology, 33, 693-704.

Dantzig, G. B. (1982). Reminiscences about the origin of linear programming. Operations Research Letters, 1, 43-48.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Darlington, R. B. (1971). Another look at "cultural fairness." Journal of Educational Measurement, 8, 71-82.

Darlington, R. B. (1978). Reduced-variance regression. Psychological Bulletin, 85, 1238-1255.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. American Psychologist, 34, 571-582.

De Corte, W. (1999). Weighting job performance predictors to both maximize the quality of the selected workforce and control the level of adverse impact. Journal of Applied Psychology, 84, 695-702.

DeMonte, E. K., & Arnold, D. W. (2000). Court upholds employee's non-linear application of test. TIP, 37, 152-153.

Dunnette, M. D. (1963). A note on the criterion. Journal of Applied Psychology, 47, 251-254.

Dunnette, M., Goldstein, I., Hough, L., Jones, D., Outtz, J., Prien, E., Schmitt, N., Siskin, B., & Zedeck, S. (1997). Response to criticisms of Nassau county test construction and validation project [On-line]. Available: http://www.ipmaac.org/nassau/zedeck.html.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. SIAM

Monograph #38. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Einhorn, H. J., & Bass, A. R. (1971). Methodology considerations relevant to discrimination in employment testing. Psychological Bulletin, 75, 261-269.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 13, 171-192.

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.

Feiring, G. R. (1986). Linear programming: An introduction. Newbury Park: Sage Publications.

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.

Fowler v. Schwarzwalden, 5 FEP 270 (1972).

Friedman, M. (1980). Differences in assessment center ratings as a function of the race and geographic location of the assessee. Internal technical report: Tennessee Valley Authority.

Gael, S., & Grant, D. L. (1972). Employment test validation for minority and nonminority telephone company service representatives. Journal of Applied Psychology, 56, 135-139.

Gael, S., Grant, D. L., & Ritchie, R. J. (1975). Employment test validation for minority and nonminority clerks with work sample criteria. Journal of Applied Psychology, 60, 420-426.

Gatewood, R. D., & Feild, H. S. (1998). Human resource selection. Fort Worth: The Dryden Press.

Ghiselli, E. E. (1956). Dimensional problems of criteria. Journal of Applied Psychology, 40, 1-4.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

Gniatczyk, L. A. (2000). An examination of career progress in a downsizing organization. Unpublished doctoral dissertation, University of Tennessee.

Gniatczyk, L. A., & Ladd, R. T. (2001). Assessment center validity in a downsizing organization. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Goldstein, H. W., Yusko, K. P., & Braverman, E. P. (1996). Why do subgroup differences vary based on type of testing method used? In M. Campion (Chr.), A symposium: New research on group differences and adverse impact in selection. Presented at the 11th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Gordon, R. A. (1986). Scientific justification and the race-IQ-delinquency model. In T. F. Hartnagel & R. A. Silverman (Eds.), Critique and explanation: Essays in honor of Gwynne Nettler. New Brunswick, NJ: Transactions Books.

Gottfredson, L. S. (1988). Reconsidering fairness: A matter of social and ethical priorities. Journal of Vocational Behavior, 33, 293-319.

Gough, H. G. (1996). CPI Manual. (3rd Ed.). Palo Alto, CA: Consulting Psychologists Press, Inc.

Grant, D. L., & Bray, D. W. (1970). Validation of employment tests for telephone company installation and repair occupations. Journal of Applied Psychology, 54, 7-14.

Griggs v. Duke Power, 401 U. S. 424 (1971).

Guilford, J. P., Lovell, C., & Williams, R. M. (1942). Completely weighted versus unweighted scoring in an achievement examination. Educational and Psychological Measurement, 2, 15-21.

Guion, R. M. (1966). Employment tests and discriminatory hiring. Industrial Relations, 5, 20-37.

Guion, R. M. (1987). Changing views for personnel selection research. Personnel Psychology, 40, 199-213.

Guion, R. M. (1998). Assessment, measurement, and prediction for personnel decisions. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Guttman, A. (2000). EEO law and personnel practices. Thousand Oaks, CA: Sage Publications, Inc.

Hartigan, J. A., & Wigdor, A. K. (1989). Fairness in employment testing. Washington, DC: National Academy Press.

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. Journal of Applied Psychology, 82, 656-664.

Hauenstein, N. M., Bess, T. L., Swartz, D. E., & Byrd, T. G. (2001). Combining predictor and criterion strategies to eliminate adverse impact in employment selection. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology. San Diego, CA.

Henderson, J. A., & Ladd, R. T. (2001). Constrained estimation: Finding the balance between competing goals. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hernstein, R. J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: Free Press.

Hoffman, C. C., & Thornton, G. C. (1997). Examining selection utility where competing predictors differ in adverse impact. Personnel Psychology, 50, 455-470.

Huck, J. R., & Bray, D. (1976). Management assessment center evaluations and subsequent job performance of black and white females. Personnel Psychology, 29, 13-30.

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. Journal of Applied Psychology, 83, 179-189.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. Journal of Applied Psychology, 81, 459-473.

Humphreys, L. (1952). Individual differences. Annual Review of Psychology, 3, 131-150.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J. E., & Schmidt, F. L. (1982). The economic benefits of personnel selection using psychological ability tests. Industrial Relations, 21, 293-308.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 86, 721-735.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. M. (1977). Fairness of psychological tests: Implications of four definitions for selection utility and minority hiring. Journal of Applied Psychology, 62, 245-260.

Ignizio, J. P. (1982). Programming in single and multiple-objective systems. Englewood Cliffs: Prentice-Hall.

Ignizio, J. P. (1985). Introduction to linear goal programming. Newbury Park: Sage Publications.

Ironson, G. H., Guion, R. M., & Ostrander, M. (1982). Adverse impact from a psychometric perspective. Journal of Applied Psychology, 67, 419-432.

Jaffee, C., Cohen, S., & Cherry, R. (1972). Supervisory selection program for disadvantaged or minority employees. Training and Development Journal, 26, 22-28.

James, L. R. (1998). Measurement of personality via conditional reasoning. Organization Research Methods, 1, 131-163.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypotheses. Behavioral and Brain Sciences, 8, 193-263.

Jordan v. City of New London, 3:97CV1012 (1999).

Katzell, R. A., & Dyer, F. J. (1977). Differential validity revived. Journal of Applied Psychology, 62, 137-145.

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

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In J. Neyman's (Ed.), Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability (pp. 481-491). Berkeley, CA: University of California Press.

Landy, F. J. (1989). Psychology of work behavior (4th ed.). Belmont, CA: Brooks/Cole.

Lawshe, C. H. (1987). Adverse impact: Is it a viable concept? Professional Psychology: Research and Practice, 18, 492-497.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). Race differences in intelligence. New York: Freeman.

Maxwell, S. E., & Arvey, R. D. (1993). The search for predictors with high validity and low adverse impact: Compatible or incompatible goals? Journal of Applied Psychology, 78, 433-437.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. L., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel Psychology, 43, 335-354.

Mooney, C. Z., & Duval, R. D. (1993). Bootstrapping: A nonparametric approach to statistical inference. Newbury Park, CA: Sage Publications, Inc.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. Journal of Applied Psychology, 75, 640-647.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. Journal of Occupational and Organizational Psychology, 66, 337-344.

Murphy, K. R. (1996). Individual differences in behavior: Much more than g. In K. R. Murphy (Ed.), Individual differences and behavior in organizations (pp. 3-30). San Francisco, CA: Jossey-Bass.

Murphy, K. R., Osten, K., & Myors, B. (1995). Modeling the effects of banding in personnel selection. Personnel Psychology, 48, 61-84.

Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. Personnel Psychology, 50, 823-854.

Myart v. Motorola, 110 Congressional Record 5662-64 (1964).

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. Personnel Psychology, 41, 517-535.

O'Connor, E. J., Wexley, K. N., & Alexander, R. A. (1975). Single group validity: Fact or fallacy? Journal of Applied Psychology, 60, 352-355.

Officers for Justice v. Civil Service Commission, 979 F.2d 721 (CA9 1992).

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.

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. Journal of Applied Psychology, 64, 569-607.

Phillips, A. J. (1943). Further evidence regarding weighted versus unweighted scoring of examinations. Educational and Psychological Measurement, 3, 151-155.

Pulakos, E. D., & Schmitt, N. (1995). Experience based and situational interview questions: Studies of validity. Personnel Psychology, 48, 289-308.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. Human Performance, 9, 241-258.

Raju, N. S., & Edwards, J. E. (1984). Note on "Adverse impact from a psychometric perspective". Journal of Applied Psychology, 69, 191-193.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. Personnel Psychology, 44, 321-332.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. Current Directions in Psychological Science, 1, 86-89.

Reilly, R. R. (1996). Alternate selection procedures. In R. S. Barrett (Ed.), Fair employment strategies in human resource management (pp. 209-221). Westport, CT: Quorum Books.

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.

Ruch, W. W. (1972). A re-analysis of published differential validity studies. Symposium paper presented at American Psychological Association, Honolulu, Hawaii, September.

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., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. Personnel Psychology, 50, 707-721.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. Personnel Psychology, 49, 549-572.

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. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 31, 699-714.

Schmidt, F. L. (1988). The problem of group differences in ability scores in employment selection. Journal of Vocational Behavior, 33, 272-292.

Schmidt, F. L. (1991). Why all banding procedures in personnel selection are logically flawed. Human Performance, 4, 265-277.

Schmidt, F. L. (1993). Personnel psychology at the cutting edge. In N. Schmitt, W. C. Borman, & Associates (Eds.), Personnel selection in organizations (pp. 497-516). San Francisco, CA: Jossey-Bass.

Schmidt, F. L., Berner, J. G., & Hunter, J. E. (1973). Racial differences in validity of employment tests: Reality or illusion? Journal of Applied Psychology, 58, 5-6.

Schmidt, F. L., Greenthal, A. L., Hunter, J. E., Berner, J. G., & Seaton, F. W. (1977). Job sample vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. Personnel Psychology, 30, 187-197.

Schmidt, F. L., & Hunter, J. E. (1974). Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. American Psychologist, 29, 1-8.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. American Psychologist, 36, 1128-1137.

Schmidt, F. L., & Hunter, J. E. (1995). The fatal internal contradiction in banding: Its statistical rationale is logically inconsistent with its operational procedures. Human Performance, 8, 203-214.

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, 262-274.

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two jobs in the petroleum industry. Journal of Applied Psychology, 66, 261-273.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on workforce productivity. Journal of Applied Psychology, 64, 609-626.

Schmidt, F. L., & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. Personnel Psychology, 24, 419-434.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. Personnel Psychology, 33, 705-724.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), International review of industrial and organizational psychology (pp. 115-140). New York: Wiley.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. P. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

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.

Sharf, J. C. (1988). Litigating personnel measurement policy. Journal of Vocational Behavior, 33, 235-271.

Society for Industrial and Organizational Psychology (2002). Principles for the validation and use of personnel selection procedures (4^{th} ed.). College Park, MD: Author.

Steffy, B. D., & Ledvinka, J. (1989). The long-range impact of five definitions of “fair” employee selection on black employment and employee productivity. Organizational Behavior and Human Decision Processes, 44, 297-324.

Steuer, R. E. (1986). Multiple criteria optimization: Theory, computation, and application. New York: John Wiley & Sons.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.) (1994). Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction. Palo Alto, CA: Consulting Psychologists Press.

Thorndike, R. L. (1949). Personnel selection: Test and measurement technique. New York: Wiley.

Thorndike, R. L. (1971). Concepts of culture-fairness. Journal of Educational Measurement, 8, 63-70.

United States v. City and County of San Francisco, 979 F.2d 169 (CA9 1992).

United States v. Detroit Edison, 6 FEP 612 (1973).

Using MATLAB (Version 5) (1998). Natick, MA: The Math Works, Inc.

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. Intelligence, 23, 15-32.

Wagner, R. K. (1997). Intelligence, training, and employment. American Psychologist, 52, 1059-1069.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213-217.

Wallace, S. R. (1965). Criteria for what? American Psychologist, 20, 411-418.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). Ability testing: Use, consequences, and controversies. Washington, DC: National Academy Press.

Wollack, S. (1994). Confronting adverse impact in cognitive examinations. Public Personnel Management, 23, 217-224.

APPENDICES

A. GLOSSARY OF TERMS

Glossary of Terms

Assessment – Any systematic process of using tests or other sources to obtain information for the purposes of drawing inferences about individual characteristics.

Assessment center – Method of selection where behavioral ratings are made using multiple assessment techniques along with the pooled judgments of multiple raters.

Battery – Group of tests administered as a unit.

Bias – Systematic error variance that differentially affects the scores of different groups of individuals.

Biographical information (biodata) – Personal history data. A type of selection instrument that relies on past and present behavior to predict future behavior.

Bootstrapping – An iterative, nonparametric technique used to make probability-based inferences about some population characteristic from an estimator derived from a sample drawn from that population. In this research, it refers to resampling the data with replacement (taking a portion of data out for each iteration) several times to generate an empirical estimate of the sampling distribution of some characteristic or statistic.

Compensating model – A higher score on one test compensating for a lower score on another test within some battery of tests used for selection.

Composite score (predictor composite) – An overall score that combines the results from several individual selection procedures according to a specified formula.

Conscientiousness – This is a facet of personality sometimes used in selection that is purported to measure individual levels of responsibility and dependability.

Consent decree – This is essentially a court ordered plan to systematically increase diversity within an organization. It is intended to remedy prior discrimination.

Contextual behavior – This refers to those behaviors on the job, such as citizenship or helping others, that are not always rewarded or noticed but remain important to the ongoing success of any organization.

Criterion – Some measure of work performance or behavior, such as productivity, accident rate, absenteeism, tenure, or supervisory ratings of job relevant behaviors, tasks, or activities.

Criterion-related validity evidence – Statistical representation of the relationship between scores on a predictor and scores on a criterion measure.

Criterion (or predictor) unreliability – Unreliability in either predictors or criterion, while unavoidable, limits the size of potential validity coefficients.

Cross-validation – The application of predictor weights empirically derived in one sample to a different sample from the same population to determine the stability of relationships based on the original weights.

Differential prediction – A case in which the use of a common regression equation results in systematic nonzero prediction errors between subgroups.

Disparate (or Adverse) impact – A substantially different rate of selection in some employment decision that works to the disadvantage of members of a particular subgroup defined by, for example, race.

Integrity test – Intended to predict a wide variety of counterproductive behaviors such as absenteeism and theft.

Job relatedness – An inference that the results from various selection procedures are relevant or related to performance or other behavior on the job; job relatedness can be demonstrated through (1) criterion-related validity coefficients, (2) showing that the content of the selection procedure is job relevant, and (3) by demonstrating that the selection instrument measures a construct deemed relevant to the job in question.

Job sample tests – Predictive tests developed from actual on-the-job samples of performance.

Leaderless group discussions – This is a selection technique where a group of individuals are provided a problem or topic of discussion and instructed to discuss the issue among themselves for a period of time while assessors rate the behavioral performance of each individual.

Meta-analysis – A statistical method of research where the results from several independent studies sharing some statistic of interest are combined to estimate that parameter over a number of differing situations.

Multiple hurdle model – Selection process where two or more procedures must be passed sequentially.

Power – The probability that a statistical test will reveal statistically significant results if a significant effect truly exists in the population.

Predictor – A measure used to predict criterion performance.

Range restriction – This refers to when the variance and range of scores on either predictors or criterion are restricted. It is important because this restriction limits the size of potential validity coefficients.

Selection ratio – The percentage of job applicants actually hired.

Top-down selection – Decision-making process whereby individuals with the highest scores within the selection system are hired.

Validity – The degree to which accumulated evidence and theory support the judgments and interpretations generated from some selection procedure.

Validity coefficient – Statistical coefficient reflecting the relationship between a selection procedure and a criterion (provides evidence about validity).

Note: These definitions were paraphrased from three sources: the 4^{th} edition of the Principle for the Validation and Use of Personnel Selection Procedures (2002), Mooney & Duval (1993), and Cascio (1991).

B. ADVERSE IMPACT EXAMPLE

Adverse Impact Example

Imagine a situation where an organization needs to fill 40 new positions. Further imagine that this organization is presented with an applicant pool of 100 individuals that includes 80 Caucasians and 20 African-Americans. If, based on the predicted performance of each individual, the organization selects 35 Caucasians and 5 African-Americans, the resultant selection ratios (SR) and adverse impact ratio (AIR) would be calculated in this manner:

Caucasian SR:	35 / 80 = .44	AIR = .25 / .44 = .57
African-American SR:	5 / 20 = .25	

The 4/5ths rule states that any AIR less than .80 represents adverse impact and deserves additional scrutiny. In this situation, the selection ratios would have to be at least those shown below to yield a showing of no adverse impact.

Caucasian SR:	33 / 80 = .41	AIR = .35 / .41 = .85
African-American SR:	7 / 20 = .35	

However, it is important to note that regardless of the results, this calculation is only a preliminary step in this process because the statistical demonstration of adverse impact does not always reflect evidence of illegal discrimination. This question would be better answered after all of the information (including information about the validity of the selection process and the job in question) had been reviewed.

C. MODEL TO MONTE CARLO COMPARISON MATRICES

Matrix of Correlations and *d* Values Associated with the Monte Carlo Generated Dataset (N = 200,000)

	d	**Subgroup Corr**	**Intercorrelations**			
			1	**2**	**3**	**4**
1. Cognitive Ability	1.0	-.40				
2. Structured Interview	.23	-.09	.24			
3. Conscientiousness	.10	-.04	.00	.12		
4. Biographical Data	.32	-.13	.19	.16	.51	
5. General Performance	.45	-.18	.30	.30	.18	.28

Bobko, Potosky, & Roth's (1999) Matrix of Correlations and *d* Values

	d	**Subgroup Corr**	**Intercorrelations**			
			1	**2**	**3**	**4**
1. Cognitive Ability	1.0	-.37				
2. Structured Interview	.23	-.09	.24			
3. Conscientiousness	.09	-.04	.00	.12		
4. Biographical Data	.33	-.13	.19	.16	.51	
5. General Performance	.45	-.18	.30	.30	.18	.28

Note: Bobko et al. did not associate a negative sign with each of the subgroup correlations, but this could only be a mistake of omission given the direction of the standardized subgroup differences reflected by *d*.

Matrix of Deviations Between the Model Matrix's Correlations and *d* Values to that of the Monte Carlo Generated Dataset

	d	**Subgroup Corr**	**Intercorrelations**			
			1	**2**	**3**	**4**
1. Cognitive Ability	.00	.03				
2. Structured Interview	.00	.00	.00			
3. Conscientiousness	-.01	.00	.00	.00		
4. Biographical Data	.00	.00	.00	.00	.00	
5. General Performance	.00	.00	.00	.00	.00	.00

Note: Negative values indicate that the Model Matrix provided smaller estimates than the Monte Carlo dataset.

D. MONTE CARLO MATRICES OF CORRELATIONS AND SUBGROUP DIFFERENCES

Matrix of Correlations and d Values for the Monte Carlo Generated Dataset with a 5% Minority Population (N = 10,000)

			Intercorrelations			
	d	**Subgroup Corr**	**1**	**2**	**3**	**4**
1. Cognitive Ability	1.0	-.23				
2. Structured Interview	.19	-.04	.23			
3. Conscientiousness	.10	-.02	.00	.12		
4. Biographical Data	.34	-.07	.18	.15	.52	
5. General Performance	.46	-.10	.28	.29	.18	.27

Matrix of Correlations and d Values for the Monte Carlo Generated Dataset with a 20% Minority Population (N = 10,000)

			Intercorrelations			
	d	**Subgroup Corr**	**1**	**2**	**3**	**4**
1. Cognitive Ability	1.0	-.40				
2. Structured Interview	.17	-.07	.25			
3. Conscientiousness	.06	-.02	.00	.12		
4. Biographical Data	.29	-.12	.18	.17	.51	
5. General Performance	.42	-.17	.30	.31	.18	.28

Matrix of Correlations and d Values for the Monte Carlo Generated Dataset with a 40% Minority Population (N = 10,000)

			Intercorrelations			
	d	**Subgroup Corr**	**1**	**2**	**3**	**4**
1. Cognitive Ability	1.0	-.49				
2. Structured Interview	.22	-.11	.26			
3. Conscientiousness	.09	-.05	.00	.13		
4. Biographical Data	.33	-.16	.20	.16	.51	
5. General Performance	.47	-.23	.31	.31	.18	.28

E. MONTE CARLO DESCRIPTIVE STATISTICS

Monte Carlo Descriptive Statistics

Descriptive Statistics Associated with the Validation Model in a 5% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	.094	.972	-.013	-.083	.101	-3.641	3.658	5107
CA	.151	.958	.094	.248	.141	-3.366	3.840	5107
SI	.029	.994	.011	-.059	.027	-3.380	3.604	5107
Con	.028	1.000	.053	-.098	.003	-3.508	3.436	5107
Bio	.059	.991	.007	.009	.073	-3.290	3.589	5107

Descriptive Statistics Associated with the Cross-Validation Model in a 5% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	.081	.992	.029	-.017	.072	-3.586	3.706	4893
CA	.152	.967	.046	.013	.136	-3.507	3.657	4893
SI	.053	1.008	-.047	.036	.052	-4.253	3.472	4893
Con	.008	.988	-.027	-.024	.017	-3.388	3.364	4893
Bio	.028	.985	-.036	-.049	.042	-3.132	3.542	4893

Descriptive Statistics Associated with the Validation Model in a 20% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	-.001	1.003	-.009	-.036	.006	-3.893	3.234	5107
CA	-.013	.995	-.024	-.016	-.008	-3.744	3.404	5107
SI	-.011	.988	.040	.027	-.020	-3.304	3.871	5107
Con	.009	1.006	-.016	.108	-.003	-4.037	4.079	5107
Bio	-.018	.990	.010	-.021	-.011	-3.286	3.591	5107

Descriptive Statistics Associated with the Cross-Validation Model in a 20% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	-.012	.990	.004	-.001	-.009	-3.701	3.379	4893
CA	-.001	.999	-.014	.104	-.009	-4.512	3.701	4893
SI	.022	.996	.046	.089	.032	-3.866	3.833	4893
Con	.021	1.003	.015	.065	.028	-3.577	4.352	4893
Bio	.007	1.000	.020	-.004	-.001	-3.435	3.783	4893

Descriptive Statistics Associated with the Validation Model in a 40% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	-.107	1.015	.010	.001	-.093	-3.737	3.238	5021
CA	-.207	1.019	.036	-.128	-.226	-3.508	3.551	5021
SI	-.059	.989	-.003	-.098	-.042	-3.705	3.339	5021
Con	-.016	1.005	.054	-.119	-.037	-3.596	3.529	5021
Bio	-.068	.992	.018	-.103	-.072	-3.630	4.213	5021

Descriptive Statistics Associated with the Cross-Validation Model in a 40% Minority Population

Variates	Mean	StdDev	Skewness	Kurtosis	Median	Minimum	Maximum	N
Perf	-.089	1.008	-.048	-.120	-.086	-3.547	3.198	4979
CA	-.188	1.035	.011	-.283	-.193	-3.571	3.024	4979
SI	-.052	1.003	.045	-.029	-.048	-4.077	3.890	4979
Con	-.010	1.005	-.035	.133	-.002	-4.211	3.337	4979
Bio	-.047	1.006	-.008	.008	-.045	-3.812	3.589	4979

F. FULL MONTE CARLO RESULTS IN A 5% MINORITY SAMPLE

Monte Carlo R-squares, Adverse Impact Ratios, Predictor Weights, and Cross-Validation Results for Various Predictor Combinations in a 5% Minority Sample

	Validation		Predictor Weights				Cross-Validation		Shrinkage	
Predictors	**R^2**	**AIR**	**CA**	**SI**	**Con**	**Bio**	**R^2**	**AIR**	**R^2**	**AIR**
CA	.08	.12	.2900				.07	.11	-.01	-.01
SI	.09	.62		.2961			.08	.93	-.01	.31
Con	.03	.88			.1740		.03	1.02	*	.14
Bio	.07	.53				.2710	.07	.60	*	.07
CA+SI										
OLS	.14	.23	.2338	.2417			.13	.22	-.01	-.01
Unit	.14	.23					.13	.22	-.01	-.01
Constrained										
AIW=0.15	.14	.31	.2091	.2631			.13	.26	-.01	-.05
AIW=0.25	.14	.31	.1802	.2823			.12	.30	-.02	-.01
AIW=0.35	.13	.35	.1268	.3049			.12	.47	-.01	.12
AIW=0.50	.09	.62	.0000	.2961			.08	.93	-.01	.31
CA+Con										
OLS	.11	.14	.2902		.17443		.10	.17	-.01	.03
Unit	.11	.24					.10	.25	-.01	.01
Constrained										
AIW=0.15	.11	.15	.2824		.1862		.10	.17	-.01	.02
AIW=0.25	.11	.15	.2729		.1981		.10	.17	-.01	.02
AIW=0.35	.11	.21	.2550		.2159		.10	.30	-.01	.09
AIW=0.50	.08	.42	.1296		.2556		.08	.44	*	.02
CA+Bio										
OLS	.13	.12	.2494			.2261	.12	.15	-.01	.03
Unit	.13	.12					.12	.19	-.01	.07
Constrained										
AIW=0.15	.13	.12	.2342			.2410	.12	.19	-.01	.07
AIW=0.25	.13	.14	.2165			.2554	.12	.19	-.01	.05
AIW=0.35	.13	.12	.1846			.2753	.11	.33	-.02	.21
AIW=0.50	.07	.53	.0000			.2710	.07	.60	*	.07
SI+Con										
OLS	.11	.64		.2788	.1395		.11	.90	*	.26
Unit	.10	.74					.10	.78	*	.04
Constrained										
AIW=0.15	.11	.65		.2781	.1408		.11	.90	*	.25
AIW=0.25	.11	.65		.2772	.1422		.11	.90	*	.25
AIW=0.35	.11	.64		.2758	.1445		.11	.90	*	.26
AIW=0.50	.11	.70		.2701	.1531		.11	.88	*	.18
SI+Bio										
OLS	.14	.53		.2606		.2309	.14	.70	*	.17
Unit	.14	.51					.14	.70	*	.19
Constrained										
AIW=0.15	.14	.54		.2667		.2239	.14	.70	*	.16
AIW=0.25	.14	.54		.2730		.2162	.14	.72	*	.18
AIW=0.35	.14	.57		.2824		.2031	.14	.70	*	.13
AIW=0.50	.13	.61		.3100		.1462	.13	.73	*	.12
Con+Bio										
OLS	.07	.59			.0432	.2482	.08	.65	.01	.06
Unit	.06	.67					.07	.75	.01	.08
Constrained										
AIW=0.15	.07	.59			.0525	.2420	.08	.65	.01	.06
AIW=0.25	.07	.54			.0624	.2349	.08	.63	.01	.09
AIW=0.35	.07	.56			.0783	.2222	.08	.65	.01	.09
AIW=0.50	.07	.66			.1362	.1597	.07	.73	*	.07

Predictors	Validation		Predictor Weights				Cross-Validation		Shrinkage	
	R^2	AIR	CA	SI	Con	Bio	R^2	AIR	R^2	AIR
CA+SI+Con										
OLS	.16	.31	.2384	.2225	.1467		.15	.22	-.01	-.09
Unit	.16	.37					.15	.34	-.01	-.03
Constrained										
AIW=0.15	.16	.32	.2101	.2428	.1534		.15	.30	-.01	-.02
AIW=0.25	.16	.37	.1770	.2609	.1583		.15	.34	-.01	-.03
AIW=0.35	.15	.42	.1164	.2816	.1608		.14	.54	-.01	.12
AIW=0.50	.11	.64	.0000	.2760	.1442		.11	.90	*	.26
CA+SI+Bio										
OLS	.18	.20	.2032	.2179		.2009	.17	.28	-.01	.08
Unit	.18	.20					.17	.28	-.01	.08
Constrained										
AIW=0.15	.18	.23	.1697	.2378		.2104	.17	.31	-.01	.08
AIW=0.25	.17	.28	.1315	.2553		.2172	.16	.39	-.01	.11
AIW=0.35	.16	.40	.0636	.2740		.2200	.15	.52	-.01	.12
AIW=0.50	.14	.59	.0001	.2870		.1961	.14	.70	*	.11
CA+Con+Bio										
OLS	.14	.14	.2570		.0772	.1841	.13	.14	-.01	*
Unit	.12	.21					.12	.30	*	.09
Constrained										
AIW=0.15	.14	.14	.2414		.0818	.1963	.13	.15	-.01	.01
AIW=0.25	.14	.14	.2233		.0862	.2081	.13	.19	-.01	.05
AIW=0.35	.13	.14	.1902		.0919	.2242	.12	.26	-.01	.12
AIW=0.50	.07	.61	.0001		.0913	.2106	.07	.67	*	.06
SI+Con+Bio										
OLS	.14	.57		.2594	.0278	.2164	.14	.70	*	.13
Unit	.12	.64					.12	.70	*	.06
Constrained										
AIW=0.15	.14	.56		.2647	.0379	.2037	.14	.70	*	.14
AIW=0.25	.14	.61		.2697	.0485	.1896	.14	.65	*	.04
AIW=0.35	.14	.62		.2764	.0657	.1655	.14	.67	*	.05
AIW=0.50	.12	.64		.2828	.1291	.0575	.12	.77	*	.13
CA+SI+Con+Bio										
OLS	.18	.21	.2098	.2140	.0583	.1696	.17	.28	-.01	.07
Unit	.17	.31					.16	.34	-.01	.03
Constrained										
AIW=0.15	.18	.26	.1763	.2338	.0600	.1782	.17	.33	-.01	.07
AIW=0.25	.17	.31	.1379	.2512	.0608	.1845	.17	.37	*	.06
AIW=0.35	.16	.45	.0695	.2700	.0598	.1877	.16	.52	*	.07
AIW=0.50	.14	.62	.0000	.2790	.0740	.1532	.14	.68	*	.06

G. FULL MONTE CARLO RESULTS IN A 20% MINORITY SAMPLE

Monte Carlo R-squares, Adverse Impact Ratios, Predictor Weights, and Cross-Validation Results for Various Predictor Combinations in a 20% Minority Sample

	Validation		Predictor Weights				Cross-Validation		Shrinkage	
Predictors	**R^2**	**AIR**	**CA**	**SI**	**Con**	**Bio**	**R^2**	**AIR**	**R^2**	**AIR**
CA	.10	.15	.3107				.08	.13	-.02	-.02
SI	.09	.84		.3017			.10	.84	.01	*
Con	.03	.92			.1869		.03	.87	*	-.05
Bio	.08	.67				.2771	.08	.62	*	-.05
CA+SI										
OLS	.15	.29	.2478	.2356			.15	.30	*	.01
Unit	.15	.30					.15	.30	*	*
Constrained										
AIW=0.15	.15	.34	.2187	.2618			.15	.33	*	-.01
AIW=0.25	.14	.39	.1841	.2851			.14	.38	*	-.01
AIW=0.35	.13	.51	.1194	.3113			.13	.49	*	-.02
AIW=0.50	.09	.84	.0000	.3018			.10	.84	.01	*
CA+Con										
OLS	.13	.21	.3091		.1843		.12	.23	-.01	.02
Unit	.12	.30					.11	.30	-.01	*
Constrained										
AIW=0.15	.13	.22	.2994		.1987		.12	.23	-.01	.01
AIW=0.25	.13	.24	.2876		.2131		.11	.24	-.02	*
AIW=0.35	.13	.26	.2649		.2341		.11	.28	-.02	.02
AIW=0.50	.03	.92	.0000		.1869		.03	.87	*	-.05
CA+Bio										
OLS	.15	.19	.2700			.2292	.14	.24	-.01	.05
Unit	.15	.21					.14	.26	-.01	.05
Constrained										
AIW=0.15	.15	.21	.2525			.2469	.14	.27	-.01	.06
AIW=0.25	.15	.23	.2320			.2639	.13	.28	-.02	.05
AIW=0.35	.14	.25	.1943			.2869	.13	.32	-.01	.07
AIW=0.50	.08	.67	.0000			.2772	.08	.62	*	-.05
SI+Con										
OLS	.11	.82		.2828	.1515		.12	.76	.01	-.06
Unit	.11	.84					.11	.77	*	-.07
Constrained										
AIW=0.15	.11	.81		.2809	.1546		.12	.76	.01	-.05
AIW=0.25	.11	.82		.2787	.1578		.12	.76	.01	-.06
AIW=0.35	.11	.82		.2752	.1629		.12	.76	.01	-.06
AIW=0.50	.11	.84		.2601	.1815		.12	.75	.01	-.09
SI+Bio										
OLS	.14	.67		.2620		.2323	.15	.66	.01	-.01
Unit	.14	.66					.15	.65	.01	-.01
Constrained										
AIW=0.15	.14	.66		.2701		.2231	.15	.67	.01	.01
AIW=0.25	.14	.66		.2781		.2128	.15	.67	.01	.01
AIW=0.35	.14	.66		.2901		.1952	.15	.68	.01	.02
AIW=0.50	.13	.74		.3206		.1147	.14	.69	.01	-.05
Con+Bio										
OLS	.08	.74			.0596	.2463	.08	.66	*	-.08
Unit	.07	.80					.07	.76	*	-.04
Constrained										
AIW=0.15	.08	.75			.0729	.2369	.08	.66	*	-.09
AIW=0.25	.08	.78			.0867	.2260	.08	.68	*	-.10
AIW=0.35	.08	.79			.1086	.2062	.08	.69	*	-.10
AIW=0.50	.06	.86			.1859	.0870	.06	.80	*	-.06

Predictors	Validation		Predictor Weights				Cross-Validation		Shrinkage	
	R^2	AIR	CA	SI	Con	Bio	R^2	AIR	R^2	AIR
CA+SI+Con										
OLS	.17	.31	.2520	.2147	.1579		.17	.31	*	*
Unit	.17	.39					.16	.38	-.01	-.01
Constrained										
AIW=0.15	.17	.37	.2172	.2400	.1668		.17	.36	*	-.01
AIW=0.25	.17	.43	.1760	.2622	.1729		.17	.43	*	*
AIW=0.35	.15	.58	.0992	.2853	.1743		.15	.58	*	*
AIW=0.50	.11	.82	.0000	.2781	.1587		.12	.76	.01	-.06
CA+SI+Bio										
OLS	.19	.27	.2190	.2086		.2026	.18	.33	-.01	.06
Unit	.19	.28					.18	.35	-.01	.07
Constrained										
AIW=0.15	.19	.33	.1784	.2337		.2143	.18	.40	-.01	.07
AIW=0.25	.18	.38	.1313	.2552		.2221	.18	.45	*	.07
AIW=0.35	.16	.54	.0463	.2762		.2223	.13	.69	-.03	.15
AIW=0.50	.14	.65	.0000	.2882		.1982	.14	.68	*	.03
CA+Con+Bio										
OLS	.15	.20	.2778		.0912	.1807	.14	.24	-.01	.04
Unit	.14	.38					.13	.37	-.01	-.01
Constrained										
AIW=0.15	.15	.21	.2595		.0976	.1948	.14	.26	-.01	.05
AIW=0.25	.15	.24	.2378		.1037	.2081	.14	.28	-.01	.04
AIW=0.35	.14	.30	.1976		.1117	.2259	.14	.33	*	.03
AIW=0.50	.08	.80	.0000		.1108	.2040	.08	.70	*	-.10
SI+Con+Bio										
OLS	.15	.69		.2602	.0466	.2085	.15	.68	*	-.01
Unit	.13	.73					.13	.66	*	-.07
Constrained										
AIW=0.15	.14	.70		.2668	.0613	.1904	.15	.67	.01	-.03
AIW=0.25	.14	.71		.2726	.0766	.1703	.15	.69	.01	-.02
AIW=0.35	.14	.71		.2793	.1009	.1356	.15	.69	.01	-.02
AIW=0.50	.11	.84		.2664	.1743	.0001	.12	.76	.01	-.08
CA+SI+Con+Bio										
OLS	.19	.29	.2266	.2039	.0751	.1632	.19	.36	*	.07
Unit	.18	.38					.17	.42	-.01	.04
Constrained										
AIW=0.15	.19	.32	.1856	.2288	.0784	.1731	.19	.38	*	.06
AIW=0.25	.18	.39	.1379	.2503	.0802	.1798	.18	.47	*	.08
AIW=0.35	.16	.58	.0514	.2712	.0786	.1805	.17	.60	.01	.02
AIW=0.50	.14	.71	.0000	.2781	.0957	.1433	.15	.69	.01	-.02

H. FULL MONTE CARLO RESULTS IN A 40% MINORITY SAMPLE

Monte Carlo R-squares, Adverse Impact Ratios, Predictor Weights, and Cross-Validation Results for Various Predictor Combinations in a 40% Minority Sample

	Validation		Predictor Weights				Cross-Validation		Shrinkage	
Predictors	**R^2**	**AIR**	**CA**	**SI**	**Con**	**Bio**	**R^2**	**AIR**	**R^2**	**AIR**
CA	.09	.17	.3000				.10	.16	.01	-.01
SI	.10	.71		.3094			.10	.76	*	.05
Con	.04	.84			.1933		.03	.96	-.01	.12
Bio	.08	.63				.2846	.08	.65	*	.02
CA+SI										
OLS	.15	.29	.2372	.2497			.16	.29	.01	*
Unit	.15	.29					.16	.28	.01	-.01
Constrained										
AIW=0.15	.15	.32	.2103	.2728			.16	.34	.01	.02
AIW=0.25	.14	.36	.1786	.2934			.16	.37	.02	.01
AIW=0.35	.13	.45	.1204	.3170			.15	.46	.02	.01
AIW=0.50	.10	.71	.0000	.3094			.10	.76	*	.05
CA+Con										
OLS	.13	.21	.2984		.1906		.13	.24	*	.03
Unit	.12	.28					.12	.30	*	.02
Constrained										
AIW=0.15	.13	.23	.2887		.2042		.13	.25	*	.02
AIW=0.25	.13	.24	.2770		.2177		.13	.27	*	.03
AIW=0.35	.12	.27	.2548		.2373		.13	.28	.01	.01
AIW=0.50	.04	.84	.0000		.1933		.03	.96	-.01	.12
CA+Bio										
OLS	.14	.23	.2549			.2358	.15	.24	.01	.01
Unit	.14	.23					.15	.24	.01	.01
Constrained										
AIW=0.15	.14	.24	.2383			.2518	.15	.25	.01	.01
AIW=0.25	.14	.26	.2190			.2672	.15	.26	.01	*
AIW=0.35	.14	.31	.1840			.2882	.14	.32	*	.01
AIW=0.50	.08	.63	.0000			.2846	.08	.65	*	.02
SI+Con										
OLS	.12	.71		.2881	.1532		.12	.78	*	.07
Unit	.11	.71					.11	.82	*	.11
Constrained										
AIW=0.15	.12	.71		.2864	.1558		.12	.78	*	.07
AIW=0.25	.12	.71		.2846	.1585		.12	.78	*	.07
AIW=0.35	.12	.71		.2816	.1629		.12	.78	*	.07
AIW=0.50	.12	.71		.2692	.1789		.12	.79	*	.08
SI+Bio										
OLS	.15	.61		.2698		.2401	.16	.63	.01	.02
Unit	.15	.60					.15	.62	*	.02
Constrained										
AIW=0.15	.15	.61		.2764		.2327	.16	.63	.01	.02
AIW=0.25	.15	.62		.2830		.2246	.16	.64	.01	.02
AIW=0.35	.15	.63		.2931		.2107	.16	.64	.01	.01
AIW=0.50	.14	.63		.3224		.1504	.15	.67	.01	.04
Con+Bio										
OLS	.08	.66			.0652	.2513	.08	.69	*	.03
Unit	.08	.73					.07	.76	-.01	.03
Constrained										
AIW=0.15	.08	.67			.0766	.2432	.08	.69	*	.02
AIW=0.25	.08	.68			.0885	.6840	.08	.68	*	*
AIW=0.35	.08	.68			.1075	.2174	.08	.70	*	.02
AIW=0.50	.07	.73			.1745	.1315	.06	.80	-.01	.07

	Validation		Predictor Weights				Cross-Validation		Shrinkage	
Predictors	R^2	AIR	CA	SI	Con	Bio	R^2	AIR	R^2	AIR
CA+SI+Con										
OLS	.17	.32	.2417	.2264	.1597		.18	.30	.01	-.02
Unit	.17	.35					.18	.35	.01	*
Constrained										
AIW=0.15	.17	.35	.2101	.2482	.1675		.18	.35	.01	*
AIW=0.25	.17	.39	.1731	.2674	.1732		.18	.39	.01	*
AIW=0.35	.15	.49	.1054	.2883	.1757		.16	.50	.01	.01
AIW=0.50	.12	.71	.0000	.2823	.1619		.12	.78	*	.07
CA+SI+Bio										
OLS	.19	.32	.2039	.2237		.2087	.20	.32	.01	*
Unit	.19	.31					.20	.31	.01	*
Constrained										
AIW=0.15	.19	.36	.1671	.2453		.2189	.20	.37	.01	.01
AIW=0.25	.18	.41	.1252	.2640		.2260	.19	.42	.01	.01
AIW=0.35	.17	.50	.0508	.2834		.2279	.17	.54	*	.04
AIW=0.50	.15	.62	.0001	.2975		.2039	.16	.65	.01	.03
CA+Con+Bio										
OLS	.15	.22	.2638		.0968	.1848	.16	.24	.01	.02
Unit	.14	.34					.14	.35	*	.01
Constrained										
AIW=0.15	.15	.24	.2463		.1033	.1971	.15	.25	*	.01
AIW=0.25	.15	.27	.2259		.1094	.2088	.15	.28	*	.01
AIW=0.35	.14	.32	.1885		.1176	.2243	.14	.32	*	*
AIW=0.50	.08	.71	.0000		.1197	.2054	.07	.71	-.01	*
SI+Con+Bio										
OLS	.15	.63		.2673	.0453	.2175	.16	.63	.01	*
Unit	.13	.63					.13	.64	*	.01
Constrained										
AIW=0.15	.15	.63		.2724	.0572	.2031	.16	.63	.01	*
AIW=0.25	.15	.63		.2771	.0647	.1873	.15	.64	*	.01
AIW=0.35	.15	.63		.2831	.0897	.1603	.15	.66	*	.03
AIW=0.50	.13	.69		.2808	.1625	.0355	.13	.75	*	.06
CA+SI+Con+Bio										
OLS	.19	.32	.2121	.2177	.0774	.1702	.20	.33	.01	.01
Unit	.18	.36					.18	.38	*	.02
Constrained										
AIW=0.15	.19	.36	.1751	.2390	.0774	.1788	.20	.37	.01	.01
AIW=0.25	.19	.39	.1328	.2575	.0793	.1847	.19	.41	*	.02
AIW=0.35	.17	.51	.0573	.2768	.0790	.1865	.18	.53	.01	.02
AIW=0.50	.14	.63	.0000	.2849	.0980	.1483	.15	.66	.01	.03

I. FULL STUDY TWO DESCRIPTIVE STATISTICS

Total Sample

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Critical Thinking Appraisal	322	36	78	65.63	7.293	-.917	.661
Manager Profile Record (MPR)							
Background (Bio)	334	13	32	24.72	3.612	-.518	.303
Judgment	334	14	32	23.09	2.912	-.043	.264
Reasoning by Inference	308	7	27	17.80	3.859	-.272	-.270
Managerial Video Simulation	313	10	99	66.79	19.514	-.293	-.950
Strategic In-Basket	298	2	4	3.04	.607	-.019	-.273
Salary 1993	340	33754	100000	69284.79	12339.63	-.390	.220

Validation Sample

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Critical Thinking Appraisal	215	36	77	65.085	7.332	-.876	.596
Manager Profile Record (MPR)							
Background (Bio)	222	13	32	24.959	3.574	-.480	.311
Judgment	222	16	32	23.279	2.877	.014	.203
Reasoning by Inference	201	7	27	17.657	4.027	-.188	-.423
Managerial Video Simulation	204	30	99	66.451	19.272	-.248	-1.085
Strategic In-Basket	194	2	4	3.041	.634	-.033	-.534
Salary 1993	227	36793	97956	69356.66	11962.93	-.376	.272

Cross-Validation Sample

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Critical Thinking Appraisal	107	41	78	65.28	7.236	-.999	.717
Manager Profile Record (MPR)							
Background (Bio)	112	14	32	24.26	3.658	-.585	.176
Judgment	112	14	30	22.71	2.958	-.130	.257
Reasoning by Inference	107	9	26	18.07	3.523	-.430	.001
Managerial Video Simulation	109	10	99	67.41	2.035	-.374	-.745
Strategic In-Basket	104	2	4	3.04	.556	.018	.320
Salary 1993	113	33754	100000	68821.98	12923.72	-.288	-.051

J. STUDY TWO MATRICES OF CORRELATIONS AND *D* VALUES FOR THE VALIDATION AND CROSS-VALIDATION SAMPLES

Study Two Matrices of Correlations and *d* Values for the Validation and Cross-Validation Samples

Validation Sample							
		Intercorrelations					
	d	**1**	**2**	**3**	**4**	**5**	**6**
1. Critical Thinking Appraisal	1.0						
2. MPR – Background	.35	.26** (211)					
3. MPR – Judgment	.69	.31** (211)	.26** (222)				
4. Reasoning by Inference	.87	.47** (189)	.16* (198)	.37** (198)			
5. Mgr. Video Simulation	.76	.18* (192)	-.03 (201)	.06 (201)	.08 (201)		
6. Strategic In-Basket	.56	.24** (182)	.11 (192)	.09 (192)	.13 (191)	.13 (194)	
7. Salary 1993	.25	.31** (215)	.46** (222)	.28** (222)	.18* (201)	.19** (204)	.16* (194)

Note: The numbers within parentheses represent the sample sizes associated with each correlation. * *p* <.05. ** *p*<.01.

Cross-Validation Sample							
		Intercorrelations					
	d	**1**	**2**	**3**	**4**	**5**	**6**
1. Critical Thinking Appraisal	1.1						
2. MPR – Background	.13	.23* (107)					
3. MPR – Judgment	.62	.24* (107)	.26** (112)				
4. Reasoning by Inference	.58	.23* (102)	.21* (107)	.21* (107)			
5. Mgr. Video Simulation	1.0	.27* (104)	.09 (109)	.20* (109)	.23* (107)		
6. Strategic In-Basket	.63	.32** (99)	.29** (104)	.06 (104)	.09 (102)	.20* (104)	
7. Salary 1993	.25	.36** (107)	.45** (112)	.22* (112)	.18 (107)	.15 (109)	. 31** (104)

Note: The numbers within parentheses represent the sample sizes associated with each correlation. * *p* <.05. ** *p*<.01.

K. STUDY TWO R-SQUARES, ADVERSE IMPACT RATIOS, AND PREDICTOR WEIGHTS FOR VARIOUS PREDICTOR COMBINATIONS IN THE TOTAL SAMPLE

Study Two R-squares, Adverse Impact Ratios, and Predictor Weights for Various Predictor Combinations in the Total Sample

	Sample		Estimates		Predictor Weights					
Predictors	**Maj**	**Min**	**R^2**	**AIR**	**CTA**	**Bio**	**SJT**	**RBI**	**Vscore**	**SIB**
CTA	296	24	.10	.16	.3221					
Bio	306	24	.21	.65		.4572				
SJT	306	24	.07	.31			.2646			
RBI	282	22	.03	.71				.1753		
Vscore	286	23	.03	.33					.1798	
SIB	273	21	.05	.74						.2130
Bio+CTA	294	22								
OLS			.26	.35	.2344	.3997				
Unit			.25	.35						
Constrained										
AIW=0.15			.26	.35	.1826	.4313				
AIW=0.25			.25	.53	.1230	.4542				
AIW=0.35			.22	.53	.0163	.4617				
AIW=0.50			.21	.53	.0000	.4587				
SJT+CTA	294	22								
OLS			.15	.35	.2741		.2125			
Unit			.15	.35						
Constrained										
AIW=0.15			.15	.35	.2717		.2152			
AIW=0.25			.15	.35	.2692		.2180			
AIW=0.35			.15	.35	.2649		.2226			
AIW=0.50			.15	.35	.2481		.2391			
RBI+CTA	269	20								
OLS			.10	.19	.2706			.0912		
Unit			.09	.38						
Constrained										
AIW=0.15			.10	.19	.2768			.0811		
AIW=0.25			.10	.19	.2828			.0702		
AIW=0.35			.10	.19	.2914			.0522		
AIW=0.50			.09	.19	.3062			.0001		
Vscore+CTA	273	21								
OLS			.11	.18	.2791				.1351	
Unit			.10	.18						
Constrained										
AIW=0.15			.11	.18	.2806				.1326	
AIW=0.25			.11	.18	.2823				.1299	
AIW=0.35			.11	.18	.2848				.1255	
AIW=0.50			.11	.18	.2935				.1084	
SIB+CTA	260	19								
OLS			.14	.20	.3028					.1482
Unit			.13	.20						
Constrained										
AIW=0.15			.14	.40	.2990					.1537
AIW=0.25			.14	.40	.2948					.1595
AIW=0.35			.14	.40	.2876					.1688
AIW=0.50			.13	.40	.2562					.2017

Predictors	Sample Maj	Min	Estimates R^2	AIR	CTA	Predictor Weights Bio	SJT	RBI	Vscore	SIB
Bio+RBI	280	21								
OLS			.19	.56		.4019		.1095		
Unit			.15	.56						
Constrained										
AIW=0.15			.19	.56		.4188		.0500		
AIW=0.25			.18	.56		.4206		.0001		
AIW=0.35			.18	.56		.4206		.0000		
AIW=0.50			.18	.56		.4206		.0000		
Bio+RBI+CTA	268	19								
OLS			.23	.40	.2042	.3672		.0574		
Unit			.19	.40						
Constrained										
AIW=0.15			.23	.40	.1729	.3991		.0224		
AIW=0.25			.22	.40	.1296	.4203		.0000		
AIW=0.35			.19	.40	.0412	.4302		.0000		
AIW=0.50			.18	.40	.0001	.4230		.0000		
SJT+Vscore+SIB	272	20								
OLS			.12	.00			.2286		.1386	.1735
Unit			.12	.00						
Constrained										
AIW=0.15			.12	.00			.2317		.1310	.1762
AIW=0.25			.12	.00			.2348		.1226	.1788
AIW=0.35			.12	.00			.2392		.1087	.1825
AIW=0.50			.11	.19			.2482		.0514	.1910
SJT+Vscore+SIB+CTA	260	18								
OLS			.19	.21	.2245		.1950		.1197	.1366
Unit			.18	.21						
Constrained										
AIW=0.15			.19	.21	.2245		.1946		.1141	.1419
AIW=0.25			.19	.21	.2243		.1940		.1082	.1474
AIW=0.35			.19	.21	.2238		.1929		.0983	.1562
AIW=0.50			.19	.21	.2182		.1855		.0592	.1872

L. STUDY TWO R-SQUARES, ADVERSE IMPACT RATIOS, AND PREDICTOR WEIGHTS FOR VARIOUS PREDICTOR COMBINATIONS WITH CROSS-VALIDATION RESULTS

Study Two R-squares, Adverse Impact Ratios, and Predictor Weights for Various Predictor Combinations with Cross-Validation Results

	Sample		Validation		Predictor Weights						Cross-Val		Sample		Shrinkage	
Predictors	**Maj**	**Min**	**R^2**	**AIR**	**CTA**	**Bio**	**SJT**	**RBI**	**Vscore**	**SIB**	**R^2**	**AIR**	**Maj**	**Min**	**R^2**	**AIR**
CTA	196	17	.09	.22	.3034						.13	.00	100	7	.04	-.22
Bio	202	17	.22	.45		.4658					.20	.55	104	7	-.02	.10
SJT	202	17	.09	.45			.2935				.05	.00	104	7	-.04	-.45
RBI	182	16	.03	.47				.1768			.03	1.33	100	6	*	.86
Vscore	185	16	.04	.24					.1992		.02	.55	101	7	-.02	.31
SIB	177	14	.03	.84						.1639	.09	.55	96	7	.06	-.29
Bio+CTA	194	15											100	7		
OLS			.24	.52	.2219	.3882					.28	.00			.04	-.52
Unit			.23	.52							.24	.00			.01	-.52
Constrained																
AIW=0.15			.24	.52	.1786	.4150					.29	.55			.05	.03
AIW=0.25			.24	.52	.1293	.4359					.30	.55			.06	.03
AIW=0.35			.21	.52	.0423	.4500					.27	.55			.06	.03
AIW=0.50			.20	.52	.0001	.4456					.20	.55			*	.03
SJT+CTA	194	15											100	7		
OLS			.16	.52	.2444		.2544				.15	.00			-.01	-.52
Unit			.16	.52							.15	.00			-.01	-.52
Constrained																
AIW=0.15			.16	.52	.2410		.2576				.15	.00			-.01	-.52
AIW=0.25			.16	.52	.2374		.2610				.15	.00			-.01	-.52
AIW=0.35			.16	.52	.2314		.2664				.15	.00			-.01	-.52
AIW=0.50			.16	.52	.2075		.2855				.15	.00			-.01	-.52

	Sample		Validation		Predictor Weights						Cross-Val		Sample		Shrinkage	
Predictors	**Maj**	**Min**	**R^2**	**AIR**	**CTA**	**Bio**	**SJT**	**RBI**	**Vscore**	**SIB**	**R^2**	**AIR**	**Maj**	**Min**	**R^2**	**AIR**
RBI+CTA	173	14											96	6		
OLS			.08	.27	.2381			.0830			.14	.00			.06	-.27
Unit			.08	.27							.14	.00			.06	-.27
Constrained																
AIW=0.15			.08	.27	.2461			.0710			.14	.00			.06	-.27
AIW=0.25			.08	.27	.2539			.0579			.13	.00			.05	-.27
AIW=0.35			.08	.27	.2648			.0362			.13	.00			.05	-.27
AIW=0.50			.08	.27	.2771			.0001			.13	.00			.05	-.27
Vscore+CTA	176	14											97	7		
OLS			.11	.27	.2469				.1771		.07	.00			-.04	-.27
Unit			.11	.27							.06	.00			-.05	-.27
Constrained																
AIW=0.15			.11	.27	.2433				.1816		.07	.00			-.04	-.27
AIW=0.25			.11	.27	.2392				.1863		.07	.00			-.04	-.27
AIW=0.35			.11	.27	.2323				.1937		.07	.00			-.04	-.27
AIW=0.50			.11	.27	.2028				.2189		.06	.00			-.05	-.27
SIB+CTA	168	12											92	7		
OLS			.11	.32	.2950					.1093	.15	.00			.04	-.32
Unit			.10	.32							.15	.00			.05	-.32
Constrained																
AIW=0.15			.11	.32	.2923					.1139	.15	.00			.04	-.32
AIW=0.25			.11	.32	.2893					.1188	.15	.00			.04	-.32
AIW=0.35			.11	.32	.2841					.1267	.15	.00			.04	-.32
AIW=0.50			.11	.32	.2614					.1550	.15	.00			.04	-.32
Bio+RBI	180	15											100	6		
OLS			.19	.51		.3964		.1175			.20	.64			.01	.13
Unit			.15	.51							.16	.64			.01	.13
Constrained																
AIW=0.15			.18	.51		.4109		.0736			.20	.64			.02	.13
AIW=0.25			.18	.51		.4166		.0251			.20	.64			.02	.13
AIW=0.35			.17	.51		.4152		.0001			.19	.64			.02	.13
AIW=0.50			.17	.51		.4152		.0001			.20	.55			.03	.04

	Sample		Validation		Predictor Weights						Cross-Val		Sample		Shrinkage	
Predictors	**Maj**	**Min**	**R^2**	**AIR**	**CTA**	**Bio**	**SJT**	**RBI**	**Vscore**	**SIB**	**R^2**	**AIR**	**Maj**	**Min**	**R^2**	**AIR**
Bio+RBI+CTA	172	13											196	6		
OLS			.20	.29	.1800	.3432		.0559			.28	.64			.08	.35
Unit			.16	.29							.24	.64			.08	.35
Constrained																
AIW=0.15			.19	.29	.1622	.3680		.0217			.29	.64			.10	.35
AIW=0.25			.19	.29	.1334	.3854		.0000			.29	.64			.10	.35
AIW=0.35			.18	.29	.0683	.3991		.0000			.28	.64			.10	.35
AIW=0.50			.15	.29	.0001	.3926		.0000			.23	.64			.08	.35
SJT+Vscore+SIB	176	13											196	7		
OLS			.12	.00			.2455		.1825	.1167	.04	.55			-.08	.55
Unit			.12	.00							.04	.55			-.08	.55
Constrained																
AIW=0.15			.12	.00			.2475		.1818	.1142	.04	.55			-.08	.55
AIW=0.25			.12	.00			.2496		.1810	.1116	.04	.55			-.08	.55
AIW=0.35			.12	.29			.2530		.1797	.1072	.04	.55			-.08	.26
AIW=0.50			.12	.29			.2645		.1739	.0903	.04	.55			-.08	.26
SJT+Vscore+SIB+CTA	168	11											92	7		
OLS			.20	.35	.1963		.2284		.1870	.0944	.08	.00			-.12	-.35
Unit			.19	.35							.08	.00			-.11	-.35
Constrained																
AIW=0.15			.20	.35	.1952		.2237		.1924	.0970	.08	.00			-.12	-.35
AIW=0.25			.20	.35	.1939		.2186		.1981	.0997	.08	.00			-.12	-.35
AIW=0.35			.20	.35	.1915		.2098		.2071	.1039	.08	.00			-.12	-.35
AIW=0.50			.19	.35	.1791		.1730		.2377	.1184	.07	.00			-.12	-.35

VITA

John Ashley Henderson was born in Booneville, Mississippi on August 27, 1970. He attended school in the Booneville School District and graduated from Booneville High School in 1988. He then attended Northeast Mississippi Community College until 1990 when he transferred to the University of Mississippi (affectionately known as Ole Miss). He graduated from Ole Miss in 1993 with a Bachelor of Liberal Arts in Psychology and worked for a year before joining the Psychology graduate program at the University of Tennessee at Chattanooga. He received a Master of Science degree with an emphasis in Industrial and Organizational Psychology in 1995 and immediately entered the Doctoral program of the same field at the main campus of the University of Tennessee in Knoxville. John officially received his Doctor of Philosophy degree in 2004.

While completing his graduate work, John has been engaged in a number of consulting efforts with projects involving employee selection, assessment centers, training, and individual development and coaching. These efforts have included working with numerous organizations such as Little Debbie Snackcakes, Ruby Tuesdays, Inc., Goody's Family Clothing, Inc., and Advanced Management, Inc. He has also conducted research in the areas of employee selection, utility analysis, adverse impact, and biographical information, presented his studies at major conferences, and published research in the Journal of Applied Psychology.