

**A TEST OF A MULTILEVEL MODEL OF PERSONNEL SELECTION  
IN A CUSTOMER SERVICE ORGANIZATION**

A Dissertation

by

MARY KATHLEEN SHEEHAN

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Psychology

**A TEST OF A MULTILEVEL MODEL OF PERSONNEL SELECTION  
IN A CUSTOMER SERVICE ORGANIZATION**

A Dissertation

by

MARY KATHLEEN SHEEHAN

Submitted to Texas A&M University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

---

Winfred Arthur, Jr.  
(Chair of Committee)

---

Adrienne Colella  
(Member)

---

Stephanie C. Payne  
(Member)

---

Steve Rholes  
(Head of Department)

---

Ludy T. Benjamin  
(Member)

December 2004

Major Subject: Psychology

## ABSTRACT

A Test of a Multilevel Model of Personnel Selection  
in a Customer Service Organization. (December 2004)

Mary Kathleen Sheehan, B.A., Purdue University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Winfred Arthur, Jr.

The objective of the current study was to provide an initial empirical test of the Schneider, Smith, and Sipe (2000) multilevel model of personnel selection. The Schneider et al. (2000) model expanded the traditional approach to validating selection systems to include the impact that selection systems have on the broader organizational system. The current project provided an empirical test of this model by extending the traditional individual-differences approach to validation research and including group- and organization-criteria (e.g., unit-level performance and customer satisfaction). Using a quasi-experimental design, archival data from a managerial development and selection program were analyzed to examine several relationships proposed in the Schneider et al. (2000) model.

The current study provided limited support for the Schneider et al. (2000) model. There were several limitations in the current study associated with the use of archival data, but the current study provides an initial indication of practical problems associated with empirically testing the model. While intuitively appealing, testing the

Schneider et al. model in applied settings may prove to be a practical challenge because of the nature and complexity of the data required to do so. Although the current study provided limited support for the model, there were some interesting findings that warranted additional examination. Findings from the current study may be informative for both researchers and practitioners. Ideas for future research related to the Schneider et al. (2000) multilevel model of personnel selection are also offered.

## **ACKNOWLEDGMENTS**

As I finish my dissertation and reflect back on my academic journey, there are scores of individuals who were instrumental in this achievement. Rather than run the risk of inadvertently omitting names, I have chosen to offer more general thoughts of gratitude to those who have supported me in achieving this personal goal. I will remain ever inspired by my professors and academic colleagues who impressed me with their intellectual capacity and I am grateful for their feedback, which allowed me to develop my research skills. I am also indebted to the professional colleagues who taught me how to translate my academic knowledge into logic that would resonate with business leaders. And last, but most certainly not least, I am eternally grateful for family and friends, who provided ongoing support and cheering along the way. I relied on the support and good will of friends and family on many occasions while I was working on my dissertation.

## TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
ACKNOWLEDGMENTS.....	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES.....	viii
LIST OF TABLES.....	ix
INTRODUCTION.....	1
Validation Research.....	5
Unitarian Model of Validity.....	7
Evidence for Demonstrating Validity.....	11
Predictor Measures.....	16
Criterion Measures.....	28
Utility Analysis.....	32
Shortcomings of the Traditional Approach to Selection Validation Research.....	38
Multilevel Model of Personnel Selection.....	45
Summary of Hypotheses and Research Questions.....	54
METHOD.....	57
Participants.....	57
Predictor Measures.....	58
Criterion Measures.....	60
Procedure.....	64
Statistical Analyses.....	67
RESULTS.....	70
Hypothesis 1. The Relationship Between the Predictors and Criterion Measures Will be Statistically Significant.....	70
Research Question 1. Do Any of the Predictors in the Current Study Display Adverse Impact? .....	87
Hypothesis 2. Managers Selected with the Predictors in the Current Study Will Have More Satisfied Employees.....	92
Hypothesis 3. Managers Selected Using the Predictors in the Current	

	Page
Study Will Outperform Managers Selected Through Alternative Selection Systems.....	94
Research Question 2. Do Managers Selected Using the Predictors in the Current Study Tend to Have More Satisfied Customers (e.g., Higher Team-Based Customer Satisfaction Scores) Than the Comparison Group? .....	95
Hypothesis 4. Organizations that have a Higher Proportion of Managers Selected with the Predictors in the Current Study will have Higher Levels of Performance than Organizations with Lower Proportions of Managers Selected with the Predictors in the Current Study.....	96
Research Question 3. Using Customer Satisfaction as an Index of Organizational Performance, Does Customer Satisfaction Increase over time as Managers are Selected Using the Predictors in the Current Study? .....	101
Research Question 4. Do Differences in Criterion-Related Validity Exist Between Organizations? .....	103
DISCUSSION AND CONCLUSIONS.....	115
Discussion of Hypotheses and Research Questions.....	116
Research and Practical Implications.....	125
Limitations.....	128
Future Research.....	131
Conclusions.....	134
REFERENCES.....	135
VITA.....	149

**LIST OF FIGURES**

FIGURE		Page
1	Binning and Barrett (1989) Unitarian Selection Validation Model.....	8
2	Schneider et al. (2000) Multilevel Model of Selection Validation.....	45



## LIST OF TABLES

TABLE	Page
1	Description of Skills Assessed in General Management Ability Battery (GMAB)..... 58
2	Description of the Information Processing Skills Assessed in the In-Basket Exercise..... 59
3	Description of Competencies Assessed in Structured Interview..... 60
4	Matrix of Business Results (‘What’) and Leadership Competency (‘How’) Ratings of Performance..... 61
5	Description of Job Performance Dimensions Included in Managerial Performance Appraisal Process..... 62
6	Criterion Measures and Sample Size for Each Measure..... 64
7	Total GMAB Converted Scores, GMAB Qualification Zones and Points Awarded for GMAB Score..... 66
8	Cutoff Scores for AMAP Measures..... 67
9	Constructs and Methods for Predictor Measures..... 68
10	Summary of Hypotheses, Research Questions, and Statistical Analyses in Current Study..... 69
11	Means, Standard Deviations, and Correlations Between all Predictor and Criterion Measures for Overall Sample..... 71
12	Regression Analysis for Merit Increases and Bonus Payments..... 85
13	Regression Analysis for Hold Time..... 86
14	Descriptive Statistics of Predictor Measures Across Sex Groups..... 87
15	Descriptive Statistics of Predictor Measures Across Race Groups..... 89

TABLE	Page
16 Differences in Predictor Scores for Race Groups as Compared to Caucasians.....	90
17 AMAP Pass Rates for all AMAP Participants, Across Sex Groups and Across Race Groups.....	91
18 Comparison of AMAP Participants and Grandparented Managers on Criterion Measures.....	92
19 Means, SD and N for Criterion Measures Across Channels/Organizations.....	97
20 Summary of Effect Sizes of Differences in Criterion Measures Across Organizations as Compared to Global Sales and Service Organization.....	100
21 Descriptive Statistics for Overall Customer Satisfaction Scores in Repeated Measures Analysis.....	102
22 Predictor-Criterion Correlations for Call Servicing, Customer Sales and Service, Global Sales and Service, and Telemarketing Organizations.....	104
23 Comparisons of Predictor-Criterion Correlations Between Organizations.....	110

## INTRODUCTION

Demonstrating the validity of selection systems has long been a cornerstone of the field of industrial/organizational (I/O) psychology. As industries grow increasingly more competitive, human resources professionals and practitioners are often charged with the task of justifying the value of all human resource management (HRM) practices, including selection systems. The primary goal of selection is to determine an optimal match between the skills and talents of candidates and the requirements of the work, with the majority of selection validation research traditionally focusing on the capabilities of selection tools to predict future job performance of the individual employee. The current project provides an empirical test of a model that expands the traditional approach to validating selection systems to examine the impact that selection systems have on the organizational system. In particular, this project extends the traditional individual-differences approach to validation research by including both group-level performance (operationalized as an aggregate of subordinate performance) and organization-level criteria (e.g., customer satisfaction). By using a more comprehensive collection of criteria, it becomes possible to assess both how effectively managers are performing on the job as well as how the selection system impacts the larger organization.

---

This dissertation follows the style and format of *Journal of Applied Psychology*.

Research on selection processes has typically focused on predicting individual differences in job performance based on the knowledge, skills, and abilities (KSAs) that are assessed during the selection process. This project makes a unique and substantial contribution to existing literature in two important ways. First, this project incorporates a more comprehensive set of criteria beyond the criterion measures typically examined in the context of validation research (e.g., individual job performance, turnover). Second, the current project examines the impact of selection systems on individual, group, and organization level criterion measures.

Many researchers regard the future of the science and practice of personnel selection optimistically (e.g., Borman, Hanson & Hedge, 1997). Meta-analytic evidence suggests that many predictor measures are encouragingly valid for predicting job performance (e.g., Schmidt & Hunter, 1998). Also, the methods available for studying predictor-criterion relationships are increasingly powerful. Personnel selection research has the possibility to play an even more important role in the future as public and private sector organizations continually reconfigure and reposition themselves to remain competitive in the expanding global environment. In Latham's 1999 presentation to the Society for Industrial and Organizational Psychology (SIOP), he reminded the society of the interplay between science and practice in the field. Journals inform the practice of psychology and the practice of I/O psychology can and does advance science (Latham, 2001). The interdependence of science and practice is essential to the future of the field, for practice leads to refinement of theory, which in turn provides improvement in guidelines for practice. Dunnette (2001) noted that

Careful observation in practice often leads to meaningful questions, which may be addressed systematically by carefully designed studies. The results of these studies can lead to significant improvements in practice. The desired outcome of research is important generalizations about the results of an organizational intervention, or predictor measure in personnel selection research. Inherent in the scientist-practitioner model in I/O psychology is the ability to make connections between generic principles and specific application (Dunnette).

Selection research has enjoyed a rich history within the field of I/O psychology, as decades of basic and applied research have been devoted to understanding and improving personnel selection strategies for organizations. The selection of qualified individuals for employment plays an important role in organizations. Hiring the most qualified person for the job while using the fewest resources is a common philosophy across a multitude of organizations. Latham (2001) also called for the need to continue to blur the distinction between the existing science and practice dichotomy.

Guion (1998a) outlined some of the fundamental assumptions that guide personnel selection. The first assumption of personnel selection is that each person has a given or inherent ability level that is relatively stable. Each person also has a habitual way of dealing with people and events in an environment. Second, people differ in any given ability. When these individual differences in abilities are relevant to a given job, they often form the basis for personnel selection decisions. The third assumption underlying personnel selection is that the relative differences in ability

between individuals remain relatively constant even after training or accrued professional experience. In general, ability in a specified domain may be enhanced by experience or training, but the relative rank order of a larger group of people will often not be changed substantially by training or experience. An exception to this assumption may occur when the criteria for training is achieving a mastery level of proficiency.

Fourth, different jobs require different knowledge and skill sets. Research from the World War II era indicated that different occupations called for different patterns of more specific abilities and selection was most effective when the needs/requirements of the job were matched closely with the knowledge and skills of the candidate (Guion, 1998a). The matching of job and person is one of the fundamental tenets of personnel selection, with the goal of maximizing the match between the needs of the job and skills of the candidate. Selection tools that do not optimally match the person's skills to the job can be costly to the organization, resulting in increased training needs and decreased levels of performance. A fifth fundamental assumption of the development of selection systems is that the abilities needed to perform a job can be measured. Researchers have devoted much time to developing tools for selection purposes, as well as trying to determine which selection techniques are most useful for a particular job and organization (Guion, 1998a).

### *Validation Research*

Research in personnel selection is relatively systematic (Guion, 1998a). The first step of the validation research process is to analyze both the needs of the organization and the specific job. Identifying the job and performance requirements establish the competencies to be included in the development of certain selection or assessment procedures (Binning & Barrett, 1989). The primary reason for job analysis is to gain an understanding of the nature and purpose of the job. The description of major work behaviors in a job serves as a foundation for determining what skills and abilities are most essential to perform effectively on the job. Then selection tools can be built to assess those competencies or personal characteristics. Analyzing both the organization and job can provide information on any expected changes in the organization in the near future that may impact the job (e.g., how quickly those changes may occur, potential problems of adapting to change).

Once the organizational and job analyses are complete, the second step is to select a criterion measure (or multiple criterion measures). The criterion choice often reflects organizational needs and values (e.g., performance appraisals are a way of communicating the values, expectations, and norms of the organization to employees). Criterion performance is complex, often attributable to multiple characteristics of the workers as well as to multiple organizational influences. For the purposes of prediction, it is likely that more than one kind of ability or personal attribute needs to be measured in the selection process if the criterion is to be predicted in all of its complexity. Even with assessing multiple attributes during selection, organizational

influences and constraints still complicate the prediction of criterion performance. Once the job and performance domains have been defined, the next step is to select the methods of measurement for assessing skills and abilities during the selection process. Historically, more research has been conducted for tests or questionnaires than for other methods, but research should not be limited to particular assessment methods. A research design strategy should be developed next. Quality research anticipates the expected manner of test use and is designed with that in mind; good administrative practice ensures that the operational use of the predictors is consistent with the conceptual goals of the research design. The next step in personnel selection research is the administration of the selection tool and collection of predictor and criterion data. Once the data have been collected, the research can be evaluated (i.e., it is time to validate the predictor trait/selection system as measured). The extent to which the quality of research and the quality of administration can be balanced contributes to the effectiveness of the selection system.

To summarize, the essential process of personnel selection involves (a) conducting a job analysis to determine the requirements of the given job; (b) identifying the performance domain, which is typically defined in terms of job behaviors or outcomes; (c) developing the selection tools; (d) collecting predictor and criterion data; and (e) validating the selection system. The validation of personnel selection processes evaluates the extent to which scores on the predictor measures reflect the meaning of the intended constructs or competencies as well as future job performance (Guion, 1998b).

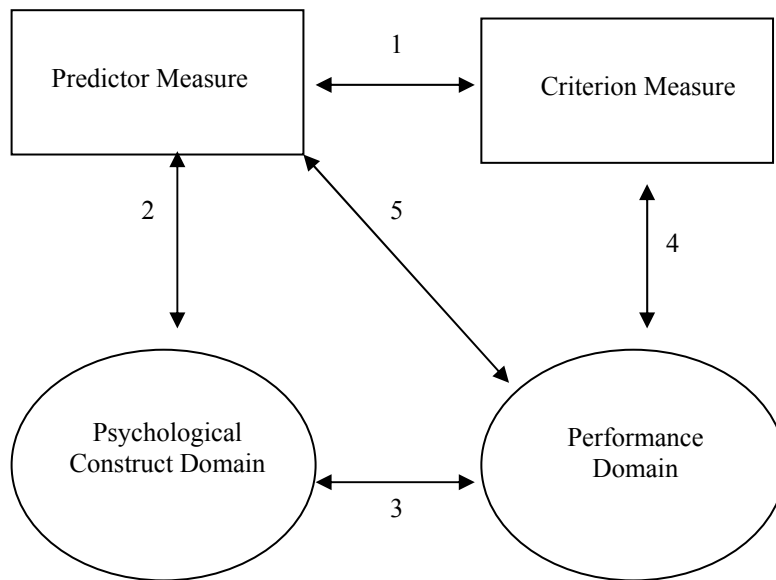


### *Unitarian Model of Validity*

Validity assesses the appropriateness of the inferences drawn from the data. Validity is often a key criterion in the evaluation of any selection tool or piece of research. Cronbach and Meehl's (1955) model, commonly referred to as the tritarian view of validity, advocated three separate types of validity – content validity, construct validity, and criterion validity. Content validity measured the extent to which a selection tool sampled the domain in which the researcher was interested. Construct validity focused on how well the selection tool assessed the concept or construct it was intended to measure. Criterion validity assessed the relationship between the predictor and criterion. Predictive validity assessed this relationship when the predictor data were collected before the criterion data, whereas concurrent validity assessed this relationship when the predictor and criterion data were collected at essentially the same time (Cronbach & Meehl, 1955).

Research in personnel selection evolved to a unitarian conceptualization of validity (e.g., Binning & Barrett, 1989; Messick, 1995, 1998). This shift in the approach to validating selection tools came with the realization that the three strategies of the tripartite model (e.g., content, criterion, construct) could not be logically separated (SIOP Principles, 2003). This new philosophy posited that although many ways of accumulating evidence for any particular inference exist, validity is generally a unitary concept. In the unitarian model, the validation process is analogous to standard hypothesis testing. Construct validity involves a judgment that a test or other

predictive measure does measure a specified construct or attribute to a significant degree, and that the predictor can promote the understanding or prediction of behavior (Messick, 1995). Binning and Barrett (1989) developed a model, which is an adaptation of Nunnally's (1978) psychometric model and illustrates the relationships among the five inferences presented in Figure 1.



*Figure 1.* Binning and Barrett (1989) unitarian selection validation model.

The essential elements of Figure 1 are as follows: (1) predictor measurements are related to criterion measurements; (2) the predictor measure is an adequate sample from a psychological construct domain; (3) the predictor construct domain overlaps

with the performance domain; (4) the criterion measure is an adequate sample from the performance domain; and (5) the predictor measure is related to the performance domain.

There are some important distinctions between Nunnally's (1978) general model of validation and Binning and Barrett's (1989) model for personnel selection validation research. First, an additional measure-construct link (inference 5 listed above) has been added to Binning and Barrett's (1989) model, illustrating that the predictor measure is related to the performance domain. Also, all links are of equal importance in Nunnally's (1978) model, whereas the additional link (inference 5) in Binning and Barrett's (1989) model takes on relative importance. In sum, this additional measure-construct link is a "dual" linkage. This inference has a theoretical meaning, as it represents what the test measures and how well it does so. At another level, which is more operational in nature, the relationship between the predictor measure and the performance domain shows the extent to which the content of the predictor reflects the content of the job. This operational interpretation of inference 5 is referred to as the content-related validity of a predictor, which will be discussed in more detail later.

As previously mentioned, the unitarian model of validity proposed by Binning and Barrett (1989) was a departure from the tritarian view. The tritarian view of validity viewed construct, content, and criterion validity to be different and distinct types of validity. In recent years, validity has increasingly been discussed as a unified construct rather than as composed of various categories (Borman et al., 1997). The

unitarian approach views construct validity to be a large umbrella for assessing what the predictor was measuring and how well it does so, with three strategies for demonstrating the construct validity of a predictor. Construct-related, content-related, and criterion-related validity are three strategies for demonstrating the validity of the predictor. An inference drawn from current available information about some aspect of future job performance (inference 5) is the single overriding inference in the unitarian model, with content-, construct-, and criterion-related validity all being quite relevant justifications for the validity of a selection tool. As such, these 3 concepts are more appropriately viewed as 3 evidential bases from which inferences about future job performance can be supported or justified. Regarding the usefulness of the unitarian model in decision situations, an applied decision maker is concerned about the extent to which test or assessment information will allow accurate predictions about subsequent job performance, which is the overriding theme of the unitarian model.

As mentioned, construct validity has three facets in Binning and Barrett's (1989) unitarian model of validity. Construct validity in its most generic sense refers to construct-construct links (link 3), construct-measure links (links 2, 4) and measure-measure links (1), as in Nunnally's (1978) model. The construct-measure links have traditionally been of particular concern to research psychologists and psychometricians, as construct validity often refers to whether a given test or measurement procedure resulted in accurate inferences about an individual's standing on a psychological construct of particular interest (Binning & Barrett, 1989).

Construct validity is primarily concerned with what a test or assessment procedure is measuring and how well it is doing so, which is the conceptual meaning behind inference 5. The three primary means of demonstrating the construct validity of a selection tool are described in more detail below.

#### *Evidence for Demonstrating Validity*

Content-related validity demonstrates the representativeness of the predictor, or extent to which the performance domain is sampled by the predictor and is represented as inference 5 in Binning and Barrett's (1989) model. The content-related validity of a predictor is a defensible rationale for the validity of a selection tool, given that predictor tests are samples of behavior from which scientists and practitioners can infer something about behavior on the job. This means of providing evidence of the validity of the predictor measure is the most widely used, and also plays an important role in the legal defensibility of the tool. Selection specialists are often called into court to explain the job relevance of a predictor; when they can rationally defend their strategy for sampling the performance domain in a given testing situation, content-related validity evidence supports inferences that scores from test are valid for predicting future performance.

There are three assumptions of content-related validity. The first assumption is that there exists a meaningful and definable universe of response. The second assumption is that the predictor (i.e., the sample of performance) can be drawn from the universe in some purposeful and meaningful fashion. The third assumption is that both the predictor and the sampling process can be defined with sufficient precision to

enable the user to judge how adequately the sample of performance exemplifies performance in the universe (Cascio, 1991).

Some research has focused on an empirical estimate of content-related validity, known as the Content Validity Ratio (CVR) (e.g., Lawshe, 1975). A CVR can be generated for each item in a predictor, based on a panel's evaluation of how essential each item is to the performance of the job. Items with CVRs that fail to meet predetermined or rational cutoffs are eliminated from the measure. Based on the mean of the retained items, the Content Validity Index (CVI) is computed. The CVI represents the extent to which perceived overlap exists between capability to function in a job performance domain and performance on the measure under investigation. However, research by Carrier, Dalessio, and Brown (1990) caution that using CVR to demonstrate the content-related validity of a selection tool may not be useful for developing empirically valid total scores in all selection situations.

Research on the content-related validity of selection tools has resulted in improved domain sampling and job analysis procedures, better behavioral measurement, as well as the importance of the role of expert judgment for confirming the appropriateness of sampling and scoring procedures and for determining the degree of overlap between content domains that have been derived separately (Cascio, 1991).

Construct-related validity is represented through inferences 2 and 3 in Binning and Barrett's model. Establishing the construct-related validity of a selection tool involves showing that the tool measures a specific construct (inference 2) that has

been determined to be critical for job performance (inference 3), thereby justifying inferences about job performance from the test scores (inference 5). Examining the process in more detail, evidence of inference 2 primarily takes the form of empirically-based relationships and judgments that are both convergent and discriminant. A multrait-multimethod matrix can be used to compare a predictor measure of interest with other measures that use different methods to capture similar constructs as well as other measures that use the same method to capture different constructs (Campbell & Fiske, 1959). Predictor measures that demonstrate construct-related validity are those measures that demonstrate convergent validity by being closely related to other measures of similar constructs as well as divergent validity by not being closely related to measures of different constructs. Evidence for inference 3 can be demonstrated by job analysis and is often justified theoretically and logically on the basis of accumulated knowledge of interrelationships between different constructs (psychological construct of interest and job performance domain in the selection context) (Binning & Barrett, 1989). However, it is important to be mindful of the fact that the psychological construct of interest and the performance domain do not share common nomological status; thus, they cannot be expected to overlap totally. To the extent that inferences 2 and 3 are supported, the use of the predictor measure to predict job performance is based on construct-related validity evidence.

Construct-related validity can be developed in concert with content-related validity. A job description, based on a thorough job analysis, includes the major work behaviors of a job. The major work behaviors can be assessed to determine the

needed KSAs to perform those activities. These KSAs often determine the constructs that are used in selection tools. The distinction between a content-related strategy and construct-related strategy is often a matter of degree, as constructs underlie all psychological measurement (Cascio, 1991).

Criterion-related validity demonstrates the effectiveness of a selection tool in predicting an individual's behavior or performance in specific situations. This strategy consists of generating direct empirical evidence that the predictor scores are related to valid measurements of job performance, which often take the form of performance appraisal ratings. A significant relationship between the predictor and criterion measures indicates that higher performance on the predictor is associated with higher performance on the job. However, the use of criterion-related validity may erroneously imply an unnecessary restriction that only correlational evidence is appropriate. The empirical linking of the predictor to a criterion measure is important, but it is only partial justification for validity of the predictor measure. Again, to have more complete confidence in the validity of inference 5, both inferences 1 and 4 must be justified. As effect sizes are a major determinant of statistical power, more criterion-related validity studies may become technically feasible if researchers base their sample size requirements on unit-weighted linear composites rather than on individual predictors. Range restriction can affect the magnitude of criterion-related coefficients, given that the size of this type of coefficient is a function of two variables. Thus, a narrowing of the range on the predictor or criterion will serve to lower the size of the resulting criterion-related validity coefficient.



The criterion-related validity of a predictor measure can be assessed through concurrent, predictive, or postdictive research designs, although postdictive research designs are rarely used in typical selection validation studies. In practice, predictive and concurrent designs are virtually the same regarding the data collected. The primary distinction between these validation study designs is the time lag between collecting the predictor and criterion data. The predictor and criterion data are collected at the same time in a concurrent design, whereas the predictor data are collected before the criterion data in a predictive design. Incumbents are typically the participants in concurrent designs. There are mixed research findings regarding which design is preferable. Cascio (1991) notes that “predictive designs for obtaining evidence of criterion-related validity are the cornerstone of individual differences measurement (p.155)”. For tests of general cognitive ability, estimates of validity obtained from predictive or concurrent validation studies can be expected to be comparable (Barrett, Philips, & Alexander, 1981; Pearlman, Schmidt, & Hunter, 1980). The SIOP Principles (2003) note, on the other hand, that this finding cannot automatically be generalized to all situations and to other types of predictors and criteria. For example, concurrent validities of paper-and-pencil measures of personality traits may not provide accurate estimates of predictive validity, as they may not take motivational or social desirability factors into account. Guion (1998a) notes that research designs utilizing incumbents have the benefit of including criterion data; however, if performance on the selection tool is not a true reflection of the skills and abilities that an incumbent may bring to the job, then examining the empirical

relationship between the selection system and performance data is meaningless.

Barrett et al. (1981) rebut these arguments by pointing out that although a few studies have found differences in motivation between predictive and concurrent designs, there is no empirical evidence that demonstrates whether differential motivation affects our ability to assess the validity of a predictor battery. Due to the minimal differences between concurrent and predictive validation studies, Barrett et al. (1981) have concluded that the effects of motivation must be minimal or relatively constant across individuals.

### *Predictor Measures*

As noted by Schmidt and Hunter (1998), 85 years of research have been dedicated to the development and validation of various selection techniques and procedures. A summary of the predictor measures used most commonly in personnel selection research is provided below. Traditional predictors tend to be standardized measures that have relatively fixed rules for scoring (e.g., standard ability tests). Alternative predictors, on the other hand, may have some fluctuations in scoring procedures (i.e., no two assessment centers will be identical).

*Cognitive predictors.* Cognitive ability and its relationship to job performance has been the subject of more personnel selection research than any other personal attribute or characteristic (Cascio, 1998; Guion, 1998a). The general term of cognitive ability can refer to a range of abilities to perceive, process, evaluate, compare, create, understand, manipulate, or generally think about information and ideas. Common work-relevant cognitive activities include reading verbal or graphic materials,

understanding the principles that make things work, planning events or procedures, solving problems, or perceiving signs of trouble in equipment or in human interactions or in contradictions in plans (Guion, 1998a). Cognitive ability is also often referred to as general mental ability, which is a general trait that is traditionally referred to as intelligence.

Different taxonomies of cognitive ability have been developed. Research has examined both general categories of cognitive ability, such as Thurstone's (1938) taxonomy, which includes verbal comprehension, word fluency, spatial ability, perceptual speed, numerical facility, memory, and inductive reasoning, as well as more specific abilities (e.g., verbal comprehension, fluency, perceptual speed, flexibility and speed of closure, spatial orientation and visualization, number facility, general reasoning, problem recognition, associative memory) (Guion, 1998a).

Sternberg (1977) developed a model that suggested that there are facets of intelligence beyond academic intelligence (e.g., practical intelligence, creative intelligence).

Given the increased competitiveness and need for innovation in many organizations, it may be useful to explore the validity of these types of intelligence in the future.

The volume of research on the role of cognitive ability in personnel selection has provided a stronger theoretical foundation for general cognitive ability than any other KSA used in personnel selection (Schmidt & Hunter, 1998). General cognitive ability has been found to be a powerful predictor of performance across a wide range of jobs, with validity coefficients often ranging from 0.43 to 0.51 (Hunter & Hunter, 1984; Schmidt & Hunter, 1998), and as high as 0.58 for professional and managerial

positions (Hunter & Hunter, 1984). The thousands of studies on cognitive ability have also provided empirical evidence that general cognitive ability is the most valid predictor of future job performance for hiring employees without previous experience in the job (Hunter & Hunter, 1984; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Ree, Earles, & Teachout, 1994; Schmidt & Hunter, 1998; Schmitt, Gooding, Noe, & Kirsch, 1984). Measures of general cognitive ability are also very cost effective for organizations.

Recent studies have demonstrated that psychometric general cognitive ability, or the common variance in a battery of cognitive ability measures, accounts for the majority of predictive power in the test battery. The remaining variance, which is often referred to as ‘specific abilities’, accounts for little or no additional variance in the criterion (e.g., Olea & Ree, 1994; Ree et al., 1994). Schmidt (1994) noted that “refinements in measurement of abilities and aptitudes are unlikely to contribute non-trivial increments to validity beyond that which is produced by good measures of general mental ability” (p. 348). However, specific abilities substantially improve the classification efficiency beyond the use of general cognitive ability alone (e.g., Scholarios, Johnson & Zeidner, 1994; Zeidner & Johnson, 1994).

*Noncognitive predictors.* In addition to cognitive ability, other personal attributes and characteristics have been investigated, including a variety of personality constructs (e.g., conscientiousness, extraversion, agreeableness, dominance, creativity, problem-solving ability, organization and planning abilities, and the ability to learn from experience). Although response distortion does not appear to have a major

impact on the validity of personality inventories in a selection context (e.g., Barrick & Mount, 1996), it is still of some concern because these measures are susceptible to social desirability. There are situational moderators to the relationship between personality inventories and performance measures. The personality-performance correlations for jobs with high levels of autonomy have lower situational strength, thus personality predicts performance better for these types of positions (Borman et al., 1997). These personal characteristics may be more relevant for some jobs rather than other jobs; hence these attributes have not been studied to the extent of general cognitive ability. Job specific knowledge and skill have also been shown to be related to job performance, so they are often assessed in the selection process. Physical and sensory competencies, psychomotor abilities, experience, and education are often used during selection as well.

*Person-organization fit.* More situationally-oriented variables, such as person-job fit, person-organization fit, and predictors for selection into teams have also been researched for their usefulness in predicting job performance. Kristof (1996) noted some confusion in the literature over the concept of person-organization fit, as person-organization fit was sometimes considered synonymous with person-environment fit. Person-environment is a broader concept than person-organization fit, encompassing person-vocation fit and person-job fit.

There are two basic assumptions of person-organization fit. First, human behavior in the workplace is a function of both the person and the organization. Second, the person and organization need to be compatible. Schneider (2001) noted

that the person and the organization are often confounded entities, in that the organizational environment is created by the people in it. Also, the degree of fit between a person and the organization is dynamic and flexible, as people adapt to the organizational environment, and also change the organizational environment (Furnham, 2001). Person-organization fit researchers rely on the classification of persons and/or organizations in categories of personality, values, and needs. However, very little is known about the specific characteristics of people and organizations that are crucial for establishing fit. No universal taxonomy has been identified as the most relevant source for operationalizing components of the person-organization fit index. Van Veenen (2001) noted reasons why fit measures often account for only a small portion of the variance in outcome measures, including neglect of the value of fit, unequal variance in the component measures of person-organization fit, and the across-person approach of person-organization fit research.

Schneider (2001) observed that two traditions have dominated research on person-organization fit: the individual differences tradition and the organizational psychology tradition. There is no reason to assume that one approach to conceptualizing fit is superior to other approaches, given certain questions, just as there is no reason to assert that operationalizing fit in one particular way is the key to measurement problems. In the individual differences tradition, personnel selection, vocational interests, and personality have been three major models for conceptualizing person-organization fit. The dominant model of personnel selection has been an implicit theory of fit. Selection research is the exception in person-organization fit,

with its focus on performance effectiveness as the outcome. Fit itself is not assessed; rather the individual differences of candidates are assessed and related directly to the criterion or criteria of interest. Individual differences predictors are often assumed to be the primary effects on individual differences in performance. Given that job attributes are constant, personal attributes become the data of interest when assessing fit. The traditional personnel selection model has essentially ignored implications of environment as a potential contributor to performance.

In the vocational interests model, research by Holland (1985) notes that people's interests yield job satisfaction and adjustment when those interests fit the environment in which the person works. The environment is indexed by the presence of people in that environment with particular vocational interests; the more people in the environment share the interests of the person joining the environment, the more likely the person is to experience satisfaction and adjustment. Holland (1985) and Schneider (1987) are among the few researchers to define the organizational environment of interest in terms of the people in it. Like most research done in the person-environment fit tradition, Holland (1985) focuses on individual affective outcomes of fit, such as job satisfaction and commitment, rather than performance or effectiveness (Van Veenen, 2001).

Schneider (1987) introduced the Attraction-Selection-Attrition (ASA) model as an alternative means for understanding organizations and the reasons behind the structures, processes, and technology of organizations. The goals of the organization yield structures and processes, which often determine the types of people who are

attracted to, are selected by, and stay with a particular organization. Over time, organizations may develop strong norms, as a result of attracting, selecting, and retaining people with high levels of fit with the goals and values of the organization, making the organization more susceptible to groupthink and inflexibility. This highlights the importance of recruiting and selection processes to bring a diverse set of candidates to an organization for the long-term success of the organization.

The personality model finds its foundation in interactional psychology, which posits that people and situations in interaction yield individual behaviors of interest (e.g., Sells, 1963). In general, the individual differences tradition has focused primarily on person variables, with the environment typically conceptualized as a moderator of person-outcome relationships. The focus centered on individual-level outcomes, such as performance, attitudes, or experienced stress. Research that indexed fit, such as vocational psychology and interactional psychology, focused on non-performance outcomes.

From the perspective of research in organizational psychology, Lewin (1935, p. 73) claimed behavior is a function of the person and the environment: “From a certain total constellation – comprising a situation and an individual – there results a certain behavior.” When the person and environment are disaggregated, then the concept of constellation is somehow violated. Schneider (2001) notes that the interaction goes both ways. The organizational psychology tradition studies individual attributes as a moderator of organizational functioning, whereas the individual differences tradition studies the environment or organization as a moderator of



individual behavior. There are both positive and negative consequences for the environment as the degree of fit between the individual person and the environment increases. Positive consequences of increased levels of fit are increased harmony, cooperation, high levels of morale, climate for well-being, and low levels of turnover and absenteeism. Negative consequences include the inability to adapt to larger environmental turbulence, yielding ineffectiveness for the organization in the long run.

As person-organization fit has been considered appropriate for selection in recent years (e.g., Adams, Elacqua & Collarely, 1994), it is subject to the same psychometrics and legal standards expected of other selection tools. A recent meta-analysis of person-organization fit as a predictor found that person-organization fit was a relatively weak predictor of job performance ( $r = .15$ ) and turnover ( $r = .24$ ), although person-organization fit did have a stronger relationship with job attitudes ( $r = .31$ ) (Arthur, Bell, Villado, & Doverspike, 2004). These results suggest that researchers and practitioners may need to refine selection tools that assess person-organization fit before using them widely to select job candidates.

*Interviews.* A variety of methods can be employed to assess job-relevant KSA or constructs in a selection process. Interviews are probably the most widely used selection tool. Although there is weak empirical support for unstructured interviews, research has demonstrated criterion-related validity evidence for structured interviews. Structured interviews generally attempt to standardize the interaction between the interviewer and interviewee, often by including a fixed series of questions that are based on a job analysis and presented to all applicants in the same order. Structured

interviews also usually have a standardized scoring process, and interviewers often provide construct-level scores as well as an overall evaluation for each applicant. Interviewer training may also precede the interviews to ensure interviewers share a similar frame-of-reference and also to minimize interviewer effects. Empirical research has shown that structured interviews can reach validity levels comparable with general cognitive ability, with validity coefficients ranging from 0.35 to 0.51 (Huffcutt & Arthur, 1994; Huffcutt, Roth, & McDaniel, 1996; McDaniel, Whetzel, Schmidt, & Mauer, 1997). Huffcutt and Arthur (1994) found that interview validities increased as structure increased, with the top 2 levels of interview structure displaying validity coefficients comparable to validities found for ability tests in Hunter and Hunter (1984). Interrater reliability (IRR) levels are also higher when interviews incorporate multiple ratings, interviewer training, and standardization of questions and response evaluation (Conway, Jako, & Goodman, 1995).

Research by Huffcutt, Conway, Roth and Stone (2001) indicates that structured and unstructured interviews in the extant literature generally do not measure the same constructs. Further, Huffcutt et al. posit that the differences in criterion-related validity between structured and unstructured interviews are not entirely due to the interview procedure itself. In fact, structured interviews tend to focus on constructs that have a stronger relationship with job performance (e.g., thinking and analytical skills, interpersonal skills).

Campion, Campion, and Hudson (1994) compared the validities of past- and future-oriented questions in employment interviews, finding that past-oriented

questions demonstrated higher levels of validity (0.51) than future-oriented questions (0.39). The past-oriented questions also demonstrated incremental validity over future-oriented questions, but the future-oriented questions did not demonstrate incremental validity over past-oriented questions. In similar research, Pulakos and Schmitt (1995) demonstrated that only past behavior questions in structured interviews were valid in predicting supervisor ratings.

Research on the combined validity of structured interviews and general cognitive ability measures has yielded mixed results. Schmidt and Hunter (1998) provided evidence that an equally weighted combination of a structured interview and an assessment of general cognitive ability has a validity coefficient of 0.63, suggesting that utilizing the combination of a structured interview and a measure of general cognitive ability can be an effective selection strategy. Cortina, Goldstein, Payne, Davison, and Gilliland (2000) suggested that interview scores incrementally contribute to the prediction of job performance beyond cognitive ability and conscientiousness to the extent that they are structured, with scores from highly structure interviews contributing substantially to the prediction of job performance.

However, Huffcutt et al. (2001) suggested that the validity of the structured interview may vary by construct. As previously mentioned, structured interviews may demonstrate higher levels of criterion-related validity in part because they often focus on constructs and dimensions that have a stronger relationship with job performance. Additional research by Huffcutt et al. (1996) suggested that structured interviews may reflect general cognitive ability, with an average correlation of 0.32 between interview

scores and scores of general cognitive ability. Campion et al. (1994) found that structured interviews correlated 0.60 with a battery of cognitive ability tests, but the interview did show incremental validity when regressed against job performance ratings. Latham and Skarlicki (1995) found that situational and behavior description interviews were resistant to same-race bias (e.g., in-group favoritism, out-group discrimination) whereas conventional interviews were not.

Herriott (1993) proposed that an interview's focus should reflect more dynamic interpersonal processes than the current prevalent psychometric perspective among academicians. By focusing on interpersonal process, the employment interview becomes more practitioner-focused and therefore more useful to organizations. Additional research has suggested shifting the importance of employment interviews away from the "prediction of performance" view that is prevalent in selection research and highlighting the usefulness of interviews for assessing person-organization fit (Adams et al., 1994). Howard and Ferris (1996) have suggested that interviewer training that considers the context of the organization might help interviewers better gauge whether applicants will be successful in the organization. Latham and Skarlicki (1995) investigated the criterion-related validity of situational and patterned interviews using organizational citizenship behaviors as the criterion. The situational interview predicted organizational citizenship behaviors, which suggests that extra-role behavior can be predicted by certain interview formats.

Dipboye (1994) reviewed the employment interview literature, citing many reasons why structured interviews show greater validities than unstructured interviews

as well as some reasons why unstructured interviews continue to be used in organizations. Dipboye (1994) noted that recruitment as well as selection is often a concern to organizations. Also, more interviewer autonomy and self-expression is possible with unstructured interviews, which is often appealing to interviewers. Dipboye (1994) also noted that the chances for good fit between job applicants and the context of a job are improved with unstructured interviews.

*In-baskets.* In-baskets are also often used in selection for managerial positions. In-baskets typically include a set of inter-related memos, appointments, and other situations applicants must consider during the exercise. In-baskets can predict job performance for managers and executives, often a difficult group of employees to select. Research on in-baskets is mixed in terms of how valid these selection tools may be in predicting job performance. Schippman, Prien, and Katz (1990) reported that the typical criterion-related validity of in-basket exercises is approximately 0.25. However, when Brannick, Michaels, and Baker (1989) adopted a construct validation approach to in-baskets, they questioned whether in-baskets were valid indicators of managerial ability. Brannick et al. (1989) found that exercise factors rather than dimension factors when they assessed the construct-related validity of in-basket scores, and suggest that behavioral consistency may be more important than previously thought by researchers. Similar to assessment centers, the same behavior is often used to infer multiple attributes when evaluating in-baskets. Brannick et al. also suggested that training or coaching may improve overall performance on an in-basket exercise.

Guion (1998a) noted that research often follows success, and the predictive value of standardized tests and questionnaires (e.g., cognitive ability tests, Big 5 personality inventories) has been demonstrated more persuasively and more frequently than for competing approaches to assessment. This success may have limited personnel selection research, rather than encouraging innovation in the development of a vast array of alternative selection tools or the exploration of additional validation strategies. Also, testing is easily standardized, enabling a more fair assessment than is possible when the method of assessment varies from one person to another. However, one substantial unintended negative consequence of the heavy reliance on tests and questionnaires in selection systems is that there is a great tendency to assess candidates on traits for which tests are available rather than to assess other attributes and characteristics that may be more relevant for a job but that may not be so easily assessed by standard testing procedures.

### *Criterion Measures*

Criteria are operational statements of goals or desired outcomes. In general terms, job performance is observable behaviors that people do that are relevant to the goals of the organization (Campbell, McHenry, & Wise, 1990). The term ultimate criterion refers the full domain of performance, including everything that ultimately defines success on the job (Thorndike, 1949). Although the ultimate criterion is strictly conceptual, clear understanding and documentation of the ultimate criterion for a job will aid in operationalizing the ultimate criterion into a relevant performance criterion measure (Astin, 1964).

Criterion measures can take on a variety of forms, although performance appraisal ratings are most commonly used in validation efforts as a measure of the job performance domain. The performance appraisal ratings should, at the very least, be aligned with the goals and objectives of the organization. However, scientific rigor is not often given to the development of criterion measures, particularly compared to the attention paid to the development of predictor measures. For criterion-related evidence to be a compelling argument for inference 5 of Binning and Barrett's (1989) model, strong evidence of both inferences 1 and 4 is required. Criterion relevance is the principal requirement for any criterion, in that the operational criterion should reflect the conceptual or ultimate criterion. Criterion deficiency occurs when there are elements of the conceptual criterion that are omitted from the operational criterion. Criterion contamination occurs when the operational criterion includes variance that is unrelated to the conceptual criterion (Cascio, 1998).

Guion (1961) developed a criterion development process, which has some similarities to predictor development. The first step is an organizational needs analysis and job analysis, which will define expected behavior on the job. Based on the job and organizational needs analysis, measures of actual behavior relative to expected behavior can be developed. The next step in criterion development is to identify the criterion dimensions underlying the job behaviors that have been identified. These dimensions can be identified using statistical procedures such as factor analysis or cluster analysis. After the criterion dimensions have been identified, reliable criterion measures should be developed. Each criterion measure should have

high construct validity; construct-valid criterion measures are just as important as construct-valid predictors. The predictive validity of the predictors should also be assessed for each criterion measure.

A predictor measure will be no better than the criterion used to establish its validity. As with predictors, anything that introduces random error into a set of criterion scores will reduce the validity of that criterion measure. It is important to establish both the relevance and the validity of a criterion measure as it is developed. A relevant criterion measure represents behavior on the job that is valued by the organization. A valid criterion measure has a theoretical relationship to the construct it is measuring. It is possible to establish criterion-related validity of a criterion measure by correlating it with other measures of performance (i.e., correlating subjective ratings of performance with productivity or turnover data). Also, it is important that the criterion measure is reliable, particularly when using performance appraisal ratings as a criterion measure. If supervisors are inconsistent in evaluating their employees' performance, criterion-related validity will suffer. Criterion contamination and criterion deficiency can also seriously impact the criterion-related validity of the predictor.

Job performance models foster a more scientific understanding of criteria, particularly models that try to understand the central latent variables that best characterize all performance requirements in work (Borman et al., 1997). In particular, Campbell, McCloy, Oppler, and Sager (1993) developed a job performance model that included eight latent factors that summarize performance requirements for



all jobs, with Project A providing empirical support for at least part of the taxonomy. The criterion domain should be carefully mapped, just as various predictor areas have been. Further development of the criterion taxonomy will help organize accumulating research findings by addressing questions about the links between predictors and individual criteria rather than predictors and overall performance (Borman et al., 1997).

McCloy, Campbell, and Cudeck (1994) divided the criterion space into declarative knowledge, procedural knowledge and skill, and motivational components, and demonstrated that declarative knowledge is predicted primarily by cognitive ability, whereas motivational elements of performance are linked to personality.

Models that examine relationships between elements of performance aids learning more about the criterion space. Differentiating criterion constructs provides increased understanding of each construct as well as the relationships among them. Research by Hunter (1983) indicated that ability has primarily a direct effect on individuals' acquisition of job knowledge, which in turn, influences employees' technical proficiency. Motowidlo and VanScotter (1994), as well as Borman and Motowidlo (1993), provided evidence that contextual performance and task performance were weighted roughly equally by supervisors in making overall performance ratings. Campbell, Gasser, and Oswald (1996) note that it is important to move beyond studying overall performance ratings as dependent, endogenous variables to examining independent dimensions of performance from substantive models. Further research has demonstrated that when the criterion space is divided up

and the separate elements are correlated with predictor scores, consistent patterns of relationships emerge between the predictor and criterion. For example, general mental ability correlates higher with criterion measures that assess maximal performance than with typical measures (Schmitt & Chan, 1998). Also, research by Organ and Ryan (1995) has provided evidence of correlations between conscientiousness and organizational citizenship behaviors (OCBs; e.g., altruism).

Operational criterion measures must be relevant, reliable, sensitive, and practical. As noted by Austin and Villanova (1992), the “criterion problem” centers on the difficulties involved in the process of conceptualizing and measuring performance constructs that are multidimensional, dynamic, and appropriate for different purposes. Even today, practitioners and researchers often resort to the most expedient criteria to use in validating predictor measures. Barrett, Alexander, and Doverspike (1992) note that problems with the definition, operationalization, and measurement of criteria in predictive validity studies are all too frequent. Greater attention should be paid to the issue of adequate design of predictive studies, including the definition and development of both predictors and criteria.

#### *Utility Analysis*

The utility of personnel selection, or the value this human resources system adds to the larger organization, is another important consideration in the evaluation of a selection system. The utility of a selection system is the degree to which its use improves the quality of individuals selected beyond what would have occurred had that system not been used (e.g., Cascio, 1991). Similar to the relationship between the

reliability and validity of a selection tool, the validity of a selection tool is directly proportional to the utility, or practical value, of the tool (Schmidt & Hunter, 1998). Utility analysis often incorporates break-even values, which are values at which the selection system's benefits equal the costs.

The utility of selection systems can be described in a variety of ways; the psychometric utility of the system and the value of one standard deviation increase in employee performance measured in dollars ( $SD_y$ ) are two common methods of assessing the utility of selection systems. One common strategy for demonstrating the utility of a selection system is to take the benefits associated with the new selection system (e.g., higher job performance) and convert them into dollar metrics. Boudreau (1991) notes that  $SD_y$  is equivalent to 40% of an employee's salary, although other researchers (e.g., Becker & Huselid, 1992) suggest this estimate may be rather conservative.

There are several utility analysis models. The Taylor-Russell model compares the validity coefficient of the new selection procedure to that of employees who have already been screened using methods other than the new selection procedure. The approach is most appropriate when individuals are divided into two groups based on test battery scores, differences in ability beyond minimum requirements do not yield differences in benefits, and differences in output are believed to occur, but are presently unmeasurable (Cascio, 1991). A limitation of this approach is that the quality of the predictor is reflected only in terms of success ratio.

Another approach to utility analysis is the Naylor-Shine model, which assumes a linear relationship between validity and utility. According to this model, utility can be defined in terms of the increase in average criterion score to be expected from use of the new selection measure with the given validity coefficient and selection ratio. This approach is most appropriate when differences in criterion performance cannot be expressed in dollar terms, but those differences assume a linear relationship between the predictor and payoff.

The Brogden-Cronbach-Gleser model is another approach to utility analysis. This model assumes only that there is a linear relationship between test scores and job performance, such that higher test scores are related to higher job performance. There have been further developments to this model, including the ability to separate the costs associated with recruitment and selection (Martin & Raju, 1992). There have been three other significant developments in this model. First, at least four alternative methods for estimating  $SD_y$  have been developed. However, Greer and Cascio (1987) note that one is left with little basis for choosing one method over another in the absence of a meaningful external criterion. The second development has been the integration of selection utility model with capital budgeting models, given similarities between the models. Capital budgeting models have also highlighted limitations that may constrain the effectiveness of the Brogden-Cronbach-Gleser utility model for capturing the benefits of selection systems within the larger firm. These limitations include the following: (a) the model does not take into account the time value of money (e.g., the discount rate); (b) the model ignores the concept of risk; and (c) the

model ignores the impact of taxation on payoffs, which may provide overly optimistic estimates of the benefits of a selection system.

Another development of the Brodgen-Cronbach-Gleser model incorporates the assessments of relative gain or loss in utility resulting from alternative selection strategies. Burke and Doran (1989) compared the payoffs associated with a number of selection tools, based on validities derived from alternative validity generalization methods. They found a sizable change in utility associated with changing from the organization's current selection procedure to an alternative procedure, regardless of the validity generalization estimation method used.

Other methods for assessing the utility of a selection system are the strategic alignment between the selection system and the organization's competitive strategy (Huselid, 1995) and the percentage of increase in output ( $SD_p$ ) (Schmidt & Hunter, 1998). Economically, the gains from increasing the validity of hiring methods can amount to millions of dollars over time, although these figures are often disregarded because the magnitude of economic benefit seems unrealistic ( $SD_y$  tends to be most problematic component of the utility function). Conversely, an organization can lose millions of dollars in reduced production by using selection systems that demonstrate low levels of validity.

The general objective of utility analysis is to determine a selection ratio that permits incremental gain associated with more valid selection to be balanced exactly by the cost associated with the new measure (Cascio, 1998). In fact, utility analysis can be a useful tool in the process of determining cutoff scores for predictor measures.

A break-even analysis could be performed to find the lowest cutoff score on the selection measure that would allow the organization to recover its fully loaded investment costs in each new employee. It is important to consider other factors, such as performance on the job, when determining cutoff scores.

Latham and Whyte (1994) found that managers expressed less support for implementing a valid selection procedure when they were presented with utility information than they did when they were presented only validity information. These findings have stimulated a good deal of discussion, given that the primary purpose of utility analysis is to communicate the value of HRM systems to managers. Cronshaw (1997) posited that Latham and Whyte's (1994) results and commentary (Whyte & Latham, 1997) may generalize to situations where practitioners are selling HRM systems, such as selection. Carson, Becker and Henderson (1998) showed more promising results, as decision makers were more receptive to utility analysis findings when utility analysis information was presented in a manner that was easy to understand; however, acceptance of utility analysis findings were disappointing regardless of how utility analysis information is presented to decision makers. Cabrera and Raju (2001) highlight the importance of involving managers in the utility analysis process, as this appears to play an important role in the extent to which managers and other decision makers accept utility analysis findings.

Research is needed to determine the decision-making processes of managers in making selection decisions and what types of information they might find useful. To this point, utility analysis needs to reflect the strategic context faced by managerial

decision makers. Profit maximization may be only one of many strategic objectives, just as maximizing performance is only one of several possible goals in selection (Russell, Colella, & Bobko, 1993). Managers may consider other factors in addition to the financial benefits of selection and other HRM systems (Macan & Highhouse, 1994). Russell et al. (1993) also suggested that many of the variables relevant to utility analysis can change over time (e.g., strategic needs, predictor-criterion relationship); therefore, it is likely that utility will need to be periodically reassessed. Huselid (1995) noted that an organization's employees can be a unique source of competitive advantage and one that cannot be easily replicated by competitors. A reliable and valid selection system can have considerable influence over the quality and types of skills new employees possess. Research by Terpstra and Rozell (1993) demonstrated evidence for a significant, positive relationship between the use of extensive recruitment and reliable as well as valid selection procedures and firm profits.

Research by Boudreau, Sturman, and Judge (1994) suggested some considerations for utility models that may more accurately reflect organizational realities. For instance, predictors may often be added to existing selection systems rather than replacing them, and organizations often do not exclusively use top-down approaches to making hiring decisions following the selection process. Utility models need to recognize that performance is multidimensional. Also, it is important to recognize that an individual's value to an organization may go beyond his/her current performance in a particular job, which may suggest that utility analysis may be

improved by incorporating considerations of person-organization fit. Although one of the primary purposes of utility analysis is to communicate the value of the selection system to managers, organizational citizenship behaviors also demonstrate the value employees add to the organization, but are not captured by utility analysis (Borman et al., 1997).

To be more useful to organizational decision makers, Russell and colleagues (Russell et al., 1993) recommend consideration of the following questions: (a) Given all other factors aside from the selection system (e.g., capitalization, resource availability), what is the expected level of performance generated by a manager as an employee of the organization? (b) How much of a gain in performance can be expected from a new performance system? (c) Are the levels of performance expected with or without the selection system adequate to meet the organization's strategic needs? (d) Is the incremental increase in performance from selection instrument "A" greater than that expected from instrument "B"? Russell et al. (1993) modified the traditional utility equation to reflect dynamic contributions of the selection system over time (e.g., validity,  $SD_y$ ) as well as changes in what is important to strategic HR decision makers, yielding a more realistic view of how firms benefit from personnel selection.

#### *Shortcomings of the Traditional Approach to Selection Validation Research*

Although the traditional approach to selection validation research has been instrumental in the development of I/O psychology, there are some limitations of this approach. Three primary limitations are the lack of criterion development, focus on



the individual level of analysis, and neglect of the broader organizational context. The lack of criterion development in personnel selection has been noted by many researchers (e.g., Austin & Villanova, 1992; Borman et al., 1997; Schmidt & Hunter, 1998). In addition, predictor development is subject to rigorous research, whereas criterion development is almost overlooked in some situations. A fair amount of time may be devoted to the conceptual nature of the criterion measure, with defining the competencies that need to be included in the performance management system, but not with operationalizing those competencies in a meaningful way. Current statistical methods may contribute to the lack of criterion development; for the structure of the multiple regression equation permits specificity in defining predictors but encourages generality in criterion measurement, as multiple predictors can be specified but only one criterion measure can be specified. Similar to many organizations relying on paper-and-pencil tests for predictors out of convenience, performance appraisal ratings are nearly de facto criteria in the validation of many selection tools. There is also a lack of research on the larger systemic effects of selection systems, which will be discussed shortly. Perhaps more effort in the arena of criterion development would allow research to explore the effects of selection systems at a team/business unit or organizational level.

Another limitation of the traditional approach to selection validation research is that this approach focuses on the individual level of analysis. As Schneider, et al. (2000) note, the study of individual differences and the application to personnel selection psychology represents an interesting and fruitful history; however, this very

fruitfulness may have stifled further progress in understanding the full effects of personnel selection and individual differences in and on organizations. Also, these authors propose that the early success of personnel selection techniques trapped researchers in a model that yielded success, but as the business environment for many organizations has changed, the development and validation of selection systems have been slow to adapt. The success of selection validation research at the individual level of analysis may have inadvertently limited the models used in demonstrating the value and validity of selection systems. The individual differences approach certainly provides a methodologically rigorous and sound way of assessing the validity of selection procedures, but it neglects the larger picture. The individual differences approach, with the individual as the unit of theory and data for both the predictor and criterion components of the model, limits the conclusions that can be reached with regard to organizational performance and organizational differences. A fundamental limitation of the individual differences model is its failure to explore relationships between individual differences and criteria at the organizational level of analysis. Other attempts to integrate the focus on individual differences with broader perspectives can be found in both the field of I/O psychology as well as the larger field of psychology. Research at the individual level does not show a clear link to the effectiveness of the overall organization (Guion, 1998a), creating some statistical challenges in empirically assessing this relationship. There are many links between an individual's score on a selection test and measures of overall firm performance, such as profits and stock price. Also, there are many variables that impact overall

organizational effectiveness that are beyond the control of an individual employee. Research has shown, that for service-focused organizations, customer satisfaction ratings are an appropriate measure of overall organizational performance (Schneider & Bowen, 1985). Many organizations collect customer satisfaction data, which would be available for selection validation research.

Schmitt and Schneider (1983) noted a relative lack of research on the contribution of effective selection and staffing practices to organizational level measures of performance and outlined a research agenda to remedy this void. To this end, organization-level criteria of importance should be defined, and the organization's use of specific selection practices and criteria should then be assessed, followed by a correlation between the criteria of importance and organizational practices. Jackson and Schuler (1995) encourage a shift from focusing on individuals to treating social systems as the target for study, allowing exploration of the systemic impact of selection systems. Messick (1995) notes that the meaning of test scores is most appropriately interpreted in the context in which they are used, which also highlights the importance of examining the broader organizational context in which the selection tools were used.

The third primary limitation of personnel selection research is the neglect of the organizational context. The fundamental cause of the 'decline' or 'marginalization' of personnel selection research by some commentators (e.g., Anderson & Herriot, 1994) appears to be a failure to integrate more situational and

contextual facts of organizational life into the individual differences tradition of personnel selection research.

Another limitation of the individual differences model of personnel selection validation is that it may not reflect the reality of organizational life. Cascio (1991) observed that much of HRM research is misunderstood and underestimated by organizations, partially because much of what we do is evaluated only in statistical or behavioral terms. “Like it or not, the language of business is dollars, not correlation coefficients” (Cascio, 1991, p. vii). Research involving the assessments related to business’ bottom line may increase the adoption rate of genuinely useful HRM practices.

Also, the individual differences model is a static research model, which does not accurately reflect the dynamic reality of life in most organizations. Guion (1998a) asserts that the traditional model assumes that the conditions of the study will remain constant. This is not reality for most organizations in corporate America, including the organization used in the current study. The dynamic nature of organizations does not diminish the value of selection systems, as Howard (1991) suggests that individual assessment remains necessary even during periods of rapid changes. Although the value of individual assessment is not the focus of the current project, this notion suggests a much broader approach to assessment and prediction than traditional thinking has recognized.

Terpstra and Rozell (1993) surveyed firms to learn more about existing staffing practices in organizations and how those selection systems relate to organizational-

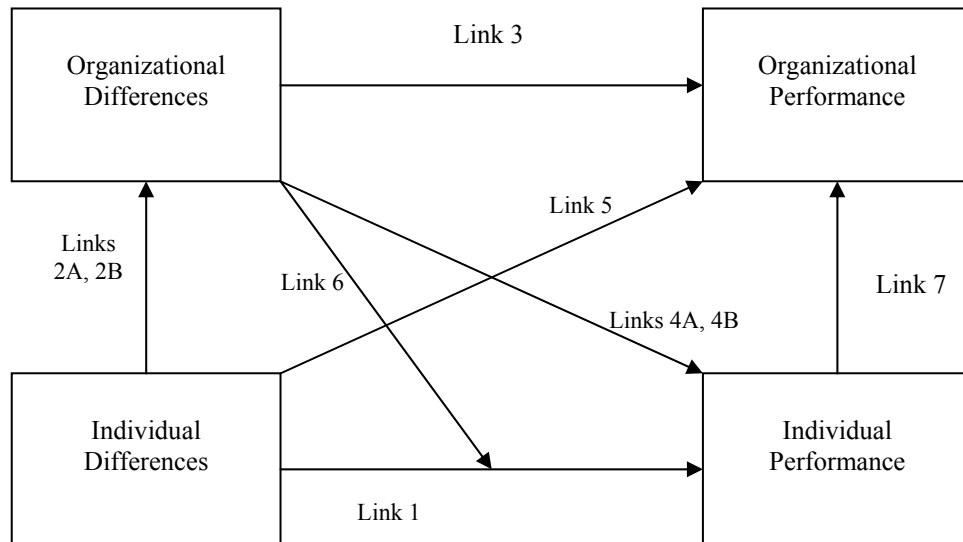
level criteria. Specifically, they focused on the relationship between validation studies, structured interviews, cognitive ability tests and organizational performance (e.g., annual profit, profit growth, sales growth). They found that larger firms were significantly more likely to use validation studies, which were significantly related to annual profit ( $r = .15, p < .05$ ) and the overall performance of the organization ( $r = .13, p < .10$ ). Larger firms were also significantly more likely to use cognitive ability tests during the selection process, but cognitive ability tests were not found to be significantly related to organizational-level criteria. Structured interviews were not significantly related to the organization's overall performance, either. The findings regarding industry type and size provide support for recent literature that emphasizes the need to consider such organizational context characteristics as significant influences upon the adoption and use of particular HRM practices. Terpstra and Rozell (1993) also suggest that longitudinal research be conducted on organizations that adopt some of these staffing practices to assess their impact on organizational performance over time. Research could also systematically examine linkages and paths from the adoption of selection and staffing practices to intermediate measures of employee performance at individual level to final organization level outcomes.

Several researchers have called for a paradigm shift in personnel selection research. Jackson and Schuler (1995) note that the field of I/O psychology has the potential to make substantial contributions to advancing knowledge about human resource management by infusing the development of many human resource systems with additional methodological rigor. For these contributions to come to fruition,

researchers need to change the way in which they approach selection research, both in terms of the content within the selection system as well as the broader organizational context in which the selection system must operate. A transition is needed from a search for the "one best way" to a search for the fundamental features that characterize the many possible ways to design and maintain effective selection systems in organizations. Another suggestion is that a shift in focus from single practices or policies to adopting a holistic approach in thinking about human resource management systems is needed. Research that compares multiple organizations and/or studies dynamic changes in organizations across times and places may be more informative than a focus on a single organization at one point in time. Additionally, a shift from viewing organizational settings as sources of error variance to attending to organizational settings as closely as individual characteristics have been attended to may be beneficial. Guion (1998a) notes that the traditional paradigm overlooks uniqueness of individuality, categorizing it as prediction error, yet organizational settings can be key influences on the impact of human resource systems. Work in the area of person-organization fit has highlighted the impact of organizational settings on employees' attitudes about their jobs and the organization. An alternative to the traditional approach to conducting validation research, recently developed by Schneider et al. (2000), is presented in the next section.

### *Multilevel Model of Personnel Selection*

A model recently developed by Schneider et al. (2000) takes a step in the direction of shifting the existing paradigm for personnel selection research. This model extends the unitarian model for selection research to include group-level and organization-level issues, as illustrated in Figure 2. Link 1 of the model represents the individual differences - individual performance relationship, which summarizes the unitarian model of validation research. In the traditional validity model, the typical personnel selection research validation study occurs in one organization at a time rather than across organizations.



*Figure 2.* Schneider et al. (2000) multilevel model of selection validation.

The individual differences - organizational differences relationship is link 2A of the model. Schneider et al.'s (2000) model suggests that some criteria of interest in

managerial and executive selection might be those criteria that are concerned with the kinds of contexts or environments that managers and executives create and maintain for their organizations. The hypothesis of the individual differences - organizational differences relationship is that managers who create and maintain the environments in which their employees work provide a basis for the organization to perform effectively. Leaders have been theorized to affect how their followers feel (George, 2000). George (1992) found that enthusiastic and energetic leaders were likely to energize their followers, as were distressed and hostile leaders likely to negatively activate their followers. Previous research by Schneider and Bowen (1985) also highlights the impact that HRM systems can have in shaping an employee's environment. Service organizations in which employees have positive perceptions of human resources practices will be those in which employees can devote their energies and resources to serving customers. In other words, when employees perceive their organization as one that facilitates performance by enhancing career opportunities and providing positive supervision, the employees are then free to do the organization's primary work of serving customers. Thus, in service organizations, the same set of human resources practices may be related to positive perceptions for not only employees but for customers as well.

Link 2B of the model is the aggregated individual differences - organizational differences relationship. Because individuals can play an important role in shaping the context that others experience, the aggregated characteristics of people in an organization can have similar effects on context. Research by George (1990, 1995)



found that work groups commonly have consistent or homogenous affective reactions, or a “group affective tone”, and also identified non-personality factors to explain why work group members tend to share moods and emotions. Common socialization experiences and common social influences (Hackman, 1992), similarity of tasks and high task interdependence (Gallupe, Bastianutti, & Cooper, 1991; Heath & Jourden, 1997), membership stability, mood regulation norms and rules (Sutton, 1991), and emotional contagion (Pugh, 2001) have been identified as impacting the collective affect of work groups. Totterdell, Kellett, Teuchmann, and Briner (1998) determined that people’s moods are influenced by the collective mood of their co-workers over time. Other research by Schneider has suggested that organizational characteristics are direct reflections of the particular types of people who are attracted to and selected by organizations, as well as those who choose to remain with particular organizations (e.g., Schneider, 1987; Schneider, Goldstein, & Smith 1995).

Link 3 of Schneider et al.'s (2000) multilevel model is the organizational differences - organizational performance relationship. Human resource systems in an organization can be examined as a way to conceptualize organizational differences and organizational performance. Human resources systems, particularly performance management systems, can be effective vehicles for communicating the values of the organization. The challenge for selection researchers is to identify the individual differences that might be related to the organizational differences, which in turn relate to organizational performance. It is important to recognize that different organizational characteristics imply very different processes with regard to linkages in

the multilevel model. One way to conceptualize organizational differences is that these differences are due to what might be called tangible or universalistic attributes of organizations, such as the structure or culture of the organization. The metaphor of core competencies could also be used to describe organizational differences; in other words, what a particular organization is uniquely competent to do. One implicit assumption underlying traditional personnel selection validation research is that if the organization contains more competent people in the aggregate (now an organizational difference), it will also be a competitively higher-performing organization. Huselid (1995) asserted that HRM practices, such as selection systems, can help create a source of sustained competitive advantage for an organization, especially when they are aligned with the organization's strategy. Specifically, recruiting procedures that provide a large pool of qualified applicants, paired with a relative and valid selection system can have a substantial effect on the quality and type of skills that new employees possess.

The organizational differences - aggregated individual performance relationship is link 4A of the model. Schneider et al. (2000) posit that aggregate individual performance, rather than individual differences in performance, is the type of performance that is of most concern to managers in organizations. Most managers are uninterested in who is the best employee; rather, the manager's concern lies in the aggregate of all of his/her employees. This linkage could certainly inform both researchers and practitioners. This reality suggests a methodological issue that has not been well explored with regard to selection validation research, and represents another

alternative to the traditional focus on individual differences in personnel selection research that needs to be explored. The organizational differences - aggregated individual performance relationship implies using an experimental or a quasi-experimental design in which the organizational differences in use of HRM practices are related to differences in aggregated individual performance. It is possible to think of personnel selection as an organizational intervention and hypothesize whether those employees selected with a new selection process outperform, in the aggregate, those employees selected in other ways (e.g., through old selection process, or through no selection process).

Link 5 of Schneider et al.'s (2000) model is the aggregated individual differences - organizational performance relationship. Even in the conceptualization of the relationship between aggregated individual differences and organizational performance, this model suggests that the intervening variable in the linkage is revealed in the relationship between individual differences and organizational differences, wherein individual differences result in organizational differences. There are simply too many factors that intervene between the personal characteristics of employees and the overall performance of the organization to assess this relationship cleanly. However, through the combination of evidence for both the individual differences - organizational differences relationship (link 1) and the organizational differences - organizational performance relationship (link 4B), we can get a sense of the relationship between individual differences and organizational performance. Rather than trying to provide empirical evidence for the relationship between

aggregated individual performance and organizational performance, a more meaningful question might be whether individual differences in the aggregate are actually related to organizational performance.

Much of the logic underlying personnel selection rests on the assumption that organizational performance is enhanced when validated personnel selection procedures are in place, and this assumption begs two questions. First, did a given organization perform better this year than last year, having implemented new or different personnel selection strategies (improvement/comparison over time)? Second, does organization A, which has validated personnel selection strategies, outperform organization B, which does not (comparison between 2 organizations – control for major differences between organizations)? These questions can be addressed in part by gathering baseline data on organization function (e.g., return on investment, return on equity, return on assets, financial analysts' ratings, customer satisfaction data), then introducing new selection procedures while tracking changes in each organization as well as the relative performance of the organizations.

For service organizations, it has been suggested that customer satisfaction data are the most appropriate indicators of overall firm performance (Schneider & Bowen, 1985). Ryan and Ployhart (2002) noted that the majority of studies of selection and customer service have predicted overall customer service performance assessed through supervisory ratings, rather than directly from customers. In many service environments, managers may not observe many of an employee's interactions with customers (Ryan & Ployhart). In fact, some employees may receive lower customer

service scores from managers by deviating from organizational norms to service the customer's needs (i.e., exceeding the preferred amount of time spent with the customer to diagnose the needs of the customer). Also, almost all of the selection research related to customer service behavior has focused on the individual-level selection and prediction of individual-level performance for employees in customer service roles. Employees may work as individuals and as a team in many service environments, including call centers. On a related note, this body of literature does not address the importance of managerial selection or the role of managers in fostering a work environment that promotes customer service behaviors. Examining the relationship between aggregated individual differences and organizational performance is critical, as it underpins the logic of personnel selection psychology.

Link 4B of Schneider et al.'s (2000) multilevel model of personnel selection is the organizational differences - individual performance relationship. It is imperative that personnel selection researchers begin to explore more fully the relative contribution of individual differences and organizational differences to individual performance and the statistical models to enable such research that already exists. An examination of intercept differences instead of slope differences is possible and could prove to be insightful. This examination could be achieved by an additive combination of the individual differences – individual performance relationship and the organizational differences - individual performance relationship. In statistical terms, studying the combined effects of individual differences and organization differences on individual performance is an issue of assessing the intercept differences

in the relationship between the predictor and criterion, as opposed to assessing the slope differences. Validity coefficients measure the magnitude of the relationship between the predictor and criterion, or the slope differences between different predictors. However, situational variables probably do have some impact on performance differences, which would be reflected in the intercept differences. Along these lines, it is reasonable to assume that organizational differences would have more impact on individual performance than individual differences have on organizational performance, as the things that define organizational differences like the structure and culture of the organization tend to be relatively stable aspects of the organization.

Organizational differences as a moderator of the individual differences - individual performance relationship is Link 6 of the multilevel model proposed by Schneider et al. (2000). This linkage seeks to assess the extent to which, if any, organizational differences moderate the validities of various selection strategies. VG research has not typically explored organizational differences as main effects on individual performance, so it would be erroneous to conclude that organizations do not have an effect on the validities of selection strategies on the basis of VG research alone. Schneider (1978) suggested that the paucity of moderator effects in the field could be attributed to the fact that moderators, when used as interaction items, require extremes. Neither extremes of situations nor extremes of individual differences are likely to be found in field research. These hoped-for moderator effects of the situation on the individual differences - individual performance relationship are unlikely. A

potentially more fruitful approach is to focus on the additive combination suggested by combining links 1 and 4b.

The aggregated individual performance - organizational performance is link 7 of the model. This linkage represents the primary assumption of the traditional validity model that individual performance combines, generally in a mechanical fashion, to yield organizational performance. When this relationship is examined in the context of the larger model, one can see many issues that affect the relationship between the aggregation of individual performance and organizational performance, and many shortcomings of this assumption become apparent. Schneider et al.'s (2000) model suggests that synergies among different aspects of an organization can yield situations in which the aggregated performance of employees exceeds an additive or more mechanical approach to showing how individual performance can be aggregated to result in organizational performance. For example, the climate created or maintained by the manager can have a profound effect on employee performance. In short, the assumption in the traditional validity model that individual performance combines neatly to yield organizational performance is not a true picture of reality in organizations.

The multilevel model developed by Schneider et al. (2000) is a substantial move in advancing research and practitioner knowledge about how selection systems contribute to the larger organization. The model addresses some limitations of the traditional validity model and suggests some ways in which those limitations might be overcome.

### *Summary of Hypotheses and Research Questions*

The current project is an initial empirical test of Schneider et al.'s (2000) model. This project employs a quasi-experimental design, as it involves a managerial development and selection program that has been implemented in a large telecommunications corporation. The previous process for selecting managers was rather unsystematic and unstructured. Consequently, it is not possible to document the details of the selection processes that were previously used for selecting managers. In the following section, each linkage of Schneider's model is summarized, along with the hypotheses and research questions that will be examined in the current study.

Link 1, the individual differences - individual performance relationship, incorporates many elements of a traditional validation study. Thus, the initial hypotheses and research questions focus on issues characteristic of the traditional validation study.

*Hypothesis 1.* The relationships between the predictors and criterion measures will be statistically significant.

*Research Question 1.* Do any of the predictors in the current study display adverse impact?

Link 2, the relationship between individual differences and organization differences, suggests that managers who create and maintain certain contexts/ environments may provide a basis for the organization to perform effectively.

*Hypothesis 2.* Managers selected with the predictors in the current study will have more satisfied employees.



Link 3, the relationship between organizational differences and organizational performance, suggests that certain organizational characteristics may be related to differences in organizational performance.

*Research Question 2:* Do managers selected using the predictors in the current study tend to have more satisfied customers (i.e., higher team-based customer satisfaction scores) than the comparison group?

Link 4A, the relationship between organizational differences and aggregated individual performance, suggests that different selection practices may result in differences in aggregate levels of performance.

*Hypothesis 3:* Managers selected using the predictors in the current study will outperform managers selected through alternative selection systems.

*Hypothesis 4:* Organizations that have a higher proportion of managers selected with the predictors in the current study will have higher levels of performance than organizations with lower proportions of managers selected with the predictors in the current study.

Link 5, the relationship between aggregated individual differences and organizational performance, suggests that organizational performance is enhanced when valid personnel selection procedures are used. Although this linkage conceptually supports the logic of selection research, it may also be the most challenging for which to provide empirical support, given the number of links between a selection system and organizational performance.

*Research Question 3.* Using customer satisfaction as an index of organizational performance, does customer satisfaction increase over time as managers are selected using the predictors in the current study?

Link 6, organizational differences and moderators of Link 1, posits that organizational differences may moderate the validities of selection strategies.

*Research Question 4.* Do differences in criterion-related validity exist between organizations?

Links 2B (aggregated individual differences – organizational differences), 4B (organizational differences – individual performance), and 7 (aggregated individual performance – organizational performance) are beyond the scope of the current study; therefore, they will not be empirically assessed in the current study.

## METHOD

### *Participants*

A total of 1380 employees from a telecommunications company were included in the current study. There were 480 men and 888 women, accounting for 34.8% and 64.3% of the sample, respectively (demographic information was not available for 12 employees). Approximately 59.8% of the sample was Caucasian, 27.5% African American, 10.3% Hispanic, 1.3% Asian, and 0.3% Native American.

Approximately 1044 employees from the total sample participated in the Associate-to-Manager Advancement Process (AMAP)'s selection process. There were 379 men and 665 women who participated in AMAP, respectively accounting for 36.3% and 63.7% of the AMAP group. Approximately 60.2% of the AMAP group was Caucasian, 26.3% African American, 11.9% Hispanic, 1.2% Asian, and 0.3% Native American.

Three hundred twenty-six managers comprised the comparison group. These managers were grandparented into the process, (i.e., hired or promoted to a first-level management position prior to the implementation of AMAP). There were 101 men and 223 women in the comparison group, respectively accounting for 31.0% and 68.4% of the grandparented managers (demographic information was not available for two managers in the comparison group). Approximately 60.1% of the grandparented managers were Caucasian, 31.9% African American, 5.5% Hispanic, 1.5% Asian, and 0.3% Native American.

### *Predictor Measures*

Employees who participated in AMAP completed the General Management Ability Battery (GMAB), an in-basket exercise, and a structured panel interview.

The GMAB included 4 subtests that assessed quantitative and verbal reasoning skills. These subtests scores were converted into an overall test score when presented to participants. Table 1 summarizes these subtests.

Table 1.

#### *Descriptions of Skills Assessed in General Management Ability Battery (GMAB)*

Proficiencies assessed in GMAB	Definition	<i>Number of Items</i>
Quantitative analysis	Ability to solve mathematical problems	50
Written communication fluency	Ability to use standard written English	20
Reading accuracy	Ability to read and comprehend words in sentences	20
Following directions	Ability to follow both written and oral directions carefully	24

The in-basket exercise assessed a variety of information-processing abilities, as described in Table 2. The in-basket activities/items generated information processing and written communication scores, as well as customer and non-customer scores for the exercise. The customer and non-customer scores were aggregated into an overall in-basket score, such that participants were presented with information processing, written communication, and overall in-basket scores. Candidates were given 90

minutes to complete the in-basket exercise. Candidates were provided with materials on a simulated company, including history, operations, and organizational chart. Each candidate assumed the role of a manager handling incoming mail that a typical manager might deal with on a daily basis, and recorded their decisions and plans in written communication (AMAP Methods & Procedures, 2000).

Table 2.

*Description of the Information-Processing Skills Assessed in the In-Basket Exercise*

Information Processing Skills Assessed in In-basket Exercise	Definition
Linking information	The ability to identify inconsistent or conflicting information and recognize the interdependencies among separate pieces of information
Developing creative solutions	The ability to combine available information and resources in order to formulate innovative solutions to problems or conflicts with client or coworkers
Meeting business commitments	The ability to follow explicit instructions, meet deadlines, and follow-up on actions when necessary
Acting on business importance	The ability to prioritize work such that customer needs are always met

The structured interview assessed a variety of communication and management skills, which are detailed in Table 3. Panels of three (3) interviewers conducted the structured interview, which typically lasted one hour. The interview included approximately eighteen questions that assessed the candidates' previous experience and skills related to 5 of the 10 competencies in the corporation's leadership

framework. Oral communication was also assessed during the structured interview. Following each interview, the panel reached agreement for consensus scores on each competency.

Table 3.

*Description of Competencies Assessed in Structured Interview*

Dimension/ Proficiency Assessed During Interview	Definition
Plans proactively	The ability to define short and long term objectives and make provisions for their achievement
Implements with excellence	The ability to prioritize and execute work plans to achieve high quality results
Learns continuously	The ability to develop, or improve upon, skills and knowledge vital for individual and business growth
Enables individual and team effectiveness	The ability to organize and manage diverse individuals and teams for the sustained accomplishment of results
Communicates openly	The ability to create an open and candid environment that promotes the sharing of information and ideas
Oral communication	The ability to describe events and ideas clearly, succinctly and without irrelevancies

*Criterion Measures*

Job performance was one primary criterion measure obtained in the current study. Annual performance appraisal ratings (i.e., supervisory ratings of job performance) were available for employees who were operating in managerial capacities from 1997 - 2001. These performance appraisal ratings were a combination of business results (i.e., 'what' was accomplished) and several leadership

competencies (i.e., ‘how’ the work was accomplished). Table 4 summarizes the combination matrix that was used to determine overall job performance ratings, such that an individual who was evaluated as ‘Above Target’ in Business Results and ‘Needs Development’ in Leadership Competencies was assigned an overall score of 7. Likewise, an individual who was evaluated as ‘Role Model’ in Leadership Competencies and ‘Significantly Below Target’ in Business Results was assigned an overall score of 5. This method for deriving overall performance management scores was based on the organization’s philosophy around the importance of rewarding employees for achieving superior business results.

Table 4.

*Matrix of Business Results (‘what’) and Leadership Competency (‘how’) Ratings of Performance*

Business Results Ratings	Leadership Competency Ratings			
	Needs Development (1)	Skilled (2)	Accomplished (3)	Role Model (4)
Above Target (4)	7	13	15	16
On Target (3)	6	11	12	14
Below Target (2)	2	8	9	10
Significantly Below Target (1)	1	3	4	5

These leadership competencies were viewed to be essential to effective management and leadership of the corporation by the executive leadership team, and

are summarized in Table 5. Turnover, another commonly used criterion measure, was included in the current project. The turnover information was obtained from corporate databases that track employee movement and tenure within the company.

Table 5.

*Description of Job Performance Dimensions Included in Managerial Performance*

*Appraisal Process*

Essential management proficiencies	Description
Quantitative and verbal reasoning skills	Ability to apply reasoning ability in mathematical and verbal contexts
Strategic thinking skills	Ability to recognize, prioritize, and use information effectively when solving problems and making decisions
Communicates openly	Ability to create an open and candid environment that promotes the act of sharing of information and ideas
Oral communication skills	Ability to describe events and ideas with clarity and economy of words, sticking to information relevant to the subject
Written communication skills	Ability to communicate in written communication with clarity, efficiency, accuracy, and good organization
Enables individual and team effectiveness	Ability to organize and manage diverse individuals and teams for the sustained accomplishment of results
Implements with excellence	Ability to prioritize and execute work plans to achieve high quality results
Learns continuously	Continuously developing or improving upon skills and knowledge vital for individual and business growth
Plans proactively	Ability to define short-and long-term objectives and make provisions for their achievement



Some supplemental performance metrics for managers were available for certain subsets of the sample in the current study. Employee satisfaction was assessed through an annually-administered survey for a portion of the sample. The 15-item survey measures the extent to which employees trust management, how involved employees feel in the decision-making process, the center's commitment/focus on quality, as well as satisfaction with job and direct supervisor. The first 10 items of the survey are consistent across all parts of the organization, and the last 5-7 items are tailored to the specific call center. Employee satisfaction data collected in October 2000 for approximately 140 managers (93 AMAP managers, 47 grandparented managers) were included in the study.

Unit performance was also available for a subset of managers in the current study. A variety of performance metrics for managers' direct reports at some call centers were gathered, including total calls, percentage of complaints resolved, adherence to schedule, and absence. Unit performance was available for approximately 45 managers (30 AMAP managers, 15 grandparented managers).

Customer satisfaction data were available for approximately 100 managers (34-65 AMAP managers, 2-40 grandparented managers over time) in the current study. A 3-question customer survey, conducted over the phone, assessed overall customer satisfaction with the corporation and the most recent interaction with a customer care representative. Table 6 summarizes the criterion measures as well as the sample size for each measure.

Table 6.

*Criterion Measures and Sample Size for Each Measure*

Criterion Measure	<i>N</i>
Performance Appraisal Ratings	523 with ratings for 1 year 478 with ratings for 2 years 386 with ratings for 3 years 233 with ratings for 4 years
Merit Increases	498 with merit increases for 1 year 451 with merit increases for 2 years 339 with merit increases for 3 years
Turnover	189
Employee Satisfaction	134
Unit Performance	33
Customer Satisfaction	138

*Procedure*

AMAP was a management development and selection process that was primarily used to identify management potential in employees in the customer care organization, which was part of the organization's consumer services business. AMAP was one of several leadership identification and development programs aligned with the company's leadership framework. The goal of AMAP was to develop more effective managers. AMAP began with a developmental process for customer care associates, whereby they sought out opportunities to gain additional skills to become successful managers. This phase of AMAP was self-paced and driven by the customer care associates, with the support of their managers. The next phase of AMAP was the formal selection process, which included the GMAB, in-

basket exercise, and interview. The current study focuses on the selection process within AMAP.

A third-party organization administered the GMAB and in-basket exercise. Mid-level managers served as panel members for the structured interview. These managers completed a two-day interviewer training, which was a combination of facilitator presentation of general interviewing techniques and procedures as well as skill practice. The presentation also focused on the importance of conducting interviews to be in compliance with legal requirements, including topics and questions to avoid (e.g., protected classes under Title VII, non-job related issues). In addition, the facilitator reviewed the scoring process by discussing behavioral anchors/ indicators for each competency as well as common rater errors. Following the presentation, interviewers had the opportunity to simulate multiple interviews and receive feedback from more experienced interviewers. Quarterly interview audits were also conducted to ensure the integrity of the structured interview was maintained (e.g., direct observation, remote observation via telephone, listening to taped interviews).

AMAP utilized a hybrid multiple-hurdle approach. The first stage of the selection process involved the completion of the GMAB and in-basket exercise. Participants completed both measures during the same testing session. The raw scores on GMAB subtests were converted to ensure equivalent weighting of the subtests. The total GMAB converted score was then mapped to the appropriate

GMAB Qualification Zone, as outlined in Table 7, which determines the number of points each candidate receives on the GMAB.

Table 7.

*Total GMAB Converted Scores, GMAB Qualification Zones and Points Awarded for GMAB Score*

Total GMAB Converted Score	GMAB Qualification Zone		AMAP Points
225 and higher	Compensatory	Q1	4
210 to 224	Standard	Q2	3
200 to 209	Intermediate	Q3	2
194 to 199	Minimum	Q4	1
193 and lower	Below Minimum	NQ	0

Table 8 summarizes the cutoff scores for each AMAP measure. If the employee surpassed the cutoff scores for both the GMAB and in-basket exercise, he/she was invited to go through the structured panel interview. If the employee did not surpass the cutoff scores at any stage during the selection process, he/she was eligible for first and second retests six months following the prior testing date. Subsequent retesting may take place twelve months following the most recent re-test. If the employee passed the interview, then he/she was placed in a talent pool for available first-line management positions. The AMAP process qualified individuals for promotion; however, successfully completing AMAP did not guarantee that the employee would be placed in a management position. In other words, successfully

completing AMAP was a necessary but not sufficient condition for promotion to a managerial position.

Table 8.

*Cutoff Scores for AMAP Measures*

AMAP Measure	Proficiency	Cutoff Score
GMAB and In-basket Exercises	Quantitative and verbal reasoning skills, strategic thinking skills	4 (out of 12)
In-basket Exercise	Written communication skills	1 (out of 2)
Structured Interview	Oral communication	1 (out of 2)
Structured Interview	Communicates openly	1 (out of 2)
Structured Interview	Enables individual and team effectiveness	1 (out of 2)
Structured Interview	Implements with excellence	1 (out of 2)
Structured Interview	Learns continuously	1 (out of 2)
Structured Interview	Plans proactively	1 (out of 2)

*Statistical Analyses*

In preparation for the statistical analyses in the current study, some of the predictor data were aggregated to allow for comparisons across constructs, rather than a combination of constructs and methods. Constructs and methods for each predictor measure are presented in Table 9.

Table 9.

*Constructs and Methods for Predictor Measures*

Construct	Method
General cognitive ability	Average of overall GMAB and overall in-basket scores
Information processing	In-basket
Written communication	In-basket
Planning and organizing	Interview (average of 'Plans proactively' and 'Implements with excellence' dimensions)
Learns continuously	Interview
Teamwork	Interview (average of 'Enables individual and team effectiveness' and 'communicates openly' dimensions)
Oral communication	Interview

Table 10 summarizes the statistical analyses that were performed to examine the hypotheses and research questions of the current study. A combination of correlation analysis, analysis of variance (ANOVA), and regression analysis were used to explore the hypotheses and research questions for this study.

Table 10.

*Summary of Hypotheses, Research Questions, and Statistical Analyses in Current Study*

Hypothesis/Research Question		Analysis Conducted
Hypothesis 1.	The relationships between the predictors and criterion measures will be statistically significant.	Correlation and regression analysis
Research Question 1.	Do any of the predictors in the current study display adverse impact?	Analysis of variance (ANOVA)
Hypothesis 2.	Managers selected with the predictors in the current study will have more satisfied employees.	ANOVA
Research Question 2.	Do managers selected using the predictors in the current study tend to have more satisfied customers (i.e., higher team-based customer satisfaction scores) than the comparison group?	ANOVA
Hypothesis 3.	Managers selected using the predictors in the current study will outperform managers selected through alternative selection systems.	ANOVA
Hypothesis 4.	Organizations that have a higher proportion of managers selected with the predictors in the current study will have higher levels of performance than organizations with lower proportions of managers selected with the predictors in the current study.	ANOVA
Research Question 3.	Using customer satisfaction as an index of organizational performance, does customer satisfaction increase over time as managers are selected using the predictors in the current study?	Repeated-measures ANOVA
Research Question 4.	Do differences in criterion-related validity exist between organizations?	Correlation analysis

## RESULTS

### *Hypothesis 1. The Relationships Between the Predictors and Criterion Measures will be Statistically Significant*

Table 11 summarizes the means, standard deviations and correlations between all predictor and criterion measures in the current study. Native Americans were included in the overall sample statistics, but were omitted from comparisons across race groups due to small sample size. There were positive, statistically significant correlations between cognitive ability, information processing, and written communication scores. There were also positive, statistically significant correlations between planning and organizing, learns continuously, teamwork, and oral communication.

Sample sizes varied greatly across criterion measures, because some criterion measures were available only for portions of the sample in the current study. All job performance correlations were positive and statistically significant. Positive, significant correlations between job performance correlations within the same year were expected, as the 'what' and 'how' job performance ratings contributed to the overall job performance rating. Positive, significant correlations between job performance ratings across years suggests that managers who received higher job performance ratings in Year One tended to receive higher job performance ratings in subsequent years. There were positive, significant correlations between merit and bonus payments within each year, suggesting that managers who received larger merit increases tended to receive larger bonus payments. There were positive, significant



Table 11.

*Means, Standard Deviations, and Correlations Between all Predictor and Criterion Measures for Overall Sample*

	1	2	3	4	5	6	7	8	9	10	11	12	13
<b>PREDICTORS</b>													
1. Cognitive ability		.74** (1042)	.13** (1011)	.04 (714)	-.02 (714)	-.01 (714)	.03 (714)	-.07 (419)	-.07 (419)	-.02 (419)	-.02 (296)	-.02 (296)	.03 (296)
2. Information processing			.19** (1011)	.00 (714)	-.01 (714)	.00 (714)	.03 (714)	.01 (410)	.01 (410)	.02 (410)	.02 (289)	.02 (289)	.07 (289)
3. Written communication				.06 (711)	.02 (711)	-.01 (711)	.03 (711)	.04 (408)	.04 (408)	.00 (408)	.04 (287)	.00 (287)	.07 (287)
4. Planning and organizing					.35** (714)	.37** (714)	.41** (714)	.07 (410)	.02 (410)	.09 (410)	.05 (289)	.01 (289)	.09 (289)
5. Learns continuously						.30** (714)	.29** (714)	.02 (410)	.02 (410)	-.02 (410)	.13* (289)	.10 (289)	.05 (289)
6. Teamwork							.41** (714)	.05 (410)	-.04 (410)	.09 (410)	-.01 (289)	-.03 (289)	.00 (289)
7. Oral Communication								.05 (410)	.03 (410)	-.03 (410)	.06 (289)	.04 (289)	.05 (289)
<b>JOB PERFORMANCE</b>													
8. JP – Year 1									.78** (704)	.67** (704)	.37** (529)	.28** (529)	.31** (529)

Table 11. (Continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13
9. What Year 1										.23** (704)	.27** (529)	.22** (529)	.20** (529)
10. How Year 1											.25** (529)	.18** (529)	.30** (529)
11. JP – Year 2												.87** (531)	.71** (531)
12. What Year 2													.42** (531)
13. How Year 2													
14. JP – Year 3													
15. What Year 3													
16. How Year 3													
Mean	1.61	6.01	1.11	1.72	1.66	1.62	1.49	11.36	3.06	2.33	11.85	3.13	2.56
<i>SD</i>	1.05	3.08	.33	.35	.47	.39	.50	1.71	.40	.56	2.07	.47	.61
Total <i>n</i>	1061	1042	1011	714	714	714	714	704	704	704	531	531	531

Table 11. (Continued)

	14	15	16	17	18	19	20	21	22	23	24	25
<b>PREDICTORS</b>												
1. Cognitive ability	.02 (158)	.08 (158)	-.05 (158)	-.01 (362)	-.08 (420)	.00 (321)	-.06 (396)	-.07 (228)	.01 (253)	-.05 (1061)	.12 (95)	.18 (30)
2. Information processing	.14 (150)	.17* (150)	.02 (150)	.09 (361)	.03 (419)	.03 (319)	-.04 (394)	.06 (220)	.08 (245)	-.02 (1042)	-.07 (93)	.19 (30)
3. Written communication	.05 (149)	.03 (149)	.04 (149)	.12* (360)	.01 (418)	-.06 (318)	-.07 (392)	.14* (218)	-.01 (243)	.02 (1011)	.09 (92)	-.21 (30)
4. Planning and organizing	.07 (150)	.05 (150)	.10 (150)	.07 (361)	.01 (419)	.03 (319)	.02 (394)	-.01 (220)	.15* (245)	.08* (714)	(.06 (87)	-.21 (24)
5. Learns continuously	.15 (150)	.07 (150)	.12 (150)	.10 (361)	.04 (419)	.11 (319)	.01 (394)	.05 (220)	.04 (245)	-.01 (714)	-.17 (87)	-.29 (24)
6. Teamwork	.10 (150)	.10 (150)	.07 (150)	.11* (361)	.13** (419)	.11 (319)	-.03 (394)	.07 (220)	.17 (245)	.03 (714)	-.05 (87)	-.15 (24)
7. Oral Communication	.03 (150)	.11 (150)	-.03 (150)	.04 (361)	.01 (419)	.11 (319)	-.01 (394)	.16* (220)	.05 (245)	.01 (714)	.05 (87)	-.23 (24)
<b>JOB PERFORMANCE</b>												
8. JP – Year 1	.25** (403)	.18** (403)	.27** (403)	.45** (384)	.19** (416)	.17** (399)	.36** (466)	.16** (459)	.19** (487)	-.10** (704)	.13 (129)	.35* (39)

Table 11. (Continued)

		14	15	16	17	18	19	20	21	22	23	24	25
9.	What Year 1	.18** (403)	.13** (403)	.17** (403)	.33** (384)	.12* (416)	.05 (399)	.29** (466)	.06 (459)	.12** (487)	-.10** (704)	-.03 (129)	.29 (39)
10.	How Year 1	.22** (403)	.15** (403)	.28** (403)	.38** (384)	.24** (416)	.21** (399)	.28** (466)	.19** (459)	.18** (487)	-.03 (704)	.16 (129)	.13 (39)
11.	JP – Year 2	.34** (371)	.26** (371)	.36** (371)	.32** (308)	.12* (321)	.49** (337)	.46** (368)	.18** (422)	.37** (433)	-.08 (531)	.19* (111)	.08 (34)
12.	What Year 2	.27** (371)	.22** (371)	.25** (371)	.27** (308)	.11* (321)	.43** (337)	.44** (368)	.12* (422)	.33** (433)	-.07 (531)	.12 (111)	.06 (34)
13.	How Year 2	.37** (371)	.23** (371)	.46** (371)	.30** (308)	.11 (321)	.42** (337)	.31** (368)	.17** (422)	.32** (433)	.01 (531)	.25** (111)	.06 (34)
14.	JP – Year 3		.85** (405)	.72** (405)	.34** (178)	.12 (179)	.34** (221)	.30** (229)	.56** (383)	.43** (392)	-.13* (405)	.28** (90)	.10 (25)
15.	What Year 3			.33** (405)	.21** (178)	.06 (179)	.20** (221)	.23** (229)	.48** (383)	.35** (392)	-.09 (405)	.23* (90)	.06 (25)
16.	How Year 3				.35** (178)	.14 (179)	.39** (221)	.38** (229)	.43** (383)	.38** (392)	-.09 (405)	.22* (90)	.10 (25)

Table 11. (Continued)

	14	15	16	17	18	19	20	21	22	23	24	25
<b>MERIT/BONUS</b>												
17. Merit Year 1					.51** (406)	.28** (315)	.29** (353)	.17* (232)	.30** (243)	-.02 (406)	.05 (87)	.33 (26)
18. Bonus Year 1						.08 (323)	.13** (384)	-.11 (234)	.23** (246)	.01 (469)	.14 (94)	-.26 (26)
19. Merit Year 2							.35** (407)	.35** (278)	.04 (293)	-.18** (407)	.13 (86)	.26 (21)
2. Bonus Year 2								.10 (287)	.36** (313)	-.24** (491)	.12 (96)	.02 (21)
21. Merit Year 3									.10 (287)	-.28** (462)	.31** (97)	.22 (26)
22. Bonus Year 3										-.31** (495)	.10 (100)	-.09 (26)
23. <b>Turnover</b>											-.10 (140)	-- <sup>a</sup> (45)
24. <b>Employee Satisfaction</b>												.14 (16)
Mean	12.10	3.17	2.67	1861.58	1610.49	2310.25	2098.61	2609.66	3495.22	.21	76.96	77.11
SD	1.86	.47	.63	1452.98	1122.96	1399.91	1583.01	1624.35	1985.61	.41	9.35	1.99
Total <i>n</i>	405	405	405	406	469	407	491	462	495	1379	140	45

Table 11. (Continued)

	26	27	28	29	30	31	32	33	34	35	36	37
<b>PREDICTORS</b>												
1. Cognitive ability	.33 (30)	-.18 (30)	.11 (30)	.37* (30)	-.02 (34)	-.16 (38)	.07 (55)	-.06 (56)	.10 (65)	.12 (50)	.28 (40)	-.02 (34)
2. Information processing	.17 (30)	-.07 (30)	.10 (30)	.38* (30)	.24 (34)	.25 (38)	-.13 (55)	-.25 (56)	.13 (65)	-.02 (50)	.12 (40)	.00 (34)
3. Written communication	-.08 (30)	.21 (30)	.03 (30)	-.04 (30)	-.07 (34)	.02 (38)	.12 (54)	.22 (55)	-.09 (64)	-.03 (50)	.10 (40)	-.20 (34)
4. Planning and organizing	-.39 (24)	-.27 (24)	-.09 (24)	-.07 (24)	.13 (34)	.12 (34)	-.07 (51)	.01 (53)	.23 (61)	-.00 (48)	-.09 (35)	.06 (29)
5. Learns continuously	-.07 (24)	-.16 (24)	-.04 (24)	.08 (24)	-.42* (34)	.16 (34)	-.13 (51)	-.00 (53)	.08 (61)	-.05 (48)	.07 (35)	.07 (29)
6. Teamwork	-.33 (24)	-.07 (24)	.02 (24)	.22 (24)	.13 (34)	.13 (34)	.21 (51)	.09 (53)	.23 (61)	-.13 (48)	-.02 (35)	.19 (29)
7. Oral Communication	.17 (24)	-.05 (24)	.38 (24)	.26 (24)	.01 (34)	.00 (34)	.15 (51)	.05 (53)	.01 (61)	-.19 (48)	.30 (35)	-.17 (29)
<b>JOB PERFORMANCE</b>												
8. JP – Year 1	-.37* (39)	-.01 (39)	.06 (39)	.18 (39)	.37* (31)	.02 (36)	.10 (82)	-.10 (84)	.14 (91)	.03 (88)	-.03 (67)	-.28* (53)

Table 11. (Continued)

		26	27	28	29	30	31	32	33	34	35	36	37
9.	What Year 1	-.33* (39)	.11 (39)	.03 (39)	.23 (39)	.34 (31)	-.06 (36)	.08 (82)	-.06 (84)	.04 (91)	.01 (88)	-.08 (67)	-.36** (53)
10.	How Year 1	-.35* (39)	-.03 (39)	.18 (39)	.02 (39)	.34 (31)	.16 (36)	.14 (82)	.02 (84)	.30** (91)	.17 (88)	-.01 (67)	-.15 (53)
11.	JP – Year 2	.12 (34)	.25 (34)	.14 (34)	.03 (34)	.15 (21)	.40* (27)	.30* (67)	.23 (71)	.08 (79)	.10 (77)	.06 (58)	-.27 (49)
12.	What Year 2	.14 (34)	.30 (34)	.21 (34)	-.01 (34)	.04 (21)	.28 (27)	.28* (67)	.08 (71)	.05 (79)	-.06 (77)	.02 (58)	-.27 (49)
13.	How Year 2	-.17 (34)	.06 (34)	-.11 (34)	.17 (34)	.34 (21)	.56** (27)	.17 (67)	.21 (71)	.05 (79)	.23* (77)	.17 (58)	-.16 (49)
14.	JP – Year 3	.18 (25)	-.22 (25)	-.22 (25)	.11 (25)	-.10 (10)	.21 (15)	.08 (60)	-.02 (61)	-.08 (69)	.10 (71)	-.03 (57)	.03 (46)
15.	What Year 3	.22 (25)	.02 (25)	.00 (25)	-.02 (25)	.19 (10)	.21 (15)	.01 (60)	-.15 (61)	-.18 (69)	-.02 (71)	-.01 (57)	-.03 (46)
16.	How Year 3	.12 (25)	-.44* (25)	-.33 (25)	.29 (25)	-.16 (10)	.10 (15)	.15 (60)	.06 (61)	.10 (69)	.30* (71)	.12 (57)	.15 (46)

Table 11. (Continued)

	26	27	28	29	30	31	32	33	34	35	36	37
<b>MERIT/BONUS</b>												
17. Merit Year 1	-.07 (26)	-.36 (26)	-.25 (26)	.24 (26)	.19 (32)	.28 (33)	.01 (51)	-.08 (52)	.14 (63)	.03 (52)	.08 (40)	-.08 (34)
18. Bonus Year 1	.01 (26)	-.00 (26)	.17 (26)	.28 (26)	.30 (34)	.39* (33)	.05 (54)	-.01 (57)	.00 (65)	-.27 (54)	-.02 (40)	.04 (34)
19. Merit Year 2	.27 (21)	.21 (21)	.18 (21)	.00 (21)	.03 (24)	.57** (28)	.23 (49)	.02 (52)	-.10 (60)	.13 (50)	.04 (40)	-.12 (33)
20. Bonus Year 2	-.12 (21)	.29 (21)	.13 (21)	.13 (21)	.00 (26)	.24 (33)	.23 (54)	-.10 (58)	.01 (66)	.02 (53)	.11 (41)	-.32 (34)
21. Merit Year 3	.12 (26)	-.20 (26)	-.18 (26)	.08 (26)	-.27 (14)	.38 (20)	.17 (62)	.01 (65)	.02 (74)	-.08 (75)	-.02 (59)	-.01 (49)
22. Bonus Year 3	.19 (26)	-.15 (26)	-.25 (26)	.08 (26)	.29 (14)	.17 (32)	-.04 (64)	-.14 (66)	-.31** (75)	-.02 (76)	.01 (60)	.04 (49)
23. <b>Turnover</b>	-- <sup>a</sup> (45)	-- <sup>a</sup> (45)	-- <sup>a</sup> (45)	-- <sup>a</sup> (45)	-.07 (36)	-.30 (41)	-.22* (86)	-.20 (87)	.03 (96)	.02 (90)	.05 (72)	.10 (58)
24. <b>Employee Satisfaction</b>	.22 (16)	-.32 (16)	-.50 (16)	-.06 (16)	.34 (17)	.25 (25)	.12 (56)	.05 (61)	-.18 (66)	.14 (60)	.24 (55)	-.10 (43)



Table 11. (Continued)

	26	27	28	29	30	31	32	33	34	35	36	37
<b>UNIT PERFORMANCE</b>												
25. % calls resolved on first attempt	.20	-.26	-.21	.30*	.58	-.46	.46	.23	-.10	-.01	-.13	-.15
	(45)	(45)	(45)	(45)	(6)	(5)	(10)	(10)	(10)	(10)	(11)	(18)
26. Hold Time		-.21	.13	.03	-.49	-.69	-.47	-.24	-.32	-.17	-.69*	.20
		(45)	(45)	(45)	(6)	(5)	(10)	(10)	(10)	(10)	(11)	(18)
27. Conformance to schedule			.48**	-.24	-.43	.39	-.15	.04	.16	.22	-.21	-.42
			(45)	(45)	(6)	(5)	(10)	(10)	(10)	(10)	(11)	(18)
28. Adherence				-.02	-.39	-.15	-.22	.19	.26	.15	-.37	-.24
				(45)	(6)	(5)	(10)	(10)	(10)	(10)	(11)	(18)
29. Absence					.40	.08	.79**	-.51	-.19	.11	.12	-.04
					(6)	(5)	(10)	(10)	(10)	(10)	(11)	(18)
<b>CUSTOMER SATISFACTION</b>												
30. Overall Customer Satisfaction (Months 1-3)						.25	.49	.58	.78	.83	.24	-- <sup>b</sup>
						(22)	(13)	(11)	(6)	(4)	(4)	(0)
31. Overall Customer Satisfaction (Months 4-6)							.62**	.37	-.09	-.27	.00	-1.00**
							(28)	(22)	(16)	(9)	(7)	(2)

Table 11. (Continued)

	26	27	28	29	30	31	32	33	34	35	36	37
32. Overall Customer Satisfaction (Months 7-9)								.43**	.30*	.05	-.01	.25
								(75)	(63)	(44)	(34)	(24)
33. Overall Customer Satisfaction (Months 10-12)									.42**	.42**	.24	.47*
									(70)	(50)	(36)	(26)
34. Overall Customer Satisfaction (Months 13-15)										.30*	-.02	.25
										(68)	(48)	(34)
35. Overall Customer Satisfaction (Months 16-18)											.56**	.08
											(62)	(41)
36. Overall Customer Satisfaction (Months 19-21)												.05
												(48)
37. Overall Customer Satisfaction (Months 22-24)												
Mean	83.29	94.34	87.37	11.12	6.83	6.88	6.93	6.90	6.86	6.75	6.78	6.86
SD	9.77	.71	2.03	.16	.21	.24	.22	.20	.20	.18	.20	.27
Total <i>n</i>	45	45	45	45	36	41	86	87	96	90	72	58

Note. Numbers in parentheses indicate the *n* for each correlation. JP – Year 1 = Overall job performance rating in Year One, JP – Year 2 = Overall job performance rating in Year Two, JP – Year 3 = Overall job performance rating in Year Three.

<sup>a</sup>Correlation not computed because at least one variable is constant. <sup>b</sup>No data available to compute correlation. \*  $p < .05$ , \*\*  $p < .01$ .

correlations between merit increases across years, implying that managers who received larger merit increases in Year One tended to receive larger merit increases in subsequent years. Many of the correlations between bonus payments across years were also positive and statistically significant, suggesting that managers who received larger bonus payments in Year One tended to receive larger bonus increases in subsequent years. (The correlation between bonus payments in Years Two and Three was positive, but not statistically significant.) There were several positive, significant correlations between the job performance and the merit/bonus measures, implying that managers who received higher job performance ratings also received larger merit increases and/or bonus payments.

There were several negative, significant correlations between job performance and turnover, suggesting that managers who received higher job performance ratings were less likely to leave the company. There were also several negative, significant correlations between merit/bonus measures and turnover, implying that managers who received larger merit and/or bonus payments were less likely to leave the company. There were a few negative correlations between turnover and customer satisfaction, with the correlation between turnover and overall customer satisfaction seven to nine months following completion of AMAP being statistically significant, suggesting that managers of employees providing better overall customer service were less likely to leave the company.

The correlations between employee satisfaction and job performance increased over time, such that managers who received higher job performance ratings in Years Two and Three tended to receive higher employee satisfaction scores.

Correlations between measures of unit performance (e.g., hold time) and other criterion measures should be interpreted with caution, given the small sample sizes. There were several statistically significant correlations between measures of unit performance, although some of these correlations were surprising.

General cognitive ability correlated negatively with job performance ratings in Years One, Two, and Three, but these correlations were very small in magnitude. Cognitive ability also displayed weak correlations with merit increases and bonus payments in Years One, Two, and Three, as well as with turnover and employee satisfaction. Cognitive ability was significantly correlated with absence, such that higher cognitive ability scores were related to higher rates of absence. Cognitive ability was also weakly correlated with customer satisfaction.

Information processing correlated positively with job performance ratings in Years One, Two and Three, but these correlations were often small in magnitude. Information processing was significantly correlated with 'what' job performance ratings in Year Three, such that higher information processing scores were related to higher 'what' job performance ratings. Information processing was weakly correlated with merit increases and bonus payments in Years One, Two, and Three, as well as with turnover and employee satisfaction. Information processing was significantly correlated with absence, such that higher information processing scores were related to

higher rates of absence among the participants' employees. Information processing was not significantly correlated with customer satisfaction.

Written communication correlated positively with job performance ratings in Years One, Two and Three, but these correlations were often small in magnitude. Written communication was significantly correlated with merit increases in Years One and Three, such that higher written communication scores were related to larger merit increases. Written communication was not significantly correlated with turnover, employee satisfaction, unit performance, or customer satisfaction.

Planning and organizing correlated positively with job performance ratings in Years One, Two and Three, but these correlations were often small in magnitude. This construct was significantly correlated with bonus payments in Year Three, such that higher scores in planning and organizing were related to larger bonus payments. Planning and organizing was also significantly correlated with turnover, such that higher scores in planning and organizing were related to increased levels of turnover among participants. Planning and organizing was not significantly correlated with employee satisfaction, unit performance, or customer satisfaction.

Learns continuously was significantly correlated with overall job performance ratings in Year Two, such that higher scores in learns continuously were related to higher overall job performance ratings. Learns continuously was not significantly correlated with merit increases, bonus payments, turnover, employee satisfaction, unit performance, or customer satisfaction.

Teamwork was not significantly correlated with job performance ratings. Teamwork was significantly correlated with merit increases and bonus payments in Year One, such that higher teamwork scores were related to larger merit increases and bonus payments. Teamwork was not significantly correlated with the other criterion measures.

Oral communication was not significantly correlated with job performance ratings. Oral communication was significantly correlated with merit increases in Year One, such that higher scores in oral communication were related to larger merit increases. Oral communication was not significantly correlated with turnover, employee satisfaction, unit performance, and customer satisfaction.

Several regression models were tested to determine the extent to which the predictors predicted performance on the criterion measures. Hierarchical regression models were tested; cognitive ability, information processing, and written communication were entered in the first step, and planning and organizing, learns continuously, teamwork, and oral communication were entered in the second step of each model. Regression analysis revealed that the predictors predicted merit increases in Years One and Three, and bonus payments in Year One, as noted in Table 12.

Table 12.

*Regression Analysis for Merit Increases and Bonus Payments*

Criterion: Merit - Year 1 (N = 360)						
Step	Predictors	$\beta$	$R$	$R^2$	Adjusted $R^2$	$\Delta R^2$
1	Cognitive ability	-.14*				
	Information processing	.17**	.18**	.03**	.02	
	Written communication	.11**				
2	Planning and Organizing	.04				
	Learns Continuously	.06	.22**	.05	.03	.02
	Teamwork	.07				
	Oral Communication	-.03				
Criterion: Bonus - Year 1 (N = 418)						
Step	Predictors	$\beta$	$R$	$R^2$	Adjusted $R^2$	$\Delta R^2$
1	Cognitive ability	-.17***				
	Information processing	.15**	.14**	.02**	.01	
	Written communication	-.00				
2	Planning and Organizing	-.03				
	Learns Continuously	.02	.19**	.04	.02	.02
	Teamwork	.14**				
	Oral Communication	-.03				
Criterion: Merit - Year 3 (N = 218)						
Step	Predictors	$\beta$	$R$	$R^2$	Adjusted $R^2$	$\Delta R^2$
1	Cognitive ability	-.18**				
	Information processing	.15*	.21**	.04	.03	
	Written communication	.13*				
2	Planning and Organizing	-.08				
	Learns Continuously	.02	.26**	.07	.04	.02
	Teamwork	.01				
	Oral Communication	.16**				

Note.  $\beta$  estimates based on final step of regression model. \*  $p < .10$ , \*\*  $p < .05$  \*\*\*  $p < .01$ .

Hierarchical regressions indicated that cognitive ability, information processing, and written communication significantly predicted both merit increases and bonus payments in Year One. Additionally, planning and organizing, learns continuously, teamwork, and oral communication significantly predicted merit

increases in Year One, but these predictors did not display significant incremental validity above cognitive ability, information processing, and written communication. Similarly, planning and organizing, learns continuously, teamwork, and oral communication did not display significant incremental validity above cognitive ability, information processing, and written communication in predicting bonus payments in Year One. Cognitive ability, information processing, and written communication also significantly predicted merit increases in Year Three. Planning and organizing, learns continuously, teamwork, and oral communication did not display significant incremental validity above cognitive ability, information processing, and written communication.

Regression analysis also revealed that planning and organizing, learns continuously, teamwork, and oral communication explained an incremental proportion of variance in hold time above and beyond the other predictors, as noted in Table 13.

Table 13.

*Regression Analysis for Hold Time*

Criterion: Hold Time (N = 24)						
Step	Predictors	$\beta$	$R$	$R^2$	Adjusted $R^2$	$\Delta R^2$
1	Cognitive ability	.88**				
	Information processing	-.7*	.42	.17	.05	
	Written communication	-.22				
2	Planning and Organizing	-.33				
	Learns Continuously	.00	.76**	.58	.40	.41**
	Teamwork	-.53*				
	Oral Communication	.30				

Note. N = 24.  $\beta$  estimates based on final step of regression model. \*  $p < .10$ , \*\*  $p < .05$  \*\*\* $p < .01$ .



The criterion-related validity estimates provide limited support of Hypothesis 1, as the predictors in the current study were weakly correlated with the criterion measures, job performance ratings in particular. Regression analyses showed that cognitive ability, information processing, and written communication significantly predicted merit increases and bonus payments in Year One, as well as merit increases in Year Three. Regression analyses also revealed that the constructs of planning and organizing, learns continuously, teamwork, and oral communication explained an incremental proportion of variance in hold time above and beyond the other predictors.

*Research Question 1. Do Any of the Predictors in the Current Study Display Adverse Impact?*

Table 14 compares the means and standard deviations of the predictor groups across sex groups. Males received significantly higher scores in cognitive ability,

Table 14.

*Descriptive Statistics of Predictor Measures Across Sex Groups*

	Males			Females			<i>d</i>
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	
Cognitive ability	1.78	1.07	382	1.51	1.03	671	0.26*
Information processing	6.27	2.99	379	5.87	3.13	663	0.23*
Written communication	1.12	0.35	369	1.10	0.32	642	0.03
Planning and organizing	1.73	0.36	270	1.72	0.35	444	0.02
Learns continuously	1.67	0.47	270	1.66	0.48	444	0.01
Teamwork	1.61	0.39	270	1.62	0.39	444	0.02
Oral communication	1.50	0.50	270	1.48	0.50	444	0.03

Note. \*  $p < .05$ .

$F(1, 1053) = 16.19$ ,  $MSE = 1.09$ ,  $p < .05$ , and information processing,  $F(1, 1042) = 4.06$ ,  $MSE = 9.47$ ,  $p < .05$ , scores than females in the current study. However, the effect sizes for differences in cognitive ability and information processing were small. Sex differences in scores on written communication, planning and organizing, learns continuously, teamwork, and oral communication were not statistically significant.

Table 15 compares the means and standard deviations of the predictor measures across race groups. There was a significant main effect for race on cognitive ability scores,  $F(1, 1053) = 23.46$ ,  $MSE = 1.02$ ,  $p < .05$ . A Tukey post-hoc test revealed that African Americans scored significantly lower than Asians, Caucasians, and Hispanics on general cognitive ability. In addition, Caucasians scored significantly higher than Hispanics on general cognitive ability. Both Caucasians and Hispanics scored significantly higher than African Americans on information processing,  $F(1, 1042) = 12.77$ ,  $MSE = 9.09$ ,  $p < .05$ . Differences among race groups on written communication, planning and organizing, learns continuously, teamwork and oral communication were not statistically significant.

Table 15.

*Descriptive Statistics of Predictor Measures Across Race Groups*

	African Americans			Asians			Caucasians			Hispanics			$\eta^2$
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	
Cognitive ability	1.15	0.92	278	1.96	1.22	13	1.84	1.04	634	1.47	1.05	124	.08*
Information processing	4.95	3.03	273	6.77	3.32	13	6.47	2.95	629	5.90	3.27	123	.05*
Written communication	1.07	0.29	258	1.00	0.00	12	1.13	0.34	619	1.13	0.34	118	.01
Planning and organizing	1.75	0.35	155	1.75	0.38	8	1.72	0.35	473	1.68	0.39	76	.00
Learns continuously	1.68	0.47	155	1.75	0.46	8	1.64	0.48	473	1.74	0.44	76	.01
Teamwork	1.62	0.40	155	1.63	0.35	8	1.60	0.39	473	1.72	0.37	76	.01
Oral communication	1.45	0.50	155	1.63	0.52	8	1.50	0.50	473	1.45	0.50	76	.01

Note. \*  $p < .05$ .

Table 16.

*Differences in Predictor Scores for Race Groups as Compared to Caucasians*

	Standardized Mean Differences		
	African Americans	Asians	Hispanics
Cognitive ability	0.70*	0.11	0.36*
Information processing	0.88*	0.17	0.32
Written communication	0.11	0.32	0.00
Planning and organizing	0.05	0.05	0.07
Learns continuously	0.06	0.16	0.15
Teamwork	0.32	0.05	0.19
Oral communication	0.07	0.18	0.07

*Note.* \*  $p < .05$ .

Table 16 presents the effect sizes for the differences in predictor scores across race groups, using Cohen's  $d$  (Cohen, 1988; Rosnow & Rosenthal, 1996). There were moderate to large effects for the differences between Caucasian and African Americans on both cognitive ability and information processing. Based on these findings, additional analyses were conducted to examine differential criterion-related validity across race groups. No statistically significant differences in criterion-related validity across race groups were revealed.

Table 17 summarizes the pass rates for all AMAP participants, as well as across sex and race groups. There were 395 individuals who completed AMAP in 1999, 258 participants who completed AMAP in 2000 and 59 individuals who completed AMAP in 2001. There were statistically significant sex differences in AMAP pass rates,  $\chi^2(1, N=1368) = 4.73, p < .05$ . Specifically, more males and fewer females passed AMAP than expected. There were also statistically significant

differences in AMAP pass rates across race groups,  $\chi^2 (4, N=1368) = 28.19, p < .05$ . This suggests that the pass rates across race groups are significantly different from the expected pass rates. In particular, more Caucasians and fewer African Americans passed AMAP than expected. Chi-square results can also be influenced by large sample sizes, which may have contributed to the statistically significant findings. A calculation of adverse impact ratios revealed that AMAP displayed adverse impact for African Americans, as the pass rate for African Americans was 75.3% of the pass rate for Caucasians. Adverse impact occurs when pass rates for subgroups are less than 80% of the pass rate of the comparison group (e.g., Caucasian).

Table 17.

*AMAP Pass Rates for all AMAP Participants, Across Sex Groups and Across Race Groups*

	AMAP pass rate	Adverse Impact Ratio	Pass number	Fail number	N
<b>All AMAP participants</b>	<b>68.20%</b>		<b>712</b>	<b>332</b>	<b>1044</b>
Males	71.00%		269	110	379
Females	66.60%	93.8%	443	222	665
Caucasian	74.90%		471	158	629
African American	56.40%	75.3%	155	120	275
Asian	61.50%	82.1%	8	5	13
Hispanic	61.10%	81.6%	77	49	126

*Note.* For adverse impact analyses, females are compared against males for sex differences and race groups are compared to Caucasians.

*Hypothesis 2. Managers Selected with the Predictors in the Current Study Will Have More Satisfied Employees*

Table 18 compares the means and standard deviations of the criterion measures between AMAP and grandparented managers. Overall employee satisfaction was not statistically significantly different for employees reporting to AMAP managers, compared to employees reporting to grandparented managers. Unit performance, measured by percent of calls resolved on first attempt, hold time, conformance to schedule, adherence and absence, was not statistically different for AMAP managers, compared to unit performance metrics for grandparented managers. The current study did not provide support for Hypotheses 2, as overall employee satisfaction was not significantly different between teams managed by AMAP managers and grandparented managers.

Table 18.

*Comparison of AMAP Participants and Grandparented Managers on Criterion Measures*

	AMAP Participants			Grandparented managers			<i>d</i>
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	
<b>JOB PERFORMANCE</b>							
Job performance – Y1	11.28	1.88	410	11.46	1.45	294	-.11
What – Y1	3.08	0.44	410	3.04	0.33	294	.10
How - Y1	2.27	0.55	410	2.40	0.56	294	-.23*
Job performance – Y2	12.01	1.99	289	11.65	2.14	242	.17*
What – Y2	3.17	0.47	289	3.09	0.46	242	.17*
How – Y2	2.60	0.60	289	2.51	0.62	242	.15

Table 18. (Continued)

	AMAP Participants			Grandparented managers			<i>d</i>
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	
Job performance – Y3	12.33	1.64	150	11.96	1.97	255	.20*
What – Y3	3.23	0.44	150	3.14	0.48	255	.19
How – Y3	2.73	0.60	150	2.63	0.65	255	.16
<b>MERIT/BONUS</b>							
Merit – Year 1	1837.95	1463.98	361	2051.11	1362.09	45	-.15
Bonus – Year 1	1550.47	1138.06	419	2113.46	840.97	50	-.51*
Merit – Year 2	2157.68	1281.43	319	2863.30	1657.90	88	-.51*
Bonus – Year 2	2084.37	1703.37	394	2156.47	953.14	97	-.05
Merit – Year 3	2306.43	1452.74	220	2885.33	1723.09	242	-.36*
Bonus – Year 3	4314.52	2318.00	245	2692.30	1117.56	250	.89*
<b>TURNOVER</b>	0.22	0.42	1044	0.17	0.38	326	.12*
<b>UNIT PERFORMANCE</b>							
% of calls resolved on first attempt	77.13	2.34	30	77.06	1.07	15	.03
Hold time	84.73	10.10	30	80.40	8.68	15	.45
Conformance to schedule	94.31	.76	30	94.42	.61	15	-.15
Adherence	87.06	2.03	30	88.00	1.93	15	-.47
Absenteeism	11.14	.15	30	11.07	.19	15	.43
<b>EMPLOYEE SATISFACTION</b>							
Employee Satisfaction (composite)	76.49	9.44	93	77.88	9.22	47	-.15
<b>CUSTOMER SATISFACTION</b>							
Post (1-3 months)	6.84	.21	34	6.61	.09	2	1.11
Post (4-6 months)	6.85	.22	38	7.28	.21	3	-1.96
Post (7-9 months)	6.92	.20	55	6.95	.25	31	-.14
Post (10-12 months)	6.89	.21	56	6.93	.20	31	-.19
Post (13-15 months)	6.82	.20	65	6.95	.16	31	-.69*
Post (16-18 months)	6.77	.18	50	6.72	.17	40	.28
Post (19-21 months)	6.83	.18	40	6.73	.21	32	.52*
Post (22-24 months)	6.91	.17	34	6.79	.36	24	.45

Note. \*  $p < .05$ .

*Hypothesis 3. Managers Selected Using the Predictors in the Current Study Will Outperform Managers Selected Through Alternative Selection Systems*

Although grandparented managers had higher overall job performance ratings in Year One after AMAP, AMAP managers had significantly higher overall job performance ratings than grandparented managers in Years Two,  $F(1, 529) = 3.90$ ,  $MSE = 4.23$ ,  $p < .05$ , and Three,  $F(1, 403) = 3.59$ ,  $MSE = 3.45$ ,  $p < .05$ , following successful completion of AMAP. AMAP managers had significantly higher ‘what’ job performance ratings in Year Two after AMAP,  $F(1, 529) = 3.76$ ,  $MSE = .22$ ,  $p < .05$ . Grandparented managers had significantly higher ‘how’ job performance ratings in Year One after AMAP,  $F(1, 702) = 9.51$ ,  $MSE = .31$ ,  $p < .05$ . Although not statistically significant, AMAP managers also had higher ‘what’ and ‘how’ job performance ratings than grandparented managers in Years Two and Three after AMAP. The effect sizes of the differences in job performance ratings were relatively small, as noted in Table 18.

Grandparented managers received significantly larger merit increases than AMAP managers in Year Two,  $F(1, 405) = 18.27$ ,  $MSE = 1879780.32$ ,  $p < .05$ , and Year Three,  $F(1, 460) = 15.08$ ,  $MSE = 2560287.98$ ,  $p < .05$ , after AMAP. Grandparented managers also received significantly larger bonuses than AMAP managers in Year One after AMAP,  $F(1, 467) = 11.48$ ,  $MSE = 1233426.65$ ,  $p < .05$ . However, AMAP managers received significantly larger bonuses than grandparented managers in Year Three after AMAP,  $F(1, 493) = 98.97$ ,  $MSE = 3290108.14$ ,  $p < .05$ . The effect size for the difference in bonus payments in Year 3 was large ( $d = .89$ ).



AMAP managers also had significantly higher turnover rates than grandparented managers,  $F(1, 1376) = 2.19$ ,  $MSE = 0.17$ ,  $p < .05$ ; however the practical significance of this difference was small ( $d = .12$ ). Turnover for the overall sample was approximately 21%, with a 22% turnover rate among AMAP managers and a 17% turnover rate among grandparented managers.

The results from the current study provided partial support for Hypothesis 3, with AMAP managers receiving significantly higher overall job performance ratings than grandparented managers in Year Two and Year Three. AMAP managers also received higher ‘what’ job performance ratings than grandparented managers in Year Two, although grandparented managers received significantly higher ‘how’ job performance ratings in Year One. Grandparented managers received larger merit increases than AMAP managers in Year Two and Year Three, and received larger bonus payments in Year One. AMAP managers received significantly larger bonus payments than grandparented managers in Year Three. Turnover was higher among AMAP managers than grandparented managers, which was counter to Hypothesis 3. Unit performance was not significantly different for AMAP and grandparented managers.

*Research Question 2. Do Managers Selected Using the Predictors in the Current Study Tend to Have More Satisfied Customers (i.e., Higher Team-Based Customer Satisfaction Scores) than the Comparison Group?*

Grandparented managers had significantly higher overall customer satisfaction than AMAP managers 13 to 15 months following completion of AMAP,  $F(1, 94) =$

9.18,  $MSE = 0.33$ ,  $p < .05$ . AMAP managers had significantly higher overall customer satisfaction than grandparented managers 19 to 21 months following completion of AMAP,  $F(1, 70) = 4.66$ ,  $MSE = 0.04$ ,  $p < .05$ . The examination of Research Question 2 revealed that grandparent managers had significantly higher overall customer satisfaction in months 13-15 and AMAP managers had significantly higher overall customer satisfaction during months 19-21 following completion of AMAP. Although not statistically significant, overall customer satisfaction was higher for AMAP managers than grandparented managers from months 16-24, suggesting that customer satisfaction levels continued to improve for AMAP managers after the completion of AMAP.

*Hypothesis 4. Organizations that Have a Higher Proportion of Managers Selected with the Predictors in the Current Study Will Have Higher Levels of Performance than Organizations with Lower Proportions of Managers Selected with the Predictors in the Current Study.*

Table 19 compares the means and standard deviations of the criterion measures across organizations. There were no statistically significant differences between organizations on job performance.

Table 19.

*Means, SD, and N for Criterion Measures Across Channels/Organizations*

	Call Servicing			Customer Sales & Service			Global Sales & Service			Telemarketing			
AMAP Managers in organization (%)	73%			77%			95%			71%			
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	$\eta^2$
<b>JOB PERFORMANCE</b>													
Job performance – Y1	11.48	.97	86	11.24	1.93	401	11.47	.52	15	11.55	1.57	194	.01
What – Y1	3.06	.24	86	3.04	.45	401	3.00	.00	15	3.11	.372	194	.01
How - Y1	2.31	.49	86	2.30	.57	401	2.47	.52	15	2.36	.56	194	.01
Job performance – Y2	11.80	1.68	71	11.77	2.13	297	12.5	2.10	14	11.94	2.10	145	.00
What – Y2	3.08	.33	71	3.12	.49	297	3.21	.58	14	3.17	.46	145	.01
How – Y2	2.62	.59	71	2.53	.63	297	2.86	.54	14	2.55	.60	145	.01
Job performance – Y3	12.20	1.52	60	12.04	2.03	227	10.88	2.03	8	12.27	1.67	105	.01
What – Y3	3.17	.38	60	3.15	.49	227	3.00	.00	8	3.26	.48	105	.02
How – Y3	2.72	.62	60	2.70	.65	227	2.38	.74	8	2.57	.57	105	.02
<b>MERIT/BONUS</b>													
Merit – Year 1	2305.66	1218.36	53	2010.08	1490.13	248	933.33	975.46	12	1317.98	1354.92	89	.06*
Bonus – Year 1	2169.80	1037.79	56	1678.10	1082.71	294	862.42	1196.83	12	1133.20	997.80	102	.09*
Merit – Year 2	3038.30	1329.52	47	2054.04	1343.49	235	2200.00	1618.20	15	2569.45	1384.78	109	.06*
Bonus – Year 2	1788.54	1306.27	61	2163.11	1695.17	284	2161.33	1381.82	15	2101.15	1476.22	128	.01
Merit – Year 3	3208.62	1870.01	58	2515.90	1648.49	243	1500.00	1135.78	15	2637.59	1448.92	141	.03*
Bonus – Year 3	3176.83	1764.56	62	37.39.3 9	2094.42	262	3279.78	2643.06	18	3230.69	1750.08	148	.02
<b>TURNOVER</b>													
	.32	.47	182	.19	.39	702	.35	.48	109	.16	.37	366	.03

There were significant differences between organizations in merit increases in Year One,  $F(3, 398) = 8.74$ ,  $MSE = 2004185.94$ ,  $p < .05$ , Year Two,  $F(3, 402) = 8.58$ ,  $MSE = 1859293.26$ ,  $p < .05$ , and Year Three,  $F(3, 453) = 5.36$ ,  $MSE = 2580439.13$ ,  $p < .05$ . Tukey post-hoc tests revealed that managers in the Call Servicing organization received significantly larger merit increases in Year One than managers in the Global Sales and Service and Telemarketing organizations. Additionally, managers in the Customer Sales and Service organization received significantly larger merit increases than managers in the Telemarketing organization in Year One. In Year Two, managers in the Call Servicing and Telemarketing organizations received significantly larger merit increases than managers in the Customer Sales and Service organization. In Year Three, managers in the Call Servicing organization received significantly larger merit increases than managers in the Customer Sales and Service and Global Sales and Service organizations.

There were significant differences between organizations in bonus payments in Year One,  $F(3, 460) = 14.40$ ,  $MSE = 1128309.01$ ,  $p < .05$ , with Tukey post-hoc tests showing that managers in the Call Servicing organization received significantly larger bonus payments than managers in the other organizations in Year One. There were also significant differences in bonus payments across organizations Year Three,  $F(3, 486) = 2.82$ ,  $MSE = 3917327.38$ ,  $p < .05$ , with Tukey post-hoc tests revealing that managers in the Customer Sales and Service organization received significantly larger bonus payments than managers in the Telemarketing organization in Year Three.

There were significant differences among organizations in turnover,  $F(4, 1362) = 9.53$ ,  $MSE = 0.16$ ,  $p < .05$ . Tukey post-hoc tests indicated that turnover in the Customer Sales and Service organization was significantly lower than turnover in the Call Servicing and Global Sales and Service organizations. Also, turnover in the Telemarketing organization was significantly lower than turnover in the Call Servicing and Global Sales and Service organizations. Between-organization comparisons were not reported for employee satisfaction, unit performance or customer satisfaction, as that data was available only for managers in the Customer Sales and Service organization.

Hypothesis 4 focused on differences in performance across organizations, based on the proportion of AMAP managers in each organization. Although there were significant differences across organizations across several criterion measures, organizations with higher proportion of AMAP managers did not tend to have higher levels of performance. However, there were not substantial differences in the proportion of AMAP managers across organizations, which may have contributed to the lack of significant differences across organizations. Even when between-organizations comparisons focused on the two largest organizations in the current study, Customer Sales and Service and Telemarketing, Customer Sales and Service performed lower on many criterion measures than Telemarketing, although Customer Sales and Service had a higher proportion of AMAP managers.

Table 20 summarizes the differences scores on criterion measures between the Global Sales and Service organization and the other organizations, given that Global

Table 20.

*Summary of Effect Sizes of Differences in Criterion Measures Across Organizations as Compared to Global Sales and Service Organization*

	Standardized Mean Differences		
	Call Servicing	Customer Sales & Service	Telemarketing
<b>JOB PERFORMANCE</b>			
Job performance – Y1	.01	-.12	.05
What – Y1	.27	.09	.31
How - Y1	-.32	-.30	-.20
Job performance – Y2	-.40	-.34	-.27
What – Y2	-.34	-.18	-.08
How – Y2	-.41	-.53	-.52
Job performance – Y3	.83	.57	.82
What – Y3	.47	.31	.56
How – Y3	.54	.49	.33
<b>MERIT/BONUS</b>			
Merit – Year 1	1.16	.73	.29
Bonus – Year 1	1.23	.75	.27
Merit – Year 2	.60	-.11	.26
Bonus – Year 2	-.28	.00	-.04
Merit – Year 3	.98	.63	.80
Bonus – Year 3	-.05	.22	-.03
<b>TURNOVER</b>			
	-.06	-.40	-.48

Sales and Service has the highest proportion of AMAP managers. There were large effects in overall job performance ratings in Year Three between Call Servicing and Global Sales and Service ( $d = .83$ ) as well as Telemarketing and Global Sales and Service ( $d = .82$ ), such that Call Servicing and Telemarketing organizations had higher job performance ratings in Year Three than Global Sales and Service. There were large effects in both merit increases and bonus payments in Year One between Call

Servicing and Global Sales and Service, such that the Call Servicing organization had higher merit increases ( $d = 1.16$ ) and bonus payments ( $d = 1.23$ ) in Year One than Global Sales and Service. There were large effects in merit increases in Year Three between Call Servicing and Global Sales and Service ( $d = .98$ ) as well as Telemarketing and Global Sales and Service ( $d = .80$ ), such that Call Servicing and Telemarketing organizations had higher merit increases in Year Three than Global Sales and Service.

*Research Question 3. Using Customer Satisfaction as an Index of Organizational Performance, Does Customer Satisfaction Increase over Time as Managers are Selected Using the Predictors in the Current Study?*

A repeated measures ANOVA was conducted to examine the changes in customer satisfaction ratings for AMAP and grandparented managers over time. Table 21 summarizes the descriptive statistics from the repeated measure analysis. Due to limited availability of consistent customer satisfaction data over time, the repeated measures analysis focused on overall customer satisfaction scores between seven and eighteen months after AMAP completion.

Table 21.

*Descriptive Statistics for Overall Customer Satisfaction Scores in Repeated Measures**Analysis*

	AMAP Managers			Grandparented Managers			<i>d</i>
	Mean	SD	N	Mean	SD	N	
Overall Customer Satisfaction (Months 7-9)	7.00	0.17	24	6.91	0.24	18	.44
Overall Customer Satisfaction (Months 10-12)	6.97	0.15	24	6.94	0.20	18	.17
Overall Customer Satisfaction (Months 13-15)	6.93	0.18	24	6.99	0.17	18	-.34
Overall Customer Satisfaction (Months 16-18)	6.77	0.18	24	6.69	0.13	18	.50

The repeated measures analysis yielded a significant within-subjects main effect for overall customer scores,  $F(3, 120) = 24.56$ ,  $MSE = 0.02$ ,  $p < .05$ , suggesting that overall customer satisfaction scores changed over time. Neither the between-groups effect nor the within-subjects interaction between overall customer satisfaction scores and AMAP status were statistically significant. Although overall customer satisfaction scores were consistently higher for AMAP managers in the repeated measures analysis, overall customer satisfaction scores decreased over time for AMAP managers, whereas overall customer satisfaction scores increased between months 7 and 15 for grandparented managers, and then decreased from months 16 to 18.



Research Question 3 focused on changes in customer satisfaction over time between AMAP and grandparented managers, with the repeated measures analysis yielding mixed results. Overall customer satisfaction was consistently higher for AMAP managers than grandparented managers over time, but overall customer satisfaction decreased over time for AMAP managers, whereas overall customer satisfaction increased over time for grandparented managers in the repeated measures analysis.

*Research Question 4. Do Differences in Criterion-Related Validity Exist Between Organizations?*

Table 22 summarizes the predictor-criterion correlations for the Call Servicing, Customer Sales and Service, Global Sales and Service, and Telemarketing organizations. Employee satisfaction, unit performance, and customer satisfaction were not included in these analyses, as these criterion measures were not available for all organizations.

Table 22.

*Predictor-Criterion Correlations for Call Servicing, Customer Sales and Service, Global Sales and Service, and Telemarketing Organizations*

Predictor	Criterion	Call Servicing		Customer Sales and Service		Global Sales and Service		Telemarketing	
		<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Cognitive ability	JP - Year 1	-0.01	45	-0.09	264	0.03	9	-0.03	98
Information processing	JP - Year 1	0.20	44	0.01	259	0.23	9	0.02	95
Written communication	JP - Year 1	0.19	44	0.03	258	-- <sup>a</sup>	9	0.06	95
Planning & Organizing	JP - Year 1	0.25	44	0.07	259	0.63	9	0.07	95
Learns continuously	JP - Year 1	0.17	44	0.02	259	0.35	9	-0.04	95
Teamwork	JP - Year 1	0.12	44	0.06	259	-0.10	9	0.04	95
Oral communication	JP - Year 1	0.00	44	0.03	259	0.10	9	0.14	95
Cognitive ability	What Year 1	-0.09	45	-0.09	264	-- <sup>a</sup>	9	0.02	98
Information processing	What Year 1	0.14	44	0.03	259	-- <sup>a</sup>	9	-0.02	95
Written communication	What Year 1	0.26	44	0.04	258	-- <sup>a</sup>	9	0.08	95
Planning & Organizing	What Year 1	0.13	44	0.00	259	-- <sup>a</sup>	9	0.07	95
Learns continuously	What Year 1	0.08	44	0.06	259	-- <sup>a</sup>	9	-0.12	95
Teamwork	What Year 1	0.15	44	-0.02	259	-- <sup>a</sup>	9	-0.13	95
Oral communication	What Year 1	0.15	44	0.02	259	-- <sup>a</sup>	9	0.08	95
Cognitive ability	How Year 1	0.02	45	-0.01	264	0.03	9	-0.09	98
Information processing	How Year 1	0.17	44	0.02	259	0.23	9	-0.05	95
Written communication	How Year 1	0.03	44	-0.01	258	-- <sup>a</sup>	9	0.02	95
Planning & Organizing	How Year 1	0.27	44	0.07	259	0.63	9	-0.01	95
Learns continuously	How Year 1	0.19	44	-0.08	259	0.35	9	0.06	95
Teamwork	How Year 1	0.03	44	0.10	259	-0.10	9	0.10	95
Oral communication	How Year 1	-0.16	44	-0.05	259	0.10	9	0.06	95

Table 22. (Continued)

Predictor	Criterion	Call Servicing		Customer Sales and Service		Global Sales and Service		Telemarketing	
		<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Cognitive ability	JP – Year 2	0.31	30	-0.05	193	-0.48	9	-0.02	63
Information processing	JP – Year 2	0.30	29	-0.01	190	-0.20	9	0.05	60
Written communication	JP – Year 2	0.38	29	0.00	189		9	0.16	60
Planning & Organizing	JP – Year 2	0.22	29	0.03	190	0.67	9	0.02	60
Learns continuously	JP – Year 2	.121.	29	0.08	190	0.80	9	0.23	60
Teamwork	JP – Year 2	-0.07	29	0.00	190	-0.09	9	-0.05	60
Oral communication	JP – Year 2	0.17	29	0.03	190	0.22	9	0.11	60
Cognitive ability	What Year 2	0.09	30	-0.04	193	-0.45	9	-0.01	63
Information processing	What Year 2	0.25	29	0.00	190	-0.18	9	0.04	60
Written communication	What Year 2	0.34	29	-0.05	189		9	0.14	60
Planning & Organizing	What Year 2	0.14	29	-0.02	190	0.63	9	-0.05	60
Learns continuously	What Year 2	0.12	29	0.06	190	0.80	9	0.16	60
Teamwork	What Year 2	-0.05	29	-0.02	190	-0.10	9	-0.09	60
Oral communication	What Year 2	0.20	29	0.01	190	0.10	9	0.10	60
Cognitive ability	How Year 2	0.29	30	-0.02	193	-0.37	9	0.07	63
Information processing	How Year 2	0.17	29	0.05	190	-0.17	9	0.13	60
Written communication	How Year 2	0.16	29	0.07	189		9	0.14	60
Planning & Organizing	How Year 2	0.16	29	0.06	190	0.50	9	0.05	60
Learns continuously	How Year 2	0.04	29	0.04	190	0.47	9	0.08	60
Teamwork	How Year 2	-0.11	29	-0.01	190	0.00	9	0.01	60
Oral communication	How Year 2	0.03	29	0.00	190	0.47	9	0.15	60
Cognitive ability	JP – Year 3	0.54	22	-0.02	113	-- <sup>a</sup>	2	-0.02	21
Information processing	JP – Year 3	0.58	21	0.08	109	-- <sup>a</sup>	2	0.45	18
Written communication	JP – Year 3	0.12	21	0.12	108	-- <sup>a</sup>	2	-0.14	18
Planning & Organizing	JP – Year 3	0.54	21	0.04	109	-- <sup>a</sup>	2	0.14	18

Table 22. (Continued)

Predictor	Criterion	Call Servicing		Customer Sales and Service		Global Sales and Service		Telemarketing	
		<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Learns continuously	JP - Year 3	0.41	21	0.15	109	-- <sup>a</sup>	2	0.24	18
Teamwork	JP - Year 3	-0.08	21	0.07	109	-- <sup>a</sup>	2	0.39	18
Oral communication	JP - Year 3	0.13	21	-0.03	109	-- <sup>a</sup>	2	0.52	18
Cognitive ability	What Year 3	0.49	22	0.07	113	-- <sup>a</sup>	2	-0.14	21
Information processing	What Year 3	0.39	21	0.15	109	-- <sup>a</sup>	2	0.20	18
Written communication	What Year 3	-0.16	21	0.10	108	-- <sup>a</sup>	2	0.08	18
Planning & Organizing	What Year 3	0.16	21	0.03	109	-- <sup>a</sup>	2	0.19	18
Learns continuously	What Year 3	0.13	21	0.05	109	-- <sup>a</sup>	2	0.24	18
Teamwork	What Year 3	-0.38	21	0.10	109	-- <sup>a</sup>	2	0.31	18
Oral communication	What Year 3	-0.21	21	0.06	109	-- <sup>a</sup>	2	0.48	18
Cognitive ability	How Year 3	0.21	22	-0.14	113	-- <sup>a</sup>	2	0.25	21
Information processing	How Year 3	0.33	21	-0.10	109	-- <sup>a</sup>	2	0.54	18
Written communication	How Year 3	0.27	21	0.08	108	-- <sup>a</sup>	2	-0.40	18
Planning & Organizing	How Year 3	0.49	21	0.07	109	-- <sup>a</sup>	2	-0.13	18
Learns continuously	How Year 3	0.36	21	0.14	109	-- <sup>a</sup>	2	-0.06	18
Teamwork	How Year 3	0.23	21	0.02	109	-- <sup>a</sup>	2	0.15	18
Oral communication	How Year 3	0.34	21	-0.10	109	-- <sup>a</sup>	2	0.10	18
Cognitive ability	Merit Year 1	-0.01	40	-0.03	232	0.26	11	0.13	78
Information processing	Merit Year 1	0.07	40	0.06	231	0.17	11	0.20	78
Written communication	Merit Year 1	0.18	40	0.12	230	-- <sup>a</sup>	11	-0.10	78
Planning & Organizing	Merit Year 1	0.11	40	0.09	231	-0.58	11	0.08	78
Learns continuously	Merit Year 1	0.08	40	0.15	231	-0.47	11	-0.01	78
Teamwork	Merit Year 1	0.10	40	0.14	231	0.00	11	0.03	78
Oral communication	Merit Year 1	-0.12	40	0.00	231	0.27	11	0.20	78
Cognitive ability	Bonus Year 1	-0.25	43	-0.11	274	0.57	11	0.08	90

Table 22. (Continued)

Predictor	Criterion	Call Servicing		Customer Sales and Service		Global Sales and Service		Telemarketing	
		<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Information processing	Bonus Year 1	0.07	43	0.01	273	0.42	11	0.09	90
Written communication	Bonus Year 1	0.07	43	-0.04	272	-- <sup>a</sup>	11	-0.10	90
Planning & Organizing	Bonus Year 1	0.17	43	0.01	273	-0.59	11	0.00	90
Learns continuously	Bonus Year 1	0.11	43	0.09	273	-0.55	11	-0.11	90
Teamwork	Bonus Year 1	-0.27	43	0.24	273	-0.24	11	0.07	90
Oral communication	Bonus Year 1	-0.08	43	0.01	273	-0.12	11	0.04	90
Cognitive ability	Merit Year 2	0.42	27	-0.04	207	0.29	10	0.00	77
Information processing	Merit Year 2	0.34	27	0.03	206	0.13	10	-0.11	76
Written communication	Merit Year 2	0.06	27	-0.12	205	-- <sup>a</sup>	10	-0.01	76
Planning & Organizing	Merit Year 2	0.42	27	0.04	206	-0.53	10	-0.09	76
Learns continuously	Merit Year 2	0.33	27	0.10	206	-0.51	10	0.15	76
Teamwork	Merit Year 2	0.22	27	0.14	206	0.29	10	-0.06	76
Oral communication	Merit Year 2	0.30	27	0.12	206	-0.15	10	0.10	76
Cognitive ability	Bonus Year 2	0.06	39	-0.12	249	-0.08	10	0.00	96
Information processing	Bonus Year 2	0.00	39	-0.07	248	-0.29	10	0.02	95
Written communication	Bonus Year 2	-0.10	39	-0.48	247		10	-0.13	95
Planning & Organizing	Bonus Year 2	-0.05	39	0.01	248	-0.28	10	0.08	95
Learns continuously	Bonus Year 2	-0.14	39	0.04	248	-0.25	10	0.06	95
Teamwork	Bonus Year 2	0.06	39	-0.03	248	0.24	10	-0.12	95
Oral communication	Bonus Year 2	-0.02	39	-0.02	248	0.35	10	0.00	95
Cognitive ability	Merit Year 3	0.23	23	-0.09	141	-0.16	9	-0.04	54
Information processing	Merit Year 3	0.32	22	0.04	137	-0.18	9	0.06	51
Written communication	Merit Year 3	0.50	22	0.00	136	-- <sup>a</sup>	9	0.07	51
Planning & Organizing	Merit Year 3	0.28	22	0.00	137	0.15	9	-0.25	51
Learns continuously	Merit Year 3	0.28	22	0.00	137	0.05	9	-0.01	51
Teamwork	Merit Year 3	0.18	22	0.07	137	0.29	9	0.03	51

Table 22. (Continued)

Predictor	Criterion	Call Servicing		Customer Sales and Service		Global Sales and Service		Telemarketing	
		<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Oral communication	Merit Year 3	0.38	22	0.12	137	0.55	9	0.06	51
Cognitive ability	Bonus Year 3	0.40	26	-0.04	152	-0.21	12	0.07	62
Information processing	Bonus Year 3	0.24	25	0.03	148	0.00	12	0.19	59
Written communication	Bonus Year 3	0.07	25	0.02	147	-- <sup>a</sup>	12	-0.06	59
Planning & Organizing	Bonus Year 3	0.33	25	0.14	148	0.37	12	0.08	59
Learns continuously	Bonus Year 3	0.26	25	-0.03	148	0.30	12	0.13	59
Teamwork	Bonus Year 3	0.24	25	0.09	148	-0.22	12	0.38	59
Oral communication	Bonus Year 3	0.14	25	-0.05	148	0.33	12	0.22	59
Cognitive ability	Turnover	-0.02	134	-0.05	549	-0.02	103	0.01	264
Information processing	Turnover	0.00	133	-0.01	543	0.02	101	-0.04	261
Written communication	Turnover	-0.20	129	0.09	543	0.17	97	-0.06	250
Planning & Organizing	Turnover	0.15	79	0.08	408	0.10	63	0.06	161
Learns continuously	Turnover	-0.01	79	-0.03	408	-0.10	63	0.08	161
Teamwork	Turnover	-0.03	79	0.07	408	-0.03	63	0.01	161
Oral communication	Turnover	-0.01	79	0.03	408	0.04	63	0.02	161

Note. . JP – Year 1 = Overall job performance rating in Year One, JP – Year 2 = Overall job performance rating in Year Two, JP – Year 3 = Overall job performance rating in Year Three.

\* denotes  $p < .05$ .

<sup>a</sup> Correlation not computed because at least one of the variables was constant.

Table 23 summarizes the differences among predictor-criterion correlations between organizations. Paired comparisons were conducted for correlations between all predictors and job performance, merit increases, bonus payments, and turnover, as these criteria were available for all four organizations. There were significant differences for approximately 6 percent of the paired comparisons. These results provided interesting results for Research Question 4, with some differences in criterion-related validity across organizations.

Table 23.

*Comparisons of Predictor-Criterion Correlations Between Organizations*

Predictor	Criterion	Paired comparisons of predictor-criterion correlations between organizations					
		CS & CSS	CS & GSSO	CS & TDS	CSS & GSSO	CSS & TDS	GSSO & TDS
Cognitive ability	JP - Year 1	0.46	-0.08	0.11	-0.27	-0.48	0.13
Information processing	JP - Year 1	1.11	-0.07	0.97	-0.53	-0.04	0.51
Written communication	JP - Year 1	0.96	0.44	0.67	0.07	-0.30	-0.15
Planning & Organizing	JP - Year 1	1.10	-1.12	1.00	-1.64	0.02	1.61
Learns continuously	JP - Year 1	0.88	-0.45	1.10	-0.83	0.48	0.95
Teamwork	JP - Year 1	0.37	0.51	0.45	0.39	0.19	-0.32
Oral communication	JP - Year 1	-0.20	-0.23	-0.74	-0.16	-0.87	-0.09
Cognitive ability	What Year 1	0.05	-0.20	-0.55	-0.23	-0.92	-0.04
Information processing	What Year 1	0.68	0.32	0.86	0.06	0.39	0.05
Written communication	What Year 1	1.39	0.62	1.01	0.09	-0.36	-0.19
Planning & Organizing	What Year 1	0.76	0.29	0.29	0.00	-0.61	-0.17
Learns continuously	What Year 1	0.12	0.19	1.10	0.15	1.54	0.29
Teamwork	What Year 1	1.06	0.36	1.51	-0.06	0.86	0.30
Oral communication	What Year 1	0.81	0.35	0.37	0.04	-0.55	-0.20
Cognitive ability	How Year 1	0.17	-0.02	0.57	-0.09	0.64	0.28
Information processing	How Year 1	0.94	-0.13	1.20	-0.52	0.55	0.67
Written communication	How Year 1	0.21	0.07	0.06	-0.01	-0.21	-0.05
Planning & Organizing	How Year 1	1.24	-1.07	1.53	-1.63	0.66	1.79
Learns continuously	How Year 1	1.58	-0.40	0.68	-1.07	-1.14	0.72
Teamwork	How Year 1	-0.38	0.31	-0.35	0.48	-0.01	-0.47
Oral communication	How Year 1	-0.64	-0.59	-1.15	-0.36	-0.89	0.10



Table 23. (Continued)

Predictor	Criterion	Paired comparisons of predictor-criterion correlations between organizations					
		CS & CSS	CS & GSSO	CS & TDS	CSS & GSSO	CSS & TDS	GSSO & TDS
Cognitive ability	JP - Year 2	1.81	1.87	1.48	1.14	-0.20	-1.17
Information processing	JP - Year 2	1.52	1.13	1.09	0.46	-0.40	-0.59
Written communication	JP - Year 2	1.89	0.87	0.99	0.00	-1.07	-0.37
Planning & Organizing	JP - Year 2	0.96	-1.28	0.87	-1.89	0.03	1.83
Learns continuously	JP - Year 2	0.22	-2.18*	-0.49	-2.49*	-1.07	2.03*
Teamwork	JP - Year 2	-0.29	0.04	-0.08	0.20	0.28	-0.09
Oral communication	JP - Year 2	0.68	-0.12	0.25	-0.48	-0.54	0.27
Cognitive ability	What Year 2	0.59	1.25	0.43	1.07	-0.16	-1.09
Information processing	What Year 2	1.21	0.96	0.91	0.44	-0.25	-0.51
Written communication	What Year 2	1.88	0.77	0.88	-0.11	-1.23	-0.33
Planning & Organizing	What Year 2	0.74	-1.34	0.77	-1.84	0.19	1.84
Learns continuously	What Year 2	0.28	-2.17*	-0.19	-2.51*	-0.70	2.18*
Teamwork	What Year 2	-0.13	0.11	0.17	0.19	0.46	-0.02
Oral communication	What Year 2	0.90	0.22	0.40	-0.22	-0.62	-0.01
Cognitive ability	How Year 2	1.55	1.53	0.99	0.89	-0.61	-1.08
Information processing	How Year 2	0.59	0.76	0.16	0.54	-0.56	-0.72
Written communication	How Year 2	0.44	0.35	0.10	0.16	-0.45	-0.32
Planning & Organizing	How Year 2	0.44	-0.87	0.45	-1.17	0.10	1.17
Learns continuously	How Year 2	0.00	-1.04	-0.14	-1.14	-0.21	1.03
Teamwork	How Year 2	-0.52	-0.25	-0.53	-0.01	-0.11	-0.03
Oral communication	How Year 2	0.12	-1.08	-0.51	-1.24	-0.97	0.86
Cognitive ability	JP - Year 3	2.53*	-- <sup>a</sup>	1.90	-- <sup>a</sup>	-0.02	-- <sup>a</sup>
Information processing	JP - Year 3	2.27*	-- <sup>a</sup>	0.49	-- <sup>a</sup>	-1.47	-- <sup>a</sup>
Written communication	JP - Year 3	-0.01	-- <sup>a</sup>	0.73	-- <sup>a</sup>	0.93	-- <sup>a</sup>
Planning & Organizing	JP - Year 3	2.23*	-- <sup>a</sup>	1.34	-- <sup>a</sup>	-0.36	-- <sup>a</sup>
Learns continuously	JP - Year 3	1.11	-- <sup>a</sup>	0.54	-- <sup>a</sup>	-0.34	-- <sup>a</sup>

Table 23. (Continued)

Predictor	Criterion	Paired comparisons of predictor-criterion correlations between organizations					
		CS & CSS	CS & GSSO	CS & TDS	CSS & GSSO	CSS & TDS	GSSO & TDS
Teamwork	JP - Year 3	-0.59	-- <sup>a</sup>	-1.39	-- <sup>a</sup>	-1.22	-- <sup>a</sup>
Oral communication	JP - Year 3	0.66	-- <sup>a</sup>	-1.24	-- <sup>a</sup>	-2.19*	-- <sup>a</sup>
Cognitive ability	What Year 3	1.90	-- <sup>a</sup>	2.08*	-- <sup>a</sup>	0.83	-- <sup>a</sup>
Information processing	What Year 3	1.04	-- <sup>a</sup>	0.62	-- <sup>a</sup>	-0.18	-- <sup>a</sup>
Written communication	What Year 3	-1.02	-- <sup>a</sup>	-0.69	-- <sup>a</sup>	0.07	-- <sup>a</sup>
Planning & Organizing	What Year 3	0.53	-- <sup>a</sup>	-0.10	-- <sup>a</sup>	-0.61	-- <sup>a</sup>
Learns continuously	What Year 3	0.31	-- <sup>a</sup>	-0.34	-- <sup>a</sup>	-0.72	-- <sup>a</sup>
Teamwork	What Year 3	-1.94	-- <sup>a</sup>	-2.06*	-- <sup>a</sup>	-0.82	-- <sup>a</sup>
Oral communication	What Year 3	-1.07	-- <sup>a</sup>	-2.13*	-- <sup>a</sup>	-1.71	-- <sup>a</sup>
Cognitive ability	How Year 3	1.42	-- <sup>a</sup>	-0.15	-- <sup>a</sup>	-1.58	-- <sup>a</sup>
Information processing	How Year 3	1.76	-- <sup>a</sup>	-0.75	-- <sup>a</sup>	-2.57*	-- <sup>a</sup>
Written communication	How Year 3	0.76	-- <sup>a</sup>	1.99*	-- <sup>a</sup>	1.81	-- <sup>a</sup>
Planning & Organizing	How Year 3	1.82	-- <sup>a</sup>	1.91	-- <sup>a</sup>	0.75	-- <sup>a</sup>
Learns continuously	How Year 3	0.95	-- <sup>a</sup>	1.24	-- <sup>a</sup>	0.69	-- <sup>a</sup>
Teamwork	How Year 3	0.86	-- <sup>a</sup>	0.24	-- <sup>a</sup>	-0.48	-- <sup>a</sup>
Oral communication	How Year 3	1.76	-- <sup>a</sup>	0.71	-- <sup>a</sup>	-0.72	-- <sup>a</sup>
Cognitive ability	Merit Year 1	0.15	-0.70	-0.67	-0.83	-1.21	0.37
Information processing	Merit Year 1	-0.34	-0.45	-1.01	-0.32	-1.07	-0.08
Written communication	Merit Year 1	0.33	0.45	1.39	0.33	1.66	0.28
Planning & Organizing	Merit Year 1	0.14	1.98*	0.15	2.08*	0.05	-2.00*
Learns continuously	Merit Year 1	-0.38	1.52	0.42	1.84	1.14	-1.37
Teamwork	Merit Year 1	-0.26	0.25	0.32	0.40	0.84	-0.09
Oral communication	Merit Year 1	-0.69	-1.02	-1.59	-0.76	-1.48	0.21

Table 23. (Continued)

Predictor	Criterion	Paired comparisons of predictor-criterion correlations between organizations					
		CS & CSS	CS & GSSO	CS & TDS	CSS & GSSO	CSS & TDS	GSSO & TDS
Cognitive ability	Bonus Year 1	-0.87	-2.32*	-1.76	-2.10*	-1.53	1.53
Information processing	Bonus Year 1	0.33	-0.97	-0.13	-1.21	-0.66	0.95
Written communication	Bonus Year 1	0.64	0.18	0.87	-0.11	0.46	0.26
Planning & Organizing	Bonus Year 1	0.97	2.21*	0.89	1.93	0.05	-1.86
Learns continuously	Bonus Year 1	0.14	1.86	1.15	1.94	1.59	-1.36
Teamwork	Bonus Year 1	-3.06*	-0.08	-1.78	1.36	1.44	-0.84
Oral communication	Bonus Year 1	-0.50	0.11	-0.60	0.36	-0.24	-0.43
Cognitive ability	Merit Year 2	2.23*	0.33	1.89	-0.88	-0.27	0.77
Information processing	Merit Year 2	1.52	0.51	1.98*	-0.28	1.01	0.62
Written communication	Merit Year 2	0.84	0.14	0.29	-0.32	-0.82	0.03
Planning & Organizing	Merit Year 2	1.88	2.41*	2.25*	1.64	0.91	-1.28
Learns continuously	Merit Year 2	1.09	2.08*	0.80	1.71	-0.35	-1.78
Teamwork	Merit Year 2	0.38	-0.16	1.23	-0.40	1.52	0.91
Oral communication	Merit Year 2	0.89	1.08	0.92	0.71	0.18	-0.63
Cognitive ability	Bonus Year 2	0.99	0.34	0.30	-0.10	-0.97	-0.20
Information processing	Bonus Year 2	0.38	0.71	-0.10	0.58	-0.70	-0.79
Written communication	Bonus Year 2	2.40*	-0.23	0.16	-1.36	-3.24*	0.32
Planning & Organizing	Bonus Year 2	-0.34	0.57	-0.64	0.77	-0.53	-0.92
Learns continuously	Bonus Year 2	-1.00	0.28	-1.01	0.77	-0.16	-0.80
Teamwork	Bonus Year 2	0.50	-0.44	0.90	-0.70	0.72	0.91
Oral communication	Bonus Year 2	0.02	-0.94	-0.11	-1.02	-0.20	0.93

Table 23. (Continued)

Predictor	Criterion	Paired comparisons of predictor-criterion correlations between organizations					
		CS & CSS	CS & GSSO	CS & TDS	CSS & GSSO	CSS & TDS	GSSO & TDS
Cognitive ability	Merit Year 3	1.35	0.86	1.07	0.19	-0.26	-0.28
Information processing	Merit Year 3	1.22	1.10	1.02	0.52	-0.13	-0.55
Written communication	Merit Year 3	2.24*	1.17	1.75	0.00	-0.45	-0.17
Planning & Organizing	Merit Year 3	1.15	0.29	2.00*	-0.35	1.54	0.94
Learns continuously	Merit Year 3	1.20	0.52	1.13	-0.12	0.06	0.14
Teamwork	Merit Year 3	0.46	-0.25	0.55	-0.55	0.22	0.61
Oral communication	Merit Year 3	1.14	-0.49	1.23	-1.22	0.32	1.30
Cognitive ability	Bonus Year 3	2.08*	1.61	1.44	0.49	-0.72	-0.78
Information processing	Bonus Year 3	0.92	0.62	0.19	0.10	-1.03	-0.54
Written communication	Bonus Year 3	0.22	0.17	0.51	0.05	0.50	0.17
Planning & Organizing	Bonus Year 3	0.91	-0.12	1.07	-0.74	0.39	0.87
Learns continuously	Bonus Year 3	1.33	-0.09	0.54	-0.98	-1.07	0.47
Teamwork	Bonus Year 3	0.71	1.18	-0.62	0.89	-2.02*	-1.74
Oral communication	Bonus Year 3	0.83	-0.50	-0.30	-1.13	-1.70	0.34
Cognitive ability	Turnover	0.32	0.00	-0.30	-0.29	-0.84	-0.27
Information processing	Turnover	0.07	-0.18	0.30	-0.28	0.33	0.47
Written communication	Turnover	-2.96*	-2.75*	-1.31	-0.73	1.94	1.90
Planning & Organizing	Turnover	0.53	0.26	0.64	-0.15	0.26	0.29
Learns continuously	Turnover	0.20	0.52	-0.61	0.47	-1.17	-1.16
Teamwork	Turnover	-0.80	-0.03	-0.26	0.68	0.68	-0.20
Oral communication	Turnover	-0.29	-0.28	-0.23	-0.09	0.04	0.11

Note. \* denotes p<.05.

<sup>a</sup> Comparison not computed because at least one of the correlations was not calculated due to constant variable.

## DISCUSSION AND CONCLUSIONS

The objective of the current study was to provide an initial empirical test of the Schneider et al. (2000) model. The Schneider et al. (2000) model expanded the traditional approach to validating selection systems to include the impact that selection systems have on the broader organizational system. The current project provided an empirical test of this model by extending the traditional individual-differences approach to validation research and including group- and organization-level criteria (e.g., unit-level performance and customer satisfaction). Using a quasi-experimental design, archival data from a managerial development and selection program were analyzed to examine several relationships proposed in the Schneider et al. (2000) model.

Overall, results from the current study provided limited support for the Schneider et al. (2000) model. At the individual level, there was a promising trend such that there were greater improvements in job performance by AMAP managers over time compared to grandparented managers, which suggests that the selection system in the current study may have contributed to improved aggregate performance of managers over time. However, inconsistent with the study's hypotheses and the extant literature, the predictors used in the current study did not display consistent patterns of criterion-related validity at the individual level.

Support was also limited for group- and organization-level outcomes. One interesting finding was that the interpersonally related constructs used in the current study's selection system were predictive of the length of time customers were kept on

hold during interactions with the managers' customer service teams (i.e., a measure of unit-level performance). In particular, fostering an environment in which openness and candor is valued along with effectively managing diverse sets of individuals may result in employees being more responsive to the needs of customers. These results are similar to findings by Schneider and Bowen (1985), as employees may perceive the interpersonal behaviors of their manager as positive supervision, thus facilitating performance of employees, such that employees are more focused on serving customers as a result. This suggests that the same set of human resources practices may be related to positive perceptions for both employees and customers. Lacking broader trends related to other group- and organization-level outcomes in the current study, additional research on this finding would be useful.

There were several limitations in the current study associated with the use of archival data, which will be discussed below in further detail. However, the limited availability of some of the criterion measures provided an initial indication of practical problems associated with empirically testing the model. While intuitively appealing, testing the Schneider et al. model (2000) in applied settings may prove to be a practical challenge because of the nature and complexity of the data required to do so.

#### *Discussion of Hypotheses and Research Questions*

Hypothesis 1 and Research Question 1 served as a test of Link 1 of the Schneider et al. (2000) model, by exploring the relationship between individual differences and individual performance. The current study provided limited support for Hypothesis 1; few predictors were significantly correlated with criterion measures.

In personnel selection research, cognitive ability is regarded as the strongest predictor of job performance with criterion-related coefficients consistently as high as .50 (e.g., Schmidt & Hunter, 1998). In the current study, however, cognitive ability was not strongly correlated with any of the criterion measures. In addition, there were few meaningful patterns of correlations across the other predictor-criterion relationships.

Schneider et al. (2000) noted that human resources systems in an organization can be effective vehicles for communicating the values of the organization. In the current study, the company's core leadership competencies were included in the selection system. These constructs were also part of the performance management system, as the leadership competencies were assessed through the 'how' performance appraisal ratings. It was disappointing that assessments of these constructs during the selection process were not predictive of future performance on the same constructs. However, it is possible that managers completed additional training in the leadership competencies once they were in managerial roles, which may explain the lack of strong relationships between these constructs in the selection system and individual-level performance measures in the current study. If this is the case, then although the results are not supportive of the hypotheses investigated in the current study, there may be positive implications regarding the training of such competencies as an avenue to improve managerial performance.

Research Question 1 assessed another aspect of Link 1, namely whether any of the predictors in the current study had the potential for displaying adverse impact. Pass rates for African Americans were less than 80% of Caucasians on the overall

selection system, thus not meeting the 4/5<sup>th</sup> rule and suggesting that some predictors may be displaying adverse impact. Specifically, Caucasians scored significantly higher on the cognitive ability measure than African Americans and Hispanics. These findings also have some practical significance, given the moderate to large effect sizes associated with the differences between Caucasians and African Americans on the cognitive ability and information processing measures.

The adverse impact associated with cognitive ability measures has been of concern to practitioners and researchers for some time. Previous research on the role of personality in selection systems has suggested that including predictors that assess interpersonal skills in addition to measures of cognitive ability in the selection system may reduce adverse impact of the overall system (e.g., Hogan, 1991), but that concept was not supported in the current study. Interestingly, Hough and Oswald (2000) highlighted meta-analytic research that counters Hogan's (1991) assertion. In particular, the 4/5<sup>th</sup> rule was almost always satisfied when cognitive ability was excluded from selection batteries, whereas the rule was almost never satisfied when selection batteries included cognitive ability alone or in a composite (Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Although the current study employed predictors that assessed cognitive ability as well as interpersonal skills, the hybrid multiple-hurdle approach required participants to achieve certain scores on the measures of cognitive ability in Phase One before they could go through the interview in Phase Two, which assessed interpersonal skills. Murphy (2002) noted that little progress has been made in identifying methods that would allow researchers and practitioners to



take advantage of the validity of cognitive ability tests in personnel selection without producing substantial adverse impact. Given the widespread use of cognitive ability measures in selection and the legal risks associated with adverse impact, additional research on reducing the adverse impact associated with measures of cognitive ability will provide valuable insight to both researchers and practitioners.

Hypothesis 2 examined Link 2A, the relationship between individual differences and organization differences, with the expectation that managers selected with the predictors in the current study would have more satisfied employees. Hypothesis 2 was not supported, such that overall employee satisfaction was not significantly different between teams managed by AMAP managers and grandparented managers. In fact, grandparented managers had more satisfied employees than AMAP managers. However, because employee satisfaction was examined at the same point in time for both AMAP and grandparented managers, managerial experience may have played a role in these results. Although not a direct test of link 2A, the interpersonally related predictors (planning and organizing, learns continuously, teamwork, oral communication) explained an incremental proportion of variance in hold time above and beyond the cognitive-related predictors (cognitive ability, information processing, and written communication). Interestingly, these results suggest that the interpersonally related constructs may be useful predictors of the length of time customers may be on hold when interacting with employees reporting to the participants in the current study. This finding is consistent with the concept that managers may be role models in establishing interpersonal norms in the

workplace, in which case managers who have strong interpersonal skills may foster a work environment that places value on employees also demonstrating strong interpersonal skills, which may transfer to improved interactions with customers.

Link 4A of the Schneider et al. (2000) model examined the relationships between organizational differences and aggregated individual performance, or whether different selection practices result in differences in aggregate levels of performance. Link 4A was tested by Hypothesis 3 and 4. Specifically, Hypothesis 3 posited that managers selected using the predictors in the current study would outperform managers selected through alternative selection systems, and Hypothesis 4 posited that organizations with higher proportions of managers selected with the predictors in the current study would have higher levels of performance than organizations with lower proportions of managers selected with the same predictors. Results provided partial support for Hypothesis 3, in that AMAP managers had greater performance improvements over time compared to grandparented managers. Although the grandparented managers displayed higher performance as a group in the first year after AMAP, the AMAP managers displayed higher job performance in the subsequent years. Trends in job performance ratings suggest that AMAP managers continued to improve their job performance after completing AMAP. It is also plausible that increased levels of professional experience could have contributed to higher job performance ratings for grandparented managers in Year One. These results suggest that the selection system results in improved aggregate levels of individual performance, which tend to be the type of performance that is of most concern to

managers in organizations (rather than individual differences in performance; Schneider et al., 2000).

In contrast to the performance improvement, grandparented managers tended to receive more financial rewards in the form of merit increases and bonus payments than managers selected with the new selection system. Although managers selected with the new selection system received higher merit increases and bonus payments each year, the grandparented managers consistently received higher financial rewards than the AMAP managers. The differences in merit increases and bonus payments may be related to salary differences between AMAP and grandparented managers. Given that AMAP managers were promoted into managerial positions during the current study, it stands to reason that AMAP managers may have had lower salaries than grandparented managers. As a result, AMAP managers may have received merit increases and/or bonus payments that were a higher percentage of salary compared to grandparented managers, yet smaller in total dollars. It would be interesting to explore the comparisons between groups on merit increases and bonus payments as a percentage of salary in future research.

Turnover was higher among AMAP managers than grandparented managers, which was counter to Hypothesis 3. Unit performance was not significantly different for AMAP and grandparented managers. External factors, such as increased competition in the telecommunications industry in the last decade, may have contributed to the increased turnover among AMAP managers, although one would expect these external factors to affect both groups of managers equally. On a related

note, the competition in the telecommunications industry also may have contributed to the lack of significant differences in unit performance between AMAP and grandparented managers.

Hypothesis 4 focused on differences in performance across organizations, based on the proportion of AMAP managers in each organization. Organizations with a higher proportion of AMAP managers did not tend to have higher levels of performance, although there were significant differences across organizations across some criterion measures. However, the proportion of AMAP managers was not drastically different across the four organizations, with proportions of AMAP managers ranging from 71% to 95%. There were also wide ranges in the number of managers in each of the four organizations in the current study. Global Sales and Service had the highest proportion of AMAP managers, yet this was the smallest organization in the current study ( $N = 25$ ). As noted by Schneider (1978), the ranges of a given group attribute available for comparing groups are often more moderate in field research than in more controlled laboratory settings.

Research Question 2 examined Link 3 of the Schneider et al. (2000) model, which focused on the relationship between organizational differences and organizational performance, by exploring whether managers selected using the predictors in the current study had more satisfied customers than the comparison group of grandparented managers. Although the results were not statistically significant, there were some promising trends in the data such that customer satisfaction continued to improve for AMAP managers after the completion of AMAP.

One implication of this finding is that AMAP managers may have served as effective role models of customer service behavior, such that employees changed and improved their interactions with customers over time.

Link 5 of the Schneider et al. (2000) model examined the relationship between aggregated individual differences and organizational performance, suggesting that organizational performance is enhanced when valid personnel selection procedures are used. Research Question 3 examined this aspect of the model by examining trends in customer satisfaction over time for AMAP and grandparented managers. The results for this component of the model were mixed; overall customer satisfaction was consistently higher for AMAP managers than grandparented managers over time, but overall customer satisfaction decreased over time for AMAP managers, whereas overall customer satisfaction increased over time for grandparented managers. These results should be interpreted with caution; only a small number of AMAP and grandparented managers were included in this analysis because few managers had complete sets of customer satisfaction data across the time periods included in the analysis. More complete customer satisfaction data would have allowed for a more robust test of the relationship between selection processes and customer satisfaction over time.

Although these results are somewhat disappointing, there are a few possible explanations for these trends in customer satisfaction. One possibility is that the grandparented managers could have received more challenging customer satisfaction goals, given their experience level – either in terms of quantitative goals or those

teams may have been asked to deal with more difficult customers. Another possibility is that there were other factors beyond the control of the managers that impacted the customers' overall satisfaction, such as the cost of the product or service they purchased. Ryan and Ployhart (2002) noted that customers may have negative perceptions of service even when employees and managers performed in an organizationally approved manner (e.g., not always providing free products or services). It is also important to note that other factors may have influenced overall customer satisfaction beyond the managerial selection process.

Link 6 of the model suggested that organizational differences may moderate the validities of selection strategies. Research Question 4 explored this portion of the model by examining differences in criterion-related validity between organizations. The results from the current provided limited support for Link 6 of the Schneider et al. (2000) model, although the results indicated some differences in criterion-related validity across organizations. These findings may be consistent with Schneider (1978), as a lack of moderator effects in field research could be attributed to the fact that moderators are most effective when extremes are available. Neither extremes of situations nor extremes of individual differences are likely to be found in field research.

Along these lines, it is possible that the current study did not provide robust support for differences in performance (Hypothesis 4) or criterion-related validity between organizations (Research Question 4) because the organizations in the current study all provided services to consumers and were part of the same division within the

company. The larger business purpose of serving consumers, whether through operator services or providing service related to a particular product or calling plan, may have superseded the differences among the organizations.

In summary, results in support of the Schneider et al. (2000) multilevel model of personnel selection were limited. However, this was the first empirical test of this model, and there were some interesting findings as previously noted. Despite the limited support for the overall model, there are several implications that can be gleaned from this study.

#### *Research and Practical Implications*

The results of the current study offer research and practical implications for consideration. First, whereas the Schneider et al. (2000) model is quite intuitively appealing, the model may need additional development. For example, the model does not take into account the investment of developing the selection system. One suggestion for the Schneider et al. (2000) model would be to include utility analysis along with individual-, group-, and organization-level outcomes. Including utility analysis findings as an indicator of the organization's return on investment may be a useful complement to group- and organization-level outcomes. From a practical standpoint, including utility analysis information along with organization-level outcomes with which decision makers are familiar (e.g., customer satisfaction), may help organizational decision makers better understand and appreciate the value of utility analysis (e.g., Carson et al., 1998).

The current study also highlights the importance of developing research questions and hypotheses prior to conducting research. Given the archival nature of the current study, it was not possible to design and obtain all of the criterion measures before the selection system was administered. However, the limited availability of some criterion measures left some questions in the current study unanswered. Thus, it would be helpful to use the Schneider et al. (2000) model as a guiding framework when developing both predictors and criteria that would be used to evaluate the impact of a selection system at individual, group, and organizational levels.

There are a number of implications for practitioners beyond the research implications previously noted, beginning with the importance of job analysis in the development and validation of selection systems. Given the archival nature of the current study, it is not clear what sort of job analysis was conducted prior to developing the selection system. Job analysis is the foundation for many human resources systems, and it is exceedingly important to conduct robust job analyses on an ongoing basis to ensure that job descriptions are current. Job analysis provides an opportunity to articulate the most important components of a given job, as the primary reason for job analysis is to gain an understanding of the nature and purpose of the job. Job analysis also informs the development of selection system, such that changes in major tasks or KSAs related to a job should be reflected in the selection process for the job. It is also important for selection systems to assess all of the key skills and attributes needed to for success in the job.



The current study also highlights the importance of linking HRM practices to organizational outcomes. Although the results of the current study may appear disappointing in that regard, it is hoped that these results will inform future efforts by practitioners to demonstrate the value of HRM practices at individual, group, and organizational levels.

Practitioners should also continue to consider how to combat the potential for adverse impact in selection tools that assess cognitive ability. In addition to the legal concerns regarding adverse impact, adverse impact has other more intangible trade-offs, such as the potential for fewer diverse perspectives to be present in an organization. Even with technological advancements related to selection systems (e.g., online selection tools), the potential for adverse impact still exists as minority groups that live in socio-economically deprived areas may have substantially lower levels of access to (and familiarity with) the Internet (Anderson, 2002).

Practitioners are also encouraged to spend adequate time on developing psychometrically sound and meaningful criteria. The limited availability of detailed criterion measures in the current study suggests that practitioners have not adequately addressed the criterion problem (Austin & Villanova, 1992). Although much of the criterion development research focuses on individual-level outcomes, the same level of rigor is needed in the development of group- and organization-level outcomes.

The current study also highlights the importance of including managers and decision makers in the development of selection processes, as well as considering outcome measures by which the selection system could be evaluated. Practitioners are

also encouraged to continually educate organizational decision makers, both within the HR organization as well as business leaders, on the importance of selection systems. Educating others on the importance of robust selection systems is an ongoing challenge for practitioners, as many organizational leaders do not understand the payoff associated with the investment to develop selection systems. Providing regular updates to managers and decision makers in ongoing discussions during the development of the selection process and outcome measures may help them better understand the importance of selection processes, as well as the importance of considering managerial performance in a more holistic manner (i.e., considering individual-, group-, and organization-level outcomes).

Another practical implication is that ongoing leadership support would be a key driver in sustaining a research effort of this magnitude in a field setting. The degree of leadership buy-in and support for the new selection system may have influenced the results of the current study. There were varying degrees of leadership support for the selection system in the current study, as well as other HRM systems, which certainly impacted the daily activities of managers within the organizations participating in the current study.

### *Limitations*

Further, there were some limitations that may have contributed to the lack of support for the model, which are discussed below. Given that the results from selection tools in the current study were processed by a third-party vendor, item-level data were not available for the predictors. It is unclear to what degree the

psychometric properties of the predictors contributed to the results obtained in the current study. Given that reliability provides the upper limit for validity estimates, it is unclear if the low criterion-related validity estimates in the current study were a result of unreliable predictor measures. Future research efforts in this area should make every effort to obtain item-level data for the predictor measures. Without item-level data for the predictor measures, it is also difficult to know if or how predictor measures could be modified or improved.

Range restriction on predictor and criterion measures may have limited the findings of the current study. Managers who were selected into managerial positions were the focus of the current study, thus attenuating the range of predictor scores. In addition, many of the predictors were assessed on a 3-point scale, and some of the job performance measures were assessed on a 4-point scale, which limited the amount of variance in these measures. Range restriction of performance appraisal ratings due to political influences in organizations have been widely researched (e.g., Murphy & Cleveland, 1995), and may have impacted the performance review scores in the current study as well. These factors may have contributed to the low validity coefficients in the current study, as narrowing the range of either the predictor or criterion lowers the size of the validity coefficient (Cascio, 1998). The current study was not able to control for tenure or experience. As a result, it was not possible to explore how experience may have contributed to the differences in performance between AMAP and grandparented managers.

It is possible that the predictors in the current study may not have addressed all of the core components of a managerial job in a service environment. The research on customer service behavior offers some good suggestions. Given the role that managers may play in creating and maintaining a certain work environment for their organizations, it may be just as important to assess service orientation when selecting managers as it is when selecting customer service representatives. It also may have been helpful to assess other personal attributes during the selection process. Research by George (1991, 1992, 2000) suggests that leaders affect how their managers feel; thus, it may be beneficial to include a predictor that assesses positive affectivity, as managers with positive outlooks may have a positive influence on interactions with their employees, as well as interactions among employees. Other research has suggested that highly empathic individuals may lead to greater responsiveness to customer needs (Rogers, Clow, & Kash, 1994). Given that relationship, it may be helpful to include empathy in selecting managers in customer service organizations, as managers often serve as role models for employees.

There were also a number of limitations in the current study that are associated with field research in general. Some criterion measures, such as metrics of unit performance, were available only for subsets of the sample in the current study. As a result, the current study was unable to provide a simultaneous statistical test of the Schneider et al. (2000) model (e.g., structural equation modeling, hierarchical linear modeling). On a related note, it is possible the amount of missing customer satisfaction data may have negatively impacted the results of analyses involving

customer satisfaction. Subsequent studies would benefit from robust customer satisfaction data.

There are reasons behind the lack of criterion data for all participants in the current study. In particular, some criterion measures were used only by a portion of the company in the current study (e.g., unit performance measures).

The records from the AMAP process were not complete, as the predictor data were often stored by local offices and business managers were responsible for entering and maintaining the database. Given multiple priorities facing managers, the reality is that they did not always provide complete and accurate data. The researcher of the current study spent a tremendous amount of time trying to reconcile potential errors in the predictor database prior to analysis, but the possibility of data entry error still exists and may have had some influence in the results. Although features of the data may have served as limitations in the current study, these limitations also offer several fruitful avenues of future research.

#### *Future Research*

Further research would be appropriate to address the limitations of the current study. As previously noted, the use item-level predictor data in future studies will allow an assessment of the reliability of predictor measures. This information will provide insight related to the validity of the predictors and may highlight items that should be revised for future administrations of the selection system.

In future research efforts, greater contrasts across organizations in terms of proportion of managers selected through a new process, along with more robust and

equivalent numbers of managers across organizations, may yield more meaningful comparison groups for analysis of this concept. This could be achieved through a pilot test of a selection system or comparing performance of portions of a company that conduct different types of work (e.g., manufacturing and sales organizations within the same company).

Future research may want to examine additional predictor measures in the selection of managers in service environments. Additional predictors may be included in future research efforts to predict managerial performance via the Schneider et al. (2000) model. For example, it may be beneficial to include a predictor that assesses positive affectivity, as managers with positive outlooks may have a positive influence on interactions with their employees, as well as interactions among employees (George, 1991). Other research has suggested that highly empathic individuals may lead to greater responsiveness to customer needs (Rogers et al., 1994). Given that relationship, it may be helpful to include empathy in selecting managers in customer service organizations, as managers often serve as role models for employees.

Given the emphasis organizations have placed on person-organization fit in recent years, it would be helpful to examine the predictive validity of person-job fit over person-organization fit in future research efforts. Organizational decision makers appreciate the intuitive appeal of selecting qualified candidates who also share values, beliefs, and interests similar to that of the organization. On a related note, as organizational flexibility in effectively utilizing employees increasingly becomes an issue (e.g., employees are moving from job to job in organizations more often), the

person-organization model may be more relevant in comparison with the traditional person-job fit approach (e.g., Kristof, 1996). However, recent research posited that person-organization fit may not clearly demonstrate job-relatedness, which is critical in the legal defensibility of selection systems (Arthur et al., 2004). Thus, additional research on person-organization fit as a selection tool will inform practical applications of this construct in selection systems.

Future research efforts should also give thoughtful consideration to the criterion measures prior to administering the selection process, to allow for a simultaneous statistical test of the Schneider et al. (2000) model. The trends in job performance, merit increases, and bonus payments for AMAP managers in the current study suggest that AMAP managers improved performance over time. It may be informative to examine trends in criterion measures at all levels of analysis over time to examine if similar patterns exist for group- and organization-level criteria. Although the customer satisfaction trends over time in the current study are promising, these analyses should be replicated with larger samples in future research efforts.

On a similar note, the analysis of employee satisfaction in the current study included only a subset of the sample in the current study; future research should examine this relationship with a larger sample. It would also be interesting to examine comparisons of employee satisfaction over time, for a longitudinal analysis may provide more insight on a manager's influence on the work environment than employee satisfaction from a singular point in time.

Future research could also explore additional measures of group and organizational performance, in the context of the Schneider et al. (2000) model. For example, it would be interesting to include employee perceptions of service climate as a component of group performance. It would also be interesting to explore additional organization-level criteria, such as sales and customer retention. These criteria would allow for a continued exploration of the relationship between service climate perceptions of employees and organizational outcomes (e.g., Schneider & Bowen, 1985). The inclusion of additional criteria would also provide a more complete picture of managerial performance, particularly as it relates to fostering a climate of customer service.

### *Conclusions*

The present study offered an initial test of a promising conceptual model in personnel selection validation research. Although the current study provided limited support for the Schneider et al. (2000) multilevel model of personnel selection, there were some interesting findings that warrant additional examination. Findings from the current study may be informative for both researchers and practitioners. Additional research and testing of this model could offer substantial contributions to both science and practice within the field of I/O psychology.



## REFERENCES

- Adams, G. A., Elacqua, T. C., & Collarelli, S. M. (1994). The employment interview as a sociometric selection technique. *Journal of Group Psychotherapy, Psychodrama, and Sociometry, 47*, 99-113.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment, 11*, 121-135.
- Anderson, N., & Herriot, P. (Eds.) (1994). *Assessment and selection in organizations*. London: Wiley.
- Arthur, W. Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2004). *The use of person-organization fit in employment decision making: Evidence of its criterion-related validity (or lack thereof)*. Manuscript submitted for publication.
- Astin, A. W. (1964). Criterion-centered research. *Educational and Psychological Measurement, 24*, 807-822.
- AT&T. (2000). *AT&T associate-to-manager advancement process: Methods and procedures*. Basking Ridge, NJ: Author.
- Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*, 836-874.
- Barrett, G. V., Alexander, R. A., & Doverspike, D. (1992). The implications for personnel selection of apparent declines in predictive validities over time: A critique of Hulin, Henry, and Noon. *Personnel Psychology, 45*, 601-617.

- Barrett, G. V., Phillips, J.S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology, 66*, 1-6.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261-272.
- Becker, B. E., & Huselid, M. A. (1992). Direct estimates of SD<sub>y</sub>, and the implications for utility analysis. *Journal of Applied Psychology, 77*, 227-233.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*, 478-494.
- 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., Hanson, M. A., & Hedge, J. W. (1997). Personnel selection. *Annual Review of Psychology, 48*, 299-337.
- Boudreau, J. W. (1991). Utility analysis for decisions in human resource management. In M. D. Dunnette and L. M Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 621-745.). Palo Alto, CA: Consulting Psychologists Press.

- Boudreau, J. W., Sturman, M. C., & Judge, T. A. (1994). Utility analysis: What are the black boxes, and do they affect decisions? In N. Anderson & P. Herriot (Eds.), *Assessment and selection in organizations* (pp.77-96). London: Wiley.
- Brannick, M. T., Michaels, C. E., & Baker, D. P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology, 74*, 957-963.
- Burke, M. J., & Doran, L. I. (1989). A note on the economic utility of generalized validity coefficients in personnel selection. *Journal of Applied Psychology, 74*, 171-175.
- Cabrera, E. F., & Raju, N. S. (2001). Utility analysis: Current trends and future directions. *International Journal of Selection and Assessment, 9*, 92-102.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258-299). San Francisco: Jossey-Bass.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.). *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.
- Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*, 313-333.

- Campion, M. A., Campion, J. E., & Hudson, J. P., Jr. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*, 998-1002.
- Carrier, M. R., Dalessio, A. T., & Brown, S. H. (1990). Correspondence between estimates of content and criterion-related validity values. *Personnel Psychology, 43*, 85-100.
- Carson, K. P., Becker, J. S., & Henderson, J. A. (1998). Is utility really futile? A failure to replicate an extension. *Journal of Applied Psychology, 83*, 84-96.
- Cascio, W. F. (1991). *Applied psychology in personnel management* (4th Ed.). Englewood Cliffs, NJ: Prentice Hall.
- Cascio, W. F. (1998). *Applied psychology in human resource management* (5th Ed.). Upper Saddle River, NJ: Prentice Hall.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565-579.
- Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325-351.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

- Cronshaw, S. F. (1997). Lo! The stimulus speaks: The insiders view on Whyte and Latham's "The futility of utility analysis". *Personnel Psychology, 50*, 611-615.
- Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. *Research in Personnel and Human Resource Management, 12*, 79-123.
- Dunnette, M. (2001). Science and practice in applied psychology: A symbiotic relationship. *Applied Psychology: An International Review, 50*, 222-225.
- Furnham, A. F. (2001). Vocational preference and P-O fit: Reflections on Holland's theory of vocational choice. *Applied Psychology: An International Review, 50*, 5-29.
- Gallupe, R. B., Bastianutti, L. M., & Cooper, W. H. (1991). Unblocking brainstorming. *Journal of Applied Psychology, 76*, 137-142.
- George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of Applied Psychology, 75*, 107-116.
- George, J. M. (1991). State or trait: Effects of positive mood on prosocial behaviors at work. *Journal of Applied Psychology, 76*, 299-307.
- George, J. M. (1992). The role of personality in organizational life: Issues and evidence. *Journal of Management, 18*, 185-213.
- George, J. M. (1995). Leader positive mood and group performance: The case of customer service. *Journal of Applied Social Psychology, 25*, 778-794.

- George, J. M. (2000). Emotions and leadership: The role of emotional intelligence. *Human Relations, 53*, 1027-1055.
- Greer, O. L., & Cascio, W. F. (1987). Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard deviation of performance in dollars with a cost-accounting- based approach. *Journal of Applied Psychology, 72*, 588-595.
- Guion, R. M. (1961). Criterion measurement and personnel judgments. *Personnel Psychology, 14*, 141-149.
- Guion, R. M. (1998a). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Guion, R. M. (1998b). Jumping the gun at the starting gate: When fads become trends and trends become traditions. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 7-16). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hackman, J. R. (1992). Group influences on individuals in organizations. In M.. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2<sup>nd</sup> ed., Vol. 3, pp. 199-267). Palo Alto, CA: Consulting Psychologists Press, Inc.
- Heath, C., & Jourdon, F. J. (1997). Illusion, disillusion, and the buffering effect of groups. *Organizational Behavior and Human Decision Processes, 69*, 103-116.

- Herriot, P. (1993). A paradigm bursting at the seams. *Journal of Organizational Behavior, 14*, 371-375.
- Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, (2<sup>nd</sup> ed., Vol. 2, pp. 873-919). Palo Alto, CA: Consulting Psychologists Press.
- Holland, J (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future – remembering the past. *Annual Review of Psychology, 51*, 631-644.
- Howard, A. (1991). New directions for human resources practice. In D. W. Bray (Ed.), *Working with organizations and their people: A guide to human resources practice* (pp. 218-251). New York: Guilford.
- Howard, J. L., & Ferris, G. R. (1996). The employment interview context: Social and situational influences on interviewer decisions. *Journal of Applied Social Psychology, 26*, 112-136.
- Huffcutt, A. I., & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184-190.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*, 897-913.

- 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.
- Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.) *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.
- Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38*, 635-672.
- Jackson, S. E., & Schuler, R. S. (1995). Understanding human resource management in the context of organizations and their environments. *Annual Review of Psychology, 46*, 237-264.
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49*, 1-49.
- Latham, G. P. (2001). The interdependence of the science and practice of industrial-organizational psychology: A rejoinder. *Applied Psychology: An International Review, 50*, 245-251.



- Latham, G. P., & Skarlicki, D. P. (1995). Criterion-related validity of the situational and patterned behavior description interviews with organizational citizenship behaviors. *Human Performance, 8*, 67-80.
- Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology, 47*, 31-46.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.
- Lewin, K. (1935). *Dynamic theory of personality*. New York: McGraw-Hill.
- Macan, T. H., & Highhouse, S. H. (1994). Communicating the utility of human resource activities: A survey of I/O and HR professionals. *Journal of Business and Psychology, 8*, 425-436.
- Martin, S. L., & Raju, N. S. (1992). Determining cut-off scores that optimize utility: A recognition of recruiting costs. *Journal of Applied Psychology, 77*, 15-23.
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493-505.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Mauer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599-616.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Messick, S. J. (1998). Alternative assessment, uniform validity standards. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 59-74). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance, 15*, 173-186.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publications, Inc.
- Nunnally, J. C. (1978). *Psychometric theory* (2<sup>nd</sup> Ed.). New York: McGraw-Hill.
- Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology, 79*, 845-851.
- Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behaviors. *Personnel Psychology, 48*, 775-802.

- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65*, 373-406.
- Pugh, S. D. (2001). Service with a smile: Emotional contagion in the service encounter. *Academy of Management Journal, 44*, 1018-1027.
- Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology, 48*, 289-308.
- Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79*, 518-524.
- Rogers, J. D., Clow, K. E., & Kash, T. J. (1994). Increasing job satisfaction of service personnel. *Journal of Services Marketing, 8*, 14-26.
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1*, 331-340.
- Russell, C. J., Colella, A., & Bobko, P. (1993). Expanding the context of utility: The strategic impact of personnel selection. *Personnel Psychology, 46*, 781-801.
- Ryan, A. M., & Ployhart, R. E. (2002). Customer service behavior. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology, Volume XII: Industrial and organizational psychology* (pp. 377-397). New York: John Wiley.
- Schippman, J. S., Prien, E. P., & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology, 43*, 837-859.

- Schmidt, F. L. (1994). The future of personnel selection in the U. S. Army. In M. Rumsey, C. Walker, and J. Harris (Eds.), *Personnel selection and classification* (pp. 333-350). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- 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.
- Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis 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.
- Schmitt, N., & Schneider, B. (1983). Current issues in personnel selection. In K. M. Rowland and G. R. Ferris (Eds.), *Research in personnel and human resources management* (pp. 85-125). Greenwich, CT: JAI Press.
- Schneider, B. (1978). Person-situation selection: A review of some ability-situation interaction research. *Personnel Psychology*, *31*, 281-297.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*, 437-453.

- Schneider, B. (2001). Fits about fit. *Applied Psychology: An International Review*, 50, 141-152.
- Schneider, B., & Bowen, D. E. (1985). Employee and customer perceptions of service in banks: Replication and extension. *Journal of Applied Psychology*, 70, 423-433.
- Schneider, B. S., Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology*, 48, 747-773.
- Schneider, B., Smith, D. B., & Sipe, W. P (2000). Personnel selection psychology: Multilevel considerations. In K. J. Klein and S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions* (pp. 91-120). San Francisco, CA: Jossey-Bass.
- Scholarios, D. M., Johnson, C. D. & Zeidner, J. (1994). Selection predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology*, 79, 412-424.
- Sells, S. B. (1963). An interactionist looks at the environment. *American Psychologist*, 18, 696-702.
- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). College Park, MD: Author.
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

- Sutton, R. I. (1991). Maintaining norms about expressed emotions: The case of bill collectors. *Administrative Science Quarterly*, *36*, 245-268.
- Terpstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, *46*, 27-48.
- Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Totterdell, P., Kellett, S., Teuchmann, K., & Briner, R. B. (1998). Evidence of mood linkage in work groups. *Journal of Personality & Social Psychology*, *74*, 1504-1515.
- Van Viewen, A. E. M. (2001). Person-organization fit: The match between theory and methodology: Introduction to the special issue. *Applied Psychology: An International Review*, *50*, 1-4.
- Whyte, G., & Latham, G. P. (1997). The futility of utility analysis revisited: When even an expert fails. *Personnel Psychology*, *50*, 601-610.
- Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 377-410). Hillsdale, NJ: Erlbaum.

## VITA

### MARY KATHLEEN SHEEHAN

6 Centre Street, 1<sup>st</sup> Floor ✧ Madison, NJ 07940

973.377.2557 ✧ mksheehan@yahoo.com

### EDUCATION

#### TEXAS A&M UNIVERSITY

Ph.D., Industrial/Organizational Psychology

Dissertation: A test of a multilevel model of personnel selection in a customer service organization

College Station, Texas

December 2004

#### TEXAS A&M UNIVERSITY

M.S., Industrial/Organizational Psychology

Thesis: An assessment of the measurement equivalence of rating sources in a multisource feedback system

College Station, Texas

August 1999

#### PURDUE UNIVERSITY

Bachelor of Arts, Psychology

3.7 GPA, Magna Cum Laude, School of Liberal Arts Honors Program

West Lafayette, Indiana

May 1996

### PUBLICATIONS AND CONFERENCE PRESENTATIONS

Woehr, D. J., Sheehan, M. K., & Bennett, W., Jr. (in press). *Assessing measurement equivalent across rating sources: A multitrait-multirater approach*. *Journal of Applied Psychology*.

Arthur, W., Jr., Sheehan, M. K., Bell, S., Tubre, T. C., & Archuleta, K. (2000, April). *The role of team process behaviors in the prediction of team performance in a dyadic team-based training and performance task*. Paper presented at the 15<sup>th</sup> annual conference for the Society of Industrial and Organization Psychology, New Orleans, LA.

Sheehan, M. K., Woehr, D. J., Bennett, W., Jr., & Bowler, M. (1999, August). *Construct-related validity evidence for measures of work ethic, personality, and cognitive ability in an Air Force sample*. Paper presented at the 103<sup>rd</sup> annual convention of the American Psychological Association, Boston, MA.

Sheehan, M. K., Woehr, D. J., & Bennett, W., Jr. (1999, April). *Understanding disagreement across rating sources: An assessment of the measurement equivalence of raters*. Paper presented at the 14<sup>th</sup> annual conference for the Society for Industrial and Organizational Psychology, Atlanta, GA.