11-4-2008

More Than Screening Tools? : an Examination of Preliminary Applicant Evaluation Methods

Matthew Eric Paronto
Portland State University

MORE THAN SCREENING TOOLS?

AN EXAMINATION OF

PRELIMINARY APPLICANT EVALUATION METHODS

by

MATTHEW ERIC PARONTO

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
in
SYSTEMS SCIENCE: PSYCHOLOGY

Portland State University
2008

DISSERTATION APPROVAL

The abstract and dissertation of Matthew Eric Paronto for the Doctor of Philosophy in Systems Science: Psychology were presented November 4, 2008, and accepted by the dissertation committee and the doctoral program.

COMMITTEE APPROVALS:

[redacted]
Donald Truxillo, Chair

[redacted]
Talya Bauer

[redacted]
Leslie Hammer

[redacted]
Wayne W. Wakeland

[redacted]
Pamela Tierney
Representative of the Office of Graduate Studies

DOCTORAL PROGRAM APPROVAL:

[redacted]
George G. Lendaris, Director
Systems Science Ph.D. Program

ABSTRACT

An abstract of the dissertation of Matthew Eric Paronto for the Doctor of Philosophy in Systems Science: Psychology presented November 4, 2008.

Title: More than Screening Tools? An Examination of Preliminary Applicant Evaluation Methods

The increased prevalence of technology in organizations has had significant impacts on the recruiting, screening, and hiring processes. However, little is known regarding whether preliminary applicant evaluation methods provide meaningful candidate information beyond possession of minimum qualifications. To address this gap in the literature, two preliminary applicant evaluation procedures used at a major utility company were examined across two separate studies.

Study 1 examined online applicant prescreening protocols across three positions. Archival prescreening data from 5,619 applicants were analyzed in terms of item characteristics that distinguished candidates at different points in the score distribution (high vs. low; highest vs. high), as well as their ability to predict key criteria (e.g., preemployment test scores). Item characteristic ratings were provided by 11 graduate students across the three positions. Items differentiating top- from bottom-scorers were expected to have higher minimum qualifications and job-relatedness ratings; items differentiating among the top-scoring candidates were hypothesized to have higher objectivity and equivalent minimum qualifications ratings. Items

predictive of key criteria were hypothesized to be more objective and verifiable. Although most results were inconclusive, items that were more objective and verifiable were found to better predict later selection stage performance across two of the three positions.

Study 2 examined the online résumé screening process and whether structuring the evaluation process would result in more consistent résumé evaluation across raters. Twelve graduate students evaluated 20 résumés for a professional position under both structured and unstructured conditions. Results suggest that a more structured rating process resulted in increased reliability in résumé ratings.

Results from the prescreening study underscore the need to take a content validity approach to the development and scoring of these protocols, as differences among candidates in terms of their performance on individual items and the assessment as a whole provided inconclusive, unsystematic results. Moreover, total prescreening scores did not predict preemployment test scores or hiring decisions, further underlining the need for content validation. Results from the résumé evaluation study also illuminated deficiencies in this screening tool. Likewise, a content validity-oriented approach to the development of a brief, structured evaluation system for résumés may dramatically increase decision-making consistency.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

CHAPTER I: INTRODUCTION

Personnel selection has a long history of research on various methods of assessing job applicants. These methods run the gamut from the most basic, cursory assessments such as responses on an application blank to more systematic and theory-based assessments such as ability tests, biographical data, and assessment centers. Although the validity and utility of ability tests have been clearly demonstrated, it has also been shown that some of these tests (i.e., cognitive ability-loaded assessments) result in adverse impact. For example, cognitive ability test score differences between Whites and non-Whites are typically far greater than the differences between these two groups in terms of actual job performance (Hunter & Hunter, 1984).

Given the potential legal ramifications for companies using selection tests that result in adverse impact, selection researchers have explored alternatives or compliments to the use of ability tests in personnel selection (e.g., Schmidt & Hunter, 1998). Examples of such non-cognitive assessments include standard application blanks, training and experience (T&E) forms, biographical data, and personality inventories. Although these methods have been utilized and advocated by personnel researchers for over 90 years (cf. Mumford & Owens, 1987; Schmidt & Hunter, 1998), some have received little if any research scrutiny over the last quarter century. These methods warrant reexamination in the context of high-tech recruitment and selection.

One consistent finding in the investigation of alternative predictors is the relatively high validity and low adverse impact associated with biographical data (Hunter & Hunter, 1984; Reilly & Chao, 1982). It has been argued that biodata

measures may be better assessments of applicant suitability for a position than cognitive ability tests because biodata inventories measure typical as opposed to maximum performance, which is more representative of how an applicant would perform in the job (Mumford & Owens, 1987). Selection tests such as biodata and other measures that assess applicants' past experiences and accomplishments are received more favorably by applicants in higher level (i.e., professional/white collar) jobs (e.g., Hough, 1984; Hough, Keyes, & Dunnette, 1983). Moreover, these assessments may be more appropriate in selecting individuals in management or leadership positions (e.g., Kuhnert & Russell, 1990; Russell, 1990).

However, training and experience (T&E) forms and application blanks have received relatively little examination in the literature, perhaps due to their low levels of validity in predicting job performance (e.g., McDaniel, Schmidt, & Hunter, 1988). Although these selection tools may have low validity in terms of predicting "ultimate" criteria, they may nevertheless demonstrate utility in early stages of a multiple hurdle selection process, where the primary goal is eliminating applicants who are clearly unqualified for a position, as opposed to determining who is best suited for a position. For example, Ash and Levine (1985) investigated the validity of four different T&E scoring methods across three different job classes in a promotional context and found that only one scoring method (the grouping method) demonstrated validity against a peer nomination criterion for two of the three job classes. Several validity studies of the weighted application blank (WAB) have shown that this selection procedure is a valid predictor of turnover and tenure (Buel, 1964; Lee & Booth, 1974; Roach, 1971;

Wernimont, 1962). However, several questions that were valid predictors of these outcomes (e.g., length of time married, age) would be clear violations of Title VII of the Civil Rights Act, should such a study be conducted today. Therefore, the relatively high validities observed in some of these studies may not be realized given today's legal guidelines for selection.

Recently, research has turned its focus to recruitment, the earliest stage of the selection process (e.g., Allen, Van Scotter, & Otondo, 2004; Breaugh & Starke, 2000). Some researchers have suggested that recruitment may be the most critical stage of the overall selection process because it necessarily limits the quality of the applicants that can be considered for a position (Carlson, Connerley, & Mecham, 2002). However, it is also one of the least understood stages of the hiring process, especially in terms of its effects on later stages (Breaugh & Starke, 2000). Moreover, few organizations systematically assess the outcomes of their applicant attraction strategies (Carlson et al., 2002), nor do they necessarily gear their recruitment strategies to key employee and organizational outcomes after hiring decisions have been made (cf. Rynes, 1991).

Research in the recruitment literature has delineated three main phases in the overall recruitment process: persuading potential applicants to apply for a position (generating applicants), persuading applicants to stay in the applicant pool until a hiring decision has been made (maintaining applicant status), and persuading applicants to accept job offers (influencing job choice; Barber, 1998). Due to the temporal sequence in which these three phases occur, Carlson et al. (2002) suggest that not only are these phases easily identifiable, but that steps can be taken by an

organization at each stage to maximize the probability that the best candidates will ultimately accept a position. The effectiveness of the overall recruitment process is reduced to the extent that (a) the best candidates are not recruited to apply for the position, (b) the best candidates leave the viable applicant pool before a hiring decision has been made, (c) the methods used to select candidates (e.g., preemployment tests) have low validity, and (d) the best candidates do not accept job offers.

Carlson et al. (2002) recommend developing quality scores on all applicants, that is, some type of index or score, possibly based on an assessment, which reflects the qualifications of the candidate related to the position, as the most appropriate manner in which to assess recruitment outcomes. Typical effectiveness measures such as assessing the quality of new hires and the volume of applicants generated from a recruiting effort are considered by these authors to be inappropriate and deficient assessments. Carlson et al. propose applicant prescreening, in which candidates answer a small set of questions related to their past work experience and qualifications, as an evaluation tool with the potential to provide an index of candidate quality, since this assessment is the only one administered to the entire applicant pool, especially in multiple hurdle selection. According to the authors, applicant prescreening is useful in the evaluation process to the extent that it provides sufficient applicant differentiation and is predictive of job performance.

Technology and Its Impact on Screening Candidates

The increased use and prevalence of technology in organizations has had a significant impact on the way in which companies recruit, screen, and hire new employees (Lievens, van Dam, & Anderson, 2002). The importance of understanding the impact of technology on the recruitment and selection processes is critical for the following reasons. First, of the most important future personnel selection trends identified by a sample of human resources representatives, issues of technological developments such as online recruitment and testing were the most frequently cited (Lievens et al., 2002). Recent estimates suggest that upwards of 90% of major US-based corporations currently use internet-based recruitment (Dineen, Ash, & Noe, 2002), and 12% of large corporations are using online screening methods (Cober, Brown, Blumental, Doverspike, & Levy, 2000). Second, companies using online recruiting and/or testing have found that these tools save the organization in terms of time and cost (Anderson, 2003; Bauer, Truxillo, Paronto, Weekley, & Campion, 2004; Chan & Schmitt, 2004; Lievens et al., 2002). For example, Harris and Dewar (2000) suggest that the use of the Internet for recruiting purposes can reduce a company's cost per hire by up to 90%. Third, there have been several calls in the literature (e.g., Anderson, 2003; Lievens & Harris, in press; Lievens et al., 2002) for research on internet-based recruitment and testing, specifically in terms of the number and quality of applicants generated, the number of job offers accepted, as well as the criterion-related validity of online selection testing.

As part of the high-tech recruitment and selection process, candidates are increasingly being required to answer a small set of prescreening questions related to their past work experience. Often, online applicant management systems, which provide recruiters the ability to organize and systematically track and evaluate candidates who apply for positions posted online, also allow or require candidates to post their résumés. Because of the increased use and availability of the internet, more and more candidates are able to apply to multiple positions in a short period of time, providing one or both sets of information on themselves, thus increasing applicant volume.

For these reasons, it is critical that more research be conducted on high-tech recruitment and selection, especially in the preliminary stages of the selection process. The present study will begin filling some of these gaps. First, online applicant prescreening will be examined in an actual hiring context across multiple job types and job levels (i.e., customer service, physical jobs at entry level and above entry level) to systematically analyze and provide recommendations regarding the most effective and appropriate content and usage of the technology. In addition, given that résumés are typically used to screen applicants for higher-level professional and management jobs, the benefits of online résumés also warrant investigation. To this end, prescreening and résumé evaluation using two different résumé evaluation procedures (e.g., structured and unstructured evaluation) will be investigated. Finally, the validity of prescreening will be assessed for entry-level and above-entry level customer service and physical jobs. Each of these areas is discussed below.

Impact of High-Tech Selection

The increasing use of the internet in the employee selection process has provided companies with the potential to streamline the hiring process and improve the quality of those candidates who are ultimately hired. The use of selection technology in the earliest stages of the hiring process has enormous potential benefits to companies because it has the ability to collect and store information on large numbers of applicants. Tools such as the internet and interactive voice response (IVR) screening (an automated form of applicant screening via touch-tone telephone) give recruiters and hiring managers the capability to develop a list of desirable applicant qualifications and experience and create corresponding questions and scoring systems to reflect desirable candidate attributes. These systems automatically collect and score the responses provided by applicants. After the data have been collected, recruiters can then sort large numbers of candidates based on their predefined scoring of the screening questions.

Although issues surrounding the effects of selection technology on the psychometric properties of tests and on applicant reactions are still being investigated (Anderson, 2003; Bauer et al., 2004; Harris, Van Hoye, & Lievens, 2003; Lievens & Harris, 2003; Ployhart, Weekley, Holtz, & Kemp, 2003), it is possible that technological innovation in selection may improve recruitment outcomes due to the shortened timeframe in which selection decisions can be made (cf. Rynes, Bretz, & Gerhart, 1991). Moreover, feedback timeliness has been identified as an important dimension influencing applicants' reactions to the selection process (Bauer, Truxillo,

Sanchez, Craig, Ferrara, & Campion, 2001; Gilliland, 1993). Consequently, if selection technology does not cause serious procedural justice violations (e.g., Bauer et al., 2004), it may have the potential to improve recruitment and selection outcomes by retaining the best applicants and reducing selection costs (Carlson et al., 2002).

Applicant Prescreening

As noted earlier, prescreening is area in the selection literature that is not well understood. It varies widely in terms of its development, content, scoring, and use in practice. Moreover, it has received little if any empirical validation (Hunt, Gibby, Hemingway, Irwin, Scarborough, & Truxillo, 2004), and only recently has it received scrutiny regarding whether it must meet EEOC or OFCCP requirements (Harris, 2004). Nevertheless, online applicant prescreening is a recruitment and selection tool that is increasingly being used by companies to manage large numbers of applicants more efficiently by reducing the number of applicants in the applicant pool as well as providing candidate scores that can be used to facilitate decisions regarding which candidates recruiters will pursue further.

Typically, applicant prescreening consists of a small set of questions (usually 20 or less) that provide an assessment of applicants in terms of minimum job requirements that they must meet in order to be considered for a position. Prescreening may also assess relevant applicant qualifications such as previous job experience, accomplishments, and task and skill experiences that are believed to relate to performance in the position for which applicants are applying (Hunt et al., 2004). Accordingly, scoring systems for prescreening, though varying widely, provide two

preliminary applicant assessments: 1) whether the applicants meet the minimum requirements for the position, and 2) a cursory assessment of the level or quality of the applicants' previous job experiences and accomplishments vis-à-vis the position.

Applicant prescreening brings many potential benefits and advantages to the overall hiring process, but it is by no means a panacea. Although prescreening may perform well in terms of eliminating large numbers of unqualified candidates from consideration, anecdotal evidence from selection researchers suggests that it does not do a good job of identifying the top candidates who will be considered in later stages of the hiring process (cf. Hunt et al., 2004). In some cases, applicants who have the highest prescreening scores may actually be overqualified for the job, or may have engaged in some form of "faking." That is, some applicants may have overstated or overestimated their qualifications, or in some cases, responded in terms of what they felt were "ideal" applicant qualifications. Regardless of the cause, the result is that candidates who may not be the best-suited for the position or who are overqualified for the position may rise to the top of the candidate pool. Moreover, the differentiation between candidate prescreening scores may hinge on one or two questions, thus calling into question the importance of these distinctions (Hunt et al., 2004). A recent study on the use of personality tests indicated that due to the presence of faking (e.g., socially desirable responding) there was greater prediction error associated with higher scores in the distribution, suggesting that the test's validity was greater at lower scores and consequently it would be more useful from a select-out rather than select-in perspective (Mueller-Hanson, Heggestad, & Thornton, 2003).

Potential Issues with Prescreening

There are also practical and legal concerns regarding the use of applicant prescreening. Applicant management systems that incorporate prescreening are often marketed as time- and cost-savers for recruiters who must decide which candidates to pass on to the next stage of the selection process. The key benefits marketed to employment groups regarding prescreening include shorter times-to-fill as well as the ability and flexibility to select questions and scoring systems, which may typically be recruiter-driven. This poses two potential problems. First, there is a potential lack of standardization across recruiters who are hiring for the same position, both in terms of the questions asked of candidates as well as the emphasis (weighting) placed on certain questions or question response options. Secondly, there is concern over whether the content of the questions ultimately chosen for these prescreening scripts address aspects of skills and experience that are required for the position (i.e., established minimum qualifications, job-relevant experience). Ultimately, if these prescreening systems are not tightly managed and systematically developed, organizations may be vulnerable to legal risk, should a candidate challenge the process. Moreover, because of the lack of research in this area (cf. Hunt et al., 2004), little is known regarding the types of questions, both in terms of the content of the items themselves as well as how they are presented to the candidate, that provide job-relevant, accurate information on candidates.

The Role of the Applicant Résumé

As noted earlier, applicant résumés are now frequently collected via online screening systems. The use of the résumé in personnel selection has received relatively little attention in the literature, possibly due to the fact that scores are rarely assigned to the information contained therein. But considering that résumés are often utilized as the sole preliminary applicant evaluation tool for professional positions in determining which applicants to pass on to later stages of the selection process, it is important to assess their usefulness and effectiveness in identifying the best applicants for a position as well.

Screening résumés may provide important information about applicants. As mentioned, in some cases and for certain job types it may be the only information available to recruiters on which to base recruitment and screening decisions. Résumé searches can also help to compliment the results from prescreening (if the two are used in tandem) as well as combat some of its limitations. For example, the résumé can be used as a check on applicant responses from prescreening. In addition, applicant résumés can aid recruiters in sorting out the best candidates once prescreening has narrowed the field.

Nevertheless, the applicant résumé is not without its limitations. A résumé potentially contains numerous important bits of information about an applicant, with little in the way of standardization in terms of its ordering or presentation. Consequently, recruiters must identify what these elements are and be able to identify, compile, and evaluate this information across relatively large numbers of candidates

and résumé formats, which can be a mentally taxing process. Research examining selection decision makers' evaluations of content from selection interviews has suggested that there is a high degree of variation in terms of what and how many pieces of information are used in their decision making process as well as whether and how these elements are combined in unique ways when evaluating applicants (Graves & Karren, 1992; Hitt & Barr, 1989).

As an example of the potential for cognitive overload in the evaluation of job candidates, Graves and Karren (1992) found that in the context of interviews, their sample of 29 interviewers displayed 6 distinct clusters of decision-making strategies for making hiring recommendations. Moreover, effective interviewers used strategies that differed from ineffective interviewers. Effective interviewers were better at identifying the job-relevant criteria from interviews and applied their decision rules more consistently than ineffective interviewers. In addition, effective interviewers' self-reports of their decision-making process mapped more closely to their actual behavior compared to ineffective interviewers.

The Impact of Selection Technology on Résumé Evaluation

Similarly, the potential for information overload in the evaluation of applicant résumés is a cause for concern, especially with the advent and proliferation of the internet in the selection process. As was the case with applicant prescreening, résumé volume has likewise dramatically increased, which means that the amount of time recruiters can realistically spend locating and evaluating relevant candidate experience has diminished. Compounding this is the fact that résumés, by their very nature, are

unstandardized documents, being created by candidates themselves, and therefore there is wide variation in terms of how candidates present and describe their past job experience.

Fritzsche and Brannick (2002) suggested that recruiters judging actual résumés in their study assessing cues contained in résumés (e.g., applicant education, work experience, etc.) may have had "difficulty in translating bits of information on the résumé into subjective cue values, combining cues into judgments, or both" (p. 167). The consequences of information overload include the possibility of advancing unqualified candidates and overlooking some qualified candidates, as well as other potential decision inaccuracies. Research in the decision-making literature has suggested an inverse relationship between the number of attributes on which to base a decision and the level of accuracy of that decision (Davis & Davis, 1996; Helgeson & Ursic, 1993).

A potential solution to this problem is to find a way in which to structure the résumé evaluation process. Research on the selection interview has demonstrated that providing structure to the interview process in terms of asking the same questions of all candidates and developing guidelines for scoring their responses results in higher-quality organizational outcomes in terms of hiring the best candidates (Campion, Palmer, & Campion, 1997). Providing structure to the evaluation of applicant résumés may hold similar benefits.

Given the wide array of tools and techniques for evaluating applicants in the early stages of the selection process and the lack of empirical research in this area,

several questions exist regarding the benefits of these applicant screening techniques and their relationships with key organizational outcomes. The present study will address some of these questions. First, applicant prescreening will be examined in an actual hiring context to provide analysis and recommendations regarding the most effective and appropriate content and usage. Effectiveness and appropriateness will be determined via assessments of the job-relatedness of questions for specific positions being assessed and the questions' ability to provide meaningful differentiation between applicants, respectively. Second, the validity of prescreening will be assessed in a sample of entry-level customer service and entry-level and above entry-level physical job applicants for predicting key organizational criteria. Finally, a standardized résumé scoring evaluation procedure will be compared to a more traditional, holistic résumé evaluation process to assess the quality and similarity of the results produced, as well as whether there are differences across these methods in terms of their reliability.

The present study incorporates a variety of literatures to inform hypotheses regarding the nature and outcomes of applicant prescreening. First, the personnel psychology literature is reviewed in terms of relevant selection methods that can be expected to have utility in understanding the nature of prescreening. Specifically, research investigating the content and validity of assessments of applicant experience is reviewed. This review focuses on application blanks, T&E forms, and biographical data. Each of these is reviewed in terms of the nature of the assessment (e.g., rationale, content), important methodological considerations regarding their use (i.e., item

development, scoring), and psychometric properties (i.e., reliability and validity). This discussion is followed by a review of selection from an organizational perspective. The organizational perspective draws upon concepts from open system theory (Katz & Kahn, 1978), the attraction-selection-attrition (ASA) framework (Schneider, 1987), and the person-organization (P-O) fit literature (e.g., Kristof, 1996) to provide context to the decision-making processes involved in employee selection. Concepts from these theories are combined with selection decision-making research and the review of selection methods to develop specific hypotheses in the context of applicant prescreening. The method, procedures, and analyses used to test the hypotheses are then elucidated. Finally, the implications of the results of the present study are discussed, as well as the study's strengths and limitations, and avenues for future research.

CHAPTER II: REVIEW OF APPLICATION BLANKS, TRAINING & EXPERIENCE FORMS, AND BIOGRAPHICAL DATA

This chapter provides a review of commonly used preliminary selection tools. A discussion of these methods is warranted in light of their similarity and relevance to understanding the nature of online applicant prescreening, as little in the way of systematic research exists regarding this burgeoning selection tool (cf. Hunt et al., 2004).

Given that all of the selection tools discussed below have the measurement of past work-related experience as their primary focus, an overview of methodological considerations regarding the work experience construct and its measurement is delineated first. This is followed by a comprehensive review of three primary methods traditionally used to assess applicants' past work-related experience: application blanks, training and experience (T&E) forms, and biographical data. Discussion of each of these tools begins with a review of the nature of the assessment, followed by a review of the empirical literature regarding the development of the assessment, as well as the assessment's psychometric properties (i.e., reliability and validity). Each section concludes with a discussion of the application and relevance of these findings in the context of online applicant prescreening.

As mentioned in the preceding chapter, online applicant prescreening is not a clearly defined selection assessment tool (cf. Hunt et al., 2004). However, the content of prescreening assessments is similar in scope to that of other selection methods that have received greater research attention. This similarity is consistent with Hunter and

Hunter's (1984) assertion that multiple selection methods may assess similar content. For example, applicants' previous job experience may be assessed using an application blank, a biographical data measure, or an interview. Although the modes of obtaining the relevant information differ, the content of the assessment (previous job experience) is the same.

The content of prescreening assessments is similar in scope to that of other selection methods that have been designated as "alternative" selection methods (e.g., Hunter & Hunter, 1984; Reilly & Chao, 1982). In this context, alternative refers to the fact that these assessments are non-cognitive in nature, and, consequently, have demonstrated significantly less adverse impact compared to cognitive ability measures. Discussion of alternative selection methods in this chapter will be confined to application blanks, T&E forms, and biographical data (biodata) because applicants' job experience and qualifications are typically assessed through these measures, which make them relevant to online applicant prescreening.

As mentioned, application blanks, T&E forms, and biodata are considered alternatives to the use of cognitive ability tests, which have been shown to result in adverse impact (e.g., Schmidt & Hunter, 1998). Although some of these methods (e.g., T&E forms, application blanks) may not always demonstrate high levels of criterion-related validity, they nevertheless are used frequently and therefore warrant discussion. Moreover, these assessments may be useful in the earliest stages of the selection process where the primary goal is to reduce the applicant pool to a more manageable number and/or to make very general distinctions among applicants in

terms of their overall quality. These objectives fit well with the stated purposes of applicant prescreening (e.g., Hunt et al., 2004).

Application Blanks and Training & Experience Forms

Nature of the Assessments

Application blanks and T&E forms are two commonly used preliminary selection tools. According to Gatewood, Feild, and Barrick (2007), these assessments are typically used early in the selection process to make broad-based assessments of applicant suitability for positions. Specifically, these tools may be used as screening devices to determine whether applicants meet minimum job requirements. They may also be scored to provide preliminary assessments of applicant suitability in multiple hurdle selection. Finally, applicant performance on these assessments may be combined with applicant scores on other selection tests to make final hiring decisions. Accordingly, they warrant discussion in the context of applicant prescreening.

Methodological Considerations

As with most selection tests, the rationale behind the use of application blanks and T&E forms is that applicants' past performance is indicative of future performance. However, the construct of work experience has lacked consistency in its assessment. In their review and meta-analysis on the work experience construct, Quiñones, Ford, and Teachout (1995) identified two primary dimensions on which the assessment of work experience varies: level of specificity and mode of measurement. The three levels of specificity identified for measuring work experience were task, job, and organization. In terms of measurement mode, the authors identified three primary

modes. First, *time-based* work experience measures may assess an employee's tenure within a particular job or organization. Second, measures of *amount* include such frequency measures as the number of times an employee has performed a given task or the number of different jobs an employee has held in an organization. Third, measures of *type* include qualitative descriptions of experience (e.g., management experience, accounting experience; Quiñones et al., 1995).

Tesluk and Jacobs (1998) have further expanded this work experience model. They incorporated additional levels of analysis (e.g., team/work group-level experience), factors that affect quantity and quality of work experience (e.g., individual differences, contextual factors), and interactions between experience types. Further, the authors propose that these factors are differentially predictive of a variety of outcomes such as work-related motivation, knowledge and skills, work-related attitudes, career development, and job performance. These models may have implications for the way in which prescreening is used in practice, which will be discussed later.

Application Blank Evaluation

The evaluation of job application information may be done through global (binary) judgments (i.e., meets or does not meet minimum requirements) or through some form of empirical scoring or weighting of questions (Levine & Flory, 1975). Much of the research on application blanks has focused on a systematic scoring of the response content, known as the weighted application blank (WAB; See Gatewood et al., 2007). Pace and Schoenfeldt (1977) addressed a variety of legal considerations in

the use of WABs. Given the prevalence of criterion-related validity studies in developing the scoring system for a WAB, the procedure runs the risk of violating EEOC guidelines regarding appropriate questions on which to base selection decisions. Moreover, WABs are typically criticized as being atheoretical (cf. Levine & Flory, 1975), and questions that are found to predict performance are sometimes not job-related. To address these criticisms, Pace and Schoenfeldt (1977) recommended use of a content validity approach to developing the relevant content of a WAB, and a criterion-related validity approach for developing the scoring system. According to the authors, this approach would not only reduce organizations' vulnerability to legal recourse, it would also ostensibly improve the quality of selection decisions based on the method, given its increased focus on job-relevant experience.

Application Blank Reliability and Validity Evidence

Much of the published research on WABs dates back to the 1970s or earlier. Consequently, some of the findings, especially in terms of specific items predictive of the focal criteria, may not be appropriate given today's legal guidelines for selection tests. Furthermore, given that biodata was an extension and improvement on the use of background measures, particularly the WAB (See Mumford & Owens, 1987), biodata has received much greater research attention and will be discussed in the next section.

Much of the research on application blanks has focused on the cross-validation of WABs (e.g., Buel, 1964; Roach, 1971). Buel (1964) found that a WAB developed to predict turnover in a sample of female clerical employees maintained its validity over a two-year period. In another study investigating the cross-validity of a WAB in

predicting tenure, Roach (1971) conducted a second cross-validation study on a WAB that had been previously cross-validated in order to determine the extent of shrinkage over longer time periods. Many of the items that had initially cross-validated showed substantial shrinkage in the second cross-validation study. Furthermore, for some items, the direction of the relationship between the items and turnover changed in sign (i.e., from positive to negative). Roach (1971) concluded that much of the WAB's loss in predictive ability was associated with "changes in personnel policies and employment conditions" (p. 160). Consequently, it was recommended that WABs be revalidated periodically to combat the decay (i.e., decrease) in validity of these measures over time. These findings coincide with an earlier study by Wernimont (1962), who found that the validity of a WAB used to predict clerical tenure greatly deteriorated within a five-year period. In revalidating the items, Wernimont also found that item weights changed dramatically, and that some items which were not predictive of tenure in the original WAB validation were found to be predictive in the revalidation study. Validity was stable over this timeframe for only a few items.

In a related vein, Lee and Booth (1974) conducted a utility analysis on a WAB in addition to cross-validating it against a tenure criterion for clerical employees. The WAB was found to be a significant predictor of tenure with little overlap in the distribution of WAB scores for long- and short-tenure employees. Moreover, based on a utility analysis, use of the WAB at an optimal cut score would save the organization approximately $250,000 over a 2-year period. Translating this into today's dollars would result in a savings of approximately $975,000.

Summary of Research on Application Blanks

Application blanks are commonly used preliminary selection tools. They are typically used to make broad-based applicant assessments (i.e., whether applicants meet minimum job requirements). The long-held maxim that past performance is the best predictor of future performance is the primary rationale behind the use of the application blank in selection. When scores are assigned to responses on the application blank, based on their ability to differentiate successful and unsuccessful employees, the assessment is referred to as a WAB. Research on the application blank has focused primarily on WABs, specifically their ability to predict important work-related outcomes (e.g., job performance, turnover, tenure). Much of this research has focused on the cross-validation of WABs over time to determine the level of shrinkage (i.e., decrease in the strength of predictor-criterion relationships). Predominately, this body of research has shown that the weights assigned to items on the WAB do not hold up well over time, with many of the relationships decreasing drastically, and in some cases, changing in sign (e.g., Roach, 1971; Wernimont, 1962). Given the sole focus on item-criterion relationships, the method has been criticized as being atheoretical.

Training & Experience (T&E) Form Evaluation

According to Gatewood et al. (2007), T&E evaluations are used to measure applicants' relevant previous education, work, and training experience. These evaluations may be made in a variety of ways. Applicants' qualifications may be evaluated by staffing professionals using a checklist to indicate whether applicants

have relevant experience based on application blank content. Training and experience may also be assessed through a separate questionnaire completed by applicants during the application process. These questionnaires may ask applicants to indicate whether they have experience with the tasks or skills in question, and in some cases these questionnaires ask applicants to rate the level of their experience in various areas. In this case, applicants' self-reported experience with job-related tasks and skills is indicated on the questionnaire, which is later evaluated or scored in some manner by staffing professionals.

There are a variety of ways in which T&E forms may be scored. As mentioned earlier, the most basic use of the T&E evaluation is as a global, overall judgment of applicant suitability based on information provided by an applicant either through an application blank or résumé. Gatewood et al. (2007) label this method of evaluation *holistic judgment*. It is the least systematic and most informal method of evaluating training and experience. As such, it has not been empirically investigated.

The *point method* employs a formula to determine point values for different levels (e.g., years) of training, education, or experience to score T&E responses. Typically, higher scores are associated with more recent and greater levels of training and experience. Applicants meeting predetermined minimum requirements for the position receive some base score (e.g., 70), with points added based on the quality or level of their past experiences beyond the baseline (Gatewood et al., 2007). McDaniel et al. (1988) consider this method "credentialistic" in that it does not focus on past

accomplishments or achievements attained during the course of the relevant experience or education.

The *grouping method* is a variation on the point method. Using this method, applicants are classified into different qualification "categories" (e.g., well-qualified, qualified, unqualified) based on their levels of training and experience (McDaniel et al., 1988). Applicants within a particular group are assigned the same score, based on potential combinations of training and experience that describe the typical applicant at a given qualification level (Gatewood et al., 2007).

Applicants are assessed based on their experience or skill with various critical job tasks using the *task-based method.* According to McDaniel et al. (1988), these self-assessed ratings may be based on such dimensions as relative amount of time spent on tasks, performance or proficiency level, amount of supervisor assistance, or amount of additional training needed to perform the task. In terms of scoring, tasks may be weighted based on job analysis information or may be given equal weights.

T&E evaluation using the *behavioral consistency method* requires applicants to describe their accomplishments along various behavioral dimensions considered to be related to job performance (McDaniel et al., 1988). The job behaviors that are used to evaluate applicants are ones that have been identified as differentiating between high- and low performers (See Gatewood et al., 2007). The scoring of these accomplishments is typically performed using anchored rating scales developed through job expert input (e.g., behaviorally anchored rating scales, or BARS). This

method was employed by Hough and colleagues (Hough, 1984; Hough et al., 1983) in developing the Accomplishment Record for selecting attorneys.

Finally, in the *KSA approach*, a job analysis (using the job element method; see McDaniel et al., 1988) is conducted where supervisors brainstorm relevant KSAs and their criticality to job performance. It is classified as a T&E measure when the results of the job element method are used to evaluate education and experience. In terms of its development and use, it is similar in nature to the task-based approach (cf. Gatewood et al., 2007).

Training and Experience Reliability and Validity Evidence

Levine and Flory (1975) developed a framework for classifying various methods used to assess job applications. These authors empirically investigated one of the job application methods within their framework (a T&E form) in terms of the reliability of the evaluations produced.

The framework focused on three main factors: job-relatedness, depth of interpretation, and general method of evaluation. *Job-relatedness* is defined here as whether the questions on the application are face valid. *Depth of interpretation* refers to whether the questions asked are intended to measure surface traits (measuring an aspect of job performance) or intended to infer some underlying trait (e.g., personality). Combining application information in a semi-unstructured manner (judgmental method) and using statistical methods to arrive at an evaluation (statistical method) are the two types of *general methods of evaluation* delineated by the authors.

In all, Levine and Flory outline eight different classifications of application evaluation methods based on the three overarching dimensions.

Levine and Flory (1975) empirically investigated the reliability of ratings of a T&E form that was high on job relatedness, assessed surface characteristics, and was scored using the judgmental method (in this case, the grouping method). They hypothesized that reliability would be affected by the experience of the raters, the standardization of the scoring system, and the standardization of the application form. Specifically, reliability was expected to be higher when raters had more experience and when the scoring system was more standardized (this variable was not manipulated in the study).

A total of 15 job classes were examined, with approximately 20 applications per job, which were analyzed by 7 personnel analysts. These analysts had substantial variability in terms of their experience in evaluating applications. The procedure used by Levine and Flory (1975) to develop the scoring system for the T&E forms was as follows. An analyst conducted a job analysis for a position, which was then checked by a second analyst. The two analysts then independently evaluated the applications and provided scores. Data were analyzed through correlations between raters and Cohen's Kappa to more accurately assess the reliability of the scoring system. Overall, the interrater correlations were high (median = .96) across the positions. Correcting this for chance agreement (Kappa) provided a median reliability of .91 (reliabilities ranged from .54–1.00). Raters' level of experience was found to have an impact over and above the lack of standardization of the applications and the scoring method. The

authors concluded that the T&E method was a robust method of evaluation. Evaluators with widely varying levels of experience had high levels of agreement in their evaluations across applicants even though the content of each evaluation by job type was unstandardized.

Ash and Levine (1985) examined the reliability and validity of four different T&E scoring methods across three different job classes in a promotional context. Data for each of the T&E methods were gathered in the following manner. The application packets for each job class contained an application form, tasks performed form, a form on which applicants indicated their job-related achievements, and a peer nomination form (used as a surrogate criterion in lieu of actual hiring decisions). The application form was used to evaluate the point and grouping methods of T&E evaluation. The tasks performed form was used to evaluate task-based T&E evaluation. Finally, the achievements form was used to evaluate the behavioral consistency method of T&E evaluation. Participants also nominated the top three individuals they felt would perform best in the supervisory position for their respective jobs. Undergraduates scored the content of the applications based on each of the T&E methods.

Results of the study suggested that the highest reliability was associated with the task-based method, followed by the point- and behavioral consistency methods (which were similar to each other), and the lowest reliability was associated with the grouping method. The reliability findings in Ash and Levine's (1985) study regarding the grouping method stand in stark contrast to the reliability findings for this method in Levine and Flory's (1975) study.

In terms of the validity associated with each method, the grouping method was the only scoring procedure that provided significant prediction of the peer nomination criterion. Its validity generalized across 2 of the 3 job classes. Based on the results of the study, Ash and Levine (1985) suggest that the grouping method of T&E evaluation is the most appropriate scoring system, given that its validity generalized across multiple job families. Furthermore, the authors recommended that "T&E evaluation procedures should generally be used as rough screening devices for positions where previous experience, education, and training are necessary" (Ash & Levine, 1985, p. 575).

Hough and colleagues (Hough, 1984; Hough et al., 1983) developed an experience measure for attorneys using the behavioral consistency approach. Taking a content validity approach to test development, the Accomplishment Record (AR) test and scoring guidelines were developed and validated on a sample of actual attorneys along seven critical attorney performance dimensions (e.g., researching/investigating, writing, oral communications). Attorneys were asked to describe major past achievements, demonstrate the knowledge, skills, and abilities related to each dimension, what specifically the attorneys did, the time period in which they did the accomplishment, any formal recognition received, and the name of an individual who could verify the information. These attorneys also completed additional assessments as part of the overall assessment battery: a background inventory, consisting of objective, verifiable biodata items (e.g., schools attended, grades, publications, LSAT scores, etc.); an interest and opinion inventory, which was described as a blend of biodata and

personality-based items (e.g., involvement in high school/college activities, leadership positions held, etc.); and a self-description inventory, consisting of standard personality items such as decision-making ability, self-assurance, and initiative, among other constructs.

The criterion developed to validate the AR consisted of a performance appraisal completed by the participating attorneys' supervisors. The evaluation contained both behavior centered and task-based performance assessments. Criterion scores were formed based on each performance dimension separately as well as an average of the two scores.

After adjusting AR scores for the number of years since the attorneys had graduated from law school (a variable found to be strongly related to AR scores), the validity of the overall AR score predicting performance criteria was relatively high (r = .25). Moreover, individual dimensions of the AR were found to have similar levels of validity both for the overall performance criterion and each of the individual performance dimensions that made up the overall performance score.

The AR was also a unique predictor of attorney performance. That is, it did not correlate with other more traditional types of background/biodata measures and aptitude tests, such as grades, honors, quality of education, the LSAT, or prior legal experience. Furthermore, these latter, more traditional predictors correlated with the overall performance measure, at best, at a level comparable to that of the AR. Measures such as the AR, therefore, although time consuming to develop, may have benefits in terms of comparable validity to other more traditional assessments and

more positive test taker reactions, given the assessment's face validity (cf. Hough, 1984).

Finally, McDaniel et al. (1988) meta-analyzed the validity associated with different T&E rating methods. Specifically, the authors analyzed the generalizability of the validities associated with the point, task-based, behavioral consistency, and KSA T&E rating methods against a measure of overall job performance. Results of the meta-analysis suggested that, in terms of validity generalization, the validity of the T&E methods is at best moderate (mean $r = .17$). However, given the large standard deviation around this point estimate, there was evidence that T&E validity varied by rating method. Specifically, the point method was found to have low validity ($r = .11$) which did not generalize. The Illinois job element method (the operationalization of the KSA method used in this meta-analysis) and behavioral consistency method showed acceptable validities that were found to be generalizable, though as a caveat it should be noted that the number of correlations in each case was quite small (e.g., 15). The Illinois job element method had a mean validity of .20 while the behavioral consistency method had a mean validity of .45. These two methods demonstrated validity generalization. Finally, the validity of the task method did not generalize, as its mean validity was quite low ($r = .15$) with a large standard deviation. These findings generally coincide with those from Ash and Levine (1985) regarding the lack of validity generalization for the point- and task-based methods of T&E evaluation. Unfortunately, McDaniel et al. (1988) did not meta-analyze T&E ratings using the

grouping method, which Ash and Levine (1985) suggested demonstrated validity generalization across jobs.

Summary of Research on Training and Experience Forms

T&E evaluations encompass multiple data collection and evaluation techniques. The goal of these techniques is to assess applicants' relevant past job experience, education, and training. These evaluations include global judgments of applicant suitability, broad-based groupings of applicants based on whether they meet or exceed minimum requirements, and highly formalized ratings of applicants along relevant job performance dimensions using behaviorally anchored rating scales. Research examining different approaches to scoring T&E forms has suggested that the grouping method performs best in terms of predictive validity generalization against promotion criteria (Ash & Levine, 1985). The behavioral consistency method has also shown acceptable levels of validity that generalizes across jobs (McDaniel et al., 1988). The predictive validity associated with these methods overall is at best moderate (see McDaniel et al., 1988), and they have been suggested as being best-suited to making broad-based distinctions among applicants (Ash & Levine, 1985).

Relevance of Application Blanks and Training and Experience Forms to Prescreening

Several themes relevant to prescreening emerge from this discussion of research on application blanks and T&E forms. First, several researchers (e.g., Ash & Levine, 1985; Gatewood et al., 2007) indicated that these methods can be used to identify clearly unqualified applicants to remove from consideration. Second, although these methods may be used to predict key performance criteria (e.g., McDaniel et al.,

1988), the obtained validities are typically relatively low, and in the case of WABs, validities of the scoring keys tend to exhibit considerable decay over time (e.g., Roach, 1971; Wernimont, 1962). Third, these assessments provide recruiters and other hiring professionals with a broad-based overview of applicants' relevant experience and qualifications to be used as an initial screening. Although the job experience construct is still not well understood (cf. Quiñones et al., 1995; Tesluk & Jacobs, 1998), it has nevertheless been demonstrated that previous job experience is predictive of later job performance. Given that prescreening is designed as a preliminary assessment of applicant qualifications and that it typically assesses applicants' past work experience, it stands to reason that the research on application blanks and T&E forms may contribute to understanding how prescreening operates in practice.

Biographical Data

Nature of the Assessment

Biodata inventories stemmed from the use of WABs (See Gatewood et al., 2007). Selection researchers sought to develop a more thorough assessment of applicants' past experiences than those obtained from the application blank. In biodata inventories, applicants receive a standardized set of multiple-choice questions concerning their past behavior and experiences. The forms are then scored according to the questions' ability to differentiate between low- and high-performing employees (Mumford & Owens, 1987).

According to Mumford and Owens (1987), the benefits of using biodata measures are threefold. First, they are more elaborate assessments of applicants'

backgrounds than other measures such as application blanks and T&E forms. Second, they are a more cost-effective method of collecting background data on applicants compared to other methods such as interviews. Third, biodata measures assess typical as opposed to maximum performance situations. Other potential selection devices, such as aptitude, ability, and achievement tests assess maximum performance, which is not representative of an individual's day-to-day job performance.

Although the rationale behind the effectiveness of biodata in predicting job performance stems from the widely-accepted axiom in the selection literature that past behavior is the best indicator of future performance, biodata's scope extends beyond assessing work-specific experiences (see Mael, 1991). Prior learning and heredity are believed to determine one's resources and limitations, which consequently make some behaviors more or less likely to occur in future situations (Mumford & Owens, 1987). Mael (1991) has elaborated on this rationale. Specifically, he discussed two primary theoretical underpinnings for the domain of biodata. The first of these, the *ecology model*, is one of the more clearly elaborated models for understanding biodata. The basic premise of this model is that individuals begin life with a set of hereditary and environmental resources and limitations that determine initial individual differences. From there, individuals attempt to maximize adaptation to their environment through learning and cognition. The model is motivational in nature in that individuals are believed to select situations to pursue based on their perceived value of the outcomes, on individuals' preexisting needs and values, as well as their worldview, in terms of beliefs regarding the way life should progress (Mael, 1991). At this stage, the model

becomes cyclical in nature, in that the choices an individual makes and the outcomes of those choices will influence later choices, and so on, forming a developmental pattern of behavior. It is this pattern that is believed to explain the utility of biodata as a predictor in the selection context.

However, the ecology model is an incomplete depiction of life experiences in that its primary focus is on behaviors in which individuals *choose* to engage. It does not account for things that happen *to* an individual. In other words, experiences outside of the individual's control are not explained by the ecology model.

Accordingly, Mael (1991) suggests that concepts from *social identity theory* (SIT) are also applicable to biodata. Briefly, SIT states that an individual's self-concept is composed of a personal identity (attributes about oneself specific to the individual) and a social identity (the self defined in terms of psychologically belonging to some social category). Each of these identities is combined to make up the individual's identity. The individual's identity, in turn, influences his or her behavior (Mael, 1991). SIT posits that people behave in accordance with how they define themselves, and these definitions are influenced by people's associations with the various social groups with which they identify. The SIT perspective is relevant to understanding biodata's purported effectiveness in that it helps to explain the effects of experiences that are outside of the individual's control. Specifically, experiences that categorize an individual have the potential to influence that individual's subsequent patterns of behavior. This falls within the purview of biodata, in that biodata items not only assess applicants' choice-based behaviors, but they may also

tap the effects of applicants' association and identification with various psychosocial groups (Mael, 1991). However, it should be noted that SIT is not deterministic in nature, only that some aspects of group membership or association may explain an individual's experiences that are non-volitional in nature.

There has been a fair amount of debate over the appropriate scope of biodata question content (see Mael, 1991, for a discussion). However, given the theoretical underpinnings discussed earlier, it is clear that biodata's content domain may extend well beyond the assessment of work-specific experiences. Biodata can be used to assess skills and abilities as well as more temperament- or personality-oriented constructs. Regardless of what specifically is being assessed through biodata, the common thread running through biodata inventories is that they assess applicants' past experiences using a quasi-longitudinal format (Mumford & Owens, 1987). That is, the types of questions used in background data measures assess relatively discrete situations in an applicant's life, and how the applicant handled these situations. The response options developed for each item are used to best describe the applicant's prior behavior and experiences in that situation. A recent study (Mount, Witt, & Barrick, 2000) investigating the relative contributions of biodata, cognitive ability, and personality (using the Big Five model) in predicting multiple performance criteria (e.g., quantity and quality of work, problem solving ability) for clerical employees found that the biodata scales, although correlated with the measures of cognitive ability and personality, predicted unique variance in each of the criteria. This finding

helps to underscore the multidimensional nature of biodata and its distinctions from other assessments.

Methodological Considerations

In terms of creating a biodata inventory, there are three main considerations outlined in the literature regarding biodata item types. These concern the various types of biodata items that may be used, methods for creating biodata content, and the scoring of biodata inventories. Each of these issues will be discussed in turn. These issues are of great importance in the development of these inventories, and have direct implications to the development and use of online applicant prescreening. In fact, many of the considerations described below are examined and directly tested in the present study.

Item types. Regarding the types of questions that fall under the domain of biodata, Mael (1991) developed a general taxonomy of item content falling into three broad categories, encompassing ten different item types. The first of these categories, *historical*, refers to the widely accepted axiom in much of the personnel research that past behavior is the best predictor of future behavior. This type of item is considered by Mael to be the cornerstone of all biodata items, as such inventories are designed to assess factual information about applicants' previous behavior and experience. Although this biodata item dimension appears similar in scope to the content assessed via T&E evaluations, biodata items may assess past behavior that is not expressly related to past work experience. That is, biodata may assess applicants' past behavior

and experiences outside of the work domain. Hypothetical or future-oriented questions are considered to be beyond the purview of biodata.

Mael's (1991) second category refers to the idea that biodata content should be, in theory, *verifiable*. Consequently, dimensions in this category consist of ideal attributes of items that should increase the likelihood that the information provided by applicants will be accurate or verifiable. Attributes of items falling into this category include external, objective, firsthand, discrete, and verifiable, in order of increasing likelihood of self-report accuracy (issues surrounding response accuracy will be addressed in greater detail later). Although thoughts, feelings, and attitudes may not be considered verifiable per se, they may be assessed in reference to actual events in an applicant's life history.

The final broad category used by Mael (1991) to classify biodata items involves *legal and ethical concerns* surrounding its use for selection purposes. Item types falling under this category include applicant behaviors and experiences that were controllable, to which all applicants had equal access, job-relevant, and non-invasive.

A fourth category on which biodata items have been classified concerns *response format* (e.g., Mumford & Owens, 1987). Examples of response formats that may be used for biodata items include binary (e.g., Yes-No), multiple-choice items with response options falling on a continuum, and multiple-response items.

Methods for creating biographical data content. As with most properly-developed selection procedures, the development of a biodata inventory should begin with an analysis of critical elements of the job that are related to successful

performance. However, in some cases, such as entry-level positions, applicants may not have directly relevant experience. Using a developmental perspective in the generation of biodata items may overcome this obstacle. Specifically, in a study by Russell, Mattson, Devlin, and Atwater (1990), biodata items were developed using a developmental-integrative approach. Using a developmental-integrative framework, "[i]nvestigators construct hypotheses about what kinds of developmental episodes are related to subsequent criterion performance" (Russell et al., 1990, p. 569), consulting such sources as the developmental psychology literature and job incumbents, among others.

Russell et al. (1990) used an essay writing approach in their development of a biodata inventory to predict the success of first-year Naval academy students. Specifically, a sample of first-year students was asked to generate life history essays that were hypothesized to relate to certain criterion dimensions with which students were expected to have had prior experience. From the content of the essays, as well as additional follow-up interviews with a subset of the students, biodata items and response options were developed.

The resulting biodata scales (developed for specific performance criteria) were empirically keyed and cross-validated on the following year's cohort, who also provided the criterion data assessed at the end of the school year. This method of biodata development showed promise in predicting first-year Naval academy student performance. Not only was the measure predictive of key performance criteria, it also

demonstrated incremental validity beyond the assessments used by the Navy to select candidates into the academy.

Scoring biographical data inventories. Turning to issues surrounding the development of scoring systems for biodata inventories, the literature has identified three main approaches. The first of these, the *rational approach*, is considered to be one of the more theoretical of the scoring systems, in that its focus is on the interpretability of biodata item variances (Mitchell & Klimoski, 1982). That is, this method employs factor analytic techniques on item responses to derive interpretable performance-related constructs. In addition to its practical benefits, the advantage of this approach is that it provides theoretically meaningful constructs related to performance, which may result in advancing theory. The main criticism of this method is that it is time-consuming to develop.

According to Mitchell and Klimoski (1982), the *empirical keying method* focuses on item-criterion relationships, with the goal of maximizing the biodata instrument's ability to predict the criterion of interest. This is considered to be the primary advantage of the empirical keying approach. However, this method is often criticized as being atheoretical, and depending on the content of the instrument, may be likened to dustbowl empiricism.

A third approach to scoring biodata, the *subgrouping method* (see Mael, 1991; Mumford & Owens, 1987), focuses on clustering individuals based on their response patterns. Based on the response profiles of those completing the assessment, individuals are grouped together in meaningful ways, indicative of common

developmental trajectories. These can then be assessed in reference to important criteria to determine which backgrounds are best suited to the position. Mael (1991) considers this approach to be more naturalistic because it combines experiences and life events to gain an understanding of how these have shaped an individual's behavior.

Biographical Data Validity

Meta-analytic evidence. Meta-analyses have indicated that biodata instruments have the potential to predict substantial amounts of variance in key performance and retention criteria. Reilly and Chao (1982) examined the validity of biodata across six different job classes (e.g., management, clerical, sales) and five criteria (e.g., tenure, productivity, performance ratings). Overall, the average validity of biodata across job classes and criteria was .35. Across four different criteria, Hunter and Hunter (1984) found biodata validities ranging from .26 (predicting tenure and promotion) to .37 (predicting supervisor ratings). Further, these authors found similar levels of validity for biodata used in entry-level jobs (.37). Finally, a more recent meta-analysis (Schmidt & Hunter, 1998) reported a mean correlation between biodata and overall job performance of .35. Overall, these studies suggest that biodata validity generalizes across a variety of job types and levels as well as multiple criteria.

Validity of specific scoring approaches. Concerning which biodata scoring methods produce the best outcomes (i.e., validity), Mitchell and Klimoski (1982) directly compared the rational and empirical keying approaches. The authors examined the cross-validities associated with each of the methods in terms of the

amount of observed shrinkage, as well as a direct comparison of the validities of each method for predicting the criterion. A sample of prospective real estate agents was used in this study, with the criterion being the obtainment of a real estate license.

The biodata instrument consisted of multiple-choice items developed with the input of job incumbents, as well as a review of job analyses to derive items that were believed to be related to the criterion. The final measure consisted of 88 items with a test-retest reliability of .84 (The authors state this is based on a very short retest interval.). Data were collected on three samples of real estate students. The first two samples were used to derive the initial validities, and the third sample was used for cross-validation. The students' data were randomly assigned to scoring method.

Standard procedures associated with each of the scoring methods were used. Specifically, empirical keying was performed by Mitchell and Klimoski (1982) according to accepted guidelines for scoring WABs. (See Gatewood et al., 2007, for a discussion of this procedure.) Data were divided based on those students who had obtained a real estate license versus those who did not. Percentage differences for each response option were computed, and appropriate weights based on the magnitude of the response differences were derived. Weights were summed across all items. Three initial searches were conducted to begin the rational scoring process. The tasks performed by real estate associates were analyzed, relevant career counseling and career development literature was examined, and a review was conducted of the psychological and sociological literatures regarding the influence of background factors on career success. These searches resulted in the development of 4 a priori life

history constructs. Item response options were then scored on a continuum (unit-weighted), and a factor analysis was performed, resulting in 6 interpretable factors. Regression equations were developed based on both scoring methods for the derivation and cross validation samples, and the predicted *Y* scores were correlated with the actual criterion in both cases to determine the validity for both methods.

The results of the Mitchell and Klimoski (1982) study supported the hypotheses. Specifically, the validity and cross-validity obtained for the rational approach did not differ ($r = .36$ in each case). Substantial amounts of shrinkage were found for the validities in the empirical keying sample ($r = .56$ and .46 for derivation and cross-validation samples, respectively). However, comparing the cross-validities of the two methods resulted in a significant difference in favor of the empirical keying method. In concrete terms, there was an 8% better prediction of the criterion using the empirical keying method.

In terms of the practical significance of the different findings across the two scoring methods, Mitchell and Klimoski (1982) created a hit/miss matrix for each of the methods based on the development of cut scores. The empirical keying method again demonstrated a 6% higher level of correct classification. The significance of this result depends on the costs associated with making an incorrect decision.

The authors conclude that the empirical keying approach, if performed correctly (i.e., using adequately large samples and item pools, using a predictive criterion-related validity strategy), may be an acceptable method of scoring a biodata inventory, especially if the items are periodically cross-validated to minimize the

amount of shrinkage over time. However, the rational keying approach is also considered an acceptable method, and due to its theory-based method of development, should have greater explanatory ability from a theory-based perspective, as well as greater generalizability. However, this method is associated with greater costs in development and has the questionable assumption of linearity. Thus, in essence, the two scoring methods are each acceptable, and, depending on whether the focus is on developing meaningful performance-related constructs or maximizing prediction of the performance criterion, the rational and empirical keying methods will be more likely to produce the desired outcomes, respectively.

Validity generalization. One criticism of biodata inventories is that, because they are developed for use within a specific organization, their validities do not transfer to similar positions in other organizations. However, this may depend on the methods used in developing the biodata instrument.

Rothstein, Schmidt, Erwin, Owens, and Sparks (1990) hypothesized that the organizational specificity of biodata validities is caused by the methods used to construct the biodata scales. To test this hypothesis, the authors conducted a meta-analysis of data obtained through the process of developing and implementing a biodata instrument in a multi-organizational sample. Items in the development process were only retained if they demonstrated desirable statistical characteristics across organizations.

The participants in the validity generalization sample contained both blue- and white-collar supervisors in various industries including utilities, automotive, and

banking. The presence of moderators to the biodata instrument's validity was tested by organization as well as by other potential moderators such as race, gender, and education. The criteria developed for the study consisted of statements about the employees' performance of specific job duties and overall job performance, and statements about supervisory abilities and overall ability to perform the supervisory job. These ratings were made by the participants' immediate supervisors and an additional evaluator who was familiar with the participants' performance.

Results for the ability rating criteria showed that the validities of the biodata instrument demonstrated little if any variation after correcting for artifacts, providing little support for the potential presence of moderators. The average validity obtained for the biodata instrument in this study was comparable to the validities that have been reported in the literature for cross-validated biodata instruments (see above). Based on these findings, Rothstein et al. (1990) suggest that biodata instruments, if developed using a multi-organizational sample to key the instrument, may demonstrate generalizable validity that is not moderated by individual difference factors.

Hinrichs, Haanperä, and Sonkin (1976) examined the generalizability of a biodata inventory's validity for sales positions in a multinational organization. Originally developed in Finland, the inventory was refined and administered to samples in Sweden, Norway, France, Portugal, and the United States. Results of the study indicated that the validity of the instrument generalized across the different international samples. The more similar the culture and occupational (e.g., tenure, experience) makeup of the international samples, the more the validity of the biodata

inventory generalized. In some cases, validity did not generalize. However, it is likely this was due to small sample sizes (e.g., the Portuguese sample consisted of only 22 employees). Overall, Hinrichs et al. (1976) demonstrated that biodata validity may generalize across national boundaries with minimal adjustments to the original scoring key. This study, as well as the Mitchell and Klimoski (1982) and Rothstein et al. (1990) studies, demonstrate some of the advantages of the rational keying approach to scoring biodata.

Relevance to Prescreening

Biographical data research has direct relevance to applicant prescreening in the following ways. First, similar to the discussion of application blanks and T&E forms, biodata may also be used to assess applicants' past work experience. Past work experience is also typically assessed with prescreening. Second, the format of biodata (i.e., use of multiple-choice questions, multiple-response questions) is similar to that used in prescreening. Third, biodata assessments may be used in a prescreening context.

A study by Pannone (1984) used a rationally developed biodata measure (using the behavioral consistency approach) to predict scores on a content-valid test for electrician applicants. Biodata was used to screen out clearly unqualified applicants before administering the selection test. Scores on the biodata questionnaire were correlated with scores on the selection test, and this relationship was compared to the relationship between the selection test and education and experience questions (Education and experience questions had been previously used to screen applicants.).

Results showed that the biodata questionnaire demonstrated a stronger relationship to the written test than did the education and experience questions. Pannone (1984) suggests that the use of more specific biographical data for screening purposes is superior to the use of broader screening criteria such as education level and years of experience. Given that applicant prescreening as defined in this proposal may be considered to be a "mini-biodata" measure, prescreening may also be a superior assessment of experience than more general T&E forms. For the reasons listed above, it appears that the biographical data literature has implications for understanding the way that prescreening operates in practice.

Comparisons and Distinctions between Methods

Based on the above discussion of application blanks, T&E forms, and biodata, a common thread emerges regarding each of these methods. Specifically, each of these applicant evaluation procedures has as their primary goal an evaluation of relevant applicant experience and qualifications. Such method-content distinctions have been articulated previously by Hunter and Hunter (1984). Specifically, the primary difference between application blanks, T&E forms, and biodata is in the method used to collect relevant applicant data. For example, although referred to as an assessment of training and experience, T&E evaluation may be based on information provided in the application blank. Biodata inventories may directly assess applicants' past work history, albeit in multiple-choice format as opposed to the typical "longhand" presentation of this information as represented in an application blank. Therefore, the primary distinction between these methods lies in the format in which the data are

collected as opposed to the content being assessed. Given that prescreening assesses the same content as these aforementioned applicant assessments, though using different methods, these assessments are relevant in the context of prescreening, as prescreening is yet another method used to collect data on applicants' relevant past work experience in the early stages of the hiring process.

CHAPTER III: A SYSTEMS PERSPECTIVE

This chapter focuses on research that takes an organizational systems perspective to understanding the nature of the hiring process. This discussion will focus on two different levels of analysis. First, a high-level analysis of the process from an organizational perspective will be discussed. This will be followed by a discussion of the concepts as applied to an individual hiring decision with a focus on the decision-making process and human information-processing limitations. This will be compared to a purely statistical approach to decision-making. Implications of these perspectives for organizational hiring practices in general, and preliminary applicant evaluation methods in particular, will be discussed.

The literature discussed in previous chapters has focused primarily on various types of selection tools and their predictive ability. Although selection research has demonstrated that these methods have the potential to provide meaningful distinctions between applicants from a purely statistical perspective, these findings do not necessarily reflect the use of selection tests in practice. One recent study (Nowicki & Rosse, 2002) that sought to understand the selection process from the manager's perspective found that attributions made by managers regarding the successes and failures surrounding a hiring endeavor had little to do with the purported benefits of the selection tests employed. Instead, successes were attributed to luck, while failures were attributed to situational constraints or inadequacies in the hiring process (e.g., not checking applicant references, inadequate information collected in interviews).

These findings are of particular concern in light of the nature of online applicant prescreening. Online applicant prescreening is a relatively new selection tool and its role in the overall hiring process is rapidly expanding, especially in large organizations (cf. Cober et al., 2000). Considering the large volumes of applicants that are processed in such organizations and the technological aspects of the prescreening process, it is critical that prescreening's operation in practice be understood. Specifically, the decision-making processes involved in prescreening need to be optimized in order to meet the multiple objectives of managing large numbers of applicants, streamlining the hiring process, hiring the best applicants, and avoiding opening the organization up to litigation.

An Open System Theory Perspective on Recruitment and Selection

Katz and Kahn (1978) applied an open system approach to understanding the function of organizations. In essence, organizations are considered to be social systems that must interact with the external environment in order to survive. That is, organizations exist because of the people in them. Generally speaking, open system theory seeks to understand the functioning of systems through their interaction with the surrounding environment. According to Katz and Kahn (1978, p. 3), "The behavior of an organization is contingent upon the social field of forces in which it occurs and must be understood in terms of the organization's interaction with that environmental field."

Using the terminology of open system theory, the acquisition of necessary resources is known as the system input. These inputs are acted upon by the

organization in order to generate a product or service, which is then used by some entity outside of the system. These processes refer to the concepts of throughput and output, respectively. These outputs are then said to provide the system with the ability to acquire additional inputs, thus perpetuating the cycle. If successful, not only does the organization reduce its entropy, it may also grow and incorporate into itself aspects of the environment.

Organizations persist because they are able to successfully acquire the resources from their environment (e.g., raw materials, human resources) necessary for them to continue functioning. The process of recruitment and selection may be considered part of this acquisition process. Recruiters operate on the boundaries of the system, interacting with the environment in order to bring in human capital from the environment necessary for the organization to complete its functions. They act to maintain the system's dynamic homeostasis (Katz & Kahn, 1978). That is, although the specific individuals that make up the organization change over time (e.g., through attrition), the functions of the organization remain the same. Hence, the role of the human resources subsystem is that of attracting and selecting individuals into the system to aid the organization in its operations and acquisition of negative entropy.

An additional key concept that Katz and Kahn (1978) take from open system theory to explain the functioning and persistence of organizations is that of maintenance energy. This concept refers to the reduction of human behavior variability that helps to produce stable patterns of activity within the system. Specifically, the authors state that "the continued existence of a social system . . .

requires not only the physical presence of human beings but their enactment of particular behavior patterns" (Katz & Kahn, 1978, p. 41). Three forces that act to reduce human variability are environmental pressures, shared values and expectations, and rule enforcement. Of these, shared values and expectations may be the most directly influenced by the recruitment and selection subsystem. Shared values and expectations refers to the fact that individuals who have common goals and shared expectations regarding the behavior required for their accomplishment results in cooperative activity. Thus, shared values and expectations aid individuals in the organization to coordinate activities in order to reach system objectives (i.e., organizational goals).

Personnel Psychology Selection Models

Two related streams of research in the personnel psychology literature directly pertain to the open system theory concept of maintenance energy as it relates to the recruitment and selection subsystem. The more general, systemic model is the attraction-selection-attrition (ASA) model first proposed by Schneider (1987). Stemming from this organizational research paradigm is research on person-organization (P-O) fit.

Attraction-selection-attrition model. The ASA model, as first articulated by Schneider (1987), takes an interactional, systemic approach to understanding organizational behavior. According to the model, contextual variables such as the structure, climate, and culture of organizations are influenced by the individuals contained within them. The traditional approach to understanding individual behavior

in organizations has been that aspects of the organization (i.e., contextual variables) are what influence individual behavior. The ASA model proposes the converse. In essence, organizational structure and culture are considered emergent properties stemming from the people contained within the system.

First articulated by Katz and Kahn (1978), organizations are social systems that are goal-directed. The behaviors engaged in by organizational members in pursuit of the organization's goals are the determining factors of organizational processes and structures (Schneider, 1987). These goals have implications for the way in which organizations select individuals into the organization. In general, people are attracted to and selected by organizations with similar values. Once selected, employees who find that they are not a good fit with the organization's culture and values tend to leave the organization. Empirical research has supported this effect (e.g., Schaubroeck, Ganster, & Jones, 1998). The overall result is a homogenization of the types of people contained within organizations. This effect may be beneficial to organizations in their nascency, but over time this effect can lead to organizational rigidity and a reduced ability of the organization to effectively adapt to changes in the environment (Schneider, Goldstein, & Smith, 1995).

Person-organization fit. Traditionally, the selection literature has focused specifically on selection test validity and utility. More recently, attention has been directed to the concept of applicant fit with the organization or position to which they are applying. These fit assessments fall under the general rubric of the P-O fit literature, and have lent direct and indirect support to the attraction and selection

aspects of Schneider's (1987) ASA model, although attraction and selection have been confounded in some of these studies (e.g., Schaubroeck et al., 1998). Broadly defined, P-O fit refers to "the compatibility between people and organizations that occurs when: (a) at least one entity provides what the other needs, or (b) they share similar fundamental characteristics, or (c) both" (Kristof, 1996, pp. 4–5). Contained within this definition is the notion that fit may be defined as compatibility or complimentarity (supplementary and complimentary fit, respectively) of an applicant and the organization in terms of the applicant's personality and the organization's culture or values. Also contained within this definition is the idea that fit may be the result of the organization satisfying applicant preferences or needs and/or it may be the result of the applicant having the requisite KSAs needed by the organization or position (needs-supplies and demands-abilities perspectives, respectively; Kristof, 1996).

The operationalization of fit can be conceptualized at different levels of analysis. Kristof (1996) identified three such levels. First, fit may be assessed in terms of the match between an individual's personality and the "personality" of an occupation (e.g., based on Holland's (1985) RIASEC typology). Assessing compatibility of the two profiles is an example of person-vocation (P-V) fit. Fit may also be assessed at the work group level (P-G fit). In this instance, compatibility between the individual and the work group is assessed. Finally, person-job (P-J) fit measures the match between an individual's KSAs and job requirements. This operationalization of fit may be the most directly applicable to the prescreening or résumé evaluation process, as evaluations at this stage of the selection process focus

primarily on whether applicants meet minimum job requirements. This contention is supported by the P-O fit literature. Specifically, Kristof (1996) suggested that P-O fit may not factor into hiring decisions until job-specific and overall fit have been assessed in the early stages of the selection process.

The empirical literature examining P-O fit has generally supported the multiple conceptualizations, unique effects, and operation of fit assessments in actual selection contexts. Much of this research has occurred in the context of interviews, which is the typical stage at which such fit assessments are made (e.g., Kristof-Brown, 2000; Rynes & Gerhart, 1990).

Rynes and Gerhart (1990) examined interviewers' assessments of fit in an actual selection context in terms of general fit for the position (P-J) as well as firm-specific employability (P-O). The goals of the study were to determine whether fit assessments differ from the assessment of general employability, to what extent recruiters' fit assessments are generalizable (i.e., non-idiosyncratic), and what applicant traits are associated with fit assessments beyond applicants' objective qualifications.

Recruiters in the sample assessed graduating MBA students who applied for positions in multiple organizations. This allowed for the determination of level of idiosyncrasy in recruiters' fit assessments. That is, recruiters within an organization should agree more in terms of an applicant's firm-specific fit than would recruiters from different organizations evaluating the same applicant. Recruiters provided three primary applicant assessments: a trait assessment of applicants along ten dimensions

(e.g., general knowledge, leadership), an assessment of the applicant's overall employability, and an assessment of the applicant's firm-specific employability.

Results of the study supported the hypothesized differences between assessments of P-J and P-O fit. First, recruiters' assessments of applicants' firm-specific fit were lower than their assessments of applicants' general employability. Within-organization recruiters demonstrated higher levels of agreement in terms of their assessment of applicants' firm-specific employability than did recruiters across organizations, demonstrating that firm-specific fit assessments extend beyond mere recruiter idiosyncrasies. Finally, controlling for general employability assessments, objective applicant qualifications did not predict variance in firm-specific employability, whereas applicant traits such as interpersonal characteristics and goal orientation were predictive of firm-specific assessments. Taken together, the findings support the contention that fit assessments differ from assessments of general employability. Moreover, fit assessments go beyond mere recruiter idiosyncratic preferences.

Cable and Judge (1997) developed and tested a comprehensive model of the operation of P-O fit assessments in the hiring process. This model tested the selection aspect of Schneider's (1987) ASA model. Incorporated into this model were physical and demographic attributes of the applicant (e.g., gender, physical attractiveness, interviewer liking of the applicant) as well as more objective aspects of applicant qualifications (e.g., work experience, GPA). These characteristics were hypothesized to influence interviewers' perceptions of applicant P-O fit. Actual (i.e., objective)

congruence between applicant and organization values were expected to influence interviewers' perceptions of applicant-organization values congruence, which in turn were also expected to influence interviewers' assessments of P-O fit. Interviewer P-O fit evaluations were expected to influence interviewers' hiring recommendations, which were hypothesized to influence organizational hiring decisions. This study used a longitudinal design, assessing the key fit variables from both the interviewee and interviewer perspectives in order to determine the extent of P-O fit.

Results of the study generally supported the linkages specified in the model. Specifically, actual values congruence between applicants and organizations significantly predicted interviewers' perceptions of congruence, which in turn significantly predicted interviewers' evaluations of P-O fit. Further, "peripheral" aspects of applicants (i.e., interviewer liking of applicant, physical attractiveness) were positively related to interviewers' perceptions of P-O fit. In turn, P-O fit evaluations were positively related to hiring recommendations, which were also positively related to actual hiring decisions. The findings of this study suggest that perceptions of fit significantly influence selection decisions, providing some support to the operation of Schneider's (1987) ASA framework.

Kristof-Brown (2000) examined the antecedents of two types of fit assessments. Specifically, aspects of applicants were expected to differentially influence recruiters' assessments of applicant P-J fit and P-O fit. Specifically, assessments of P-J fit were expected to be more influenced by applicant KSAs than by applicant values. Conversely, assessments of P-O fit were hypothesized to be

influenced more by applicant values and personality than by applicant KSAs. Actual recruiters viewed mock interviews of applicants that varied in values, personality, and qualifications (as demonstrated in résumés). Results of the study suggested that, although all recruiters in the study used KSAs in both P-J and P-O fit assessments, there were significant differences in the frequency with which KSAs were mentioned as indicators of each type of fit. Specifically, KSAs were reported with greater frequency in P-J fit assessments than in P-O fit assessments. Values and personality were also used to assess both types of fit. However, they were mentioned as indicators of fit more frequently for P-O fit than for P-J fit. The results demonstrate that assessments of job-specific and organizational fit, although having overlapping antecedents, are most influenced by different applicant characteristics.

Based on these findings, Kristof-Brown (2000) examined the relative influence of fit perceptions on actual hiring recommendations. Recruiters conducted interviews with actual applicants, after which they completed measures regarding their perceptions of the applicants' P-O and P-J fit as well as their hiring recommendations. Results showed that although P-O and P-J fit were highly correlated, they are factorially distinct. Moreover, each fit type predicted unique variance in hiring recommendations over and above the other fit type, with P-J fit contributing more to the prediction of hiring recommendations.

Decision-Making Approaches in Selection

Based on the above discussion of issues of applicant fit, it is clear that the applicant hiring process consists of more than simple test-performance relationships.

Applicants are not only evaluated based on their objective qualifications, they are also assessed in terms of their match to organizational goals, whether these goals are defined based on job requirements or broader organizational values and objectives (cf. Kristof, 1996). Thus, a consideration of various decision-making strategies in the selection process may be useful in understanding the ways in which organizations identify applicants who are best suited for positions.

Image theory and prospect theory. In her discussion of the importance of P-O fit in selection, Kristof (1996) draws upon image theory and prospect theory in order to understand the decision-making processes involved in the selection process. Using prospect theory (Kahneman & Tversky, 1979), the selection process can be conceptualized as a decision-making task occurring under conditions of risk. Risk, in the selection context, can be operationalized as the costs associated with selecting the wrong individuals for positions. That is, the applicants selected are either unsuccessful in carrying out the requirements of the position or do not fit the overarching organizational culture, and thus are either removed from or voluntarily leave the organization, respectively. In the early stages of the selection process, as in the case of prescreening, decision-makers try to avoid losses (see Kristof, 1996).

Image theory (e.g., Beach, 1990; Beach & Potter, 1992) also has direct relevance to the selection decision-making process. According to this theory, decision-making is conceptualized as a two-step process. In the first stage, options are first screened out if they do not meet relevant selection criteria. Beach (1990) labeled this stage the compatibility test. Once undesirable options have been removed, the second

phase of the decision-making process entails choosing from among the remaining alternatives. This phase is known as the profitability test.

These models have direct application to prescreening. Given that one of the primary goals is to reduce the viable applicant pool, prescreening can be conceptualized as a test of applicant compatibility with job requirements. Once applicants have been assessed in terms of whether they meet minimum requirements, they are then evaluated based on additional information (e.g., job-relevant experience, résumé content, interview) to arrive at a final hiring decision.

Mechanical vs. judgmental assessments of applicant information. Applicant information may be measured and evaluated in a variety of ways to arrive at selection decisions. These methods vary along two primary dimensions as outlined by Gatewood et al. (2007). The first of these, labeled *mechanical*, refers to the measurement and/or assessment of applicant information that does not rely on human judgment. In contrast, *judgmental* procedures involve the collection and/or evaluation of applicant information based on human judgment.

Based on these distinctions, Gatewood et al. (2007) outlined eight methods of collecting and evaluating applicant information that vary along these two broad dimensions. These methods include the purely judgmental collection and evaluation of applicant information (pure judgment), judgmental collection with mechanical evaluation (trait rating; e.g., judgmental interview ratings that are evaluated statistically), pure statistical methods, in which applicants complete selection tests which are scored according to statistical algorithms, and mechanical composites,

whereby judgmental and mechanical data are combined statistically (e.g., combining judgmental interview ratings and test scores). These decision-making strategies have direct applications to prescreening as investigated in the present proposal. As will be discussed in greater detail in subsequent chapters, applicant prescreening can be conceptualized as an instance of profile interpretation. That is, applicant prescreening data are collected mechanically, and points are assigned to various responses based on recruiter judgment, which are then combined mechanically by the prescreening program.

In terms of which methods are best, Gatewood et al. (2007) reviewed the extant literature examining various strategies. In general, mechanical/statistical methods were found to perform at or above the level of judgmental methods. The primary reasons for the superiority of mechanical methods outlined by Bass and Barrett (1981; cited in Gatewood et al., 2007) include more appropriate weightings of relevant information provided by mechanical or statistical methods; the ability of these methods to produce more accurate prediction models because they do not suffer from human cognitive limitations; human decision makers, at best, could only be expected to match the precision of mechanical methods; and finally, when combining subjective and objective data, human decision-makers incorporate a substantially greater amount of error. Issues of human decision-making limitations will be revisited in the next chapter in the context of résumé evaluation.

CHAPTER IV: HYPOTHESES

Several important questions regarding prescreening and résumés warrant empirical research. In this chapter, specific hypotheses and research questions will be developed based on the personnel selection and decision-making literatures. (See Table 1 for an overview of proposed hypotheses and research questions.) The primary goals of this study are to provide empirical evidence that may inform best practices regarding the use of applicant prescreening in terms of the most effective types of prescreening questions, the nature and quality of the distinctions that may be made between applicants, and the ability of prescreening methods to reliably sort applicants and the nature of the distinctions made between applicants.

Two primary screening methods will be examined. The first of these is online prescreening, the content of which has been discussed in terms of its similarity to the content of application blanks, T&E forms, and biodata. The second screening method that will be investigated is the applicant résumé. It is important to understand how résumés are evaluated by recruiters, especially since some practitioners have considered the résumé to be of relatively little value to the selection process (see Handler, 2002). The relative merits and limitations of each of these methods are empirically investigated in the present study. Examining online prescreening and résumé evaluation is critical given the rapidly increasing use of online screening assessments and online testing in general in personnel selection (e.g., Cober et al., 2000; Lievens et al., 2002; Ployhart et al., 2003), and the fact that applicant résumés are frequently used by organizations as part of the preliminary applicant evaluation

process, and, in some cases, as in the present study, they may be the sole source of information used by decision-makers to determine which applicants will advance to the interview stage of the selection process.

The second broad-based consideration of prescreening in the present study will be an investigation of the validity of prescreening assessments as they are used in practice. These hypotheses are informed by research on the validity of relevant selection methods (i.e., application blanks, T&E forms, biodata) and will be tested using samples of applicants for entry-level and above entry-level positions. This may be the most appropriate sample on which to test the validity of prescreening due to the large applicant volume and the likelihood that prescreening will be the only basis on which decisions will be made regarding who will advance to the next stage in the selection process.

The hypotheses outlined below will be tested across two separate studies. The first study, which will be referred to as the Prescreening Study, will focus on assessing the most appropriate usage of online applicant screening. Specifically, issues of the nature and type of meaningful applicant differentiation that can be expected from use of this tool are explored. Further, questions regarding the ability of online applicant prescreening to predict key criteria and the characteristics of the items that are predictive of these outcomes are examined in the entry-level and above entry-level physical jobs and the entry-level customer service job applicant samples.

The second study, referred to as the Résumé Study, will explore whether providing evaluation guidelines and a standardized evaluation format improves the

résumé evaluation process. Given that, in many higher level jobs, the résumé is the basis for preliminary selection decisions, the role of the résumé and the manner in which this information is evaluated become critical. Thus, the Résumé Study will evaluate the effectiveness of different résumé evaluation techniques using a sample of applicant résumés for a professional position. The hypotheses involved within each of these studies are outlined below.

Prescreening Study

Prescreening Validity

An important issue regarding online applicant prescreening that has not been investigated concerns the validity that can be expected from these assessments. Given the brevity and nature of the assessment, it is expected that its validity would be similar to that of a T&E evaluation. Validity is defined in the present proposal as the relationship (i.e., correlation) between the predictor measure (prescreening) and key criteria (e.g., preemployment written test scores, hiring decisions, employee tenure). This conceptualization of validity is typically referred to in the personnel psychology literature as predictive or criterion-related validity.

One study that meta-analyzed the relationship between T&E scores and performance ratings found an overall validity of .17 (McDaniel et al., 1988). Similar levels of validity may be expected for prescreening. Meta-analyses of the criterion-related validity of biodata measures predicting supervisor ratings of performance and tenure have found mean correlations of .35 and .32, respectively (Mumford & Owens, 1987; Reilly & Chao, 1982; Schmidt & Hunter, 1998). Similar levels of validity

against tenure were found by Hunter and Hunter (1984). Studies investigating the WAB have found significant relationships with turnover. For example, Lee and Booth (1974) found a WAB used in the selection of clerical employees significantly predicted employee tenure. Given the purported similarities between prescreening and T&E and biodata measures discussed earlier, it is expected that prescreening will likewise demonstrate validity in predicting key criteria.

Issues of Faking, Biodata Item Characteristics, and Their Relationship to Validity

Of greater importance than merely demonstrating a correlation between online applicant prescreening scores and performance outcomes is to determine the characteristics of the questions that 1) contribute to applicant prescreening score differentiation and 2) are predictive of key criteria. These issues are of great importance because having a clearer understanding of what types of items demonstrate predictive utility will aid in the development of better-quality items, and hence, more effective prescreening protocols. The quality of prescreening assessments in terms of applicant score differentiation (i.e., variance) may depend on the types of questions asked. Evidence from the biodata literature regarding the potential for applicant faking provides some indirect support for this potential effect. Aspects of Mael's (1991) biodata item classification system also have implications in the context of prescreening. These studies are reviewed below.

There is an abundance of research evidence that suggests that the more objective or verifiable a question is, the greater likelihood that responses to that question will be accurate. For example, Weiss and Dawis (1960) examined the

verifiability of objective information (e.g., employment history) collected in questionnaire format in a sample of participants with physical disabilities. Data provided by these individuals were verified against multiple sources, including employers, relatives, and agency records. Results suggested that the more objective and less socially desirable the question content was, the less faking or inaccuracy was associated with that item. Mosel and Cozan (1952) examined the accuracy of work history information provided by applicants on an application blank. Specifically, they sought to verify applicants' responses to weekly salary, duration of employment, and job duties. Responses to these questions were verified against previous employer responses to a recommendation questionnaire. The relationship between applicant and previous employer responses to these questions was very high (i.e., correlations were in the .90s), suggesting a high degree of accuracy. Goldstein (1971) also compared responses to an application blank verified against most recent previous employer responses and found a substantial amount of inaccuracies. However, Goldstein operationalized inaccuracy via an absolute agreement index, whereas studies mentioned earlier used correlation analysis. Therefore, results of this study may have been more in line with other research on response verifiability if correlation analysis had been used.

Researchers have also examined response accuracy in biodata inventories. Specifically, Cascio (1975), using a sample of police officers, examined the response accuracy associated with a subset of biodata items that were considered historical and verifiable (e.g., number of siblings, marital status). Responses to these items were

compared for accuracy against the responses officers gave in their employment application. Correlations between the two reports were quite high, and there was little evidence that moderators (e.g., gender, age, length of employment) affected the accuracy of these data.

Beyond merely investigating whether and to what extent faking on biodata measures occurs, research has also attempted to tie faking behavior to item characteristics and to explore whether certain item characteristics systematically vary with item validity. Shaffer, Saunders, and Owens (1986) used a multimethod approach in assessing biodata accuracy. Biodata items were assessed along a continuum of objectivity-subjectivity (i.e., objective, moderately subjective, subjective), and accuracy was assessed using both test-retest and external observer input. High school students and their parents served as the participants. Students completed the biodata inventory during their freshman year and again five years later. Parents of the students also completed a shortened version of the inventory. Results of the study showed that test-retest reliabilities for the biodata inventory at the factor level were generally high, with more objective factors having greater response stability over time. In addition, there was a trend toward greater student – parent agreement as biodata items became more objective.

Kluger and Colella (1993) examined the effects of including a warning against faking on a biodata measure, and how this warning would alter item means and variance in relation to the level of the items' transparency. Beyond demonstrating that faking does occur in practice on biodata measures (using an applicant sample), the

results of the study suggest that the effect of providing a warning against faking is most effective on items that are transparent in nature (based on SME ratings). Specifically, this effect was most noticeable for items that were deemed job desirable (i.e., job related, which predicted incremental variance above ratings of social desirability, the other of the two indicators of item transparency). The faking warning greatly affected item means and variances on transparent items in the group of applicants who were warned about faking. Because past studies that investigated social desirability did not use a job-specific desirability rating in assessing the effects of faking on biodata validity, it is possible that the extent to which applicant faking affects selection test validity may be underestimated. Accordingly, if prescreening assessments contain highly transparent questions, and these questions are highly job desirable, issues of applicant dissimulation may become important, as these questions would allow applicants to portray themselves as being more qualified for a position than they truly are, assuming that applicant faking behavior in practice is not uniform across all applicants (i.e., some applicants are better "fakers" than others), which some past research has suggested is the case, especially for noncognitive assessments such as personality measures and biodata (McFarland & Ryan, 2000; Mueller-Hanson et al., 2003).

In a similar attempt to address issues of candidate response distortion, Schmitt and Kunce (2002) examined the effects of requiring elaboration on a subset of items in a pilot biodata measure administered to examinees for a federal civil service job, examining the effects of elaboration on mean item scores. Using a 2 (elaboration

required or not required) X 2 (elaboration required on first half vs. second half of the instrument) design, the authors found support for the effects of requiring elaboration. Specifically, means for the items requiring elaboration were 0.6 standard deviations lower than the same items when elaboration was not required. Moreover, these effects carried over, albeit to a lesser extent, to items where elaboration was not required when these items were located in the same section of the measure as the items that required elaboration. (Elaborated items were interspersed within the first half, second half, or both halves of the measure. One condition required no elaboration.) However, the authors did not tie these effects to characteristics of the biodata items or to characteristics of the participants (e.g., social desirability, impression management).

Other studies have investigated the interrelationships between faking behavior, biodata item characteristics, and their relationship to biodata validity using Mael's (1991) framework. In one of the first studies to use this framework, Becker and Colquitt (1992) examined the nature of biodata items that were most susceptible to faking using a subset of Mael's biodata item typology categories. Based on undergraduate ratings of a subset of biodata items using definitions of Mael's dimensions, items that were more likely to be faked were less historical, objective, discrete, verifiable, and external, and more job-related. However, it should be noted that these results were based on a subsample of only three biodata items that were found to be faked in practice.

In a comprehensive examination of Mael's (1991) biodata typology, McManus and Masztal (1999) tested the relationship between biodata item attributes (referred to

as methodological attributes) and their relationship to both item validity and socially desirable responding. Based on previous research (e.g., Becker & Colquitt, 1992; Mael, 1991), McManus and Masztal hypothesized that historical, external, objective, verifiable, and discrete items would exhibit greater validity compared to items that were on the opposite end of the spectrum (e.g., nonhistorical, internal, etc.), as the former items should demonstrate a stronger association with applicants' prior behavior (which is considered the best predictor of future behavior). Non-historical, internal, subjective, non-verifiable, secondhand, summative, job relevant, and invasive items were expected to demonstrate a stronger relationship to a social desirability index, compared to biodata items reflecting the opposite poles of these dimensions. Items from two biodata inventories (one for entry level sales positions and one for managerial positions) used in the study were validated against employee tenure (defined as one-year survival in the position). Three experts (PhDs in Industrial Psychology) rated the items along Mael's (1991) 10 dimensions using Likert-type scales (cf. Becker & Colquitt, 1992). The relationship between item-level validity and item attributes was tested in the entry-level sales sample. A personality-based social desirability measure was administered to the managerial sample along with the managerial biodata measure in order to assess the relationship between biodata item attributes and social desirability. Greater item-level validity was associated with biodata questions characterized by the five dimensions hypothesized to be predictive of the criterion. It should be noted that there was a fair amount of covariance among dimension ratings, which could have accounted for some of the unexpected results in

this study (e.g., job relevance was negatively related to item-level validity, potentially due to the fact that job relevance covaried with the subjective and non-verifiable dimensions; equal access was negatively related to item-level validity, and was unrelated to or negatively correlated with all the other item-type dimensions). No support was found for the item attribute – social desirability relationship.

In contrast, Lefkowitz, Gebbia, Balsam, and Dunn (1999) examined Mael's (1991) taxonomy in relation to biodata item validity and found somewhat different results. The authors took a different approach in operationalizing Mael's (1991) constructs, fleshing out in operational terms the dichotomous nature of most of the dimensions, and, in the case of four of the dimensions, developed midpoints or trichotomous items (e.g., "contemporary" as a midpoint in the historial vs. future/hypothetical dimension). Five expert raters used this classification scheme to categorize the 160 biodta items, which were then administered and concurrently validated against a 7-dimension supervisory rating criterion on a sample of university clerical workers. Item validity was associated with items that were indirect (i.e., secondhand) and non-controllable, with some support for the relationship between item validity and verifiability and job relevance. Regressing these four dimensions onto item validity resulted in the dimension of directness (firsthand) was the most (uniquely) predictive, being inversely related to item validity. However, issues of social desirability and applicant dissimulation were not investigated in this study.

Finally, in a replication and extension of Schmitt and Kunce's (2002) investigation of the impact of requiring elaborated responses to a subset of biodata

items, Schmitt, Oswald, Kim, Gillespie, Ramsay, and Yoo (2003) investigated the relationship of response elaboration to social desirability and impression management. The authors found that there was a negative relationship between the objectivity and verifiability of the biodata items (based on Mael's [1991] classification) and the items' relationship to participants' social desirability and impression management scores (based on Paulhus' [1991] measure). In other words, items that are more objective or verifiable may be less susceptible to faking, defined as either socially desirable responding or impression management.

These studies suggest that prescreening items that are more objective and verifiable in nature are less susceptible to applicant faking. Further, items that are more transparent in nature (i.e., more visibly job-related) may be more susceptible to faking (cf. Schrader & Osburn, 1977). Therefore, it is possible that some objective prescreening items may be faked in practice. Mueller-Hanson et al. (2003) found that a personality test showed greater validity at lower scores than at higher scores. According to the authors, faking may have different effects on validity at various points in the score distribution. In essence, applicants at the top of the distribution may have high scores either because they are truly high on the characteristic being assessed, or because they have successfully faked their performance on the test. In the former case, these applicants would be expected to perform well on the job, whereas the latter group of applicants would be expected to perform less well on the job. At the low end of the distribution, applicants' scores may be low because applicants truly are low on the characteristic being assessed, or because they were unsuccessful at faking

to improve their scores. In this case, neither group would be expected to perform well on the job.

Ultimately, these scenarios suggest that validity may be more negatively impacted at the higher end of the score distribution than at the low end, because more error variance—presumably due to faking—is introduced at the high end of the distribution. Therefore, it is possible that prescreening items that distinguish among the highest-scoring applicants may be those that are most objective in nature, since these would be the most difficult for applicants to fake. However, these questions are less likely to assess critical minimum job requirements, as applicants in the high end of the prescreening score distribution would be expected to have met minimum qualifications or to have successfully faked their qualifications. In other words, the items that distinguish between the highest-scoring applicants may be more objective in nature, but less clearly job-related.

In contrast, items that distinguish between high- and low-scoring candidates should be more job-related in nature, since low-scoring applicants are expected to lack minimum qualifications for the position or were unsuccessful at faking their qualifications. Based on the above discussion of online applicant prescreening validity and prescreening item characteristics that contribute to applicant differentiation and predict performance and retention criteria, the following hypotheses are proposed.

Hypothesis 1a: Prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings.

Hypothesis 1b: Prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications.

Hypothesis 2a: Applicant prescreening scores will be predictive of key new hire job performance criteria (e.g., test scores, hiring decisions).

Hypothesis 2b: Prescreening items predictive of key new hire performance criteria will be high on objectivity and verifiability.

Another potentially useful tool for informing best practices regarding the use of applicant prescreening and understanding the nature of distinctions between applicants is to examine the overall score distribution for meaningful "breaks" or sharp drop-offs in the frequency of applicants receiving a given score as one moves from the low end to the high end of the score distribution. Using a histogram analysis in this fashion is analogous to the scree plot number of factors decision typically used in exploratory factor analysis. Given the lack of firm research evidence upon which to make specific predictions regarding the nature of the differences between applicants at various "break" points, the following research question is proposed.

Research Question 1: Are there meaningful distinctions between applicants that can be made at points in the applicant prescreening score distribution where there are sharp increases or decreases in the number of applicants receiving a particular score? What is the nature of these distinctions?

Another key issue to address concerns whether there is systematic covariation in applicants' responses to prescreening items that can be used to predict whether

applicants receive job offers, and if so, whether they will be successful in the position. Given the lack of firm evidence on which to base predictions, the following research question is proposed.

> *Research Question 2:* Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?

Résumé Study

Little is known about the quality of the résumé evaluation process, although some practitioners consider the résumé to be of little value (cf. Handler, 2002). Research on résumés has examined recruiter perceptions of résumé content. Brown and Campion (1994) investigated recruiters' perceptions of the presence of biodata content in the résumé and found that recruiters interpreted biodata items in résumés reflecting abilities and other attributes (e.g., leadership, motivation) and these inferences were made reliably. Moreover, these items were interpreted differently depending on the job in question. That is, recruiters were able to determine that résumé content reflected a specific type of ability, as well as which abilities were more or less attractive (e.g., job-related) for a position. Fritzsche and Brannick (2002) conducted a policy capturing study investigating the comparability of screening résumés based on profiles (e.g., profiles developed based on graduate student criticality ratings of résumé elements such as education, work experience, etc.) versus actual résumés. Recruiters provided ratings of applicant suitability based on profiles or actual résumés. Results indicated that ratings based on profiles were significantly

higher than ratings for actual résumés, and cues were emphasized differently across the two conditions, as evidenced by significant differences in the beta weights derived from regressing suitability ratings onto the résumé cues. Moreover, recruiter evaluations were more consistent when based on evaluation of profiles. Taken together, the results of these studies suggest that although recruiters may be able to identify critical information contained within résumés, their evaluation of this information, especially across large numbers of résumés (i.e., 50 or more) is inconsistent.

Résumés can be evaluated in multiple ways (See Table 2 for an overview of each résumé evaluation method). The most basic way this can be accomplished is through recruiters scanning through applicant résumés (referred to hereafter as manual résumé evaluation or scanning). This procedure involves recruiters reading through each applicant's résumé, evaluating the content based on the recruiter's implicit assumptions of what constitutes an ideal applicant. The evaluation process amounts to a global impression of each applicant's suitability. Due to the potential for information overload associated with this task, a substantial amount of information that could differentiate between applicants may be lost. The consequences of information overload include the possibility of advancing unqualified candidates and overlooking some qualified candidates, as well as other potential decision inaccuracies. Research in the decision-making literature has suggested that as the number of attributes on which to base a decision increases, the level of decision accuracy decreases (e.g., Davis & Davis, 1996; Helgeson & Ursic, 1993).

A potential solution to this problem is to find a way in which to structure the résumé evaluation process. Research on the selection interview has demonstrated that providing structure to the interview process, in terms of asking the same questions of all candidates and developing guidelines for scoring candidates' responses results in higher-quality organizational outcomes in terms of hiring the best candidate (Campion, Palmer, & Campion, 1997). Standardizing the selection interview has been associated with a greater tendency for rulings in favor of the organization in disparate treatment and disparate impact court cases (Williamson, Campion, Malos, Roehling, & Campion, 1997). Pulakos and Schmitt (1995) found that a standardized interview using experience-based questions demonstrated a significant relationship with performance ratings. Based on these results, providing structure to the evaluation of applicant résumés may improve decision outcomes.

Another potential solution is to automate the résumé scanning process. Automated systems vary in terms of the richness of their information processing capabilities. These range from the use of keyword searches to programs that recognize the structure of résumés and can draw inferences between various elements within that structure. The strengths of keyword searching include the amount of time saved by recruiters in terms of reducing the amount of résumés they must examine (i.e., résumés that do not contain the relevant search terms are excluded), and their ease of use. However, there are a number of weaknesses associated with keyword searches as well. First, keyword searches can only examine a few key applicant qualifications at a time. Second, given that applicants may use different, synonymous terms that may be

of interest to the recruiter, the recruiter may not search on each of these, therefore potentially eliminating qualified candidates from consideration. Third, simple keyword searches cannot combine multiple terms in meaningful ways, and consequently, recruiters will still need to sift through a number of candidates who are not qualified for the job. Finally, keyword searches cannot capture the context in which a particular skill or applicant attribute may actually be used in a particular industry or job, resulting in recruiters sifting through numerous unqualified candidates and/or omitting candidates who would have been qualified for the position.

Fortunately, systems have been developed that can address some of the limitations of the keyword search. However, these are relatively new and consequently they are still being refined and evaluated. These systems blend keyword searching with a structured evaluation process, based on semantic recognition of résumé content. In other words, these programs not only recognize the structuring of résumés, they can also make connections between different phrases or terms contained in the résumé in meaningful ways. Many of the limitations of keyword searching can be addressed by automated résumé scoring systems. For example, these programs can recognize and understand the usage of various terms within different industries and jobs. However, development of these systems is both time-intensive and expensive.

Briefly, proprietary automated résumé scoring programs have the capability of matching the content of résumés to job requirements as they would appear in a job description or requisition (DeSanto & Crow, 2004). To produce such a system, a hierarchically structured knowledge base must be developed. This knowledge base can

be organized around job roles, which are structured hierarchically to the most specific job class or job title. Incorporated into this structure are synonymous terms or titles that allow the program to recognize alternative descriptions of the same position. This format is also used to describe and recognize skills and experience related to these positions (DeSanto & Crow, 2004). Further, the knowledge base can also be structured by market or industry, which provides additional context on which to evaluate applicant résumé content. Once the résumé scoring program extracts the relevant content from the résumé, skill concept scores are produced by the program based on a formula that may take into consideration recency of skill use, duration of the use of relevant skills, and any related skills as part of the scoring algorithm, the end result being an overall score for each relevant skill across applicants (e.g., Crow & DeSanto, 2004). These are referenced against the job description in order to determine which applicants are the best matches to the position. Matches are determined by the proximity of an applicant to the requisition job description (conceptualized as the "ideal" candidate), as mapped in the concept space (e.g., in a 3-dimensional space; Crow & DeSanto, 2004), using Euclidean distances (similar to a cluster analysis).

A potential alternative to the use of automated résumé scoring systems is to develop a system for manually scoring résumés. This methodology is analogous to the way in which T&E forms and structured interviews are scored. That is, based on predetermined critical job experiences, applicants may be assigned scores based on the presence and/or quality of their work experience based on the content of their résumés. The potential benefits of this procedure include reducing the cognitive demands

associated with scanning multiple résumés and removing the reliance on global assessments of résumé quality, the latter potentially resulting in lowered overall decision quality. The potential disadvantage of manually scoring résumés is that this process may not directly reflect the ways in which recruiters combine résumé elements to arrive at overall decisions. Specifically, Hitt and Barr (1989), in their investigation of managers' decision-making based on résumé information, found evidence that managers engaged in configural cue processing. That is, managers in this study combined résumé information in non-linear ways to arrive at overall assessments of applicant quality. Other studies, however, (e.g., Graves & Karren, 1992) have not found evidence of configural cue processing in evaluating applicant information.

Based on the relative merits and liabilities associated with these different methods, an important question to answer is to determine which method(s) provide the greatest correspondence to the results from the applicant evaluation process as it currently exists in the focal organization. That is, which applicants would be passed on to the interview based on each of the résumé evaluation methods? Based on the advantages and disadvantages of each applicant evaluation method (See Table 3), manual résumé scoring and automated résumé scoring have two primary advantages over the current résumé evaluation procedure, manual résumé scanning. First, each of these methods is standardized in terms of the content of the résumé that is evaluated. Therefore, each applicant is evaluated on predetermined qualifications that are considered the most important or desirable. Second, this information receives a rating in terms of the quality of the applicants' experience, which results in an overall score

for each applicant or degree of match to the "ideal" candidate for the position for each applicant. However, these standardized procedures may not capture the way in which recruiters evaluate applicant qualifications based on the résumé. That is, recruiters may combine applicant qualifications in non-linear and/or compensatory ways (i.e., configural cue processing; Hitt & Barr, 1989). This advantage may be counterbalanced, however, by the fact that manual résumé scanning is more susceptible to human error and fatigue.

In addition to examining the comparability of results across multiple applicant evaluation methods, it is also important to determine the degree to which these evaluations are made reliably for methods that involve human judgment. As mentioned previously, there is research evidence suggesting that evaluators use distinct strategies in deriving their final selection decisions (Graves & Karren, 1992), based on the ways in which they combine and assess the information. Therefore, when structure in the evaluation process is lacking, as in manual résumé evaluation, recruiters may operate on their tacit assumptions regarding what constitutes an ideal applicant. However, as more structure is infused into the process, as in a manual structured résumé scoring procedure, recruiters are more likely to focus on the same information across applicants in their evaluations, which should lead to greater interrater reliability. Based on the above discussion of the benefits of structure and standardization, the following research question and hypothesis are proposed.

Research Question 3: What is the extent of the overlap between the results from an unstructured vs. a structured résumé evaluation system?

Hypothesis 3: Reliability of a structured résumé evaluation system will be greater than the reliability of an unstructured résumé evaluation system.

CHAPTER V: STUDY 1 METHOD

Study 1: Prescreening Study

Participants

Job applicants. Prescreening data from actual applicants were collected from a major utility company located in the Western US. Data were sampled from three positions across the organization's geographical service area to allow for a representative examination of prescreening across these job classes. (See Tables 1 and 2 for specific samples used.) One entry-level customer service job, one entry-level physical job, and one above-entry-level physical job which contained a significant customer service component were investigated. The entry-level customer service job, Customer Service Representative (CSR), is characterized by such tasks as responding to customer telephone inquiries (e.g., regarding billing, payment, etc.) and performing necessary documentation. Utility Worker (UW), the entry-level physical job investigated in the present study, involves such core job activities as digging trenches, raising and lowering tools and equipment, transporting vehicles and equipment, and flagging traffic. Finally, Gas Service Representative (GSR), the focal above-entry-level physical job, includes key tasks such as interacting with residential and business customers, troubleshooting and re-lighting various gas appliances, and investigating reported gas leaks. A total of 5,619 unique candidates made an initial application for any one of the three positions investigated.

Utility Worker. The UW position applicant pool contained an initial number of 1,443 applications. Of these, 1,296 (89.8%) were unique applications to the UW

position (candidates could screen multiple times to the same or to different requisitions if they chose to do so). The sample was further reduced based on codes assigned to candidates that indicated what stage in the process they were in or where they were no longer considered for the position. In total, there were 8 codes used to classify applicants. Depending on the stage in the selection process, these codes were assigned either by the screening vendor or by members of the organization involved in the selection process. The stages/statuses which correspond to these 8 major classifications were as follows. First, a candidate may have completed the online screening but failed to call the number provided to take the telephone screen. (It is also possible, though rarely the case, that a candidate did not pass the online screening.) Candidates may also have taken the telephone screen (1), but did not pass it, or, alternatively, they may have decided they were no longer interested in the position, so they were coded as not qualified (2) or not interested (3), respectively. A candidate may have passed the screening, was scheduled for a testing appointment, and failed to attend the test session (4). A candidate may have attended the test session, but did not qualify on the preemployment test (5). Or, the candidate may have qualified on the screening but there were no open appointments for the candidate to be scheduled to take the preemployment test (6). Candidates may have taken the preemployment test and qualified on the test (7). Finally, of those candidates who qualified on the preemployment test, some also successfully interviewed and passed the background check and drug screen were coded as being hired by the organization (8). Of the 1,296 applicants to the UW position, 263 (20.3%) completed the online screen but did not

call to take the telephone screen or failed the online screen, 148 (11.4%) indicated at some point in the process that they were no longer interested in the position, 104 (8.0%) completed and passed the screening but were coded as in progress because no testing appointments were available, 207 (16.0%) completed and passed the screening, were scheduled for a testing appointment, and did not attend the test session ("no shows"), 223 candidates (17.2%) were classified as not qualified, either because they failed the telephone screen or because they failed the preemployment test, and the remaining 351 candidates (27.1%) passed the screening or passed the preemployment test. It should be noted that of these, 72 indicated they had previously tested and qualified on the preemployment test. Because this was self-report data, not all of these candidates previously qualified on the preemployment test, nor did they all necessarily previously take the test. It should also be noted that these codes were subject to human error and could not be verified in all cases. Therefore, in the results to follow in the next chapter, total *N*s may deviate somewhat from those reported here. Nevertheless, based on the frequencies noted above, candidates who took the online screen only and/or failed the online screening, were coded as being in progress, or indicated they were not interested in the position were removed from the sample. In all, this resulted in the removal of 509 candidates. In addition, 36 additional candidates who had discrepant data were also removed. Examples of such discrepancies included candidates who were coded as passing the preemployment test but either did not pass the test or could not be matched up in the organization's testing database. In all, the final useable sample contained 751 candidates. In terms of demographic make-up,

2.8% of the candidates were female, 73.2% were male, and the remaining candidates did not disclose their gender. White candidates made up 32.1% of the sample, 19.3% were Hispanic, 11.3% identified themselves as African American, Asian/Pacific Islander candidates comprised 6.5% of the sample, 1.1% were American Indian, and the remaining candidates did not specify their race/ethnicity.

Customer Service Representative. The CSR applicant pool contained an initial number of 3,051 applications over the course of a 5-month period. Of these, 2,874 (94.2%) were unique applications. Because data on hired candidates were collected throughout an entire 1-year period, candidates who qualified on the preemployment tests throughout the year were also added to the dataset for comparison purposes. This resulted in the inclusion of 215 additional candidates, for an initial candidate sample of 3,089. As described in the UW sample, the CSR sample was further refined based on a candidate coding system that reflected the final stage or result the candidate obtained in the selection process. Due to discrepancies between candidate codes and the other data available on candidates in the dataset, 16 candidates were removed from the sample. Of the 3,073 remaining candidates to the CSR position, 586 (19.1%) completed the online screen but did not call in to take the telephone screen or failed the online screen, 191 (6.2%) indicated at some point in the process that they were no longer interested in the position, 90 (2.9%) completed and passed the screening but were coded as in progress because no testing appointments were available, 270 (8.8%) completed and passed the screening, were scheduled for a testing appointment, and did not attend the test session ("no shows"), 1,152 candidates (37.5%) were classified as

not qualified, either because they failed the telephone screen or because they failed the preemployment test, and the remaining 784 candidates (25.5%) passed the screening and passed the preemployment test. Of these candidates, 82 (10.5%) indicated they had previously tested and qualified on the preemployment test battery. A total of 70 candidates in the sample (2.3%) were hired.

Based on the frequencies described above, candidates who took the online screen only and/or failed the online screening, were coded as being in progress, or indicated they were not interested in the position were removed from the sample. In total, this resulted in the removal of 867 candidates. Of the remaining 2,206 candidates for the CSR position, 1,188 (53.9%) were female, 531 (24.1%) were male, and the remaining 487 candidates (22.0%) did not provide this information. In terms of race/ethnicity, 589 candidates (26.7%) did not provide data, 542 (24.6%) were Hispanic, 445 (20.2%) were White, 407 (18.4%) were African American, 202 (9.2%) were Asian/Pacific Islander, and the remaining 21 (1.0%) were American Indian.

Gas Service Representative. The GSR applicant pool contained an initial sample of 1,751 applications over the course of a 1-year period. Of these, 1,449 (82.8%) were unique applications. As was the case with the UW and CSR positions, this sample was further reduced based on codes assigned to candidates indicating where in the selection process a decision was made on each candidate. After removal of 20 candidates with discrepant codes which could not be reconciled, the remaining 1,429 were broken down into the following classifications. There were 409 candidates (28.7%) who completed the online screen but did not call to take the telephone screen

or failed the online screen, 125 (8.7%) indicated at some point in the process that they were no longer interested in the position, 97 candidates (6.8%) completed and passed the screening but were coded as in progress because no testing appointments were available, 199 candidates (13.9%) completed and passed the screening, were scheduled for a testing appointment, but did not attend a test session ("no shows"), 373 candidates (26.1%) were classified as not qualified, either because they failed the telephone screen or because they failed the preemployment test battery, and the remaining 226 candidates (15.8%) passed the screening and passed the preemployment test battery. Of these 226 candidates, 51 (22.6%) indicated they had previously tested and qualified on the preemployment test battery for physical jobs.

As described in the discussion of the UW and CSR samples, candidates in the GSR sample who took the online screen only and/or failed the online screening, were coded as being in progress, or indicated they were no longer interested in the position were removed from the sample. This resulted in the removal of 631 candidates and a final useable sample of 798 candidates. Female candidates comprised 27.8% of the sample (n = 222), 374 (46.9%) were male, and the remaining 202 candidates (25.3%) did not report their gender. In terms of the ethnic or racial background of the final GSR candidate sample, 203 candidates (25.4%) were African American, 157 (19.6%) were White, 109 (13.6%) were Hispanic, 70 (8.8%) were Asian/Pacific Islander, 4 (0.5%) were American Indian, and the remaining 256 candidates (32.1%) did not report their race/ethnicity.

Graduate student subject matter experts. A sample of 11 graduate students in industrial/organizational psychology who were knowledgeable about employee selection served as subject matter experts (SMEs) in this study. These students participated in a rating exercise designed to classify the content of the prescreening scripts (i.e., the prescreening questions) along 13 question classification dimensions. (See definitions and sample surveys in Appendix A.) A list of potential participants was compiled with the assistance of a faculty member who identified all potential graduate students who had classroom and/or practical experience related to employee selection. All potential participants were informed about the study and were invited to participate via email. Interested students contacted the faculty member who compiled the potential participant list, who indicated how many students had expressed interest in participating. Surveys were sent to the faculty member via US mail, and a pre-addressed, postage-paid return envelope was provided for each participant to mail back their completed surveys. These materials were obtained by the graduate student participants from the faculty member. Surveys were self-administered by the graduate students and completed surveys were returned over a 2-month period in the pre-addressed, postage-paid envelopes provided. Demographic data on this graduate student sample was not collected. However, the demographic make-up of this sample most likely reflects that which is reported in Study 2, given that these students were identified for both studies from the same pool of potential graduate student participants, and, it is possible that a portion of the students who participated in Study 1 also participated in Study 2.

Materials

Prescreening data. Archival prescreening data were culled from the organization's screening vendor databases for the three key high-volume positions. As shown in Table 4, these screening protocols were broken into two sections, with applicants typically completing the first half of the screening online. If applicants passed the first half of the screening, they took a second screening via telephone with a live screener from the screening vendor organization. Across the entire screening protocol, there were a total of 9 questions in the UW script, 13 questions in the CSR script, and 8 questions in the GSR script. The online portion of the screen was comprised of 4 to 5 questions focused primarily on legally required and other perfunctory "gateway" questions (i.e., minimum requirements) which cut across all jobs for which a screening assessment was put in place in the organization. The remainder of the questions tapped basic job experience, work history, job-relevant skills, and motivation.

Preemployment test batteries. Qualified candidates (based on their performance in the screening phase of the selection process) took a cognitive ability-based, paper-and-pencil preemployment test as part of the selection process. Candidates for the UW and GSR positions took a 2-component test battery which was used as an indicator of ability to successfully perform in physical jobs. Each of these test components came from the Psychological Services, Inc. (PSI) Employee Aptitude Survey (EAS) series of cognitive ability tests. The entire EAS series consists of 10 individual tests designed for preemployment, promotional, and training and

development purposes. These tests were developed and validated based on extensive research and have been normed on nearly 100 different occupational and educational classifications. The two test components that make up the preemployment test battery for physical jobs were identified based on a thorough job analysis of the positions for which they are used, employing a validity transportability approach. Each test component employed a multiple choice format. Test battery validities based on meta-analysis of over 160 studies (corrected for criterion unreliability) for technical and mechanical/production jobs are reported as .46 and .35, respectively (Psychological Services, Inc., 2008). These batteries included the two test components of the preemployment test battery for physical jobs, but it should be noted that the technical and mechanical/production test batteries employed additional test components.

The first of the two test components administered for the physical jobs was a measure of basic mathematical ability. The test consisted of a total of 75 items, divided equally into 3 sections, which measured the ability to perform arithmetic calculations. The first section (2 minutes) measured candidates' ability to work with whole numbers, section two (4 minutes) required candidates to make calculations using decimals and percents, and the third section (4 minutes) required candidates to perform calculations using fractions. Candidates marked their answer on the test form from among the possible answers. If their answer was not provided, they selected the "none of the above" option. The total testing time for the mathematical ability test was 10 minutes.

The second component of the physical jobs test battery was a measure of spatial ability. Candidates were presented with a series of diagrams of block piles, five of which were labeled with a letter (A through E). For each of the lettered blocks, candidates were asked to indicate how many other blocks in the pile the lettered block touched along its sides, tops, bottoms, or ends (not corners). The test contained a total of 10 block piles with 5 lettered blocks in each block pile, for a total of 50 test items. The time allotted for the spatial ability test was 5 minutes. To qualify on the test battery for physical jobs, candidates had to obtain a qualifying score on each of the two test components. In other words, the preemployment test battery for physical jobs employed a multiple cutoff scoring format.

Candidates for the CSR position took a 5-component test battery which was used as an indicator of their clerical ability. The 5 components of the preemployment test battery for clerical jobs came from the PSI Basic Skills Tests (BST) series of cognitive ability tests. These tests were developed for use in customer service, clerical, and administrative positions. The entire BST series consists of 15 tests. These tests were developed and validated based on extensive research involving over 17,000 employees in more than 60 organizations. Generalized validities reported by PSI for the BST series range from .27 to .34 for individual test components (Psychological Services, Inc., 2008). The generalized validities provide estimates of the validity that would be obtained when using these tests to predict performance in jobs similar to those on which the tests were validated in the PSI testing database. The 5 components used for the CSR position were identified through a thorough job analysis and validity

transportability study. As was the case with the two test components in the physical jobs test battery, the test format for each test component in the CSR test battery was multiple choice.

The first test component in the clerical ability test battery was a measure of basic language skills. Candidates were presented with a series of sentences with a word or portion of each sentence underlined. Candidates had to determine whether there was an error in spelling, punctuation, grammar, or usage contained in the underlined word or portion of the sentence, or whether there was no error. This test component consisted of 25 items with a 5-minute time limit.

The second test component assessed basic arithmetic, working with whole numbers, decimals, percents, and fractions. Unlike the arithmetic test component for the physical jobs described above, the mathematical ability test for the CSR position was not divided into sections. Candidates selected their answer from among the choices listed, and if their answer was not provided, they chose the "none of the above" option. The total test contained 40 multiple-choice items with a total testing time of 5 minutes.

The third test component assessed candidates' problem solving ability. The test was comprised of 25 word problems. Candidates read a problem, discerned what information in the problem was relevant to determining the solution, performed basic arithmetic calculations to arrive at their answer, and recorded their answer on the test form based on the available response options. If their response was not one of the

options provided, they selected the "none of the above" option. The total time allowed for this test component was 10 minutes.

The fourth test component assessed candidates' ability to code information. Candidates were presented with a table of codes along 4 categories (e.g., department, item, etc.). For a set of items, candidates were provided with information (e.g., Main Street, paper clips, $800) related to the categories and had to code the information accordingly. The test consisted of a total of 72 items (18 items with four coding tasks per item) with a 5-minute time limit.

Finally, the fifth clerical ability test battery component assessed visual speed and accuracy. Each item consisted of two sets of alphanumeric data (e.g., 791 79!) and candidates had to indicate whether the two were exactly the same or whether they differed in any way. The test consisted of 150 items with a 5-minute time limit. To qualify on the preemployment test battery for clerical jobs, candidates had to obtain a passing score based on the unit-weighted sum of their scores on each of the 5 test components. In other words, this preemployment test battery used a compensatory scoring model.

Hiring data. The organization's internal hiring database was queried to identify those candidates who had been hired into the focal jobs. Data were collected on whether the candidate was hired and whether the employee had been terminated during the 6-month probationary period. Although these data were available for employees in each of the jobs investigated in the present study, the GSR position, due to its low volume of hires during the data collection period, was not included in the

analyses involving candidates who were ultimately hired into the organization. Of the candidates who qualified on the preemployment test battery, candidates were coded 1 if they were hired by the organization and 0 if they were not hired. Hired candidates were coded 0 if they were not terminated from the position during the 6-month probationary period, and 1 if they were terminated.

Prescreening item taxonomy rating survey. Graduate students in industrial/organizational psychology were administered 3 surveys assessing the content of the prescreening scripts for each of the positions investigated (see Appendix A). Using a methodology similar to that which was used in Becker and Colquitt (1992), definitions of the 10 biodata dimensions as outlined in Mael (1991) were developed, along with the addition of 3 other dimensions (minimum qualifications, social desirability, and susceptibility to faking). The surveys presented participants with a description of the position for which they would be evaluating prescreening items as it would appear to an applicant applying on the company's website. This was followed on the next page by the definition for the first prescreening taxonomy dimension, which concluded by asking the participant, "Please indicate the extent to which the following questions are _____." The entire screening protocol for the position in question followed. This format was followed throughout the entire prescreening item taxonomy rating survey. On a 5-point scale, participants indicated the extent to which each prescreening item was indicative of that item taxonomy dimension (1 = not at all; 5 = completely). This process was repeated for each of the 13 dimensions across the three focal jobs. That is, once participants completed ratings

on the first dimension, they were then presented with the definition of the second dimension, and made ratings on all the prescreening questions in relation to that dimension.

Procedure

Selection process. As mentioned previously, interested parties visit the company's website to view job openings within the organization and express their interest in a position by providing some basic information about themselves. If an interested party decides to make an application for a position, the process continues on to the first phase of the screening process which is completed online. In the online screening stage, the candidate is presented with a small set of prescreening questions. The initial online screen is primarily geared toward asking questions that are required for legal purposes (e.g., authorization to work in the US) and other minimum qualification (MQ) questions. The questions asked in this phase of the screening are primarily in yes/no format, with the CSR script also containing an experience-based question with a multiple-choice response format (see Table 4). If the candidate does not select a response to any of the initial online screening questions designated as a disqualifier, he or she is provided a number to call to take a further screening with a live screener via telephone. The candidate does not continue in the selection process unless he or she calls the screening organization to take the telephone screening.

The telephone screen is more detailed than the initial online screen. It is conducted with a live screener (as opposed to an automated system such as IVR), and has a greater focus on job-related experience and skills. Similar to the online screen,

many of the telephone prescreening questions contain disqualifiers (knockouts). Another primary difference between the online and telephone screens is the inclusion of some questions that are open-ended and screener-rated. These questions appear in the CSR and GSR scripts and focus on motivation to join the department and organization and the candidate's overall communication skills (see Table 4). For each of these questions, the telephone screener uses anchored rating scales to assess the candidate's performance. (See Table 5 for examples of these anchors.) Specifically, in the case of the motivation to join the department (CSR) or organization (GSR) question and the communication skills question (CSR and GSR), the screener evaluates the candidate's performance and assigns a score based on the anchored rating scales.

If a candidate chooses a disqualifying response to any of the questions in the telephone screen, the telephone screen ends. However, if a candidate provides satisfactory responses to all questions, he or she is scheduled to take a paper-and-pencil preemployment test. At the test location, candidates also complete a job application blank and authorize a background check to be conducted. Candidates receive cursory feedback on their performance on the preemployment test via mail (qualified/not qualified). Based on a recruiter's review of the candidate's application blank and test performance, competitive candidates who qualified on the preemployment test are invited to interview. Based on performance in the interview, job offers are made to candidates, pending the results from a standard background check and drug screen.

Power Analysis

A power analysis (UCLA Department of Statistics, 2002) was conducted to determine sample sizes needed to detect a correlation of .10 with 80% power. The magnitude of correlation used for this analysis was determined based on the nature of the analyses conducted to test the hypotheses. Specifically, some of the tests required an assessment of prescreening item-total and prescreening item-criterion relationships. Given that it was expected that prescreening total scores would have, at best, moderate relationships to key performance and retention criteria (e.g., $r = .20$), individual items were expected to have even smaller relationships. Relationships lower than $r = .10$ have little if any practical significance, as the variance accounted for by such items would be less than 1%. Based on this analysis, the minimum sample size needed for each job type investigated in the study to detect a correlation of .10 with 80% power (using a 2-tailed test) was 781 (616 1-tailed). To detect a correlation of .20 with 80% power, a minimum sample of 194 applicants was needed (153 1-tailed).

To determine the sample of graduate student raters needed to detect significant differences in mean ratings of prescreening items that differentiated applicants based on Mael's (1991) biodata dimensions, a power analysis was conducted using the following parameters. Based on mean differences in biodata item ratings from previous research (Becker & Colquitt, 1992), the sample size needed to detect a mean difference of 0.80 (on a 5-point scale), given a pooled standard deviation of 1.00, with 80% power, was 25. Given the repeated measures design of this portion of the study, a minimum sample of 13 graduate student raters was needed. It should be noted,

therefore, that given the smaller participant sample that was obtained, the prescreening item comparison statistical tests may suffer from low statistical power.

Data Preparation and Coding

Data for each job investigated were provided by the organization's third party prescreening vendor in Excel files. Manipulations to the data were performed to derive meaningful values for use in analysis. Specifically, for single-response prescreening questions whose response categories fell along a continuum (e.g., years of experience), data were recoded to reflect the continuum. For example, if there were five response options to a question regarding years of experience in the utility industry, the shortest span of time received a code of 1, and the longest span of time received a code of 5. All questions of this type were recoded in this manner for the sake of consistency (and also to explore alternate ways of scoring the questions for purposes of the present study). For multiple response questions, each response option was treated as a single, binary item. For example, a multiple response question regarding experience with Microsoft Office software with 5 possible responses was recoded into 5 separate binary questions. Responses were coded 0 for no (did not choose the response option) and 1 for yes (did choose the response option). This is a standard approach to analyzing these types of items in biodata research (e.g., Schrader & Osburn, 1977). Data, once recoded and formatted, were imported into SPSS for analysis.

Utility Worker. There were 9 questions in the UW screening scripts that were coded for use in the present study (see Table 4). The online screening contained 4 questions. The first 3 of the online screening questions assessed minimum

qualifications (at least 18, high school diploma/GED, and possession of a valid [State] driver's license). The fourth question asked candidates whether they possessed a Class A driver's license. Each of these questions was binary in nature and were coded 0 for No and 1 for Yes.

If the candidate passed the initial online screen (i.e., did not select a disqualifying response to any of the minimum qualification questions) and called the screening vendor number provided at the end of the online screen, they were administered the remainder of the screening questions via telephone by a live screener. Several of the questions in the telephone screen were binary in nature. Each of these were coded 0 for No and 1 for Yes. These questions asked whether the candidate was interested in work that required exposure to a variety of weather conditions, whether they had ever worked for a utility company, and whether they could meet the overtime work requirement in emergency situations. Candidates for the UW positions were also asked how many years of construction experience they had. This question was treated as a continuous variable with 8 levels ranging from 0 (No experience) to 7 (5 or more years). Lastly, candidates were asked (if they answered Yes to the previous work experience with a utility company question) what type of utility they had worked for. This question contained 5 different response options (e.g., Electric, Phone, etc.) and was coded as 5 separate binary items, coded 0 if the candidate did not have experience with a particular type of utility, and 1 if they indicated they did have previous work experience with the type of utility company in question. Note that the way in which this question was created in the screening script precluded candidates from indicating

that they had worked in more than one type of utility company. The candidates' total screening score was comprised of the sum of their scores on each of the screening questions across both screens.

Customer Service Representative. In terms of the substantive content of the candidate screening questions for the CSR position, there were a total of 13 questions, 5 of which were self-administered by the candidate during the online screening phase. The remaining 8 questions were administered by a live screener via telephone if the candidate passed the initial screen and called the toll free number provided to take the second portion of the screening (See Table 4). The first 4 questions of the online screen (age 18 requirement, high school diploma/GED, previous computer experience, and ability to meet overtime requirements) were coded 0 for No and 1 for Yes. The final online screening question, which asked about previous customer service experience in a call center, contained six different response options, which were coded as a continuous variable, with the lowest level of experience (No experience) coded 0 and the highest level of experience (4 years or more) coded 5. As shown in Table 4, each of the first 4 questions in the online screen contained disqualifying responses. In each case, a No response prevented the candidate from proceeding to the second phase of the screening stage.

The telephone screening for the CSR position contained two multiple response questions (software experience and shift/work schedule availability). The software (e.g., word processing, spreadsheets, etc.) experience question was recoded into a series of 6 binary variables, reflecting each of the choices (each of these was coded 0

for no experience, 1 for experience; if the candidate selected None, they were assigned a score of 0). The total number of endorsed software programs was also summed to provide a total software experience score. The shift/work schedule availability question was coded as a simple binary item because the response options fell into two general categories: either the candidate indicated they could work any shift (coded 1) or they indicated they could not work one or up to seven days out of the week (coded 0). The typing skills question contained three levels which were coded along a continuum, ranging from less than 15 (coded 0) to 50 or more words per minute (coded 2). Anticipated tenure as a CSR, if hired, consisted of 6 different response options. These were coded as a continuous variable ranging from 1 (less than three months) to 6 (3 years or more). Candidates' customer service experience with the general public was coded in a similar fashion to the customer service experience in a call center question from the online screening. This variable contained 6 levels, ranging from No experience (0) to 4 years or more (5). The remaining 3 questions were scored by the screener. The first of these questions asked the candidate to provide reasons why he or she wanted to be part of the call center. The screener selected the one performance anchor that best described the candidate's response. Although there were 5 anchors, 3 of these anchors did not constitute a clearly discernable continuum (i.e., unable to provide a reason, prolonged hesitation in providing a reason, providing a negative reason). These anchors were coded 0 as they clearly did not reflect positive performance regarding this question. The most positive anchor (Immediately responded with enthusiastic reasons) was coded 2. The second

open-ended response question required the candidate to articulate the meaning of good customer service. The screener selected the appropriate behavioral anchor reflecting the candidate's performance. This variable was coded as a continuous variable ranging from Poor (1) to Excellent (5). Finally, the screener assessed the candidate's overall communication skills. The rating scale anchors were coded as a continuous variable, ranging from 1 (Poor) to 5 (Excellent). A total screening score was computed based on the sum of all the prescreening item scores across the online and telephone screen.

Gas Service Representative. The 8 questions contained in the GSR screening were split into 3 questions in the online screen, with the remaining 5 questions administered to qualified candidates in the telephone screening phase. Similar to the items described previously in the UW and CSR scripts, the online screen primarily assessed minimum qualifications, with each of the questions containing a disqualifying response option. Each of these questions (age 18 requirement, high school diploma/GED, possession of a valid [State] driver's license) had a yes/no response format, which were coded 0 for No and 1 for Yes.

Two of the questions in the telephone screening phase addressed work schedule considerations. The first, whether the candidate could meet the requirement of being able to arrive at the work location during an emergency with 30 minutes notice, was coded 0 for No and 1 for Yes. The same coding and response options also applied to the second work schedule-related question, whether the candidate could meet the flexible work schedule (i.e., nights, weekends) with occasional overtime requirement. The remaining 3 questions were screener-rated. The first asked

candidates to provide the screener with reasons why they wanted to join the organization. The anchors were identical to the ones for the similar question that was asked of CSR candidates regarding reasons CSR candidates wanted to join the call center. Because of the lack of a logical continuum for the three lowest anchor ratings (e.g., prolonged hesitation in providing a response vs. providing a negative response), each of these was coded 0. The remaining two anchors did form a logical continuum, and consequently, they were coded 1 and 2, with the highest anchor indicating the candidate responded with enthusiastic reasons. The remaining questions were identical to the ones asked of CSR candidates. The first asked candidates to describe what good customer service meant to them, which contained 5 rating anchors on a continuous scale. These were coded 1 (Poor) to 5 (Excellent). Lastly, screeners provided a rating of the candidate's overall communication skills. This variable was also coded 1 (Poor) to 5 (Excellent). The candidates' total prescreening score consisted of the sum of their scores to all the screening items in both the online and telephone screen.

Analysis of prescreening item content. As noted earlier, a sample of graduate students in industrial/organizational psychology served as SMEs in evaluating the prescreening questions across the 3 positions investigated, using Mael's (1991) taxonomy framework. Mael's biodata taxonomy consists of a total of 10 dimensions: verifiability, job-relatedness, objectivity, historical, external, firsthand, discrete, controllable, equal access, and noninvasive. In addition to these 10 dimensions, 3 additional dimensions were also assessed. Because one of the aims of prescreening is to assess whether candidates possess minimum qualifications for a position, graduate

students were asked to evaluate the prescreening protocols along this dimension. Also, because responses to some of the items in the prescreening protocols may be susceptible to faking, due to their potentially transparent nature, susceptibility to faking was the second additional dimension included in the prescreening taxonomy survey. Finally, as has been investigated in previous research (Becker & Colquitt, 1992), the level of each prescreening question's potential to elicit a socially desirable response was the third dimension added to the original 10 taxonomy dimensions from Mael (1991). Thus, in all, the graduate student SMEs made ratings of the prescreening items along a total of 13 dimensions. (See Appendix A for all taxonomy dimensions and definitions of each that were included in the surveys.) The ratings provided by the graduate students were used to assess the item content that differentiated applicants at various points in the prescreening score distribution, as well as the screening questions' relationships to key outcome criteria. Although only a subset of the item dimensions measured in the survey were tied to specific hypotheses in the study, the remaining dimensions were included and analyzed for exploratory purposes to determine whether any additional systematic relationships existed between these item characteristics and prescreening scores as well as later selection stage criteria and hiring outcomes.

Job types and descriptions. The hypotheses in Study 1 were tested using 3 high-volume jobs within the focal organization (See Tables 1 and 2 for specific samples used). The entry-level customer service job, Customer Service Representative (CSR), is characterized by such tasks as responding to customer telephone inquiries

(e.g., regarding billing, payment, etc.) and performing necessary documentation. Utility Worker (UW), the entry-level physical job investigated in the present study, involves such core job activities as digging trenches, raising and lowering tools and equipment, transporting vehicles and equipment, and flagging traffic. Finally, Gas Service Representative (GSR), the focal above-entry-level physical job, involves key tasks such as interacting with residential and business customers, troubleshooting and re-lighting various gas appliances, and investigating reported gas leaks.

Original Analysis Plan

The hypotheses and research questions proposed in Study 1 were tested using multiple methods and analyses. Table 1 outlines each hypothesis and research question, samples used, and originally-proposed analyses. To the extent possible, these tests were carried out as originally planned, but in some cases, due to data restrictions, modified versions of the analyses originally proposed were used. These are noted in the next chapter. Each proposed hypothesis test is outlined in detail below.

Hypothesis 1a (Prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings.) was tested by first dividing applicant samples (within specific positions) into two groups based on applicant score distributions. High-scoring applicants were defined as those applicants who scored at or above +1 standard deviation (SD; based on *Z*-scores) from the mean applicant prescreening score. Low-scoring applicants were defined as those applicants who scored at or below –1SD from the mean prescreening score. The mean score for each prescreening item within each group was compared

using *t*-tests or chi-square (depending on the nature of the item). Items where significant differences were found were grouped and compared to items that did not demonstrate significant group differences, using the graduate student sample's prescreening question content taxonomy survey, described below.

The mean ratings of prescreening items provided by the graduate student sample using Mael's (1991) biodata item type dimensions was used to analyze the nature of the items that were found to differentiate the two applicant groups mentioned above. The mean rating on each dimension was averaged across all items found to differentiate between high- and low-scoring applicants and was compared to the mean rating on each dimension averaged across all items found not to differentiate between high- and low-scoring applicants, using repeated measures *t*-test analysis. Due to multiple comparisons involved in this procedure (i.e., separate *t*-tests for each of the item type dimensions), significance levels (alpha) were adjusted to correct for familywise error. This procedure has been used in previous biodata research (Becker & Colquitt, 1992).

A similar procedure was used to analyze differences among the highest-scoring applicants. Specifically, to test Hypothesis 1b (Prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications.), applicant samples (within positions) were divided into two groups based on applicant score distributions. Due to the expected smaller number of applicants (based on a normal distribution of scores and the removal of 84% of the sample for each position; see

below) in the high end of the score distribution, dividing the sample into applicants scoring +1SD to +2SD above the mean prescreening score and applicants scoring above +2SD above the mean prescreening score may not be feasible. Therefore, a "median split" of applicants scoring at or above +1SD above the mean prescreening may be used in order to achieve more equivalent groups. In terms of percentile scores, this equated to a comparison of applicants scoring at the 84^{th}–91^{st} percentile to applicants scoring at the 92^{nd} percentile and above.

The mean score for each prescreening item within each applicant group was compared using *t*-tests. Items where significant differences were found were aggregated and compared to the grouping of items that did not demonstrate significant inter-group differences.

The mean ratings of prescreening items provided by the graduate student sample using Mael's (1991) biodata item type dimensions were used to analyze the nature of the items that were found to differentiate the two applicant groups. The mean ratings on each dimension were averaged across all items found to differentiate among the two high-scoring applicant groups and were compared to the mean rating on each dimension averaged across all items found not to differentiate among the two groups, using repeated measures *t*-test analysis. Due to multiple comparisons involved in this procedure (i.e., separate *t*-tests for each of the item type dimensions), significance levels (alpha) was adjusted to correct for familywise error.

Hypothesis 2a (Applicant prescreening scores will be predictive of key new hire performance criteria.), was tested by correlating applicants' total prescreening

scores with criterion data collected by the organization for each job. Due to the unionized nature of each of the positions and the current organizational climate (i.e., organization-wide culture and business transformation), disseminating a research-only performance survey was prohibitive. Candidate hiring data, where available, were also used to assess the relationship of prescreening scores to more distal key job-related criteria. Finally, data were also collected on whether candidates who were hired survived the initial 6-month probationary period.

Hypothesis 2b (Prescreening items predictive of key new hire performance criteria will be high on objectivity and verifiability) was tested through examination of the beta weights derived through multiple regression analysis (against performance criteria) associated with each prescreening item. The items with significant beta weights were grouped together and compared to items that were not predictive of key criteria, based on graduate student ratings of the items' objectivity and verifiability (averaged across items as described in previous analyses). Mean differences between the two groups of items were analyzed using repeated measures *t*-tests. Hypotheses 2a and 2b were tested for each of the three jobs.

Research Question 1 (Are there meaningful distinctions between applicants that can be made at points in the applicant prescreening score distribution where there are sharp increases or decreases in the number of applicants receiving a particular score? What is the nature of these distinctions?) was tested by developing histograms of prescreening scores for each of the three jobs investigated. Points in the score distribution were identified where there were sharp drop-offs or increases in the

frequency of applicants obtaining a given score. These differences were analyzed in terms of the content of the questions that differentiated applicants above and below the score distribution break point (based on student evaluation of prescreening question content as described earlier).

Finally, to examine to Research Question 2 (Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?), discriminant analysis was proposed as a way to determine whether prescreening response patterns are predictive of employee outcomes. For example, participants may be categorized into three groups: applicants who were hired and did not turn over, applicants who were hired and turned over, and applicants who were not hired. Discriminant analysis of prescreening responses may identify which questions classify applicants into each of these groups. Further, results of the analysis provide information regarding the accuracy of these classifications. This information may be informative in developing prescreening question scoring algorithms. Research Question 2 was examined in the UW and CSR applicant samples. It should also be noted that in addition to the specific item characteristic dimensions tested in the hypotheses, the remaining item characteristic dimensions on which specific hypotheses were not posited were also explored in each analysis.

CHAPTER VI: STUDY 1 RESULTS

Utility Worker

Descriptives and Zero-order Relationships

Means, standard deviations, and intercorrelations of the prescreening items used in the UW position are presented in Table 6. Note that due to the severe lack of variance in the preliminary screening questions (i.e., the questions from the online screen; see Table 4) and the fact that many of these questions required a yes response in order for the candidate to proceed in the screening process, these variables were not included in Table 6. Each of the screening questions, with the exception of the overtime requirement question, correlated significantly with the total prescreening score. The most likely explanation for the lack of correlation for this item is the severe restriction of range associated with this question. The candidates' total screening scores failed to demonstrate a significant relationship with test pass/fail ($r = -.07$, *ns*). However, whether candidates indicated they possessed a Class A driver's license was inversely related to passing the preemployment test battery ($r = -.11, p < .05$). However, it should be noted that this relationship was significant at the less stringent alpha level. Likewise, total screening score failed to demonstrate a significant relationship with the more distal outcomes of hiring decisions ($r = .04$, *ns*) and terminations. This most likely was a reflection of how hires were coded. Namely, only those candidates who qualified on the preemployment test battery for physical jobs were coded as hired or not hired. Although the total screening score was not significantly related to employee terminations (again, most plausibly due to the

severely restricted range in candidate scores, i.e., only candidates who had been hired by the organization were coded as having been or having not been terminated), there was a significant relationship between the previous experience in a utility company question and involuntary turnover ($r = .25, p < .01$). This association was even stronger for those candidates who more specifically indicated they had previous experience in a cable utility ($r = .57, p < .01$). However, given that the sample of termed employees was extremely small, these relationships are most likely spurious in nature, or at the very least, should be interpreted with caution.

SME Prescreening Item Taxonomy Ratings

Table 7 displays the mean prescreening item taxonomy ratings across each of the 13 dimensions investigated, their intercorrelations, and intraclass correlations (ICCs). In calculating the ICCs, raters were treated as fixed effects, and given that these raters were considered "experts," they were the only population of interest. The ICC values reflect the consistency measure of reliability as opposed to absolute agreement (cf. Shrout & Fleiss, 1979). Overall, the ICCs were high (.55–.98), suggesting satisfactory reliability, with the exceptions of job relatedness (.23) and the noninvasive dimension (−.06). Given the small sample of expert raters, few of the prescreening dimension correlations were statistically significant, although most were non-trivial. The controllable dimension was inversely related to most of the other dimensions, with significant relationships found between the controllable dimension and verifiability ($r = -.54, p < .10$), external ($r = -.64, p < .05$) and discrete ($r = -.59, p < .10$). In addition, the more verifiable, objective, historical, and external the items

were, the less equally accessible they were. Finally, objective items were also more job related ($r = .69, p < .05$).

Hypothesis Tests

Hypothesis 1a posited that prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings. Candidate total screening scores were first converted to *Z*-scores. Candidates with a *Z*-score greater than or equal to 1.00 were compared to candidates with a *Z*-score less than or equal to −1.00 on each of the following 5 questions: possession of a Class A driver's license, number of years of construction experience, previous experience in a utility company (further breakdowns by type of utility endorsed were explored), and ability to meet the overtime requirement. Due to the multiple comparisons made in testing this hypothesis, a more stringent alpha of .01 was used. Results of the *t*-tests appear in Table 8. As can be seen in Table 8, each of the comparisons by question was significant, with the exception of prior experience in a water utility company. Ability to meet the overtime requirement could not be tested because there was no variance in either group. Although the one non-difference between high- and low-scoring candidates (given the .01 alpha level) involved a response option to the utility company experience question as opposed to the overall question endorsement rate itself, for the purpose of exploring Hypothesis 1a in the UW sample, the question was considered to not differentiate the highest and lowest scoring candidates. To test for differences among the highest and lowest scoring candidates in terms of the questions' taxonomic classification (as provided by graduate

student SME ratings), the mean dimension ratings across the screening questions differentiating high- and low-scoring candidates was compared to the mean dimension ratings across the questions which did not differentiate between high- and low-scoring candidates. Thus, the mean screening taxonomy ratings across the Class A driver's license and years of construction experience questions (averaged) were compared to the mean screening taxonomy ratings for the previous experience in a utility company question. Results of the analysis appear in Table 9. The questions differentiating between the top- and bottom-scoring candidates were rated higher on minimum qualifications, $t(10) = 3.27, p < .01$. Although not hypothesized, the question which did not differentiate between high- and low-scoring candidates, previous experience in a specific type of utility company, was significantly more discrete than the average of the three items which differentiated candidates, $t(10) = -4.20, p < .01$. It is also interesting to note that, although not statistically significant, ratings of job relatedness and objectivity trended in the direction hypothesized. That is, of the prescreening items that differentiated high-scoring from low-scoring candidates, the SME expert ratings on these dimensions were higher than the same ratings made on the question which did not differentiate among the highest- and lowest-scoring candidates. These relationships may have been borne out in a larger sample of expert raters. Thus, Hypothesis 1a received partial support in the UW sample.

Hypothesis 1b posited that prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications. Due to the range

restriction in screening scores, a reflection of the way in which the screening was administered to candidates, comparing candidates with *Z*-scores greater than or equal to 2.00 to candidates with *Z*-scores between 1.00 and 1.99 could not be conducted (*Z*-scores in the sample ranged from −1.78 to 1.70). Therefore, to test Hypothesis 1b, percentile scores were examined as an alternative method. However, this approach was likewise unfeasible because the screening score values at both the 75th and 90th percentile were identical (raw score = 12). Upon examination of the screening score distribution (See Figure 2), there appeared to be a distinct split between candidates with a prescreening score of 12 and candidates who scored above 12 (the maximum score in the distribution was 14). Thus, the methodology proposed to explore Research Question 1 was employed as a means to test Hypothesis 1b. That is, candidates who received a prescreening score of 12 were compared to candidates who scored above this value. The mean item score comparisons are shown in Table 10.

Results of the item score differences between the highest scoring candidates differed somewhat from what was found in the comparison of high- vs. low-scoring candidates. Unlike the results from the test of Hypothesis 1a, the two groups of high-scoring candidates did not differ in terms of their previous construction experience, $t(250) = -1.88$, *ns*. Similar to the previous analysis, certain response options to the question regarding the type of utility company in which candidates had previous work experience (i.e., endorsement rates of previous work experience in a water utility company) did not differentiate between the two groups and, because each of these response options could not be assessed separately, the entire question was considered

to not differentiate between the highest-scoring candidates. Thus, Hypothesis 1b was tested by comparing the screening item taxonomy dimensional ratings for the Class A license and previous utility experience questions to the mean dimensional ratings across the years of previous construction experience and type of utility company in which the candidate had previous work experience questions. Results of these comparisons appear in Table 11.

In total, screening item characteristics differed between questions that distinguished the highest-scoring candidates from high-scoring candidates across 5 dimensions. Items that distinguished among the highest-scoring candidates were rated by the SME sample as being more verifiable ($t(10) = 3.51, p < .01$) and more objective ($t(10) = 3.61, p < .01$). The prescreening questions differed along 2 additional dimensions at the .05 level, but it should be noted that due to the number of comparisons, these differences may be an artifact of familywise error. Nevertheless, the results suggest that in addition to being more verifiable and objective, questions that differentiated among the highest-scoring candidates were also more discrete ($t(10) = 2.82, p < .05$) and less susceptible to faking ($t(10) = -2.47, p < .05$). Contrary to Hypothesis 1b, the prescreening items that differentiated the highest-scoring candidates were also rated higher in terms of their assessment of minimum qualifications ($t(10) = 2.61, p < .05$). Again, it should be noted that this comparison was significant using a less stringent alpha to account for the multiple comparisons made. In all, then, Hypothesis 1b received partial support: the prescreening questions that differentiated between the two high-scoring groups were more objective than the

questions on which the highest-scoring candidates did not differ. However, these questions were also considered to measure minimum qualifications to a greater extent than the questions where differences were not found.

Hypothesis 2a posited that applicant prescreening scores will be predictive of key new-hire criteria. Given that the criteria available on which to "validate" prescreening scores in the organization were sparse, screening scores were examined for their potential relationships to later stages in the selection process, such as preemployment test outcomes, hiring decisions, and whether the candidates who were hired survived their probationary period (defined as 6 months). Based on the zero-order correlations between total screening score and each of the two test components comprising the preemployment test battery for physical jobs ($r = -.07$, *ns* and $r = .00$, *ns* for the mathematical ability and spatial ability tests, respectively; $r = -.07$, *ns* with preemployment test battery pass/fail), Hypothesis 2a was not supported. Similarly, prescreening scores failed to demonstrate a significant relationship to the more distal outcomes of hiring decisions and whether the candidate was terminated during the initial 6-month probationary period (rs = .05 and −.02, *ns* for the hired and termed criteria, respectively). It should be noted that these relationships suffered from even greater restriction of range than the relationships tested in Hypotheses 1a and 1b. That is, these relationships were based on the select sample of candidates who attended a preemployment test session and subsequently qualified on the preemployment test battery ($N = 327$ for candidates who took the preemployment test, and of these, 70 were subsequently hired). Likewise, restriction of range and small sample size

truncated any potential relationship between terminations and prescreening scores (N = 69 in the termination sample, and of these, 3 were terminated within the 6-month probationary period).

Hypothesis 2b stated that prescreening items predictive of key new hire performance criteria would be high on objectivity and verifiability. Although the total screening score failed to demonstrate a relationship with key outcomes, two significant zero-order relationships were found between individual screening questions and the outcomes. Specifically, the possession of a Class A license question was significantly correlated with test battery pass/fail ($r = -.11, p < .05$), as well as two of the response options to the question regarding the specific type of utility company in which candidates had previous work experience ($r = .09, p < .05$, for the relationship between experience in an electric utility and the mathematical ability test component, and $r = -.10, p < .01$ for the relationship between experience in a cable utility and the mathematical ability test component). It is interesting to note that two of these relationships were negative, which was contrary to what was expected. That is, assuming these types of experience or qualifications were considered relevant to success in the position (or later stages of the selection process), then the demonstrated relationships, should any exist, would be expected to be positive. It is possible that due to the severe restriction of range in prescreening scores encountered in this dataset, the absence of the lowest-scoring candidates on this item (due to the fact that they were knocked out based on questions preceding this one) may have caused this relationship to be negative. There is also the possibility that a greater proportion of candidates

scoring the highest on this question faked their response, thus introducing error, with the end result again being that candidates who scored lower on this item performed better on the preemployment test. Nevertheless, the Class A license and which type of utility questions were compared to the construction experience and previous utility company experience questions along the prescreening taxonomy dimensions. Results are displayed in Table 12. At the more conservative alpha level (i.e., .01), two prescreening taxonomy dimensions demonstrated significant differences. First, screening questions that demonstrated a significant zero-order correlation with preemployment test pass/fail or with the mathematical ability test component score were rated significantly higher in terms of verifiability compared to the questions that did not demonstrate significant relationships to the preemployment test criterion, $t(10) = 6.21, p < .01$. Likewise, these questions were also rated as being significantly more objective, $t(10) = 3.61, p < .01$. Thus, Hypothesis 2b was supported in the UW sample.

Two additional prescreening item taxonomy dimensions demonstrated significant differences at the less conservative .05 alpha. Specifically, prescreening questions that were significantly correlated with test scores/outcomes were rated higher on minimum qualifications and were more discrete compared to questions with non-significant test criterion relationships, $ts(10) = 2.48$ and 3.09, respectively, $p < .05$. Finally, although not statistically significant at conventional alpha levels, screening questions with non-significant test criterion relationships trended toward being significantly more socially desirable, compared to those questions that

demonstrated a significant relationship to the preemployment test criterion, $t(10) = -2.17, p < .10$.

Research Question 2 (Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?) was modified slightly in terms of the analysis conducted. Given that the UW script contained only 4 questions on which a profile could be developed, instead of discriminant analysis, the approach taken to test the preceding hypotheses was used. In addition, much in the way of candidate "profiles" was examined in the main hypotheses. That is, mean scores on individual prescreening items were compared across those candidates who performed well on the screen and those who did not perform well. Another limitation of the data in the present study that affected the exploration of this research question was the fact that the sample of termed employees was insufficient to allow for a discriminant analysis. Thus, to examine Research Question 2, candidates who were hired were compared to candidates who were qualified on the screening and preemployment test battery but were not ultimately hired to determine what, if any differences existed along the prescreening protocol between these two candidate groups.

Results of the screening item mean comparisons between hired and not hired (but considered qualified based on performance in the prescreen and preemployment test battery) candidates are presented in Table 13. As shown in the table, there were no

significant differences along any of the prescreening items between these two candidate groups. Thus, Research Question 2 was not testable in the UW sample.

Customer Service Representative

Descriptives and Zero-order Relationships

Means, standard deviations, and intercorrelations of the CSR prescreening items are presented in Table 14. As was the case in the UW sample, a number of questions in the preliminary online screening lacked variance. Specifically, 3 items had no variance: age 18 or older, high school diploma/GED, and overtime and weekend work schedule requirements questions. Thus, these variables were not included in the correlation matrix. Each of the screening questions presented in Table 14 demonstrated a significant item-total correlation with total prescreening score (*r*s ranged from .17 to .61, *p*s < .01). The candidates' total prescreening scores failed to demonstrate significant relationships to any of the testing or hiring outcome variables, again, most likely due to severe restriction of range among the subsample of candidates who made it to these stages in the selection process. However, a number of the individual prescreening items did exhibit significant bivariate relationships to key outcomes. Years of customer service experience in a call center demonstrated significant inverse relationships to total preemployment test battery score ($r = -.10, p < .01$), test battery pass/fail ($r = -.07, p < .05$), and whether the candidate was terminated during the 6-month probation period ($r = .18, p < .01$; the correlation is positive based on the way it was coded in the dataset). The sum of computer software programs endorsed by candidates was positively related to total test score ($r = .06, p <$

.05) and test pass/fail ($r = .08, p < .01$). Likewise, typing words per minute exhibited similar, positive relationships (rs = .12 and .13 with total test score and test battery pass/fail, respectively, ps < .01). Finally, anticipated tenure in the call center was inversely related to the test outcome variables ($r = -.10, p < .01$ with total test score; $r = -.07, p < .05$ with test battery pass/fail).

SME Prescreening Item Taxonomy Ratings

Table 15 displays the mean prescreening item taxonomy ratings across each of the 13 dimensions investigated, their intercorrelations, and intraclass correlations (ICCs). Overall, the ICCs were high, with the exception of the equal access dimension. Given the small sample of expert raters, few of the prescreening dimension correlations were significant. Ratings of verifiability were positively related to ratings of objectivity ($r = .57, p < .05$), and job related items were positively related to items measuring minimum qualifications ($r = .58, p < .10$). External items were also more verifiable and objective (rs = .63 and .72, respectively, ps < .05). Finally, noninvasiveness was positively associated with item verifiability ($r = .91, p < .01$) and discreteness ($r = .75, p < .01$).

Hypothesis Tests

Hypothesis 1a predicted that prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings. Figure 3 shows the score distribution for the CSR screening scores. As can be seen, the distribution was highly skewed. This was likewise reflected in the Z-score distribution (scores ranged from −3.67 to 1.68). Thus the

approach taken to explore Hypothesis 1a reflected what was proposed in Research Question 1. That is, instead of looking at differences based on the original score distribution, the distribution was truncated and *Z*-scores recalculated so that Hypothesis 1a could be tested on candidates with the most complete data. Based on the score distribution displayed in Figure 3, the score distribution "normalizes" at around the screening score of 22. Thus, candidates scoring below this value were filtered out of the dataset and the *Z*-scores were recalculated on screening scores ranging from 22 to 37. This resulted in the removal of 341 candidates from the analysis. *Z*-scores based on this smaller distribution ranged from −1.94 to 2.53. Screening item scores of candidates with *Z*-scores greater than or equal to 1.00 (n = 398) were compared to the screening item scores of candidates with *Z*-scores less than or equal to −1.00 (n = 397) based on the smaller candidate sample on each of the following 8 questions: (1) years of customer service experience in a call center, (2) software experience (treated as 6 separate variables as well as a seventh, total software programs endorsed variable), (3) typing words per minute, (4) anticipated tenure in the position, (5) reasons the candidate stated he/she wanted to join the call center, (6) their description of what customer service meant to them, (7) shift availability, and (8) communication skills. Due to the multiple comparisons made in testing this hypothesis, a more stringent alpha of .01 was used. Results of the *t*-tests by prescreening item are displayed in Table 16. As shown in Table 16, each of the *t*-test comparisons was significant. Therefore, Hypothesis 1a was not testable in the CSR sample: each of the screening questions significantly differentiated between high- and

low-scoring candidates and therefore questions could not be compared for differences in terms of the prescreening item classification taxonomy.

Hypothesis 1b stated that prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications. Because an insufficient number of candidates had a *Z*-score greater than or equal to 2.00, percentiles based on the raw total prescreening scores were used. Candidates at the 75th percentile but lower than the 90th percentile (n = 353) were compared to candidates scoring at the 90th percentile and above (n = 239). In raw score terms, this equated to a comparison of candidates with a screening score of 31 or 32 against candidates scoring 33 or above. Mean item scores across the same 8 prescreening questions analyzed in Hypothesis 1a were compared across the two high-scoring groups using a series of *t*-tests. Results of the mean prescreening item tests appear in Table 17.

Based on the results of the *t*-tests displayed in Table 17, 2 questions failed to distinguish between the top-scoring groups: years of customer service experience with the general public, and shift availability. Therefore, to test Hypothesis 1b, the mean screening taxonomy ratings for these two questions were compared against the mean screening taxonomy ratings for the rest of the CSR screening protocol questions. Results of these comparisons appear in Table 18. At the more conservative alpha level (.01), one prescreening item taxonomy dimension demonstrated a significant difference among the high-scoring candidates. Specifically, the questions that differentiated the highest-scoring candidates from the high-scoring candidates were

significantly less objective, $t(10) = -3.52$, $p < .01$. The questions that differentiated the two high-scoring candidate groups did not differ in terms of the extent to which they assessed minimum qualifications, $t(10) = -1.25$, *ns*. Thus, results in the CSR sample did not support Hypothesis 1b. No relationship was found for the minimum qualifications dimension, and results for the objective dimension were in the direction opposite what was hypothesized.

Using a less stringent alpha level (.05, .10), there were some other trends in terms of prescreening item taxonomy differences between the two high-scoring groups in the CSR sample. First, the questions differentiating the highest scoring candidates may be less job-related, $t(10) = -1.94$, $p < .10$, less historical, $t(10) = -2.58$, $p < .05$, less external, $t(10) = -1.96$, $p < .10$, more noninvasive, $t(10) = 2.16$, $p < .10$, more highly susceptible to faking, $t(10) = 2.05$, $p < .10$, and more likely to elicit a socially desirable response, $t(10) = 1.82$, $p < .10$, which would coincide with the results indicating the questions were less objective. Overall, these results were in the expected direction, as it was expected that among the top-scoring candidates, the questions that would differentiate would be less likely to be job-relevant, as all top-scoring candidates would be expected to possess relevant job qualifications and experience. Moreover, it was expected that there would potentially be more faking or socially desirable responding at the top of the prescreening score distribution, which was in line with the trends for the social desirability and faking dimensions.

Hypothesis 2a explored whether applicant prescreening scores would be predictive of key new hire criteria. As mentioned in the discussion of the UW sample,

the criteria investigated were preemployment test score (the test battery for the CSR position used a compensatory scoring model so the sum of all test components was also investigated, unlike the UW and GSR positions where scores on individual components was investigated), preemployment test pass/fail, hiring decisions, and whether the candidates who were hired survived the 6-month probationary period. Based on the zero-order correlations between total prescreening score and the preemployment test scores ($r = -.03$, *ns*) and test battery pass/fail ($r = .02$, *ns*), Hypothesis 2a was not supported. Moreover, total prescreening score did not demonstrate a significant relationship to hiring decisions ($r = .06$, *ns*). Total screening score did correlate with survival beyond the first 6 months in the job at a less stringent alpha level ($r = .18$, $p < .10$). Overall, there was no support for the validity of applicant prescreening in relation to later stages in the selection process, hiring decisions, or post-hire outcomes. That is, total prescreening scores, as they were conceptualized in this study, failed to predict performance in later selection stages.

Hypothesis 2b predicted that prescreening items that were related to key new hire performance criteria would be high on objectivity and verifiability. Although the total prescreening score failed to demonstrate a significant relationship to the key outcome criteria, a number of the individual screening items demonstrated significant bivariate relationships to total preemployment test battery score and test battery pass/fail. Specifically, 4 of the CSR prescreening questions correlated with test score (see Table 14): years of customer service experience in a call center ($r = -.10$, $p < .01$ with total test score; $r = -.07$, $p < .05$ with test battery pass/fail), software experience

(operationalized as total number of software programs endorsed; $r = .06$, $p < .05$ with total preemployment test battery score; $r = .08$, $p < .01$ with preemployment test battery pass/fail), typing words per minute (rs = .12 and .13 with total test score and test pass/fail, respectively, ps < .01), and anticipated tenure as a CSR ($r = -.10$, $p < .01$ with total test score; $r = -.07$, $p < .05$ with test pass/fail). The means on these questions along the prescreening taxonomy dimensions were compared to the means of the questions which did not exhibit a significant correlation with preemployment test battery score. Results are shown in Table 19.

Results of the t-tests comparing prescreening items with a significant correlation to preemployment test score to those without a significant relationship along the taxonomy dimensions showed that the "valid" items were significantly more verifiable and objective (ts(10) = 4.09 and 6.85, respectively, ps < .01), supporting Hypothesis 2b. In addition, significant differences along 8 additional prescreening item taxonomy dimensions were also found. The 4 prescreening items with significant relationships to preemployment test score were significantly more historical ($t(10) = 6.64$, $p < .01$), external ($t(10) = 9.91$, $p < .01$), firsthand ($t(10) = 3.21$, $p < .01$), discrete ($t(10) = 4.51$, $p < .01$), and controllable ($t(10) = 3.28$, $p < .01$). At the less stringent alpha of .05, these items were also more noninvasive ($t(10) = 3.04$, $p < .05$) and less susceptible to socially desirable responding ($t(10) = -2.87$, $p < .05$).

Research Question 2 (Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?) was tested in the

same manner as it was tested in the UW sample. That is, mean scores on each of the CSR prescreening items between hired candidates and candidates who were considered qualified but who were not ultimately hired were compared. Results of these mean comparisons appear in Table 20.

Similarly to the results for Research Question 2 in the UW sample, there were no meaningful differences between hired and qualified but not hired candidates along the prescreening questions explored in the CSR sample. One comparison, namely, the percentage of candidates in each group who indicated they had PowerPoint experience, did differ at the .05 alpha level, $t(782) = -2.27$, $p < .05$. However, in terms of the total number of software programs with which candidates indicated they had experience, there was no difference, $t(782) = -1.20$, *ns*. Thus, given the fact that only one response option to this question demonstrated significant a significant mean difference between hired and qualified but not hired candidates, and no other questions in the prescreening protocol differentiated between these two groups, this one difference was considered to be spurious and thus follow-up analyses were not warranted.

Gas Service Representative

Descriptives and Zero-order Relationships

Means, standard deviations, and intercorrelations of the GSR prescreening items, preemployment test scores, and preemployment test outcomes (pass/fail) are presented in Table 21. Note that, due to the severe range restriction for many of the preliminary prescreening questions (endorsement rates exceeded 99%), they are not

included in the table. The questions removed from the table include all of the online screening/minimum qualification questions (i.e., at least 18, high school diploma/GED, valid [State] driver's license, and ability to arrive at work site within 30 minutes during an emergency), as well as one of the telephone screening questions (ability to work a flexible work schedule). Reasons provided by the candidate regarding why he/she would like to join the organization also suffered from considerable range restriction, but it was not so great that it was removed entirely from the correlation matrix.

Each of the three prescreening questions that did not suffer from extreme range restriction correlated significantly with total prescreening score. In addition, two of the prescreening questions demonstrated modest relationships ($p < .10$) with elements of test performance. The candidates' description of what good customer service means to them correlated with preemployment test outcomes ($r = .10$). This question demonstrated a somewhat stronger relationship to the spatial ability test score ($r = .12$, $p < .05$). The screeners' ratings of candidate communication skills was significantly related to scores on the mathematical ability test component ($r = .09$, $p < .10$). As was the case in the UW and CSR samples, total screening score failed to correlate with testing outcomes ($r = -.05$, *ns*).

SME Prescreening Item Taxonomy Ratings

Means, standard deviations, and intercorrelations among the 13 prescreening taxonomy dimensions are presented in Table 22. Overall, the reliabilities of SME ratings were acceptable, with the exception of the ICC for the noninvasive dimension.

Ratings of objectivity correlated with both verifiability ($r = .78, p < .01$) and job relatedness ($r = .55, p < .10$). Discrete items were also higher on verifiability in the GSR sample ($r = .64, p < .05$). Prescreening items which were considered to be equally accessible to all candidates were also more historical ($r = .69, p < .05$) and external ($r = .87, p < .01$). Finally, there were significant relationships between prescreening item social desirability and minimum qualifications ($r = .75, p < .01$) and the controllable dimension ($r = .61, p < .05$).

Hypothesis Tests

Hypothesis 1a predicted that prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings. Candidate prescreening scores were first converted to *Z*-scores. As was the case in the previous *Z*-score transformations, the range of the scores did not allow for a test of this hypothesis (*Z*-scores in the sample ranged from −3.52 to 0.55). Thus, the methodology suggested for the exploration of Research Question 1 was once again employed. Based on an examination of the prescreening score distribution (see Figure 4), Hypothesis 1a was tested in the following manner. Given the somewhat bimodal nature of the distribution, candidates scoring 14 and above were compared to candidates scoring 11 and below. Candidates in each score group were compared along the following 3 questions: reasons to join the company, what good customer service means, and communication skills. Results are displayed in Table 23. Based on the results of the *t*-tests, each of the questions significantly differentiated high- and low-scoring candidates (the communication skills question

could not be tested because the mean in the low-scoring group was 0). Thus, Hypothesis 1a could not be tested in the GSR sample.

Hypothesis 1b predicted that prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications. As was the case with Hypothesis 1a, a distributional analysis approach was taken in testing Hypothesis 1b. That is, candidates who obtained the highest score in the distribution (17) were compared to candidates receiving the next-highest score (16) because these two scoring groups had relatively large numbers of candidates. Results of the prescreening item comparisons are displayed in Table 24.

Based on the results of the *t*-tests, 2 of the 3 questions significantly differentiated the highest-scoring candidates: their description of good customer service and their communication skills. In each case, the highest-scoring candidates outperformed the high-scoring candidates. Although a *t*-test could not be computed on the reasons candidates wanted to join the company question, given that the means were identical, this item was treated as not differentiating candidates. Thus, the means of the two differentiating questions were compared to the means of the non-differentiating item across the prescreening item taxonomy questions. Results of these comparisons appear in Table 25.

Based on the results presented in Table 25, Hypothesis 1b did not receive support in the GSR sample. Although the trend was toward greater objectivity in the differentiating screening items, the mean comparison test was not statistically

significant ($t(10) = 1.66$, *ns*). The hypothesis was not supported in relation to the minimum qualifications dimension. In fact, the trend was in the direction opposite that which was hypothesized. That is, prescreening items differentiating among the two high-scoring candidate groups were rated as being more related to minimum qualifications compared to the question which did not differentiate the candidate groups ($t(10) = 1.88, p < .10$). Overall, then, Hypothesis 1b was not supported in the GSR candidate sample.

Although the prescreening taxonomy dimensions hypothesized to show significant mean differences failed to do so, there were a number of other dimensions which did demonstrate significant mean differences. Most notably, prescreening items that differentiated the highest-scoring candidate groups were significantly less susceptible to faking ($t(10) = -3.75, p < .01$) and significantly more verifiable ($t(10) = 3.71, p < .01$), although the susceptible to faking means were both well above the midpoint of the scale and both verifiability means were well below the midpoint of the scale, suggesting on the whole, the three prescreening items compared in these analyses were highly susceptible to faking and difficult to verify. Similarly, at a less stringent alpha level, the items that differentiated the highest-scoring groups were significantly less likely to elicit a socially desirable response ($t(10) = -2.71, p < .05$). Again, both sets of items were extremely high on the social desirability dimension. Finally, the prescreening item that did not differentiate the two high-scoring candidate groups was significantly more firsthand in nature and was rated higher in terms of equal access ($ts(10) = -2.32$ and -2.28, respectively, $ps < .05$).

Hypothesis 2a posited that prescreening items will be predictive of key new hire criteria. Given that there were insufficient numbers of hires on which to conduct such an analysis, the relationship of prescreening score and preemployment test battery outcome (pass/fail) was examined instead. Based on the zero-order correlation between these two variables ($r = -.05$, *ns*; see Table 21), Hypothesis 2a was not supported.

Hypothesis 2b stated that prescreening items predictive of key new hire performance criteria will be high on objectivity and verifiability. One of the three prescreening items examined, candidates' description of what good customer service means to them, demonstrated a significant relationship with preemployment test battery pass/fail, albeit at more relaxed alpha ($r = .10$, $p < .10$). However, this same item also demonstrated a significant zero-order correlation with scores on the spatial ability test component ($r = .12$, $p < .05$). Thus, the prescreening dimensional means of this item were compared to the prescreening dimensional means of the other two prescreening items. Results are displayed in Table 26.

Based on the results of the prescreening item taxonomy comparisons between the prescreening item with a significant item-criterion relationship and the items that did not demonstrate a significant item-criterion relationship, Hypothesis 2b was not supported. The valid prescreening item was neither more verifiable ($t(10) = -1.17$, *ns*) nor objective ($t(10) = 1.17$, *ns*). Nevertheless, three prescreening item classification dimensions did demonstrate significant differences, albeit at the less stringent alpha level. Specifically, the valid prescreening item was rated significantly higher on the

firsthand dimension ($t(10) = 3.07, p < .05$), was considered less discrete ($t(10) = -2.89, p < .05$) and was higher on social desirability ($t(10) = 2.71, p < .05$) compared to the prescreening items that did not exhibit significant relationships to preemployment test battery pass/fail.

CHAPTER VII: STUDY 1 DISCUSSION

In Study 1, I sought to understand potential best practices in the area of applicant prescreening protocols. The specific prescreening system examined in the present study was, on the whole, typical of what is generally understood in the personnel selection field as prescreening (cf. Hunt et al., 2004). That is, as part of the initial application process for the three positions examined in Study 1, candidates were required to answer a short series of questions which pertained to legal requirements, minimum qualifications, and a cursory measurement of candidates' relevant previous work experience, skills, and attitudes about the position to which they were applying. However, the modalities under which applicant prescreening is administered may vary across organizations from what was examined here.

In the present study, the screening protocol was administered using two separate media: online and via telephone with a live screener. Although this is a representative administration process, some screens may be completed entirely via the internet, with a live screener, or may be automated and administered via telephone (i.e., IVR technology; e.g., Bauer et al., 2004). In addition, prescreening administration systems may also vary in terms of the data collected on candidates. In other words, although the administration of the prescreening in the organization from which data were collected may have ended once a candidate chose a "knockout" response, other prescreening systems may collect a complete set of data on all applicants across the entire screening protocol. Thus, the results of Study 1 must be interpreted with these caveats in mind.

Employing a taxonomic framework originally presented by Mael (1991), the present study examined whether such a classification scheme could be applied to prescreening content, given its potential overlap with the types of questions asked in a biodata inventory. Moreover, this framework was used to examine whether there were systematic relationships between prescreening item classifications and the items' ability to predict outcomes in later stages of the selection process. If such relationships were obtained, it would logically follow that such questions would be of greater use in the development of prescreening scripts, as such protocols would provide organizations the ability to not only screen out or deselect candidates who clearly do not possess the minimum qualifications required for the position, but also to make meaningful distinctions among the remaining viable candidates. Such an outcome would be particularly useful for positions which have extremely large applicant volumes, and would also allow recruiters to more effectively target candidates for résumé screens or telephone interviews, two of the most common next stages in a multiple hurdle selection process.

To examine these potential relationships, prescreening data from actual job applicants were collected across three high-volume positions in a major utility company, as well as preemployment test data, and hiring and termination data, where available. To link relationships among prescreening items and key organizational criteria to the biodata item taxonomy framework, a sample of expert graduate student raters evaluated the prescreening questions along Mael's (1991) ten biodata

dimensions as well as three additional dimensions added for the present study: minimum qualifications, susceptibility to faking, and social desirability.

A summary of the results from Study 1 appears in Table 27. In addition to the specific taxonomy dimensions where relationships were hypothesized in Study 1, all 13 item characteristic dimensions measured in the study were examined and results of those exploratory tests are also discussed in this chapter. As shown in Table 27, results across the hypotheses and the jobs investigated were mixed. Specifically, in the UW sample, results comparing the high-scoring candidates to low-scoring candidates were inconclusive. Although the differentiating questions were evaluated by the SME sample as being significantly higher on the minimum qualifications dimension as hypothesized, they were not found to be higher on job-relatedness. The differentiating items were also rated as being more discrete. In other words, these questions were more likely to assess a single, unique behavior or simple count of unique instances of events compared to the prescreening items that did not differentiate candidates at opposite ends of the prescreening score spectrum. Results in the CSR and GSR samples were inconclusive because all questions in the screening protocol differentiated between high- and low-scoring candidates. Overall, the results of the tests of Hypothesis 1a were inconclusive; there was little in the way of insight into what distinguished between the high- and low-scoring candidates across the three positions investigated.

There are three potential explanations for the lack of consistent findings for Hypothesis 1a in the UW sample. First, the simplest explanation is that the job-

relatedness ratings provided by the graduate student sample suffered from low interrater reliability, as evidenced by the .23 intraclass correlation obtained on this dimension, which would be reflected in larger standard deviations for the mean ratings, and therefore greater difficulty in detecting a statistically significant difference, given the small sample of raters. This lack of reliability also suggests that it may have been unclear to the graduate students what constituted a job-related question across the CSR prescreening protocol. Second, the high mean ratings on the job-relatedness dimension ($M = 4.40$ or above in the test of Hypothesis 1a) suggests a potential ceiling effect in the item ratings, which would also greatly reduce the ability to detect a significant difference between the differentiating and non-differentiating prescreening item clusters. Finally, it should be noted that the removal of the online portion of the prescreening protocol, which assessed the most basic of qualifications, was tied to this ceiling effect. That is, the mean job-relatedness rating across the entire screening protocol was 3.60. The low intraclass correlation may have been an artifact of the disagreement among the graduate students regarding whether minimum qualifications, such as being 18 years of age or older, should also be considered job-related.

Stronger results were obtained from comparisons of the high-scoring candidates across the three positions. In the UW sample, prescreening questions that differentiated the highest-scoring candidates from high-scoring candidates differed significantly from the prescreening questions which did not differentiate the two high-scoring candidate groups along five dimensions. The hypothesized relationships in the

UW sample were partially supported. Although the prescreening questions differentiating the two top-scoring candidate groups were indeed found to be more objective than those prescreening questions which failed to differentiate, they were also found to be rated higher on the extent to which they assessed minimum qualifications. The questions distinguishing the top candidates were also rated higher on verifiability and discreteness and lower on susceptibility to faking, although the latter two relationships were significant at a less stringent alpha level. This hypothesis was also partially supported in the CSR sample: the subset of prescreening questions that differentiated among the highest-scoring candidates did not significantly differ in terms of the extent to which they assessed minimum qualifications. However, the differentiating questions were also significantly *less* objective than the prescreening questions which did not differentiate the two high-scoring candidate groups. Somewhat conflicting results were likewise found in the GSR sample. Although the trend was such that prescreening questions that differentiated the top-scoring candidates were more objective, the difference was not statistically significant. Moreover, the items that differentiated candidates were rated significantly higher in terms of minimum qualifications. Potential avenues for future research designed to clarify the equivocal results are discussed below.

In terms of the validity of prescreening items as a whole, no support was found for any of the three jobs investigated. Total prescreening scores did not correlate with test outcomes (pass/fail) or test scores, hiring decisions, and tenure (defined as survival of the 6-month probationary period). Furthermore, for the latter criterion,

sample size was insufficient to explore the relationship in any meaningful way. In contrast, some of the individual prescreening items across the three jobs did demonstrate significant zero-order correlations with some of these criteria. Test scores, either at the test component level (UW and GSR positions) or in terms of total test score (CSR position) were not predicted by applicant prescreening scores. Nevertheless, in the UW sample, two of the individual prescreening items were correlated with preemployment test battery outcome. Compared to the remaining UW prescreening questions which did not correlate with testing outcomes, these questions were significantly higher in terms of verifiability and objectivity. At the .05 alpha level, these questions were also rated higher by the SME sample on minimum qualifications, and they were considered more discrete. Significant relationships between prescreening items and the prescreening item taxonomy were also found in the CSR sample across the four prescreening items which correlated significantly with the preemployment test for clerical jobs. These questions had significantly higher ratings on verifiability and objectivity, as well as several additional dimensions, lending support to the hypothesis. In terms of the additional dimensions, these items were also significantly more historical, external, firsthand, discrete, and controllable. And, to a lesser extent (alpha = .05), they were also more noninvasive and less susceptible to socially desirable responding.

Finally, in the GSR sample, results were more tenuous, as the one valid prescreening item was significantly correlated to testing outcomes at the .10 alpha level. However, these relationships were still tested, and the results failed to support

the hypothesized relationships; this item was neither more verifiable nor more objective than the prescreening items that did not correlate with testing outcomes. It should be noted that the standard deviations associated with the prescreening item taxonomy dimensions for these prescreening items were considerable, in some cases exceeding 1.00, which may provide an explanation for the lack of support for this hypothesis in the GSR sample. This also points to a potential lack of agreement among the graduate student raters on these dimensions among the three items examined to test this hypothesis. Because each of the three prescreening items involved a screener's rating of the candidate's responses and communication skills, some graduate students may not have felt that these items were objective. Also, candidates' descriptions of their attitudes (i.e., meaning of good customer service) may have been interpreted by the graduate students as not being highly verifiable. The mean verifiability and objectivity ratings were 2.00 or lower on a 5-point scale across differentiating and non-differentiating items, which supports this interpretation. Further issues surrounding the SME ratings are discussed below. Although the hypothesized relationships were not demonstrated in the GSR sample, at the .05 alpha level, the valid prescreening item was rated significantly higher on the firsthand dimension, was significantly less discrete, and was significantly higher in terms of social desirability. Given that the valid prescreening item involved candidates' description of what good customer service means to them, this result makes sense.

There are a couple of potential explanations for this for the lack of total prescreening score relationships with any of the key criteria investigated. First, it

could be the case that the restricted range of scores limited the ability of the total prescreening score to correlate with any of the criteria. Secondly, it this could have occurred because the screening responses contained a high amount of error variance which would have masked any potential relationships to the criteria investigated. And finally, prescreening and preemployment tests may be completely distinct in terms of what they measure (assuming each predict a unique component of job performance), and as a consequence, would not correlate. Regardless, prescreening protocols should be developed using a content validity approach in order to ensure that they can be legally defended as a selection tool, especially if the scores obtained do not demonstrate any criterion relationships.

Implications for Research

Tying the results from the present study to prior research, the present study has begun to extend the biodata item characteristics framework to the prescreening arena. Because there is some potential overlap between the content of prescreening and biodata, both in terms of question content and response formats, it seems logical to extend this framework in order to develop a greater understanding of best practices in the realm of applicant prescreening. On the whole, to the extent that items are more verifiable, objective, external, and job-related, the quality of the items is better. Specifically, Becker and Colquitt (1992) examined the biodata classification model in relation to faking behavior and found that items which were faked in their study lacked these critical attributes. The present study indirectly corroborates this finding, in that the items that were found to be predictive of later stages in the selection process

(i.e., testing outcomes) were rated significantly higher in terms of the items' verifiability and objectivity (in the UW and CSR samples). If such items had a high incidence of or susceptibility to faking behavior, this would introduce substantial amounts of error variance and thus these items would be far less likely to correlate with other constructs. Indeed, in one such study that examined the effects of requiring elaboration to biodata item responses as a check on faking behavior (Schmitt et al., 2003), it was found that biodata item objectivity and verifiability was inversely related to the participants' scores on both a social desirability and impression management ("faking") measure. Moreover, the methodology employed by Becker and Colquitt to assess the characteristics of their biodata items was also used in the present study.

A more direct parallel in terms of the results from the present study to those found in past research comes from a study conducted by McManus and Masztal (1999). The authors specifically examined the validity of biodata items for an entry-level sales job, linking these items to Mael's (1991) dimensions (assessed in a manner similar to that of Becker and Colquitt, 1992, and the present study) and found that the valid items were indeed those which were rated higher on the objective and verifiable dimensions, as well as the historical, external, and discrete dimensions. Thus, on the whole, similar results were obtained in the arena of applicant prescreening. However, further research is still needed, as other studies using differing methods of classifying biodata items have found different results. In particular, Lefkowitz et al. (1999) used a dichotomization (and in some cases, trichotomization) technique to classify biodata items in their study, with the intention of linking items that predicted supervisor

ratings to the biodata taxonomic framework. In their study, the valid items were not more verifiable and objective, but instead were more indirect (i.e., secondhand as opposed to firsthand) and non-controllable. Finally, it should be noted that in each of these previous studies, different jobs were investigated, and by extension, the biodata inventories investigated differed in terms of the specific item content. This was also the case in the present study. Nevertheless, it would stand to reason that, regardless of the job investigated, the item characteristics which roll up into Mael's meta-category of verifiability (i.e., external, objective, firsthand, discrete, and verifiable) should be higher-quality items, compared to those which are low on this higher-order factor. Further research is needed, both in terms of refining the measurement of Mael's (1991) item characteristics as well as in determining whether job type moderates the relationship of item characteristics and validity.

Recommendations for Practice

Based on the results from the present study, online prescreening is most likely better-suited to deselecting rather than selecting or otherwise differentiating candidates (cf. Hunt et al., 2004). However, steps can be taken to collect additional information on candidates that may not only allow recruiters to sort candidates in a meaningful manner in order to determine which candidates to pursue further, but to also provide them with a means by which to evaluate their recruiting efforts. There have been calls in the literature for such evaluation efforts (e.g., Carlson et al., 2002), and to the extent that applicant prescreening systems can be refined so that they employ questions with the greatest chance of being predictive of later selection stages (i.e., items that are

more verifiable and job related but less susceptible to conscious applicant dissimulation or less likely to elicit socially desirable responses), these outcomes can begin to be realized.

Based on the results of the present study, one key recommendation for practice is to develop prescreening questions using a content validity approach. A thorough job analysis and identification of which KSAs are absolutely critical for employees to have at the time of hire would aid in identifying the minimum qualifications for the position and would justify the designation of certain response options as knockouts in terms of the scoring of these questions. In addition, based on the results from the present study, efforts should be made to ensure that the prescreening questions developed for a given job are not only clearly job-related, but that they are also highly objective in nature and the responses provided by applicants are verifiable. This approach to item development should not only increase legal defensibility of prescreening, it should also minimize the potential for applicants to engage in successfully faking their responses.

Moreover, unlike the present study which scored many of the prescreening item response options along a continuum, scoring of prescreens could be informed by SMEs in the form of a criticality survey or an Angoff-like approach, whereby responses expected of a minimally qualified applicant could be identified. Scoring could also be developed and/or refined through use of an empirical keying approach, similar to what has typically been done in the biodata arena. Finally, based on the lack of strong criterion relationships obtained in the present study, as well as the potential

that applicant faking may inflate candidate scores (assuming that, in the development of the prescreening protocol, questions with great susceptibility to applicant dissimulation were not removed) it may be advisable to interpret the scores from a prescreen along broad categories as opposed to taking a strict top-down approach. This could be accomplished either through a banding technique or through simple score categories, similar to what is sometimes done in T&E evaluations.

Ultimately, the results of this study contribute to the body of literature on so-called alternative selection procedures by underscoring the importance of developing content based on job requirements, and developing scoring systems based on input from job incumbents and others familiar with the position. Prescreening scoring systems should not be recruiter-driven, as is marketed by several prescreening vendors. Instead, other well-established methods such as content validity-oriented and empirical keying approaches (as are often used in the development of biodata inventories) may result in better-quality prescreening protocols. It remains to be seen to what level of detail these brief questionnaires can differentiate candidates, and, based on the present study, whether prescreening can reliably predict candidate success, either in later selection stages or in terms of job performance.

Potential Limitations

The present study suffers from several potential limitations. First, the applicant prescreening scores suffered both in terms of lack of variance in the question responses themselves (a reflection of the fact that candidates were not administered the entire screening protocol if they selected a knockout response) and in terms of the

breadth of questions. That is, due to the fact that most if not all the online screening questions contained no variance, the entire script of questions investigated in the present study was halved. Having the full range of applicant scores across the entire prescreening protocol and larger question sets would have allowed for a more thorough investigation of the relationships explored in the present study, as well as avoiding the removal of large numbers of candidates. However, steps were taken to ensure that the candidate data used in the study were as complete as possible, and, in terms of the number of questions, the CSR sample was not as affected by the online prescreen truncation, compared to the other two jobs investigated. More complete data would have allowed for better tests of the distributional analyses, particularly those involving high- and low-scoring candidates, as these could not be tested in the manner originally proposed, and, in the CSR and GSR samples, these tests could not be conducted at all because all of the questions differentiated the two applicant groups. However, it is likely that this approach is not uncommon to screening systems. That is, it is fair to assume that most organizations would not want to spend additional resources on screening candidates who they have already determined have not met the minimum requirements outlined for the position.

A second limitation of the present study concerns the expert sample rating the prescreening items. Specifically, given the small sample size, tests of the hypotheses were limited to a series of *t*-tests. If the sample had been larger, this would have allowed for a more robust test of prescreening item differences across the 13 dimensions investigated (e.g., via MANOVA). This limitation was addressed by

employing a more stringent test of significant mean differences. In spite of this limitation, there were some patterns found in the results, particularly across the UW and CSR samples in terms of prescreening item validity and the characteristics of these items, which helps to temper this concern. Moreover, past studies have also employed rather small rater samples, and in one case (Becker & Colquitt, 1992), a relatively small sample of naïve raters was used. Nevertheless, the relationships found in the present study should be tested again in future studies, using a larger SME pool, if possible, to rate the item dimensions.

Finally, another concern and potential limitation relative to the graduate student SME sample involved the poor reliabilities obtained for some of the ratings of prescreening item dimensions, although this occurred rarely. This further suggests that refinements need to be made to the way in which the dimensions are assessed and/or defined, and larger samples of raters could likewise alleviate this problem. Even in light of these limitations, however, patterns of significant results were obtained, and the use of more reliable ratings in a future study should only strengthen and potentially broaden the nature of the relationships obtained in the present study. In addition, the use of graduate students with an I/O background may not have been as significant a limitation as it may initially appear. Previous research in the area of assessment centers has found that psychologists are better raters than those intimately familiar with the focal position (i.e., job incumbents and managers; Woehr & Arthur, 2003).

Future Research

Some potential avenues for future research have been alluded to above. First, future research should continue to refine the measurement of Mael's (1991) biodata dimensions. In particular, it is important to determine whether measurement via Likert-type scales or binary classifications produces more consistent and meaningful results. Future research could either explore items at the factor level (the three higher-order constructs identified by Mael, namely, verifiability, historical, and ethical/legal), or, if the binary classification technique is used, items should be "clustered" along multiple dimensions. This may allow a clearer picture to emerge in terms of what the best item "profiles" are. One of the drawbacks of the present study is that if valid items differ from non-valid items along a particular dimension, for example, verifiability, and if both items are rated below the midpoint of a scale, one cannot conclude definitively that items that are higher on verifiability are more valid.

Another potential avenue for future research would be to take the examination of prescreening into the lab setting, allowing for more control over the way in which prescreening data are collected, as well as allowing for the manipulation of the way in which items are presented to participants (e.g., two items could be constructed for the same work-related construct, with one being high on the desirable characteristics, and the other being low on those characteristics). However, the drawback of such an approach is the potential lack of generalizability of the results.

Future research should also investigate whether ratings of Mael's dimensions vary based on the level of expertise of those providing the ratings. Some past studies

have used modest samples of naïve undergraduate students, some have used extremely small samples of experts with PhDs in Industrial/Organizational psychology, and, as in the present study, a slightly larger sample of graduate students with coursework and research experience in personnel psychology was used. If the definitions of the taxonomy dimensions can be standardized and refined, this may allow for non-experts to provide ratings, which would allow for a more robust dataset. Ultimately, the tradeoffs in this are of research may be between complete agreement in a small set of raters (e.g., Lefkowitz et al., 1999) and stability of mean rating estimates for use in statistical tests (using a larger number of less-knowledgeable raters; e.g., Becker & Colquitt, 1992).

In a related vein to the previous recommendation, the examination of prescreening should be taken to the lab setting, using an undergraduate student sample, which would allow for greater control as well as the ability to collect complete screening data as well as later "selection" stage data such as employment test scores and/or interview data on all participants. One study that examined IVR as a screening tool in relation to applicant reactions employed such a study design (i.e., Bauer et al., 2004). Although the benefits of added control and the elimination of severe range restriction as experienced in the present study would be removed in the lab, the greatest challenge would be to create sufficient motivation in a lab study with students who have little to nothing at stake. Consequently, the generalizability of the results from such a study to a real-world selection context may be dubious at best, depending upon the level of motivation investigators are able to generate in the student sample.

Finally, assuming the measurement of the item taxonomy dimensions can be perfected, and more complete prescreening data on actual applicants or study participants can be collected, future research should also attempt to link item characteristics to applicant reactions. Although it would seem logical that items high on the taxonomy dimensions of job-relatedness and noninvasiveness should be received more favorably by applicants, demonstrating these links, as well as identifying which dimensions are related to negative applicant reactions, could potentially allow for a reduction and consolidation of Mael's original model. Moreover, given that prescreening is one of the first, if not the first interaction an applicant has with a hiring organization, understanding and maximizing positive applicant reactions is crucial. Applicant reactions to prescreening assessments has not received extensive scrutiny in the literature (cf. Hausknecht, Day, & Thomas, 2004).

CHAPTER VIII: STUDY 2 METHOD

Study 2: Résumé Lab Study

Participants

Two recruiters from the same public utility company from which data were acquired for the physical and customer service positions examined in Study 1 served as participants in identifying critical résumé elements in Study 2. A matched sample of graduate and advanced undergraduate students (repeated measures design) in industrial/organizational psychology ($N = 12$) served as participants in the portion of the study exploring the role of résumé evaluation in the overall screening process through evaluating actual applicant résumés under two separate rating conditions: structured and unstructured evaluation. The details of each rating condition will be outlined in greater detail below. The student SMEs were evenly split in terms of gender, with a mean age of 26.83 ($SD = 7.80$). The majority of the students (75.0%) were Caucasian, with the remainder identifying themselves as Asian/Pacific Islander (16.7%) or "Other" (8.3%). Three-fourths of the participants had at least one year of graduate school experience, and 25% indicated they had previous work-related résumé evaluation experience.

Materials

Job description. A detailed job description as it would appear to applicants on the company website was provided to participants in the résumé review portion of the study. The research questions in Study 2 were tested on a professional job: Business Account Representative. Broadly speaking, this position involves effective use of

consultative sales skills and interpersonal skills, and the ability to interact with small- and mid-size business customers to sell them energy-related products and services. Business Account Representatives are also expected to provide customers with effective energy solutions and to advocate process improvements that better meet customer needs. This job description was provided to participants across both evaluation conditions. (See Appendix B.) They could refer to the job description at any time during the résumé evaluation process.

Critical résumé elements summary. The student SMEs were provided with a summary document outlining critical résumé elements as identified by the two company recruiters familiar with the Business Account Representative job. A list of résumé evaluation "best practices" in the organization in general was likewise provided to student SMEs. In total, nine critical résumé elements were delineated, along with descriptions of what to look for in candidates' résumés for each of the dimensions as described by the recruiters. This document was provided to student SMEs across both résumé evaluation conditions. (See Appendix C.)

Résumé evaluation surveys. Two different versions of the résumé evaluation survey were created for use in the study, which varied in the level of structure associated with the rating task (rating detail that was asked of participants). In the unstructured condition, a single, "global" rating scale was presented at the top of each candidate résumé, followed by the résumé content. In the structured condition, a candidate résumé was presented, followed by a detailed résumé rating sheet containing the nine critical résumé dimensions as outlined in the critical résumé elements

summary sheet, as well as an overall rating, similar to the "global" rating in the unstructured condition. (See Appendix D for initial invitation to graduate students to participate in the study, and Appendices E and F for a sample cover letter and unstructured and structured résumé evaluation surveys, respectively.)

Procedure

Applicant résumés from actual job applicants submitted for the Business Account Representative position were utilized in Study 2. The application process for this position differed from the physical and customer service positions described in Study 1 (See Figure 1). Similar to the application processes in Study 1, professional job opportunities are posted on the organization's website and interested candidates apply for these positions by providing some basic information about themselves (e.g., name, contact information) and uploading their résumés. Unlike the selection process described in Study 1, candidates for professional positions in the organization do not undergo online prescreening (although candidates with competitive résumés are phone screened by recruiters). Thus, the preliminary applicant screen for professional jobs in the organization is a review of their résumés by a recruiter.

Identification of critical résumé elements. An interview was conducted with a recruiter from the organization who had experience reviewing résumés for the focal position and the manager of recruitment for professional positions within the organization. Given the small recruiter staff and lack of resources (i.e., recruiter time), a detailed focus group interview was not feasible. Instead, these two SMEs were asked to review a relatively small number of résumés of varying quality which were

identified and selected by the recruiter through an examination of archival résumé data in the staffing department's résumé database. The SMEs were asked to discuss how they reviewed résumés and to indicate which elements of each résumé contributed to their designation of them as high- or low quality. A brief interview protocol was developed and provided ahead of time to guide the discussion (See Appendix G). This information was recorded as an initial insight into recruiters' decision-making processes. The outcome of the interviews was a detailed list of critical résumé elements, including broad descriptions of any relevant education, experience, skills, and other applicant characteristics of interest, as well as other key substantive or stylistic considerations for use in the résumé evaluation study. In addition, these interviews provided a general understanding of how recruiters at this organization determined the quality of applicant résumés. This process aided in identifying the number and types of résumé information that were viewed as most critical to the résumé evaluation process (e.g., minimum requirements, candidate desirables).

Based on the results from SME interviews, a list of applicant desirables (in terms of applicant experience and qualifications) was compiled and the list was presented to the recruiter to determine the importance of each of the résumé elements identified by the initial recruiter and recruiting supervisor interviews and résumé review. The recruiter was asked to review the list to verify that each of the elements identified in the initial interview had been accurately captured, making any necessary edits, as well as providing short descriptions of each dimension, including general guidelines regarding what to look for to determine whether a candidate has met the

acceptability threshold for that dimension. The outcome of this process was a final résumé review guide used in the study. The 9 résumé dimensions identified by the staffing SMEs were (1) statement of objective; (2) summary of relevant skills; (3) current employment status; (4) work history; (5) advancement; (6) education, experience, or compensatory training; (7) continuing education; (8) volunteerism; and (9) résumé format. The detailed descriptions provided by the SMEs for each of the 9 résumé evaluation dimensions appear in Appendix C.

Selection of focal job and résumés. One requisition for the high-volume Business Account Representative professional job was selected by the professional recruitment SMEs to test the effectiveness of the two résumé screening methods investigated in Study 2. Résumés posted for the focal job were culled from the organization's résumé database by the recruiting department SMEs. A total of 50 résumés was identified, which reflected a recent population of applicants for the Business Account Representative position. This was considered a typical number of applicant résumés that recruiters would be expected to examine in order to determine which candidates they would pursue further. Of these, a sub-sample was selected for the purpose of data collection for the present study, given that the amount of time that would be required for the graduate student participants to evaluate 50 résumés was considered too time-intensive, especially in light of the repeated measures design. The 20 résumés chosen for the study were selected based upon a review of all the résumés by the investigator, who made attempts to select résumés that varied sufficiently in quality and were proportionate to the total sample of 50 résumés in terms of whether

the candidates—in the actual selection process—were considered to have "passed" or "failed" the résumé screen. In the actual selection context, 18 applicants (36%) "passed" and the remaining 32 applicants (64%) "failed" the initial screen. In an effort to reflect, proportionately, the outcomes from the actual selection context, 7 of the "pass" résumés (35%) and 13 of the "fail" résumés (65%) were selected for inclusion in this study. Given that résumé quality might be associated with résumé length, the average word counts of the "pass" and "fail" résumés selected for the study were compared using a *t*-test. Results of the *t*-test failed to show a statistically significant difference in the average length of "pass" and "fail" résumés, $t(18) = 1.37$, *ns* (M_{pass} = 301.71; M_{fail} = 255.62).

Study design. As previously noted, this study employed a repeated measures design, with participants completing the résumé evaluation exercise under two different conditions. In terms of the presentation of résumés and ordering of conditions, each was randomized in the following ways. First, evaluation condition was counterbalanced such that approximately half the participants (41.7%) completed the structured résumé evaluation first and the global evaluation second, while the other half (58.3%) of the participants completed both evaluation conditions in the reverse order. The order in which participants completed each of the two conditions was completely random. Second, a random number generator was used to randomly order the presentation of the résumés within each résumé packet, such that no participant evaluated the résumés in exactly the same order, either within or across conditions. Moreover, to further counteract any potential order or memory effects, conditions

were spaced approximately 4 to 6 weeks apart. The first administration was conducted in person, allowing participants to ask the investigator any questions they may have had about the process or what they were being asked to do. The second administration was administered via US mail, with participants completing the second condition independently and sending their completed résumé evaluation packets back to the investigator using a pre-addressed, postage-paid envelope. In the event that participants had questions as they completed the second portion of the study, they were provided with the researcher's contact information. Each of the résumé evaluation conditions is outlined in detail, below.

Unstructured résumé evaluation. In this condition, considered to be the control condition because in many respects it mirrored standard résumé evaluation practices, graduate students in industrial/organizational psychology and advanced or recently-graduated undergraduates with relevant research and/or coursework experience were presented with the job description for the Business Account Representative position as it appears on the organization's website in order to familiarize them with the nature of the job for which they would be evaluating the résumés. They were also provided with descriptions of critical résumé elements as identified by the recruiter SMEs as a separate handout in evaluating each résumé (the 9 dimensions discussed earlier), to which they could refer at any time during the evaluation exercise. Student SMEs reviewed each of the résumés and provided one overall rating for each applicant's résumé using a 3-point scale (1 = Does Not Meet Requirements; 2 = Meets Requirements; 3 = Exceeds Requirements; see Appendix E). This "tiered" approach

was commonly used in the organization from which the résumés in the present study originated, based on the initial interviews with recruiter SMEs. Given potential issues with individuals outside of the organization viewing sensitive applicant information, potential candidate identifiers were removed from the résumés. This condition was considered to be unstructured because, although the participants received descriptions of each of the 9 résumé evaluation dimensions described previously, they were not explicitly instructed to consider these dimensions in their rating process, nor did they make ratings on each of the dimensions separately. In other words, the participants were asked to consider each résumé in its entirety and make a single, global rating of the candidate's résumé quality in relation to the Business Account Representative position.

Structured résumé evaluation. In similar fashion to the previous condition, the same student SMEs were presented with the same job description, 20 candidate résumés, and descriptions of the 9 recruiter-identified relevant aspects of the résumé to evaluate. As in the previous condition, participants could refer to any of these materials at any time during the evaluation process. The participants rated all résumés on the relevant attributes as they appeared in the rating form for each résumé, using the same 3-point rating scale as in the unstructured evaluation condition. (The wording of each anchor on the rating scale differed slightly, given that the participants were considering résumé dimensions as opposed to the entire résumé; see Appendix F.) In addition to rating the résumés on the 9 critical dimensions, graduate student SMEs were also asked to provide a tenth, overall rating of each résumé, which appeared at

the end of each résumé evaluation sheet. (The evaluation sheets appeared on the page immediately following each candidate résumé; see Appendix F for an example.) In this condition, the ratings corresponded to the recruiter-identified résumé dimensions, except for the overall rating, which referenced the candidate's qualifications overall, similar to the unstructured evaluation condition. Across the two résumé evaluation conditions, the only aspect of the exercise which differed was the level of detail at which the résumé evaluations were made.

Data Preparation

Missing values. A substantial proportion of rating dimensions in the structured résumé rating condition contained missing values, making rating reliability analyses problematic. To address this issue, the missing values were investigated to determine whether they were systematically related to specific rating dimensions. Preliminary investigations of the missing data by participant demonstrated that 58.3% of the participants in the structured résumé evaluation condition had at least one missing value. Thus, removal of participants with missing data was not feasible, as this would result in retention of only 5 of the 12 matched participants. Moreover, the percentage of missing values out of the total number of 180 dimension ratings (9 dimensions each for 20 résumés; the overall rating did not contain any missing values across participants and was thus excluded from the missing data analysis) did not exceed 10% for any participant. (The highest percentage for any participant was 9.4%.)

A closer inspection of the missing data at the résumé dimension level suggested that there was a systematic association between certain résumé dimensions

and a larger proportion of missing data. Specifically, for 35% (7) of the résumés, 25% (3) of the participants failed to provide a rating on the Objective dimension (i.e., a rating of the candidates' stated résumé objective). A smaller proportion of participants failed to provide a rating on the Volunteerism dimension (55%, or 11, of the résumés had at least 1 missing value on this dimension).

Proposed solution. As indicated above, missing data were not so pervasive that they warranted removal of résumés or participants. Nevertheless, it was enough of a concern that the following steps were taken to address the issue. First, each participant's résumé dimension missing value was replaced with the mean for that rating dimension (for that particular candidate's résumé, as opposed to the mean rating on that dimension across all résumés). The resulting value was used to replace any missing values on that dimension for that candidate résumé. Thus, if Participant 6 had a missing value for the Work Experience dimension on Résumé A, and the mean value on that dimension was 1.76 across the participants who provided a rating of Work Experience for Résumé A, the missing Work Experience data point for Participant 6 on Résumé A was replaced with this value. This mean replacement procedure was used to replace all missing values for each dimension within a candidate résumé (i.e., grand means across résumés for each dimension or means across all dimensions for a particular résumé were not used).

Analyses

Research Question 3 (What is the extent of the overlap between the results from an unstructured vs. a structured résumé evaluation system?) was explored

through an examination of the cross-condition rating correlations. Hypothesis 3a (Evaluation of résumés using a structured scoring system will result in greater interrater agreement compared to résumé evaluation using an unstructured scoring system.) was explored through examining the correlations across the two conditions.

To examine Hypothesis 3 (Reliability of a structured résumé evaluation system will be greater than the reliability of an unstructured résumé evaluation system.), intraclass correlations (ICCs) using the consistency model were computed within each résumé evaluation method. These were compared to determine which of the two evaluation methods resulted in greater consistency of applicant evaluation across different evaluators. Both the overall ICC estimate across all raters within condition as well as the estimate for a single rater within condition were examined. Because of the small sample of raters, the 95% confidence intervals for each estimate were compared across the two résumé evaluation conditions.

CHAPTER IX: STUDY 2 RESULTS

Study 2: Résumé Study

Means, standard deviations, intercorrelations, and intraclass correlations (ICCs) are reported in Table 28 across all 20 résumé targets. These descriptive statistics are also reported broken down by whether the résumé target was deemed to have "passed" (Table 29) or "failed" (Table 30) the résumé screen in the actual selection context in order to examine whether the relationships change as résumés go from lower to higher quality. In general, dimension ratings in the structured résumé evaluation condition exhibited strong interrelationships. Interestingly, the Format dimension, which included things such having an organized chronology, proper spelling, grammar, and punctuation, and consistency in terms of presentation style (e.g., consistent use of italics, capitalization, etc.), demonstrated the most consistent set of relationships in the matrix, being strongly associated with ratings of Skills (r = .81, p < .01), Employment Status (r = .72, p < .01), Work Experience (r = .69, p < .01), Education/Compensatory Experience (r = .84, p < .01), and Continuing Education (r = .52, p < .10). Considering the aforementioned résumé components would appear, intuitively, to be the most significant in rendering a decision on whether to pursue a candidate further, the way in which this information is presented by a candidate—and how conscientiously the information is reviewed by the candidate for accurate spelling, punctuation, and other typographical errors—is nearly as important as the information itself. The relationship of Format to overall ratings of résumé quality held across rating conditions as well as within the résumé "pass" and "fail"

sub-samples. Thus, even among the top candidates, formatting played a role in the evaluation of résumé quality. Work Experience, as would be expected, demonstrated significant, positive relationships with ratings of résumé Objective ($r = .65, p < .05$), Skills ($r = .81, p < .01$), Education/Compensatory Experience ($r = .74, p < .01$), and the previously-discussed Format dimension.

Turning to the relationships among dimensional and composite or overall ratings in the structured résumé evaluation condition, the composite structured rating (a data-driven, unit-weighted average of the 9 résumé dimensional ratings) exhibited different relationships to the dimensional ratings compared to the overall (global) rating within the structured condition. Specifically, for 4 of the 9 résumé dimensions (Employment Status, Advancement, Continuing Education, and Volunteerism), the relationship between these dimensions and the composite rating was significant, whereas the same relationship to the overall résumé rating made in the structured condition was not. A similar phenomenon was also found in the pass/fail breakdown (See Tables 29 and 30), albeit along somewhat different dimensions. In the case of the subset of "pass" résumés, the composite structured rating correlated significantly with each of the 9 dimensional ratings, whereas the overall structured rating did not correlate significantly with the Education/Compensatory Experience, Continuing Education, and Volunteerism dimensions. In the subset of "fail" résumés, the composite structured rating again demonstrated significant relationships across each of the 9 dimensions, whereas the overall rating in the structured condition did not

demonstrate a significant relationship to ratings of Employment Status, Advancement, Continuing Education, or Volunteerism.

It is likewise noteworthy that the overall rating in the unstructured condition demonstrated several significant relationships to the dimensional ratings in the structured evaluation condition. The unstructured condition (global) rating was significantly related to ratings of Skills ($r = .78, p < .01$), Work Experience ($r = .76, p < .01$), and Format ($r = .55, p < .10$) from the structured résumé evaluation condition. In addition, the unstructured condition rating correlated significantly with the corresponding overall rating made in the structured condition ($r = .67, p < .05$), as well as with the structured evaluation condition unit-weighted composite rating (at the .10 alpha level; $r = .55$). However, in neither case did the shared variance between any of these pairs of ratings exceed 45% across the 20 résumé targets (44.9% between the overall ratings in each condition; the shared variance between overall résumé ratings across conditions increased slightly to 53.3% when considering the "fail" résumés only). This suggests that other factors are at play which may be influencing the overall ratings. One possibility may be rating errors. Thus, the results suggest that the evaluation process differs between unstructured and structured résumé evaluation. That is, the evaluation criteria and/or the way in which they are evaluated and "weighted" (e.g., configural cue processing; Hitt & Barr, 1989) vary based on the way in which the résumé evaluation task is presented to raters.

Exploring Résumé Rating Reliability

To examine the level of reliability associated with the structured and unstructured résumé evaluation conditions, and how reliability may vary based on evaluation condition (Hypothesis 3), ICCs were calculated across conditions. The ICC statistic may be calculated in various ways, depending on the question asked. Although a detailed exposition on interrater reliability and ICCs in particular is beyond the scope of this dissertation, some discussion of the various ICC models and the rationale behind the model used in the present study is warranted.

According to a seminal article on the topic of interrater reliability, Shrout and Fleiss (1979) discuss six different ICC models which vary along two dimensions. The first involves raters and targets. The three models vary on whether the targets are rated by different judges (Model 1) or all the targets are rated by the same set of judges (Models 2 and 3), and whether the judges are a random sample of possible judges (Model 2) or whether they are the only judges of interest (Model 3). The second dimension involves whether the ICC reflects the reliability associated with the entire set of (k) judges or a single or subset of judges. Thus, if a researcher is interested in generalizing the reliability derived from a sample of judges who rated all the targets of interest to the broader population of potential raters, he or she would employ the ICC (2, k) model. These models were further expounded upon by McGraw and Wong (1996), with particular focus on agreement and consistency models, and statistical tests of obtained ρ estimates.

To explore Hypothesis 3, whether résumés are evaluated more reliably using a structured or unstructured rating system, the ICC (3, *k*) and ICC (3, 1) models were used, employing the model terminology from Shrout and Fleiss (1979). The rationale and questions to be answered with this ICC model were as follows. First, these models treat raters as fixed effects, and in the present study, raters were considered "experts" and thus the only population of interest. Second, consistency as opposed to absolute agreement was of primary interest, given that in an actual selection context, it would be of greater importance that recruiters were consistent in how they evaluated candidates as opposed to whether they agreed in an absolute sense (i.e., provided the exact same rating of each candidate résumé), and given that scores are not typically assigned to résumés, it was less critical to demonstrate that résumés received the same "score" across evaluators than it was to demonstrate that they were ranked in the same relative order. Although the ICC statistic is reported for both the entire set of raters as well as for a single rater, in some ways, the single rater estimate is of greater relevance in the present case since it is a closer reflection of how résumé evaluation is performed in an actual selection context. That is, résumés are typically reviewed by a small number of evaluators, and in some instances, may be reviewed by only one individual in an organization who makes the decision whether to pursue the candidate further.

Results of the ICC analyses are presented in Table 31. In general, the ICC (3, 12) estimates across experimental conditions across the entire set of raters were essentially identical. However, there were slight differences in terms of the results of the single measures (1 rater) estimates. Although not statistically significant (as

evidenced by the overlap in the 95% confidence intervals), the ICC (3, 1) estimates were much higher for the unit-weighted composite score in the structured condition (.60) compared to the unstructured condition (.46). Moreover, this estimate for the structured global rating (.39) was similar to that found for the global rating in the unstructured condition. This finding suggests the possibility that adding structure to the résumé evaluation process results in greater rating consistency and thus fewer raters would be needed to reliably evaluate candidate résumés, but only if the dimensional ratings are combined mechanically. That is, providing dimensional rating prior to making an overall ("clinical") assessment of résumé quality does not result in the same increase in reliability. Although the present study lacked the power necessary to test this difference statistically (beyond examination of the confidence intervals of the obtained estimates), such a large difference suggests that adding structure may significantly improve the quality (consistency) of résumé evaluation. Thus, Hypothesis 3, which explored whether interrater reliability across résumé evaluation would vary as a function of the level of structure employed, although not supported statistically in the present study, may warrant further exploration.

CHAPTER X: STUDY 2 DISCUSSION

Study 2 was an exploratory investigation into the effects of adding structure to the résumé evaluation process. The results of the study suggest that structuring the résumé evaluation process may have benefits, but future research is needed to identify specifically what those benefits are, how they relate to decisions and meaningful outcomes (e.g., hiring decisions, job performance), and how those effects may be maximized.

Résumés were evaluated by graduate students in I/O psychology under two conditions, structured and unstructured, in a repeated measures experimental, counterbalanced design. Although there was a strong relationship between overall résumé ratings produced across the two conditions, there was also an indication that the evaluative process differed across the two conditions. The cross-condition correlations between the unstructured overall résumé rating and the dimensional ratings from the structured résumé rating condition demonstrated significant relationships with three of the nine dimensions: skills, work experience, and format, and the variance shared between the overall ratings across conditions was less than 50%. Within the dimensional ratings in the structured evaluation condition, résumé format, skills, and continuing education/compensatory experience demonstrated the strongest relationships to the other dimensions.

Given that there was no true criterion against which to "validate" the resulting résumé scores across each rating condition, there was no realistic way in which to test whether the potentially differing evaluative processes involved across the two rating

conditions are truly meaningful. If the results of the two conditions were not expected to differ, then the correlation between the overall ratings should have approached 1.00. Although the correlation was rather high, it did not in any way approach unity ($r = .67$ between the overall rating in the structured condition and the unstructured condition overall rating). In fact, the findings indicate that the two approaches share only about 45% of their variance. This suggests that infusing the evaluation process with structure (or lacking structure in the evaluation process) is associated with differing rating outcomes.

Nonetheless, there was no way in which to test the significance of these differences against an external criterion. So the question remains as to whether these differences are meaningful in some way. The only proxy available to explore this in the present study is to look at the interrater reliability results across the two conditions. Although there was insufficient power in the sample used to evaluate the reliability of résumé ratings, some notable trends were observed. Adding structure to the résumé evaluation process may result in more reliable ratings, defined in the present study as relative agreement among a group of raters. The intraclass correlation coefficients obtained in the present study showed that, across the entire set of 12 raters, reliability was nearly identical between the structured and unstructured résumé evaluation conditions. However, when examining the reliability obtained from each condition for a single rater, the structured résumé evaluation condition composite rating, may be more reliable than an unstructured (i.e., global) résumé rating. A global rating made after providing dimensional ratings did not have the same effect. Unfortunately, the

present study, given the small sample of expert raters employed, was unable to demonstrate this effect statistically. The results of the present study also lend possible credence to the finding from previous research that suggests mechanical (statistical) decision-making models outperform clinical judgments (cf. Gatewood et al., 2007). The reliability of the overall (global) résumé evaluation rating in the structured evaluation condition single measures ICC analysis was essentially identical to the same overall résumé evaluation rating from the unstructured condition. In contrast, as mentioned, the single measures ICC result from the unit-weighted composite overall rating was considerably higher.

The results of the single measures ICC analysis across conditions in many ways is of greater importance than the results of the ICC analysis for the entire sample of raters. First, it is highly unlikely in an actual organizational context that 12 individuals would review a résumé at the initial stages of the selection process, if ever. Typically, in the early stages of the selection process, a single recruiter may be responsible for filtering candidates down to the select few who will be pursued further. Thus, if a reasonable amount of reliability can be obtained in rendering decisions across many more candidate résumés than raters in the present study were asked to evaluate, this would be an encouraging finding. Because of the nature of this task, the brevity with which résumés are typically reviewed, and the sheer volume of résumés that recruiters sift through in order to derive preliminary selection decisions, obtaining a reliability coefficient above .60 would be a significant improvement. The typical reliability of a résumé review process, where recruiters make overall

acceptable/unacceptable judgments, most likely reflects, at best, that which was found in the unstructured evaluation condition (.46). Assuming this estimate is a representative one, it suggests that single-evaluator résumé reviews are rife with inconsistencies and error. It is extremely disconcerting to think that a pervasive selection tool such as the résumé would demonstrate such abysmal reliability. Moreover, in the present study, the "control" condition included several evaluation tools such as a job description and descriptions of the résumé evaluation dimensions that were rated in the structured condition which ostensibly would have oriented the student SMEs in their evaluation process, and yet the estimated single measures ICC was .46. And, given the fact this was an experiment, it is reasonable to assume that the student SMEs were much more thorough in their review of the résumés than recruiters would be in an actual selection context, and even in this context the reliability of .46 in the unstructured résumé evaluation condition was surprisingly low. Given the fact that research in the area of assessment centers has demonstrated that psychologists typically make better raters (Woehr & Arthur, 2003), the ICC estimates obtained may be an upper bound estimate.

Prior research in the résumé evaluation and interviewing literature provide some insights. Although it was not tested in the present study, it is possible that in the absence of structure one of two things may be occurring. First, previous research has suggested that raters may engage in what is known as configural cue processing (Hitt & Barr, 1989). In general, this refers to the possibility that decision-makers, faced with a multitude of variables on which to focus in order to arrive at a final decision, may

combine these variables (cues) in non-linear ways. According to the authors (Hitt & Barr, 1989, p. 53), "When dealing with complex tasks involving uncertainty, complexity, or ill-structured problems, such as selection decisions, people try to simplify the decision process by developing and relying on heuristics." The addition of several interaction effects in the Hitt and Barr study explained variance in managerial decision-making beyond that explained by the main effects. Although that study did not focus on résumé reviews, but rather videotaped interviews with candidates for upper level management positions, the results lend support to the notion that implicit interactional models among job relevant—and in the Hitt and Barr (1989) study, job-irrelevant—variables predicted managerial decisions better than a simple linear combination of the variables. Future research in the area of applicant résumés should explore the potential effects of interactional models.

Another potential explanation is that, in the absence of a structured process for decision-makers to follow, they rely on their own implicit decision-making models. However, this is not to say that effective decisions or evaluations of candidates cannot be obtained from multiple decision-making approaches. There is research in the context of interviewing that suggests decision-makers may employ disparate strategies in arriving at their final selection decisions (Graves & Karren, 1992). In other words, out of the multitude of job-relevant information available, decision-makers may rely on a small subset of variables in arriving at their final decisions. These implicit models vary in terms of their effectiveness depending on the choice of information and awareness of its relevance and understanding of the decision strategy, itself. In the

study, effective interviewers were more likely to emphasize job relevant criteria in their decision-making process compared to ineffective interviewers. Moreover, effective interviewers also had greater awareness and understanding of their decision-making strategies than did ineffective interviewers. It is plausible to assume that a similar phenomenon could exist in the résumé review process, which would lend further support to adding some form of structure to the review process to increase consistency of decision-making models and the evaluation process. The fact that the reliability of the overall ratings in the structured condition were no better than the reliability in the unstructured condition, suggests that these ratings are more influenced by decision-makers' implicit evaluation models. If this were not the case, then the dimensional ratings made in the structured condition should have influenced the overall résumé rating. Alternatively, it may be the case that evaluators, short of calculating an average score themselves, are not able to synthesize their dimensional ratings to arrive at an overall (clinical) rating, at least not as consistently as that which can be obtained by taking a mechanical combination (arithmetic average) approach.

One additional notable finding from this study concerns the effects of résumé format on the evaluation of résumé quality. Ratings of résumé format, that is, the quality of the résumé in terms of its structure, organization, and use of proper spelling, grammar, and punctuation, was strongly associated with overall ratings of résumé quality. This was the case across both "pass" and "fail" résumés as well as across experimental conditions (i.e., ratings of format from the structured condition correlated with overall ratings in the unstructured condition). This finding raises some

interesting questions regarding what drives the overall assessment of a candidate's résumé. Although there was no way in which to test this in the present study, there are two possible explanations for this result. First, it could be the case that the quality of a candidate's résumé in terms of its formatting correlates with an overarching factor of mental ability or extensive work experience (i.e., highly successful employees may be better able to construct their résumés to highlight their qualifications and experience). Over the entire set of candidate résumés, format was significantly related to the assessment of candidates' skills, employment status, work experience, education/compensatory experience, and continuing education, lending some support to the idea that there is an underlying general mental ability or "career success" factor that is driving the significance of résumé formatting. Alternatively, format could be related to self-presentation processes, such as impression management. In other words, candidates who are savvier in terms of their ability to present themselves in a positive light may be more adept at constructing their résumés in such a way as to make them appear more desirable. Either they are better at presenting their job experience or alternatively they may be better "embellishers" and have a better understanding of what employers are looking for when they review candidate résumés. Given that there was no criterion against which to "validate" the résumé scores, this could not be tested. Based on the résumé rating correlations, it appears that the former explanation of the effects of format is more plausible, although it is also possible that both general mental ability and impression management explain this phenomenon.

One additional potential explanation for the effect of format on résumé ratings comes from the perspective of the evaluator. It is possible that the positive association between ratings of résumé format and overall résumé quality ratings is driven by a form of the halo effect. That is, because the résumé is well-constructed, is devoid of typographical and grammatical errors, evaluators may evaluate the résumé more favorably, even if the candidate may not be competitive for the position based on his or her job-related qualifications. Positive associations between format and overall ratings of résumé quality, whether this was operationalized as a global rating (unstructured condition) or a mathematical average of the dimensional ratings (structured condition), were found for both "pass" and "fail" résumés. Future research should attempt to gain a better understanding of whether format's relationship to résumé quality ratings constitutes a halo effect or some sort of "career success" factor.

Conclusions

In general, the résumé is an imperfect instrument for selection decisions in that it is inherently unstandardized, subjective, and rife with self-presentation effects on the part of candidates, and therefore it potentially requires several inferences to be made by those who seek to make decisions based on the information contained within them. This is not to say that it cannot be a useful tool, especially during the preliminary stages of the selection process, where the primary objective is to narrow the field down to a manageable number of competitive candidates (i.e., the compatibility test; Beach, 1990). However, to the extent that the process can be standardized and simplified, the outcomes that result should improve. Standardization of the selection

interview has demonstrated these effects, being a more valid and legally defensible selection tool as a result (Campion et al., 1997; Pulakos & Schmidt, 1995; Williamson et al., 1997). Previous research by Fritzsche and Brannick (2002) also suggests that recruiters are able to evaluate résumé profiles more reliably than they were able to evaluate actual résumés. A similar profile methodology was also employed in the Graves and Karren (1992) study. Such profiles distill the information from résumés into a more user-friendly format, allowing recruiters to quickly and more reliably identify and evaluate relevant candidate qualifications. Even more cursory standardization techniques, such as résumé checklist evaluation forms (similar to the T&E method) may likewise provide sufficient standardization to improve the decision-making process. The present study used a rather high-level, broad-based evaluation form which measured a relatively small set of dimensions and yet produced substantial improvement in rating consistency. It took a relatively small amount of time and effort to develop, and could apply to a wide variety of jobs, particularly at the professional (experienced) level.

Potential Limitations

The present study suffers from several potential limitations, some of which have been stated previously. First, the small sample size of evaluators precluded the possibility of conducting any sort of rigorous statistical comparisons between the two résumé evaluation conditions. However, the desire was to have a sample of "experts" and, although graduate students may not be the ultimate experts when it comes to recruiting and reviewing résumés, the better alternative, using actual recruiters in the

focal organization, was not feasible. Moreover, the use of actual recruiters would have limited the sample size even further. The organization had a recruiting team of only four recruiters for professional positions, only one of whom was particularly familiar with the position investigated. Moreover, based on the results of a meta-analysis in the area of assessment center ratings showed that, in terms of potential rater types, psychologists were better raters than job incumbents or managers (Woehr & Arthur, 2003). Therefore, it is possible that the graduate student ratings provided an upper bound estimate of résumé evaluation reliability compared to what might have been obtained from a sample of recruiters.

The standardized résumé evaluation form developed may also have been a potential limitation. The nine résumé dimensions were developed with the input of only two recruiting professionals. It is possible that, had a broader set of SMEs been employed in the development of the standardized evaluation form, the final model may have differed substantially. However, one previous study in résumé evaluation developed a similar evaluation model containing six dimensions: career objective, education, work experience, activities, references, and format (Fritzsche & Brannick, 2002). This model is quite similar to the one that was developed for the present study.

In a similar vein, the study could have taken a more task- or competency-oriented approach to the development of the standardized résumé evaluation form. Although a more detailed scoring form might have increased the amount of variance in overall résumé scores and further reduced any potential ambiguity or reliance on personal inferences into the résumé content, this approach was not feasible for the

following reasons. First, as previously mentioned, recruiter resources at the focal organization were incredibly limited, and the amount of face-time with recruiters that would have been required (not to mention with other SMEs such as hiring managers and job incumbents) in order to develop a competency-based résumé evaluation form was prohibitive. Further, the addition of cues to the résumé evaluation tasks could have had the opposite effect from the one intended. In other words, the evaluation task may have been more cognitively taxing on participants had they been required to evaluate candidates along 20 different dimensions as opposed to the nine they were asked to evaluate. And, practically speaking, recruiters would not spend the amount of time required to evaluate candidates using such a model, particularly when the objective at this stage of the selection process is to narrow the candidate field down to those candidates who are competitive to determine which candidates should proceed to the next selection stage. Although the present study was rather modest in terms of its scope, the results suggest that the résumé evaluation process, as it currently exists in most organizations, may be of little practical value. This sentiment has been expressed by other practitioners in the field (cf. Handler, 2002), and much more research needs to be conducted to better understand the limitations of the résumé evaluation process and to identify ways in which to improve it.

Future Research

There are several avenues for future research on the role of the résumé in the selection process. In line with what was investigated in the present study, future research should further investigate the potential benefits of a structured résumé

screening process. There are a few different paths which could be pursued. First, a study very similar to the one discussed here should be conducted, with the primary differences being the sample size and composition. One option would be to conduct the study using a sample of expert recruiters, preferably in significantly greater numbers than in the present study. Given the potential challenges in obtaining such a sample, especially assuming a repeated measures design similar to the one employed in the present study, an alternative would be to use a large sample of naïve raters (e.g., undergraduate students). Using a larger sample would allow for the ability to regress overall ratings onto dimensional ratings. Although using an inexperienced sample of participants would raise questions concerning the generalizability of the results, the tradeoff would be a larger sample on which to conduct more rigorous statistical tests of hypothesized differences between conditions. Alternatively, to gain greater confidence in the results from a study using inexperienced raters, a sample of recruiters could be used as a "check" on the results, perhaps in an independent measures study where two separate samples of recruiters complete the résumé evaluation exercise under one of the two rating conditions only. Given that the idea behind adding structure to the résumé evaluation process is to standardize it so that "anyone" could effectively and reliably carry it out, the expectation might be that the results are quite similar across sample types in the structured evaluation condition, whereas there may be differences across the two samples in the unstructured evaluation condition. The question would be whether these differences were systematic in any way. In a related vein, a more comprehensive study into rater type

could be conducted, whereby results from recruiters, naïve raters, and I/O psychologists can be fully crossed in a factorial design. Ideally, such a study would also include some sort of criterion data so that not only could the reliability of résumé evaluation across rater type and evaluation condition be examined, these ratings could also then be tied to meaningful performance or organizational outcomes.

Second, future research should tie the outcomes of different résumé evaluation methods to meaningful, job-relevant criteria. This would require identifying a job to examine that has a high hiring volume, which the present study lacked. Simple criteria such as hiring decisions and retention could be used, or, depending on availability and willingness of an organization to provide such data, research-only job performance data or other objective performance criteria, should any exist, could provide a deeper understanding of the benefits of a structured résumé evaluation process, similar to what has already been investigated in relation to the structured interview.

Finally, future research should investigate résumé evaluation methods that eliminate most of the human judgment element. This would involve feeding résumé data for a particular job into an automated résumé scoring program. There are several challenges and considerations involved in carrying such a study out. First, because these programs semantically interpret résumé content, a hierarchically structured knowledge base would first need to be developed. Moreover, in order to populate the "criterion space" which is required in order for such programs to plot the résumé "scores" in multidimensional space the job for which such a study would be conducted would need to have thousands of résumés available. As evidenced by the foregoing

discussion, conducting such a study would be a serious undertaking. Then, of course, the question becomes whether and how the results derived from a proprietary résumé scoring program could be compared to other résumé evaluation conditions such as those investigated in the present study. At a minimum, the human participants involved in such a study could not possibly be expected to evaluate hundreds of résumés. Thus, there would be several critical study design elements that would need to be thoughtfully addressed in order to make direct comparisons. And, should the proprietary résumé scoring system outperform methods involving human judgment, the question becomes whether the time and financial investment required to develop such a system is worthwhile. This solution could have great benefits for large companies with high applicant volumes. And, for smaller organizations, a structured evaluation system involving human judgment, although flawed, may also provide substantial benefits. Ultimately, because the reliability of a predictor necessarily limits its validity in predicting a criterion, establishing a consistent, reliable manner in which to evaluate résumés would serve as a first step toward increasing their utility in the selection process.

CHAPTER XI: GENERAL DISCUSSION

Across the two studies presented in this dissertation, I have attempted to better understand and identify ways in which preliminary applicant evaluation procedures may be improved upon in order to increase the likelihood that organizations focus on, and ultimately hire, the best candidates. The first study focused on online applicant prescreening. An item taxonomy originally developed as a means to classify biodata items was applied to the questions asked in prescreening protocols used in three different positions at a utility company. The prescreening items were examined both in terms of how they differentiated among candidates, based on the distribution of prescreening scores in the candidate sample, as well as in terms of their ability to predict variance in later stages of the selection process and hiring decisions. The second study sought to improve the consistency of the résumé evaluation process through increasing the level of structure involved. This was accomplished through the development of a rudimentary résumé evaluation form that required raters to evaluate the résumé along 9 critical dimensions. This structured evaluation process was compared to an unstructured process where raters provided a single, overall assessment of each résumé's quality. The overarching results of each of these studies and their implications are reiterated below.

Prescreening Study (Study 1)

Study 1 sought to link differences in applicant responses to prescreening questions at various points in the prescreening score distribution and significant prescreening item-criterion relationships to an item taxonomic framework used to

describe biodata items (Mael, 1991), to assess whether the characteristics of these items (as evaluated by a sample of graduate student SMEs) would be likely to result in meaningful applicant differentiation. For example, if items predicting performance in later stages of the selection process were found to be rated significantly lower in terms of their verifiability and job-relatedness and higher in term of their susceptibility to faking behavior, these would not be considered quality items. Results of the prescreening score distribution analysis comparing the top and bottom of the prescreening score distribution were inconclusive. In the UW sample, the subset of prescreening items that distinguished candidates at the ends of the score distribution were higher on the minimum qualifications dimension as predicted, but they were not higher on job-relatedness. Although not hypothesized, these items were also more discrete in nature. This hypothesis was not testable in the other candidate samples because all prescreening items exhibited significant differences between candidates in the top and bottom of the score distribution. Results of the prescreening score distributional analyses involving the top-scoring candidates also were somewhat mixed. In the UW sample, the questions distinguishing among the top candidates were higher on the minimum qualifications dimension, contrary to what was expected. However, these questions were also higher on objectivity, as hypothesized. In contrast, in the CSR candidate sample, the questions differentiating among the top-scoring candidates were not rated higher on minimum qualifications, as hypothesized, but these questions were also significantly *less* objective than the questions which did not differentiate candidates. Results in the GSR sample were also somewhat conflicting.

Prescreening questions that differentiated the high scoring candidates trended in the direction of being significantly more objective and they were also significantly higher on the minimum qualifications dimension, contrary to what was hypothesized. Across all three jobs, total prescreening scores failed to correlate with preemployment test battery scores, test battery pass/fail, and hiring decisions. However, some of the individual prescreening questions did demonstrate significant item-criterion relationships. Questions that predicted preemployment test outcomes were significantly higher on verifiability and objectivity in both the UW and CSR samples. In the CSR sample, these items were also more historical, external, firsthand, discrete, and controllable. The relationships in the GSR sample were more tenuous, primarily due to the fact that there was only three questions investigated, and only one of these was related to testing outcomes at the .10 level. Nevertheless, overall, the hypothesized prescreening item – criterion relationships received support in terms of the characteristics of these items. Finally, Research Question 2, which sought to distinguish prescreening response profiles of candidates who were hired by the organization from candidates who were considered qualified for the position based on the selection hurdles investigated, but who were not ultimately hired, was not testable across the jobs investigated.

Recommendations and Future Research

The results of this study lend themselves to preliminary recommendations in terms of the way in which practitioners may partner with organizations in developing improved applicant prescreening protocols. As mentioned in Chapter VII, in its current

state, assuming that the prescreening question sets investigated in the present study are representative of those used in organizations as a whole (cf. Hunter et al., 2004) applicant prescreening is more effective as a de-selector of candidates as opposed to a candidate differentiator. However, refinements in terms of the development and scoring of prescreening protocols should result in an increased ability and confidence in making judgments of candidate quality. First, because the prescreening scores calculated across the positions investigated in the present study failed to predict performance in later selection stages, a content validity approach to prescreening item development is crucial. In addition, items should be constructed in such a way that they are more verifiable and objective, as questions with these characteristics are less likely to be susceptible to applicant dissimulation. And, although highly job-related questions are at times more susceptible to faking, it would stand to reason, from a legal defensibility standpoint, that questions should also be based on a detailed analysis of the position in question. Empirical keying or the use of an Angoff-like method of developing the scoring system associated with the prescreening question set may also result in the ability to statistically validate the prescreening protocol. At a minimum, a more detailed question writing process should allow recruiters to separate candidates into broad prescreening score categories.

Future research in the arena of applicant prescreening should continue to explore the item classification taxonomy outlined in Mael (1991). This framework demonstrated some utility in understanding applicant prescreening protocols in the present study. Fruitful avenues include a continued refinement of the way in which the

dimensions are defined and operationalized, and testing the model in a controlled setting. The added control in a laboratory study would allow researchers to manipulate the way in which questions tapping a single work-related construct are written (e.g., one version may be objective, the other subjective) to increase the internal validity of the relationships obtained. Finally, prescreening item characteristics should be examined in relation to applicant reactions. To date, no study has employed an item classification structure in relation to reactions to preliminary applicant screening systems (although screening, using IVR technology, has been investigated in terms of applicant reactions; Bauer et al., 2004).

Résumé Study (Study 2)

Study 2 explored whether adding structure to the résumé evaluation process would result in more consistent evaluation of candidates. Although not definitive from a statistical standpoint, the results of the study suggest that adding a modest amount of structure to the evaluation process may result in significant improvements in terms of the ordering of candidates, compared to a control condition where an overall assessment of candidate résumés was made. Graduate student SMEs provided ratings of résumé quality in this study. Although ratings across the two conditions were highly correlated, the ratings were not interchangeable, sharing only 45% of the variance, suggesting that the underlying evaluation process differs depending on whether structure is present or absent. Intraclass correlation estimates in the two conditions were essentially identical across the entire set of raters, but the single measures estimates—the reliability of one rater—differed substantially, but because of the small

sample of "expert" raters used in this study, this difference was not statistically significant. These results are in line with the generally accepted finding that mechanical decision-making models outperform clinical judgment models (cf. Gatewood et al., 2007). Simply requiring evaluators to provide structured ratings prior to rendering an overall résumé quality rating was insufficient in raising the reliability of the evaluation process. Only when the dimensional ratings were combined statistically did the reliability of the résumé evaluation process improve. The results of the single measures ICC analysis were of greater importance in the context of the present study than the average measures ICC because in practice, résumés at the initial stages of the selection process may only be reviewed and evaluated by a single member of the organization. The results of this study suggest that the typical résumé review process in organizations is rife with inconsistencies and errors. One additional finding of interest in this study involved the strong relationships of résumé format to many of the other dimensional ratings in the structured condition, as well as the overall résumé ratings made across conditions. The results of the present study could not disentangle the possible GMA and self-presentation processes effects which may serve as potential explanations for these relationships. Future research into the applicant résumé should investigate the effects of different raters (similar to what has been done in the performance appraisal literature; Woehr & Arthur, 2003) and how the presence or absence of structure in the rating task affects evaluation consistency. In addition, different structured approaches should be examined. One such structuring approach would be to develop a KSA-based structured evaluation format, similar to

T&E evaluation. Finally, future research should investigate automated résumé evaluation programs and their potential benefits.

Conclusion

In an era being described as a war for talent, with organizations fighting for the same small pool of candidates, organizations' ability to attract, select, and retain top talent is more important than ever. Preliminary applicant screening and evaluation procedures, the information collected on candidates, and the ways in which this information is collected, have likewise increased in importance. With the rise in prevalence of so-called high-tech selection methods, including online applicant screening (e.g., Cober et al., 2000; Dineen et al., 2002; Lievens et al., 2002), the need to effectively and efficiently collect and evaluate applicant data and make quick decisions is paramount. Moreover, these methods must do double duty, not only serving as an initial assessment of applicants but also as a recruiting tool. Simultaneously, there has been a call in the literature for increased understanding and systematic assessment of the effectiveness and quality of recruitment outcomes (Carlson et al., 2002) as recruitment is one of the least understood processes in terms of its effects on later stages of the selection process (cf. Breaugh & Starke, 2000). Both of the preliminary selection tools examined in the two studies discussed in this dissertation have the potential to provide meaningful information on the quality of candidates who apply to positions in organizations. However, many improvements must be made, and further research conducted, in order to realize these potential benefits.

According to image theory, a decision-making model first put forth by Beach (1990; Beach & Potter, 1992), selection decisions occur in two stages. The first stage, the compatibility test, involves the process of screening out options that do not meet the desired criteria. In their current form and typical usage, both online applicant prescreening and résumé screening procedures currently do an adequate job of narrowing down the candidate field. However, in order to better evaluate the outcomes of recruitment efforts, as well as to begin to decide who, among the remaining alternatives, should proceed in the process (the profitability test), more must be done with these preliminary applicant evaluation procedures. The results outlined across the two studies discussed earlier provide some preliminary insights and potential avenues for future research in order to realize additional benefits from these methods. Specifically, in terms of online applicant prescreening, steps can be taken to improve the process by using informed, theory-based decisions as to the content of the questions asked in screening, and making attempts to link the results, however modest the relationships might be, to performance in later stages of the selection process and/or after candidates have been hired. Likewise, adding an element of structure to the way in which résumés are screened and evaluated, requiring a specific yet brief structured evaluation may result in more consistent résumé evaluation, in line with previous discussions in the literature of the superiority of mechanical (statistical) decision-making models over judgmental models (cf. Gatewood et al., 2007).

Organizations are social systems which must interact with and obtain resources from the environment in order to survive (Katz & Kahn, 1978). One of the ways in

which they accomplish this is by attracting, selecting, and retaining the human capital needed in order to maintain negative entropy and dynamic homeostasis. In other words, the recruitment function aids the organization in continuing to survive and grow by placing or replacing (i.e., selecting) talent into the organization's critical roles or functions. More specifically, in the personnel psychology literature, the success of organizations striving to meet these environmental demands and system needs hinges on their ability to attract, select, and retain key talent (Schneider, 1987). Asking the right questions, via an online prescreen, may aid in attracting candidates to the organization. Changing the way in which preliminary evaluations of applicants are made may improve the organization's ability to attract and select the right people. If screening content is relevant to the job, objective, verifiable, and, ideally, provides candidates with an idea of the type of work involved in the position for which the organization is hiring, this can aid in reducing the field to those candidates the organization feels are worth pursuing. Approaching the problem from the opposite pole, prescreening candidates can also allow the organization to spend little time on candidates who clearly do not fit the organization's needs.

In sum, the two preliminary applicant evaluation methods examined across the two studies provide organizations with critical information on candidates' potential person-job (P-J) fit (e.g., Kristof, 1996). In both prescreening and résumé evaluation, the primary objective is to determine whether candidates possess the necessary skills and/or prior job experience needed to be successful in the role. Thus, these methods can be conceptualized as measures of P-J fit. The studies presented here did not

capture the person-organization (P-O) element of fit. This assessment takes place via telephone or on-site interviews, where detailed information about the candidate beyond their job-related skills is assessed. Preliminary applicant evaluation procedures are not designed to make such candidate evaluations. Moreover, the P-O literature suggests that P-O fit does not factor into selection decisions until both overall (e.g., vocation) and job-specific fit have been assessed, which ideally occurs during the initial stages of the selection process (Kristof, 1996). In fact, Kristof-Brown (2000) found that P-O and P-J fit have different antecedents.

Much additional research is needed, both in the arena of applicant prescreening and résumé evaluation, in order to maximize the utility of each of these selection tools in the preliminary stages of the employee selection process. The studies outlined above have begun to apply a structured, theoretical approach to understanding how to optimize each. Although neither of these tools will ever render a final decision on candidates, they serve as the foundation upon which organization can build in their quest for finding the best candidates.

TABLE 1

Study Hypotheses and Proposed Analysis Techniques

Hypothesis	Job Type	Test/Analysis
Study 1: Prescreening Study		
H1a: Prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings. *H1b:* Prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications.	UW, CSR, & GSR	• Divide sample into high- and low-scoring groups (H1a) and into 2 high-scoring groups (H1b). • *t*-tests to identify items differentiating groups. • Divide items into 2 groups; items differentiating applicants and items not differentiating applicants; *t*-tests of mean differences between item groups based on graduate student item ratings.
H2a: Applicant prescreening scores will be predictive of key new hire performance criteria (e.g., test scores, training performance). *H2b:* Prescreening items predictive of key new hire performance criteria will be high on objectivity and verifiability.	UW, CSR, & GSR	• Correlation of applicant prescreening scores with available criterion data (H2a). • Regression analysis to identify items with significant beta weights. • Group items predictive/not predictive into separate groups; *t*-tests of mean differences between item groups based on graduate student ratings of objectivity and verifiability.
RQ1: Are there meaningful distinctions between applicants that can be made at points in the applicant prescreening score distribution where there are sharp increases or decreases in the number of applicants receiving a particular score? What is the nature of these distinctions?	UW, CSR, & GSR	• Histogram analysis to identify dramatic increases/decreases in frequency of applicants obtaining given score; identification of questions providing differentiation and nature of items (based on graduate student item ratings).
RQ2: Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?	UW, CSR, & GSR	• Discriminant analysis; assess accuracy of classification indices (hired/not hired)
Study 2: Résumé/Lab Study		
RQ3: What is the extent of the overlap between the results from an unstructured vs. a structured résumé evaluation system? *H3:* Reliability for a structured résumé evaluation system will be greater than the reliability of an unstructured résumé evaluation system.	BAR	• Absolute agreement/correlation matrix between methods (RQ3). • Comparison of the intraclass correlations across methods; adjustment for sample size for direct comparison of reliability (H3).

Note. UW = Utility Worker; CSR = Customer Service Representative; GSR = Gas Service Representative; BAR = Business Account Representative.

TABLE 2

Applicant Evaluation Method, Participants, and Procedures Summary

Evaluation Method	Description	Participants	Job Type	Procedures/Comparisons
Applicant Prescreening	Job-related questions assessing applicant past experience.	Applicants; Graduate Students	UW CSR GSR	Data provided by screening vendor in Excel files that are formatted, manipulated, and imported into SPSS for analysis; graduate students provide ratings of prescreening items based on Mael's (1991) biodata taxonomy for use in interpreting meaning of applicant prescreening differentiation
Unstructured Résumé Evaluation	Evaluation of résumé content based on global impressions of applicant suitability	Graduate Students	BAR	Present graduate students with sample of 20 résumés and job description, key elements to look for; graduate students evaluate résumés and provide overall rating
Structured Résumé Evaluation	Structured procedure for scoring résumés developed for the study, similar to T&E and application blank scoring systems.	Recruiters; Graduate Students	BAR	Recruiters identify important elements of résumés on which they base decisions; ratings made by recruiters of the criticality of these elements used to develop structured résumé scoring system; graduate students presented with sample of 20 résumés (same résumés as in previous scenario), job description, and scoring system; score résumés based on key elements from scoring system

Note. UW = Utility Worker; CSR = Customer Service Representative; GSR = Gas Service Representative; BAR = Business Account Representative.

TABLE 3

Advantages and Disadvantages of Preliminary Applicant Evaluation Techniques

Advantages	Online Applicant Prescreening	Unstructured Résumé Evaluation	Structured Résumé Evaluation	Automated Résumé Scoring
Allows for verification of applicant responses		✓	✓	
Content evaluated is standardized	✓		✓	✓
Standardized Scoring	✓		✓	✓
Electronically stores applicant data/scores	✓			✓
Allows for combining applicant qualifications in non-linear way		✓		✓
Available 24 hrs a day	✓			
Procedure easily explained to applicants	✓	✓	✓	

Disadvantages	Online Applicant Prescreening	Unstructured Résumé Evaluation	Structured Résumé Evaluation	Automated Résumé Scoring
Difficult to verify applicant responses	✓			
Content evaluated is unstandardized		✓		
Scoring susceptible to human error/fatigue		✓	✓	
Time consuming to set up			✓	✓
Assumes linear combination of applicant qualifications	✓		✓	
Requires human judgment		✓		
Difficult to explain procedure to applicants				✓

TABLE 4

Prescreening Questions across Utility Worker, Customer Service Representative, and Gas Service Representative Positions, Study 1

Prescreening Questions				Job		
Screening Phase	Question Content	Response Format	Disqualify	UW	CSR	GSR
Online/MQ	At least 18 years of age	Yes/No	✓	✓	✓	✓
	High school diploma or GED may be required	Yes/No	✓	✓	✓	✓
	Possession of or ability to obtain a valid [State] driver's license may be required	Yes/No	✓	✓		✓
	Possession of a Class A driver's license	Yes/No		✓		
	Ability to work in various weather conditions may be required	Yes/No	✓	✓		
	Previous computer experience may be required	Yes/No	✓		✓	
	Ability to meet overtime requirement as needed on weekdays/weekends	Yes/No	✓		✓	
	Years of customer service experience in a call center environment	Multiple Choice			✓	
	Ability to arrive at work location in 30 minutes in an emergency situation	Yes/No	✓			✓
	Ability to work flexible schedule, including overtime on weekends/holidays	Yes/No	✓			✓
Telephone	Number of years of construction experience	Multiple Choice		✓		
	Previous experience working for a utility company	Yes/No		✓		
	Utility company work experience (which type of utility)	Multiple Choice		✓		
	Ability to meet overtime requirement in emergency situations	Yes/No	✓	✓		
	Ability and willingness to travel and transfer for Title 300 construction work	Yes/No	✓	✓		
	Software programs/applications experience	Multiple Response	✓		✓	
	Typing skills (words per minute)	Multiple Choice	✓		✓	
	Anticipated tenure in role	Multiple Choice	✓		✓	
	Provide reasons [candidate] would like to join the department/company team	Open-ended	✓		✓	✓
	Years of customer service experience working with general public	Multiple Choice	✓		✓	
	Describe what good customer service means to [candidate]	Open-ended	✓		✓	✓
	Shift/work schedule availability	Multiple Response	✓		✓	
	Rating of overall communication skills	Screener-rated	✓		✓	✓

Note. MQ = minimum qualifications. Question content contains a paraphrased version of the actual screening content. Open-ended and Screener-rated require the telephone screener to rate the candidate's answers/performance using 5-point anchored rating scales. Disqualify indicates the question contains at least one response option that would disqualify the candidate. UW = Utility Worker; CSR = Customer Service Representative; GSR = Gas Service Representative.

TABLE 5

Examples of Anchored Rating Scales for CSR and GSR Screener-evaluated Telephone Prescreening Questions, Study 1

	Anchors				
Question	1	2	3	4	5
Description of Good Customer Service	• Does not consider customer satisfaction • Does not exhibit empathy for customer need	• Basic interest in understanding customer • Desire to achieve customer satisfaction • Not always able to meet expectations	• Recognizes customer needs • Satisfactorily meets customer specific needs • Understands methods to assist customers	• Responds promptly to customer requests • Listens to customer concerns	• Wants to exceed customer expectations • Initiative to "go the extra mile"
Communication Skills	• Inarticulate answers to questions • Makes grammatical errors • Tone is harsh, rude, and/or belligerent	• Does not communicate clearly • Problems with pronunciation throughout call • Responds with improper answers	• Answers with proper pronunciation • Uses appropriate grammar and no slang • Responds with understanding to most questions	• Most answers are reasonable and logical • Responses are polite and courteous • Speaks in pleasant tone of voice	• Clear, organized, and logical answers • Highly professional and polite responses • Pleasant tone of voice, very courteous

Note. Not all behavioral examples for each answer are provided.

TABLE 6

Means, Standard Deviations, and Intercorrelations for Utility Worker Screening Questions, Preemployment Test Scores, and Selection Outcomes, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Class A License	0.15	0.36	—							
2. Yrs. Construction Exp.	3.87	2.72	.09*	—						
3. Utility Co. Experience	0.09	0.28	.16**	.26**	—					
4. Gas Utility Experience	0.01	0.11	.18**	.13**	.37**	—				
5. Phone Utility Experience	0.02	0.13	−.02	.11**	.41**	−.02	—			
6. Electric Utility Experience	0.03	0.18	.09*	.15**	.60**	−.02	−.02	—		
7. Water Utility Exp.	0.01	0.10	.03	.06	.33**	−.01	−.01	−.02	—	
8. Cable Utility Exp.	0.01	0.12	.07*	.10**	.39**	−.01	−.02	−.02	−.01	—
9. Overtime Requirement	1.00	0.04	.02	−.04	.01	.00	.01	.01	.00	.00
10. Total Screen Score	9.11	2.87	.22**	.99**	.37**	.19**	.14**	.22**	.10**	.14**
11. Mathematical Ability Test	26.56	10.39	−.05	−.07	.06	.03	.08	.09*	−.01	−.10**
12. Spatial Ability Test	26.02	10.24	−.05	.01	−.02	.04	.06	−.06	−.02	−.03
13. Test Battery Pass/Fail	0.60	0.49	−.11*	−.05	−.01	.06	.06	−.04	−.04	−.02
14. Hired	0.21	0.41	.08	.05	−.05	−.02	−.02	−.05	−.05	.03
15. Termed	0.04	0.21	−.09	−.03	.25*	−.03	−.03	−.03	—	.57**

Table continues

TABLE 6, CONTINUED

Means, Standard Deviations, and Intercorrelations for Utility Worker Screening Questions, Preemployment Test Scores, and Selection Outcomes, Study 1

Variable	9	10	11	12	13	14	15
1. Class A License							
2. Yrs. Construction Exp.							
3. Utility Co. Experience							
4. Gas Utility Experience							
5. Phone Utility Experience							
6. Electric Utility Experience							
7. Water Utility Exp.							
8. Cable Utility Exp.							
9. Overtime Requirement	—						
10. Total Screen Score	−.02	—					
11. Mathematical Ability Test	—	−.07	—				
12. Spatial Ability Test	—	.00	.36**	—			
13. Test Battery Pass/Fail	—	−.07	.56**	.65**	—		
14. Hired	—	.05	.06	.07	—	—	
15. Termed	—	−.02	−.05	.00	—	—	—

Note. *N*s range from 69 to 751.
* $p < .05$, ** $p < .01$.

TABLE 7

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations for Utility Worker SME Screening Item Taxonomy Ratings, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Verifiability	3.66	0.38	(.98)												
2. Job Rel.	3.60	0.60	−.26	(.23)											
3. Objectivity	3.84	0.35	−.08	.69*	(.98)										
4. Min. Qual.	3.27	0.72	−.01	.26	.34	(.86)									
5. Historical	3.56	0.35	.02	.18	−.01	.25	(.97)								
6. External	3.55	0.45	−.09	.27	.30	−.11	.47	(.94)							
7. Firsthand	4.39	0.53	−.34	.16	.24	−.49	.05	.36	(.89)						
8. Discrete	3.74	0.43	.49	−.11	.31	.40	.37	.33	.04	(.95)					
9. Controllable	3.65	0.45	−.54†	.13	−.20	.11	−.26	−.64*	−.09	−.59†	(.96)				
10. Eq. Access	3.15	0.36	.11	−.54†	−.63*	−.18	−.52†	−.56†	−.32	−.48	.16	(.55)			
11. Noninvasive	4.19	0.50	.61*	−.35	−.24	−.15	−.26	−.10	−.48	.12	−.41	.47	(−.06)		
12. Faking	2.58	0.46	−.16	−.06	−.04	−.38	−.17	.22	−.18	−.26	−.16	.09	.55†	(.98)	
13. Soc. Des.	2.96	0.80	−.36	.17	.57†	.33	−.20	−.22	−.06	.08	.26	−.35	−.29	.14	(.94)

Note. $N = 11$.

† $p < .10$, * $p < .05$.

TABLE 8

Utility Worker Mean Prescreening Item Score Comparisons between High- and Low-scoring Candidates, Hypothesis 1a, Study 1

Screening Question	+1 *SD*	−1 *SD*	Difference	*t*
Class A Driver's License	0.22	0.10	−0.12	−3.43**
Years of Construction Experience	6.94	0.12	−6.82	−256.97**
Previous Utility Co. Experience	0.21	0.00	−0.21	−6.99**
Which Utility?	—	—	—	—
Gas Utility	0.04	0.00	−0.04	−2.79**
Phone Utility	0.04	0.00	−0.04	−2.64**
Electric Utility	0.08	0.00	−0.08	−3.92**
Water Utility	0.02	0.00	−0.02	−1.95
Cable Utility	0.04	0.00	−0.04	−2.64**

Note. *N*s for the −1 *SD* group ranged from 189–190; $N = 252$ for the +1SD group.
** $p < .01$.

TABLE 9

Utility Worker Comparison of Top Scoring versus Bottom Scoring Candidates as a Function of Screening Question Type, Hypothesis 1a, Study 1

Item Classification Dimension	*M* of Differentiating Items	*M* of Non-differentiating Items	Difference	*t*
Verifiability	4.40	4.45	−0.05	−0.45
Job Relatedness	3.73	3.36	0.36	1.40
Objectivity	4.51	4.36	0.15	1.24
Minimum Qualifications	2.97	2.27	0.70	3.27**
Historical	4.61	4.64	−0.03	−0.15
External	4.33	4.27	0.06	0.52
Firsthand	4.82	4.72	0.09	0.90
Discrete	4.18	4.91	−0.73	−4.20**
Controllable	3.48	3.27	0.21	1.00
Equal Access	2.76	2.55	0.21	0.75
Noninvasive	4.45	4.55	−0.09	−0.82
Faking	1.76	2.09	−0.33	−1.02
Social Desirability	2.33	2.30	0.03	0.24

Note. $N = 11$. The mean of differentiating items was comprised of 3 items: Class A driver's license, years of construction experience, and whether candidate had previous utility company experience. The mean of non-differentiating items was comprised of 1 item: which utility in which the candidate had previous work experience.

** $p < .01$.

TABLE 10

Utility Worker Mean Prescreening Item Score Comparisons among the Highest-scoring Candidates, Hypothesis 1b, Study 1

Screening Question	High Score	Highest Score	Difference	*t*
Class A Driver's License	0.03	0.68	−0.64	−15.64**
Years of Construction Experience	6.93	6.99	−0.06	−1.88
Previous Utility Co. Experience	0.04	0.61	−0.57	−13.17**
Which Utility?	—	—	—	—
Gas Utility	0.00	0.14	−0.14	−5.25**
Phone Utility	0.00	0.12	−0.12	−4.95**
Electric Utility	0.01	0.23	−0.22	−6.43**
Water Utility	0.01	0.04	−0.03	−1.52
Cable Utility	0.02	0.08	−0.06	−2.52*

Note. $N = 178$ for the High Score group; $N = 74$ for the Highest Score group. High Score was defined as candidates receiving a screening score of 12; Highest Score was defined as candidates receiving a screening score of greater than 12.
* $p < .05$, ** $p < .01$.

TABLE 11

Utility Worker Comparison of Top Scoring Candidates as a Function of Screening Question Type, Hypothesis 1b, Study 1

Item Classification Dimension	*M* of Differentiating Items	*M* of Non-differentiating Items	Difference	*t*
Verifiability	4.65	4.18	0.47	3.51**
Job Relatedness	3.68	3.59	0.09	0.27
Objectivity	4.68	4.27	0.41	3.61**
Minimum Qualifications	3.23	2.36	0.86	2.61*
Historical	4.55	4.68	−0.14	−0.71
External	4.27	4.36	−0.09	−0.45
Firsthand	4.86	4.73	0.14	1.94
Discrete	4.82	3.91	0.91	2.82*
Controllable	3.59	3.27	0.32	1.88
Equal Access	2.95	2.45	0.50	1.45
Noninvasive	4.50	4.45	0.05	0.32
Faking	1.59	2.09	−0.50	−2.47*
Social Desirability	2.10	2.55	−0.45	−2.17

Note. $N = 11$. The mean of differentiating items was comprised of 2 items: Class A driver's license and whether candidate had previous utility company experience. The mean of non-differentiating items was comprised of 2 items: years of previous construction experience and type of utility in which the candidate had previous work experience.
* $p < .05$, ** $p < .01$.

TABLE 12

Utility Worker Comparison of Screening Questions as a Function of Relationships to Preemployment Test Scores, Hypothesis 2b, Study 1

Item Classification Dimension	*M* of Questions with Significant Criterion Relationships	*M* of Questions with Non-significant Criterion Relationships	Difference	*t*
Verifiability	4.72	4.10	0.62	6.21**
Job Relatedness	3.50	3.77	−0.27	−0.76
Objectivity	4.68	4.27	0.41	3.61**
Minimum Qualifications	3.18	2.41	0.77	2.48*
Historical	4.55	4.68	−0.14	−0.71
External	4.23	4.41	−0.18	−0.71
Firsthand	4.86	4.73	0.14	1.94
Discrete	4.86	3.86	1.00	3.09*
Controllable	3.64	3.23	0.41	2.04
Equal Access	2.95	2.45	0.50	1.98
Noninvasive	4.50	4.45	0.05	0.32
Faking	1.73	1.95	−0.23	−1.46
Social Desirability	2.10	2.55	−0.45	−2.17

Note. $N = 11$. The mean of questions with significant criterion relationships was comprised of 2 items: Class A driver's license and type of utility company in which candidate had previous work experience. The mean of questions with non-significant criterion relationships was comprised of 2 items: years of previous construction experience and whether candidate had previously worked in a utility company.

* $p < .05$, ** $p < .01$.

TABLE 13

Utility Worker Mean Prescreening Item Score Comparisons between Hired Candidates and Qualified Candidates Not Hired, Research Question 2, Study 1

Screening Question	Not Hired	Hired	Difference	*t*
Class A Driver's License	0.10	0.16	−0.06	−1.42
Years of Construction Experience	3.68	4.01	−0.33	−0.90
Previous Utility Co. Experience	0.09	0.06	0.04	0.96
Which Utility?	—	—	—	—
Gas Utility	0.02	0.01	0.01	0.29
Phone Utility	0.02	0.01	0.01	0.29
Electric Utility	0.04	0.01	0.02	0.89
Water Utility	0.01	0.00	0.01	0.91
Cable Utility	0.01	0.01	−0.01	−0.51

Note. $N = 257$ for the Not Hired group; $N = 70$ for the Hired group. Values in the Difference column may appear discrepant due to rounding.

TABLE 14

Means, Standard Deviations, and Intercorrelations for CSR Candidate Screening Data and Preemployment Test and Hiring/Termination Outcomes, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Yrs. CS Exp. in Call Center	2.99	1.95	—										
2. Database Experience	0.43	0.50	.14**	—									
3. Word Processing Experience	0.92	0.27	.03	.17**	—								
4. PowerPoint Experience	0.59	0.49	.09**	.30**	.18**	—							
5. Spreadsheet. Experience	0.83	0.38	.13**	.25**	.26**	.29**	—						
6. Visio Experience	0.08	0.27	.09**	.28**	.08**	.19**	.10**	—					
7. Other Software Experience	0.31	0.46	.04	.15**	.03	.09**	.11**	.11**	—				
8. Total Software Experience	3.16	1.35	.16**	.69**	.43**	.66**	.58**	.46**	.49**	—			
9. Typing Words per Minute	1.28	0.46	.07**	.09**	.04	.08**	.09**	.07**	.04	.13**	—		
10. Anticipated Tenure	4.53	0.80	.13**	.00	−.03	−.02	.02	.02	.00	.00	.00	—	
11. Reasons Join Call Center	1.59	0.53	.04	.07**	.03	.07**	.07**	.05*	.15**	.15**	.01	.06**	—
12. Yrs. CS Exp. General Public	4.61	0.91	.33**	.08**	.00	.03	.11**	.00	.01	.08**	.02	.15**	.09**
13. Describe Good CS	3.03	0.59	.00	.04	.01	.07**	.07**	.03	.11**	.11**	.03	.06**	.31**
14. Shift Availability	0.91	0.29	.03	.01	.01	−.01	.02	−.01	.01	.01	−.03	.04	.07**
15. Overall Comm. Skills	3.14	0.47	.07**	.04	.02	.03	.08**	.00	.10**	.09**	.01	.07**	.26**
16. Total Screen Score	26.33	6.35	.51**	.31**	.34**	.29**	.40**	.17**	.23**	.51**	.19**	.36**	.44**
17. Preemp. Test Battery Score	84.71	17.09	−.10**	.00	.02	.05	.06*	−.01	.06*	.06*	.12**	−.10**	.06
18. Preemp. Test Pass/Fail	0.63	0.48	−.07*	.02	.05	.06*	.09**	.03	.04	.08**	.13**	−.07*	.03
19. Language Skills Test	7.39	3.61	−.07*	.02	−.02	.03	.03	−.02	.11**	.06*	.11**	−.15**	.03
20. Basic Arithmetic Test	14.13	5.65	−.09**	.00	.02	.05	.06*	.01	.04	.06*	.08**	−.12**	.04
21. Problem Solving Test	7.43	3.40	−.12**	.01	.03	.02	.05	−.01	.04	.04	.08**	−.12**	.04
22. Coding Test	48.78	10.29	−.06*	.00	.04	.04	.06	−.01	.03	.05	.08**	−.05	.05
23. Vis. Speed & Accuracy Test	94.09	17.85	−.04	−.01	.00	.04	.02	−.01	.01	.02	.09**	.04	.03
24. Hired	0.15	0.36	.01	.00	.02	.08*	.02	−.01	.02	.04	.01	.05	.06
25. Termed	0.21	0.41	.18*	.10	−.02	.13	.02	.14	.06	.15	−.07	.06	.02

Table continues

TABLE 14, CONTINUED

Means, Standard Deviations, and Intercorrelations for CSR Candidate Screening Data and Preemployment Test and Hiring/Termination Outcomes, Study 1

Variable	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Yrs. CS Exp. in Call Center														
2. Database Experience														
3. Word Processing Experience														
4. PowerPoint Experience														
5. Spreadsheet. Experience														
6. Visio Experience														
7. Other Software Experience														
8. Total Software Experience														
9. Typing Words per Minute														
10. Anticipated Tenure														
11. Reasons Join Call Center														
12. Yrs. CS Exp. General Public	—													
13. Describe Good CS	.08**	—												
14. Shift Availability	.00	.10**	—											
15. Overall Comm. Skills	.09**	.53**	.01	—										
16. Total Screen Score	.61**	.53**	.35**	.36**	—									
17. Preemp. Test Battery Score	.01	.04	−.02	.02	−.03	—								
18. Preemp. Test Pass/Fail	.04	.05	.03	.04	.02	.78**	—							
19. Language Skills Test	−.01	.05	−.01	.02	−.03	.62**	.47**	—						
20. Basic Arithmetic Test	−.01	.06	−.01	.01	−.03	.78**	.62**	.43**	—					
21. Problem Solving Test	.00	.05	.01	.01	−.04	.70**	.58**	.41**	.71**	—				
22. Coding Test	.02	.02	−.01	.03	−.01	.78**	.60**	.37**	.41**	.38**	—			
23. Vis. Speed & Accuracy Test	.03	−.01	−.03	.00	.00	.68**	.51**	.20**	.26**	.19**	.55**	—		
24. Hired	.05	.02	.01	.04	.06	.02	—	.02	−.02	−.02	.05	.03	—	
25. Termed	−.02	.04	−.18	.01	.18	.04	—	−.22*	−.06	.01	.04	.20**	—	—

Note. *N*s ranged from 119 to 2,206. CS = Customer Service; Exp. = Experience; Comm. = Communication; Preemp. = Preemployment; Vis. = Visual.

* $p < .05$, ** $p < .01$.

TABLE 15

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations for Customer Service Representative SME Screening Item Taxonomy Ratings, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Verifiability	3.00	0.37	(.97)												
2. Job Rel.	3.65	0.53	.30	(.84)											
3. Objectivity	3.03	0.43	.57†	.15	(.96)										
4. Min. Qual.	3.16	0.61	−.05	.58†	−.26	(.89)									
5. Historical	2.88	0.29	.53†	.02	.21	.19	(.96)								
6. External	3.05	0.38	.63*	−.01	.72*	−.18	.43	(.96)							
7. Firsthand	4.36	0.54	.00	−.16	.53†	−.20	.00	.27	(.76)						
8. Discrete	2.85	0.78	.66*	.42	.34	.13	.40	.33	.24	(.88)					
9. Controllable	3.26	0.45	−.33	.06	.01	.22	−.29	−.26	.06	.09	(.89)				
10. Eq. Access	3.14	0.52	.33	−.18	.09	.08	.47	−.07	.02	.39	.28	(−.46)			
11. Noninvasive	3.82	0.64	.91**	.38	.50	−.02	.39	.39	−.10	.75**	.03	.49	(.84)		
12. Faking	3.31	0.57	−.46	−.44	−.22	−.10	−.03	−.37	.16	−.45	.33	.00	−.39	(.86)	
13. Soc. Des.	3.43	0.56	−.32	−.18	−.40	.02	−.29	−.65*	−.37	−.60†	.12	.25	−.21	.40	(.92)

Note. $N = 11$. Rel. = Relatedness; Min. Qual. = Minimum Qualifications; Eq. = Equal; Soc. Des. = Social Desirability.
† $p < .10$, * $p < .05$, ** $p < .01$.

TABLE 16

Customer Service Representative Mean Prescreening Item Score Comparisons between High- and Low-scoring Candidates, Hypothesis 1a, Study 1

Screening Question	Low Score	High Score	Difference	*t*
Years CS Experience in Call Center	1.93	4.76	−2.83	−29.24**
Total Software Experience	2.58	4.35	−1.76	−23.76**
Databases	0.24	0.77	−0.53	−17.50**
Word Processing	0.90	0.99	−0.10	−6.05**
PowerPoint	0.46	0.85	−0.39	−12.66**
Spreadsheets	0.74	0.98	−0.24	−10.24**
Visio	0.03	0.21	−0.18	−8.10**
Other	0.22	0.56	−0.33	−10.20**
Typing WPM	1.20	1.46	−0.27	−8.25**
Anticipated Tenure in Call Center	4.33	4.88	−0.55	−11.52**
Reasons Join Call Center	1.45	1.87	−0.42	−13.96**
Years of CS with General Public	4.53	4.98	−0.45	−11.31**
Describe Good Customer Service	2.71	3.39	−0.68	−17.29**
Shift Availability	0.78	1.00	−0.22	−10.48**
Communication Skills	2.92	3.38	−0.46	−11.98**

Note. $N = 795$.

** $p < .01$.

TABLE 17

Customer Service Representative Mean Prescreening Item Score Comparisons among the Highest-scoring Candidates, Hypothesis 1b, Study 1

Screening Question	High Score	Highest Score	Difference	*t*
Years CS Experience in Call Center	4.48	4.87	−0.39	−6.09**
Total Software Experience	3.66	4.62	−0.97	−11.74**
Databases	0.59	0.83	−0.24	−6.48**
Word Processing	0.96	1.00	−0.04	−3.14**
PowerPoint	0.72	0.89	−0.16	−4.91**
Spreadsheets	0.95	1.00	−0.05	−3.19**
Visio	0.10	0.27	−0.17	−5.65**
Other	0.34	0.64	−0.30	−7.51**
Typing WPM	1.34	1.51	−0.17	−4.17**
Anticipated Tenure in Call Center	4.78	4.92	−0.15	−3.58**
Reasons Join Call Center	1.80	1.91	−0.10	−3.45**
Years of CS with General Public	4.96	4.98	−0.03	−1.66
Describe Good Customer Service	3.24	3.49	−0.25	−6.37**
Shift Availability	0.99	1.00	−0.01	−1.43
Communication Skills	3.20	3.47	−0.27	−6.66**

Note. $N = 592$. WPM = words per minute.

** $p < .01$.

TABLE 18

Customer Service Representative Comparison of Top Scoring Candidates as a Function of Screening Question Type, Hypothesis 1b, Study 1

Item Classification Dimension	*M* of Differentiating Items	*M* of Non-differentiating Items	Difference	*t*
Verifiability	2.61	3.00	−0.39	−1.80
Job Relatedness	3.56	3.95	−0.40	−1.94
Objectivity	2.60	3.23	−0.63	−3.52**
Minimum Qualifications	2.66	3.05	−0.38	−1.25
Historical	2.56	2.82	−0.26	−2.58*
External	2.75	3.18	−0.43	−1.96
Firsthand	4.25	4.36	−0.12	−0.59
Discrete	2.62	2.45	0.17	1.01
Controllable	3.35	3.09	0.26	1.41
Equal Access	3.25	3.00	0.25	0.98
Noninvasive	3.83	3.41	0.42	2.16
Faking	3.73	3.32	0.41	2.05
Social Desirability	3.90	3.36	0.53	1.82

Note. $N = 11$. The mean of differentiating items was comprised of 7 items: years of customer service experience in a call center, total software experience, typing words per minute, anticipated tenure in a call center, reasons the candidate wants to join the call center, candidate description of good customer service, and communication skills. The mean of non-differentiating items was comprised of 2 items: years of customer service experience with the general public and shift availability.
* $p < .05$, ** $p < .01$.

TABLE 19

Customer Service Representative Comparison of Screening Questions as a Function of Their Relationships to Preemployment Test Scores, Hypothesis 2b, Study 1

Item Classification Dimension	*M* of Questions with Significant Criterion Relationships	*M* of Questions with Non-significant Criterion Relationships	Difference	*t*
Verifiability	3.25	2.25	1.00	4.09**
Job Relatedness	3.82	3.51	0.31	1.56
Objectivity	3.50	2.13	1.37	6.85**
Minimum Qualifications	3.05	2.51	0.54	2.45*
Historical	3.34	2.04	1.30	6.64**
External	3.73	2.15	1.58	9.91**
Firsthand	4.61	4.00	0.61	3.21**
Discrete	3.14	2.15	0.99	4.51**
Controllable	3.75	2.93	0.82	3.28**
Equal Access	3.18	3.20	−0.02	−0.06
Noninvasive	4.00	3.53	0.47	3.04*
Faking	3.43	3.80	−0.37	−1.13
Social Desirability	3.45	4.04	−0.58	−2.87*

Note. $N = 11$. The mean of questions with significant criterion relationships was comprised of 4 items: years of customer service experience in a call center, software experience, typing words per minute, and anticipated tenure as a CSR. The mean of questions with non-significant criterion relationships was comprised of 5 items: reasons the candidate wanted to join the call center, years of customer service experience with the general public, candidates' description of what good customer service means to them, shift availability, and overall communication skills rating.

* $p < .05$, ** $p < .01$.

TABLE 20

Customer Service Representative Mean Prescreening Item Score Comparisons between Hired Candidates and Qualified Candidates Not Hired, Research Question 2, Study 1

Screening Question	Not Hired	Hired	Difference	*t*
Years CS Experience in Call Center	3.04	3.07	−0.03	−0.16
Total Software Experience	3.39	3.54	−0.15	−1.20
Databases	0.46	0.46	0.00	0.05
Word Processing	0.96	0.97	−0.01	−0.51
PowerPoint	0.62	0.73	−0.11	−2.27*
Spreadsheets	0.89	0.91	−0.01	−0.48
Visio	0.09	0.08	0.01	0.28
Other	0.36	0.39	−0.02	−0.51
Typing WPM	1.34	1.34	−0.01	−0.17
Anticipated Tenure in Call Center	4.54	4.64	−0.10	−1.28
Reasons Join Call Center	1.70	1.78	−0.08	−1.78
Years of CS with General Public	4.79	4.86	−0.07	−1.28
Describe Good Customer Service	3.22	3.24	−0.02	−0.54
Shift Availability	0.99	0.99	0.00	−0.22
Communication Skills	3.19	3.24	−0.04	−0.99

Note. $N = 665$ for the Not Hired group; $N = 119$ for the Hired group. WPM = Words per Minute. Some values in the Difference column may appear inconsistent due to rounding.

* $p < .05$.

TABLE 21

Means, Standard Deviations, and Intercorrelations of Gas Service Representative Prescreening Scores and Testing Outcomes, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Reasons Join Company	1.92	0.37	—						
2. Good Customer Service	4.73	0.50	—	—					
3. Communication Skills	4.90	0.30	—	.00	—				
4. Total Screening Score	15.24	3.19	.66**	.88**	.75**	—			
5. Mathematical Ability Score	28.95	11.38	−.03	−.03	.09†	.02	—		
6. Spatial Ability Score	23.13	11.50	−.03	.12*	.04	.02	.45**	—	
7. Test Pass/Fail	0.55	0.50	−.05	.10†	.03	.03	.52**	.74**	—

Note. *N*s ranged from 414–794.
† $p < .10$, * $p < .05$, ** $p < .01$.

TABLE 22

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations for Gas Service Representative SME Screening Item Taxonomy Ratings, Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Verifiability	2.99	0.57	(.97)												
2. Job Rel.	3.75	0.58	.28	(.81)											
3. Objectivity	3.05	0.49	.78**	.55†	(.96)										
4. Min. Qual.	3.26	0.46	−.26	.51	.22	(.91)									
5. Historical	2.77	0.44	−.12	−.22	−.20	−.07	(.98)								
6. External	2.89	0.26	.39	−.06	.22	−.15	.50	(.98)							
7. Firsthand	4.15	0.53	−.25	−.32	−.40	−.39	−.27	−.30	(.93)						
8. Discrete	3.03	0.49	.64*	.42	.51	−.16	−.21	−.12	−.13	(.98)					
9. Controllable	3.32	0.37	.21	.28	.30	.53†	.37	.38	−.30	.27	(.87)				
10. Eq. Access	3.40	0.89	.07	−.34	−.25	−.30	.69*	.87**	−.18	−.31	.24	(.56)			
11. Noninvasive	3.76	0.69	.29	−.21	−.11	−.66*	−.34	.09	.44	.47	−.12	.09	(−.02)		
12. Faking	3.18	0.39	−.40	−.24	−.37	.08	.21	−.10	.34	.14	.38	.01	.22	(.97)	
13. Soc. Des.	3.28	0.64	−.28	.26	−.07	.75**	.22	.16	−.37	−.17	.61*	.18	−.47	.34	(.96)

Note. $N = 11$. Rel. = Relatedness; Min. Qual. = Minimum Qualifications; Eq. = Equal; Soc. Des. = Social Desirability.
† $p < .10$, * $p < .05$, ** $p < .01$.

TABLE 23

Gas Service Representative Mean Prescreening Item Score Comparisons between High- and Low-scoring Candidates, Hypothesis 1a, Study 1

Screening Question	Low Score	High Score	Difference	*t*
Reasons Join Company	1.66	2.00	−0.34	−11.54**
Good Customer Service	3.88	0.34	−1.05	−40.70**
Communication Skills	—	4.90	—	—

Note. *N*s for the low score group ranged from 144–175; $N = 619$ for the high score group.

** $p < .01$.

TABLE 24

Gas Service Representative Mean Prescreening Item Score Comparisons among the Highest-scoring Candidates, Hypothesis 1b, Study 1

Screening Question	High Score	Highest Score	Difference	*t*
Reasons Join Company	2.00	2.00	0.00	—[a]
Good Customer Service	4.59	5.00	−0.41	−18.99**
Communication Skills	4.42	5.00	−0.58	−26.70**

Note. $N = 95$ for the high score group; $N = 520$ for the highest score group.

** $p < .01$.

[a] *t* could not be computed because there was no variance in mean item scores.

TABLE 25

Gas Service Representative Comparison of Top Scoring Candidates as a Function of Screening Question Type, Hypothesis 1b, Study 1

Item Classification Dimension	*M* of Differentiating Items	*M* of Non-differentiating Items	Difference	*t*
Verifiability	1.86	1.36	0.50	3.71**
Job Relatedness	3.45	3.27	0.18	0.80
Objectivity	2.00	1.55	0.45	1.66
Minimum Qualifications	2.05	1.73	0.32	1.88
Historical	1.50	1.36	0.14	0.90
External	1.50	1.09	0.41	2.17
Firsthand	2.86	3.64	−0.77	−2.32*
Discrete	1.50	1.55	−0.05	−0.29
Controllable	2.91	3.45	−0.55	−1.75
Equal Access	3.59	3.91	−0.32	−2.28*
Noninvasive	3.82	4.18	−0.36	−1.55
Faking	4.14	4.82	−0.68	−3.75**
Social Desirability	4.18	4.82	−0.64	−2.71*

Note. $N = 11$. The mean of differentiating items was comprised of 2 items: candidates' description of what good customer service means to them and overall communication skills. The mean of non-differentiating items was comprised of 1 item: reasons candidate wants to join the organization.
* $p < .05$, ** $p < .01$.

TABLE 26

Gas Service Representative Comparison of Screening Questions as a Function of Their Relationships to Preemployment Test Scores, Hypothesis 2b, Study 1

Item Classification Dimension	*M* of Question with Significant Criterion Relationship	*M* of Questions with Non-significant Criterion Relationships	Difference	*t*
Verifiability	1.55	1.77	−0.23	−1.17
Job Relatedness	3.36	3.41	−0.05	−0.32
Objectivity	1.73	1.91	−0.18	−1.17
Minimum Qualifications	1.91	1.95	−0.05	−0.29
Historical	1.45	1.45	0.00	0.00
External	1.27	1.41	−0.14	−1.40
Firsthand	3.82	2.77	1.05	3.07*
Discrete	1.36	1.59	−0.23	−2.89*
Controllable	3.00	3.14	−0.14	−0.42
Equal Access	3.82	3.64	0.18	0.67
Noninvasive	4.09	3.86	0.23	0.96
Faking	4.64	4.23	0.41	1.63
Social Desirability	4.82	4.18	0.64	2.71*

Note. $N = 11$. The mean of items with significant criterion relationships was comprised of 1 item: candidates' description of what good customer service means to them. The mean of items with non-significant criterion relationships was comprised of 2 items: reasons candidate wants to join the organization and overall communication skills.
* $p < .05$.

TABLE 27

Summary of Study 1 Findings by Hypothesis/Research Question by Job Type

	Job		
Hypothesis/Research Question	UW	CSR	GSR
H1a: Prescreening items that differentiate high- from low-scoring candidates will have significantly higher job-relatedness and minimum qualifications ratings.	~	N/A	N/A
H1b: Prescreening items that differentiate between the highest-scoring and high-scoring candidates will have significantly higher objectivity ratings, but will not differ in terms of minimum qualifications.	~	~	~
H2a: Applicant prescreening scores will be predictive of key new hire performance criteria (e.g., test scores, training performance).	X	X	X
H2b: Prescreening items predictive of key new hire performance criteria will be high on objectivity and verifiability.	✓	✓	X
RQ1: Are there meaningful distinctions between applicants that can be made at points in the applicant prescreening score distribution where there are sharp increases or decreases in the number of applicants receiving a particular score? What is the nature of these distinctions?	—	—	—
RQ2: Are there response profiles that differentiate between applicants who are hired versus not hired, and applicants who turn over versus those who do not? What are the characteristics of these response profiles?	X	X	N/A

Note. UW = Utility Worker; CSR = Customer Service Representative; GSR = Gas Service Representative.
✓= Fully-supported hypothesis; ~ = partially supported hypothesis; X = hypothesis not supported; N/A = hypothesis not testable; — = research question not tested.

TABLE 28

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations of Overall and Dimensional Résumé Ratings across All Résumés, Study 2

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Objective	1.61	0.30	(.91)											
2. Skills	1.74	0.25	.67*	(.80)										
3. Employment Status	1.91	0.37	.44	.56†	(.86)									
4. Work Experience	1.77	0.30	.65*	.81**	.63*	(.86)								
5. Advancement	1.63	0.32	.31	.08	.55†	.27	(.84)							
6. Education/Comp. Exp.	1.89	0.18	.36	.67*	.71**	.74**	.29	(.93)						
7. Continuing Education	1.54	0.30	.40	.29	.56†	.38	.62*	.46	(.95)					
8. Volunteerism	1.48	0.34	.35	−.03	.35	.18	.67*	.22	.87**	(.94)				
9. Format	1.80	0.35	.39	.81**	.72**	.69*	.25	.84**	.52†	.20	(.89)			
10. Overall Structured	1.67	0.29	.63*	.73**	.49	.92**	.31	.67*	.39	.29	.60*	(.88)		
11. Composite Structured	1.70	0.22	.69*	.71**	.84**	.79**	.64*	.77**	.79**	.60*	.80**	.74**	(.95)	
12. Global	1.73	0.35	.46	.78**	.39	.76**	.14	.44	.31	−.05	.55†	.67*	.55†	(.91)

Note. $N = 12$. Comp. Exp. = Compensatory Experience
†$p < .10$, *$p < .05$, **$p < .01$.

TABLE 29

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations of Overall and Dimensional Résumé Ratings, "Pass" Résumés Only, Study 2

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Objective	1.75	0.37	(.93)											
2. Skills	1.80	0.37	.43	(.55)										
3. Employment Status	2.01	0.38	.56†	.32	(.86)[a]									
4. Work Experience	2.04	0.31	.63*	.44	.50†	(.78)								
5. Advancement	1.77	0.36	.43	.10	.53†	.60*	(.81)							
6. Education/Comp. Exp.	2.28	0.29	.25	.25	.86**	.35	.46	(.72)						
7. Continuing Education	1.74	0.32	.43	.38	.50	.46	.76**	.39	(.97)					
8. Volunteerism	1.30	0.39	.42	.26	.24	.52†	.69*	.14	.87**	(.95)				
9. Format	1.74	0.38	.28	.48	.69*	.49	.49	.79**	.61*	.32	(.91)			
10. Overall Structured	1.88	0.40	.58*	.58*	.57†	.95**	.56†	.44	.45	.44	.59*	(.78)		
11. Composite Structured	1.82	0.26	.68*	.57†	.79**	.76**	.77**	.68*	.83**	.69*	.78**	.79**	(.87)	
12. Global	2.10	0.37	.16	.53†	.31	.52†	.28	.22	.30	.05	.56†	.56†	.45	(.89)

Note. $N = 12$. *N*s range from 7–12 for the intraclass correlation estimates because of a lack of variance in some SME ratings for some dimensions. Comp. Exp. = Compensatory Experience.
†$p < .10$, *$p < .05$, **$p < .01$.

TABLE 30

Means, Standard Deviations, Intercorrelations, and Intraclass Correlations of Overall and Dimensional Résumé Ratings, "Fail" Résumés Only, Study 2

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Objective	1.54	0.29	(.88)											
2. Skills	1.71	0.27	.70*	(.86)										
3. Employment Status	1.85	0.38	.40	.58*	(.86)									
4. Work Experience	1.62	0.38	.59*	.71**	.60*	(.88)								
5. Advancement	1.56	0.32	.22	−.01	.54†	.15	(.85)							
6. Education/Comp. Exp.	1.68	0.18	.60*	.54†	.46	.69*	.13	(.93)						
7. Continuing Education	1.44	0.30	.37	.11	.55†	.29	.49	.45	(.93)					
8. Volunteerism	1.57	0.31	.25	−.18	.37	−.03	.59*	.17	.84**	(.95)				
9. Format	1.83	0.36	.41	.75**	.62*	.67*	.08	.55†	.49	.14	(.90)			
10. Overall Structured	1.55	0.28	.65*	.60*	.40	.89**	.07	.75**	.34	.10	.56†	(.90)		
11. Composite Structured	1.64	0.22	.71*	.67*	.84**	.75**	.53†	.70*	.74**	.51†	.77**	.68*	(.96)	
12. Global	1.53	0.40	.64*	.67*	.42	.78**	.11	.60*	.33	−.10	.55†	.73**	.64*	(.87)

Note. $N = 12$. *N*s range from 11–12 for the intraclass correlation estimates because of a lack of variance in some SME ratings for some dimensions. Comp. Exp. = Compensatory Experience.
†$p < .10$, *$p < .05$, **$p < .01$.

TABLE 31

Comparison of Intraclass Correlations (ICCs) across Résumé Evaluation Conditions, Hypothesis 3, Study 2

Target Condition or Dimension	ICC Average Measures	95% Confidence Interval	ICC Single Measures	95% Confidence Interval
Overall Ratings	—	—	—	—
Unstructured Evaluation	.91	.84–.96	.46	.30–.66
Structured (Global) Evaluation	.88	.79–.95	.39	.24–.60
Structured Composite Evaluation	.95	.90–.98	.60	.44–.77
Dimensional Ratings	—	—	—	—
Objective	.90	.83–.96	.44	.29–.65
Skills	.80	.64–.91	.25	.13–.45
Employment Status	.86	.74–.93	.33	.19–.54
Work Experience	.86	.76–.94	.35	.21–.56
Advancement	.84	.71–.93	.30	.17–.51
Education/Compensatory Exp.	.93	.88–.97	.53	.37–.72
Continuing Education	.95	.91–.98	.62	.47–.79
Volunteerism	.94	.90–.97	.58	.43–.76
Format	.89	.81–.95	.41	.26–.62

Note. $N = 12$. The ICC Average Measures value reflects the level of reliability associated with 12 raters. The ICC Single Measures value reflects the level of reliability associated with 1 rater. Structured (Global) Evaluation was an overall résumé rating, similar to the Unstructured Evaluation condition. Structured Composite Evaluation was the unit-weighted composite score based on the dimensional ratings. Exp. = Experience.

FIGURE 1

Overview of the selection process for professional and entry-level/above entry-level jobs.

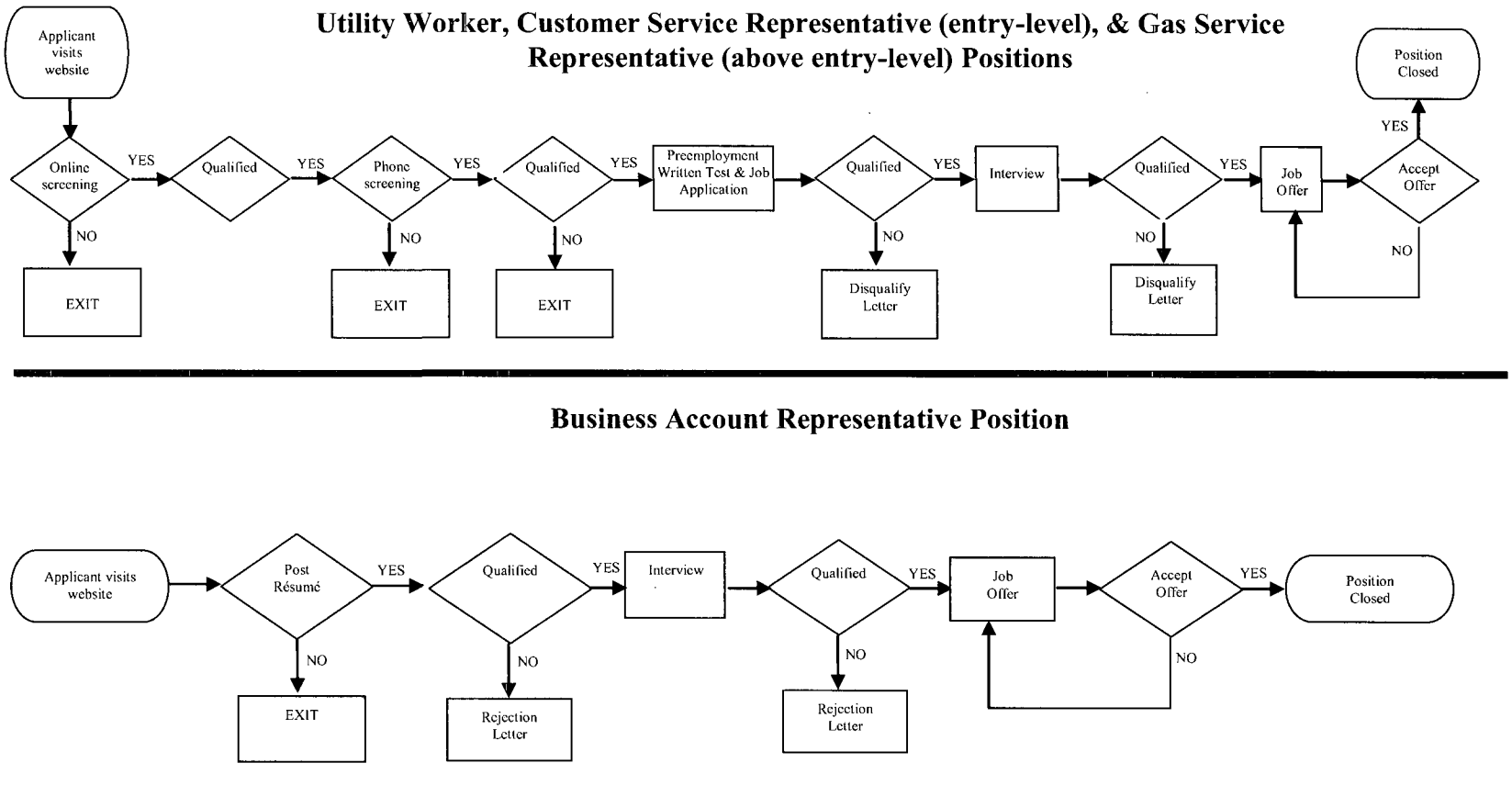
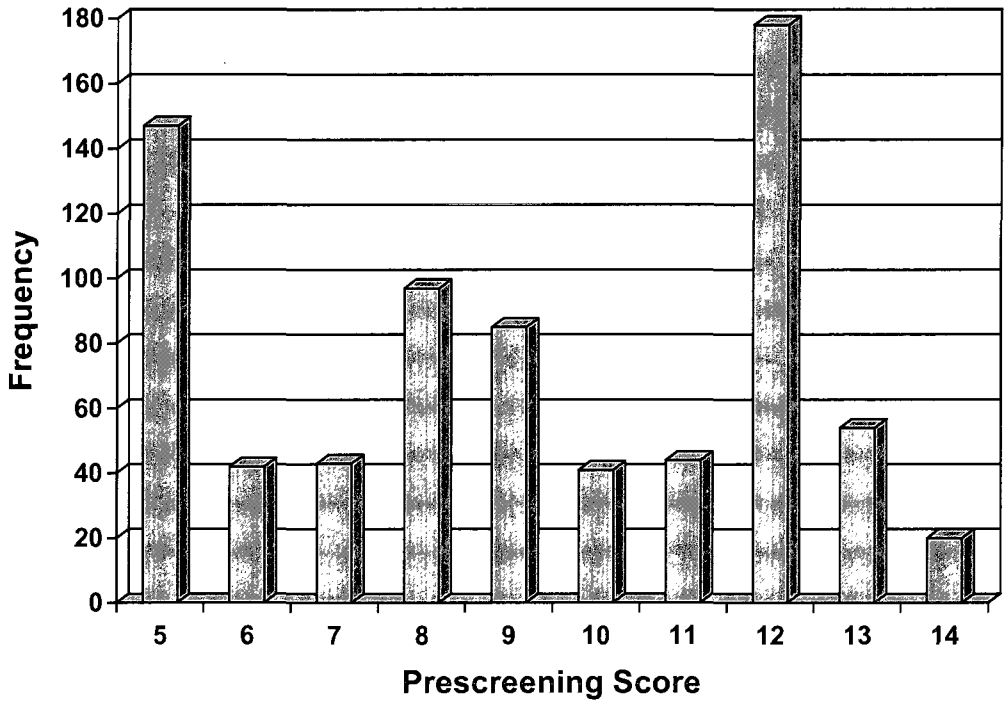


FIGURE 2

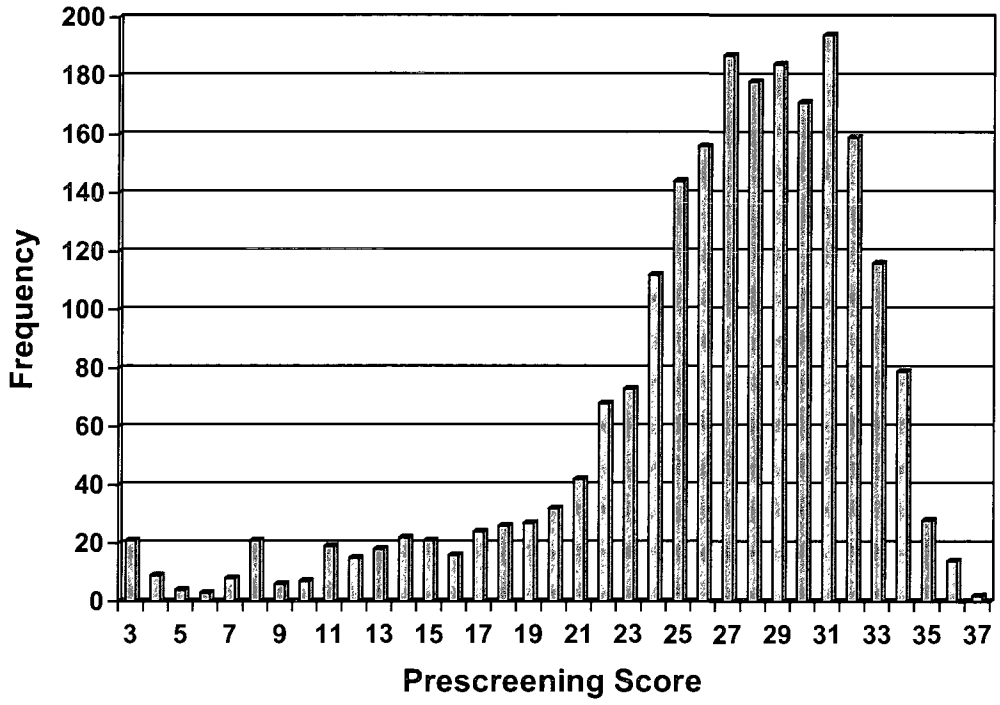
Utility Worker screening score distribution, Study 1



Note. $N = 751$.

FIGURE 3

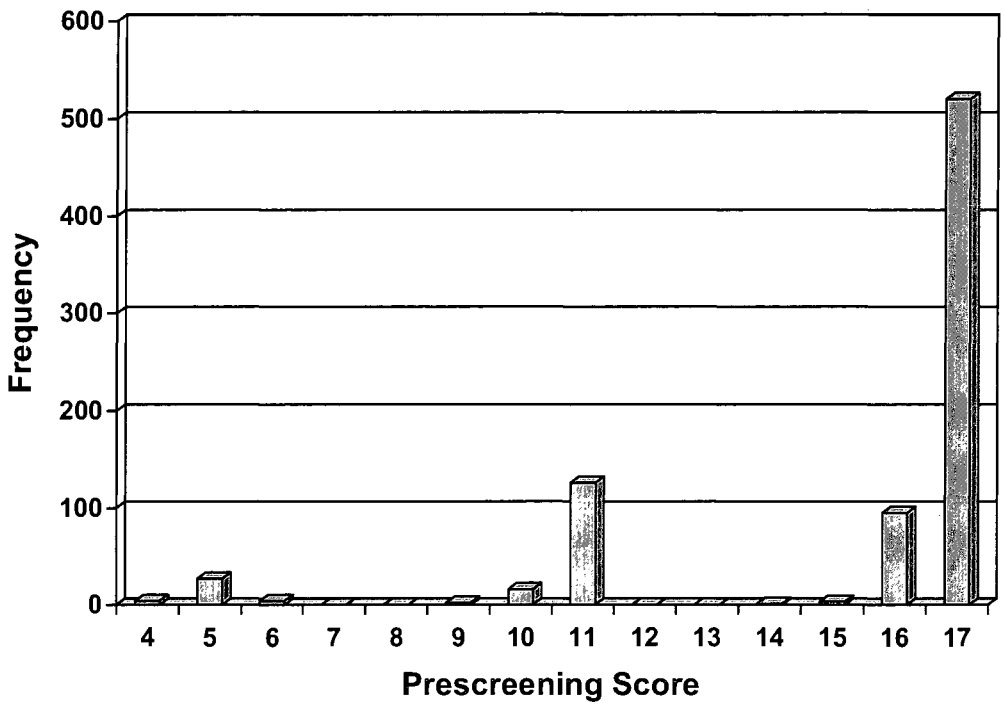
Customer Service Representative screening score distribution, Study 1



Note. $N = 2{,}206$.

FIGURE 4

Gas Service Representative screening score distribution, Study 1



Note. $N = 798$.

REFERENCES

Allen, D. G., Van Scotter, J. R., & Otondo, R. F. (2004). Recruitment communication media: Impact on prehire outcomes. *Personnel Psychology, 57,* 143–171.

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,* 121–136.

Ash, R. A., & Levine, E. L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology, 70,* 572–576.

Barber, A. E. (1998). *Recruiting employees.* Thousand Oaks, CA: Sage Publications.

Bass, B. M., & Barrett, G. V. (1981). *People, work, and organizations* (2nd ed.). Boston: Allyn & Bacon.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology, 54,* 387–419.

Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment, 12,* 135–148.

Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts.* Chichester, UK: Wiley.

Beach, L. R., & Potter, R. E. (1992). The pre-choice screening of options. *Acta Psychologica, 81,* 115–126.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45,* 389–406.

Breaugh, J. A., & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management, 26,* 405–434.

Brown, B. K., & Campion, M. A. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in résumé screening. *Journal of Applied Psychology, 79,* 897–908.

Buel, W. D. (1964). Voluntary female clerical turnover: The concurrent and predictive validity of a weighted application blank. *Journal of Applied Psychology, 48,* 180–182.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82,* 546–561.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50,* 655–702.

Carlson, K. D., Connerley, M. L., & Mecham, R. L., III. (2002). Recruitment evaluation: The case for assessing the quality of applicants attracted. *Personnel Psychology, 55,* 461–490.

Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology, 60,* 767–769.

Chan, D., & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment, 12,* 9–23.

Cober, R. T., Brown, D. J., Blumental, A. J., Doverspike, D., & Levy, P. (2000). The quest for the qualified surfer: It's time the public sector catches the wave. *Public Personnel Management, 29,* 479–494.

Crow, D., & DeSanto, J. (2004). A hybrid approach to concept extraction and recognition-based matching in the domain of human resources. Paper presented at the 16th annual IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Boca Raton, FL.

Davis, C. E., & Davis, E. B. (1996). Information load and consistency of decisions. *Psychological Reports, 79,* 279–288.

DeSanto, J., & Crow, D. (2004). Utilizing a domain specific knowledge base to improve résumé matching in human resources. Paper presented at the 16th annual IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Boca Raton, FL.

Dineen, B. R., Ash, S. R., & Noe, R. A. (2002). A web of applicant attraction: Person-organization fit in the context of web-based recruiting. *Journal of Applied Psychology, 87,* 723–734.

Fritzsche, B. A., & Brannick, M. T. (2002). The importance of representative design in judgment tasks: The case of résumé screening. *Journal of Occupational and Organizational Psychology, 75,* 163–169.

Gatewood, R. D., Feild, H. S., & Barrick, M. (2007). *Human resource selection* (6th ed.). Fort Worth, TX: The Dryden Press.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18,* 694–734.

Goldstein, I. L. (1971). The application blank: How honest are the responses. *Journal of Applied Psychology, 55,* 491–492.

Graves, L. M., & Karren, R. J. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology, 45,* 313–340.

Handler, C. (2002, March 22). Screening vs. résumés: What will be the weakest link? Retrieved November 10, 2003, from http://www.erexchange.com/Articles/default.asp?CID={02893B2B-9038-4C05-9288-F7DBA2862FFB}.

Harris, M. (2004, June 21). OFCCP, EEOC, and internet applicants: An update on the Proposed OFCCP Definition of Internet Applicant. Retrieved March 5, 2008, from http://www.ere.net/2004/06/21/ofccp-eeoc-and-internet-applicants/.

Harris, M., & Dewar, K. (2000). Understanding and using Web-based recruiting and screening tools: Key criteria, current trends, and future directions. Workshop presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Harris, M. M., Van Hoye, G., & Lievens, F. (2003). Privacy and attitudes towards internet-based selection systems: A cross-cultural comparison. *International Journal of Selection and Assessment, 11,* 230–236.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57,* 639–683.

Helgeson, J. G., & Ursic, M. L. (1993). Information load, cost/benefit assessment and decision strategy variability. *Journal of the Academy of Marketing Science, 21,* 13–20.

Hinrichs, J. R., Haanperä, S., & Sonkin, L. (1976). Validity of a biographical information blank across national boundaries. *Personnel Psychology, 29,* 417–421.

Hitt, M. A., & Barr, S. H. (1989). Managerial selection decision models: Examination of configural cue processing. *Journal of Applied Psychology, 74,* 53–61.

Holland, J. L. (1985). *Making vocational choices: A theory of careers* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hough, L. M. (1984). Development and evaluation of the "Accomplishment Record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69,* 135–146.

Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three "alternative" selection procedures. *Personnel Psychology, 36,* 261–276.

Hunt, S. T., Gibby, R. E., Hemmingway, M. A., Irwin, J., Scarborough, D. J., & Truxillo, D. M. (April 2004). Internet prescreening: Does it lead to better hiring decisions? Panel discussion presented at the 19th Annual Society for Industrial and Organizational Psychology conference, Chicago, IL.

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York: John Wiley & Sons.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49,* 1–49.

Kristof-Brown, A. L. (2000). Perceived applicant fit: Distinguishing between recruiters' perceptions of person-job and person-organization fit. *Personnel Psychology, 53,* 643–671.

Kluger, A. N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology, 46,* 763–780.

Kuhnert, K. W., & Russell, C. J. (1990). Using constructive developmental theory and biodata to bridge the gap between personnel selection and leadership. *Journal of Management, 16,* 595–607.

Lee, R., & Booth, J. M. (1974). A utility analysis of a weighted application blank designed to predict turnover for clerical employees. *Journal of Applied Psychology, 59,* 516–518.

Lefkowitz, J., Gebbia, M. I., Balsam, T., & Dunn, L. (1999). Dimensions of biodata items and their relationships to item validity. *Journal of Occupational and Organizational Psychology, 72,* 331–350.

Levine, E. L., & Flory, A., III. (1975). Evaluation of job applications—A conceptual framework. *Public Personnel Management, 4,* 378–385.

Lievens, F., & Harris, M. M. (in press). Research on internet recruitment and testing: Current status and future directions. *The International Review of Industrial and Organizational Psychology*, Chichester, England: John Wiley.

Lievens, F., van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review, 31,* 580–601.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44,* 763–792.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41,* 283–314.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812–821.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46.

McManus, M. A., & Masztal, J. J. (1999). The impact of biodata item attributes on validity and socially desirable responding. *Journal of Business and Psychology, 13,* 437–446.

Mitchell, T. W., & Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology, 67,* 411–418.

Mosel, J. N., & Cozan, L. W. (1952). The accuracy of application blank work histories. *Journal of Applied Psychology, 36,* 365–369.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53,* 299–323.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., III. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88,* 348–355.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11,* 1–31.

Nowicki, M. D., & Rosse, J. G. (2002). Managers' views of how to hire: Building bridges between science and practice. *Journal of Business and Psychology, 17,* 157–170.

Pace, L. A., & Schoenfeldt, L. F. (1977). Legal concerns in the use of weighted applications. *Personnel Psychology, 30,* 159–166.

Pannone, R. D., (1984). Predicting test performance: A content valid approach to screening applicants. *Personnel Psychology, 37,* 507–514.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. L. Shaver, L. Wrightsman, & F. M. Andrews (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56,* 733–752.

Psychological Services, Inc. (2008). Basic Skills Tests. Retrieved August 24, 2008, from http://corporate.psionline.com/pdfs/BST_jatc.pdf.

Psychological Services, Inc. (2008). Employee Aptitude Survey. Retrieved August 24, 2008, from http://www.ppicentral.com/Pdf/Employee_Aptitude_Survey.pdf.

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

Quiñones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48,* 887–910.

Roach, D. E. (1971). Double cross-validation of a weighted application blank over time. *Journal of Applied Psychology, 55,* 157–160.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1–62.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75,* 175–184.

Russell, C. J. (1990). Selecting top corporate leaders: An example of biographical information. *Journal of Management, 16,* 73–86.

Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology, 75,* 569–580.

Rynes, S. L. (1991). Recruitment, job choice, and post-hire consequences. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 399–444). Palo Alto, CA: Consulting Psychologists Press.

Rynes, S. L., Bretz, R. D., & Gerhart, B. (1991). The importance of recruitment in job choice: A different way of looking. *Personnel Psychology, 44,* 487–521.

Rynes, S., & Gerhart, B. (1990). Interviewer assessments of applicant "fit": An exploratory investigation. *Personnel Psychology, 43,* 13–35.

Schaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. *Personnel Psychology, 39,* 791–809.

Schaubroeck, J., Ganster, D. C., & Jones, J. R. (1998). Organization and occupation influences in the attraction-selection-attrition process. *Journal of Applied Psychology, 83,* 869–891.

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., & Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology, 55,* 569–587.

Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88,* 979–988.

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437–453.

Schneider, B., Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology, 48,* 747–773.

Schrader, A. D., & Osburn, H. G. (1977). Biodata faking: Effects of induced subtlety and position specificity. *Personnel Psychology, 30,* 395–404.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personnel Psychology, 51,* 321–355.

UCLA Department of Statistics (2002, May 22). Power calculator. Retrieved June 13, 2004, from http://calculators.stat.ucla.edu/powercalc.

Weiss, D. J., & Dawis, R. V. (1960). An objective validation of factual interview data. *Journal of Applied Psychology, 44,* 381–385.

Wernimont, P. F. (1962). Re-evaluation of a weighted application blank for office personnel. *Journal of Applied Psychology, 46,* 417–419.

Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82,* 900–912.

Woehr, D. J., & Arthur, Jr., W. A. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29,* 231–258.

APPENDIX A: PRESCREENING COVER LETTER AND SURVEYS[1]

Dear Participant:

My name is Matthew Paronto. I am a graduate student in the Psychology Department at Portland State University conducting my doctoral dissertation study on preliminary job applicant evaluation procedures such as prescreening and résumés. You may not receive any direct benefit from participating in this study. However, this research may help others in the future, as the information you provide may help organizations set best practices regarding the use of these preliminary applicant evaluation methods.

You are being asked to take part so that we can start to better understand how organizations can best leverage the use of online applicant selection procedures to select job applicants. This is extremely important for organizations both from a legal as well as from a financial standpoint, as hiring employees who are not a good fit with the organization due to poor screening procedures results in delays in getting the human capital needed for an organization to be successful.

As part of this study, I am interested in leveraging your knowledge of employee selection to categorize and analyze the content of applicant prescreening. If you decide to participate, you will be presented with prescreening questions typically used to screen applicants for a given job and will be asked to evaluate them along a taxonomy of question types. To assure the anonymity of your responses, please do not place your name anywhere on the survey materials. These surveys may take up to 2 hours to complete, so it is advisable to complete the surveys over the course of a couple of days to avoid fatigue.

If for any reason you feel uncomfortable, you may stop participating at any time. Also, you may skip any questions on the questionnaire to which you feel uncomfortable responding. However, I assure you that your responses will be kept confidential, and there will be no way in which to match responses to the individuals who provide them, since your name will not be on the survey. Also, only group data will be reported, and surveys will be securely stored and will not be seen by individuals not affiliated with this research.

Participation is entirely voluntary. Your decision to participate or not will not affect your relationship with the researcher or with Portland State University in any way. Please keep a copy of this letter for your records.

If you have any concerns or questions about your participation in this study or your rights as a research subject, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, 503.725.4288. If you have any questions about the study itself, contact me at 503.381.2338 or Dr. Donald Truxillo, of the Psychology Department at 503.725.3969.

Sincerely,

Matthew E. Paronto
Department of Psychology
Portland State University

---

[1] Prescreening questions have been altered and response options removed due to the proprietary nature of the prescreening questions.

Prescreening Item Taxonomy Survey
ENTRY-LEVEL PHYSICAL JOB

Please carefully read the description of an entry-level physical job below as it appears to a job applicant. After reading the job description, you will be presented with a series of prescreening question classification definitions and the questions applicants must answer when they are screening for this position. Beginning with Dimension 1, first review the definition of the dimension, and then read and rate each of the prescreening questions that candidates are required to answer based on the extent to which each individual prescreening question reflects the classification dimension. Please rate all the screening questions on a dimension before proceeding to the next dimension.

Job Description/Posting: UTILITY WORKER

Responsibilities: Assists journey level gas and electric department employees. Digs trenches (using shovels, picks, or pneumatic tools); raises and lowers tools and equipment up to 100 pounds utilizing ropes, cables, rollers, slings, and blocks; transports vehicles and equipment; and flags traffic. Travel within the service territory may be required.

Requirements: High school diploma or equivalent, a valid [State's] driver's license and successful completion of a pre-employment test, drug analysis, and background check. Construction experience is desired. Some overtime required. We seek individuals with a strong customer focus, safety consciousness, a team orientation, and high work standards. A qualifying score on the pre-employment [written test] is required.

For each prescreening question classification dimension below, please read the definition of the question dimension to be assessed. For each question, please indicate the extent to which the question reflects the dimension, using the scale below. Note: please rate all items on one dimension before proceeding to the next dimension.

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 1: Verifiability

Verifiable questions tap aspects of applicant experience that can be externally confirmed by an independent source. Please indicate the extent to which the following questions are verifiable.

1.	At least 18 years of age	1	2	3	4	5
2.	High school diploma or GED may be required	1	2	3	4	5
3.	Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
4.	Possession of a Class A driver's license	1	2	3	4	5
5.	Ability to work in various weather conditions may be required	1	2	3	4	5
6.	Number of years of construction experience	1	2	3	4	5
7.	Previous experience working for a utility company	1	2	3	4	5
8.	Utility company work experience (which type of utility)	1	2	3	4	5
9.	Ability to meet overtime requirement in emergency situations	1	2	3	4	5
10.	Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 2: Job-Relatedness

Job-related questions assess applicant experience and knowledge, skills, or abilities that are considered essential to successful performance on the job. Please indicate the extent to which the following questions are job-related.

11.	At least 18 years of age	1	2	3	4	5
12.	High school diploma or GED may be required	1	2	3	4	5
13.	Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
14.	Possession of a Class A driver's license	1	2	3	4	5
15.	Ability to work in various weather conditions may be required	1	2	3	4	5
16.	Number of years of construction experience	1	2	3	4	5
17.	Previous experience working for a utility company	1	2	3	4	5
18.	Utility company work experience (which type of utility)	1	2	3	4	5
19.	Ability to meet overtime requirement in emergency situations	1	2	3	4	5
20.	Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 3: Objectivity
Objective questions ask applicants to recall factual information about their past experiences. Objective events would be those that could be corroborated by an external observer. Please indicate the extent to which the following questions are objective in nature.

21. At least 18 years of age	1	2	3	4	5
22. High school diploma or GED may be required	1	2	3	4	5
23. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
24. Possession of a Class A driver's license	1	2	3	4	5
25. Ability to work in various weather conditions may be required	1	2	3	4	5
26. Number of years of construction experience	1	2	3	4	5
27. Previous experience working for a utility company	1	2	3	4	5
28. Utility company work experience (which type of utility)	1	2	3	4	5
29. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
30. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 4: Minimum Qualifications/Knockout
Minimum qualifications questions assess skills, abilities, experience, or other attributes an applicant must possess in order to be considered further for a position; otherwise, the candidate is no longer considered. Please indicate the extent to which the following questions assess minimum qualifications.

31. At least 18 years of age	1	2	3	4	5
32. High school diploma or GED may be required	1	2	3	4	5
33. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
34. Possession of a Class A driver's license	1	2	3	4	5
35. Ability to work in various weather conditions may be required	1	2	3	4	5
36. Number of years of construction experience	1	2	3	4	5
37. Previous experience working for a utility company	1	2	3	4	5
38. Utility company work experience (which type of utility)	1	2	3	4	5
39. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
40. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 5: Historical
Historical questions assess behaviors that have taken place or continue to take place. Historical questions do not assess behavioral intentions. Please indicate the extent to which the following questions are historical.

41. At least 18 years of age	1	2	3	4	5
42. High school diploma or GED may be required	1	2	3	4	5
43. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
44. Possession of a Class A driver's license	1	2	3	4	5
45. Ability to work in various weather conditions may be required	1	2	3	4	5
46. Number of years of construction experience	1	2	3	4	5
47. Previous experience working for a utility company	1	2	3	4	5
48. Utility company work experience (which type of utility)	1	2	3	4	5
49. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
50. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 6: External
External questions assess past behavior and experiences that occur in specific, real-life situations. External questions do not assess unexpressed thoughts, attitudes, opinions, or reactions to events. Please indicate the extent to which the following questions are external.

51. At least 18 years of age	1	2	3	4	5
52. High school diploma or GED may be required	1	2	3	4	5
53. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
54. Possession of a Class A driver's license	1	2	3	4	5
55. Ability to work in various weather conditions may be required	1	2	3	4	5
56. Number of years of construction experience	1	2	3	4	5
57. Previous experience working for a utility company	1	2	3	4	5
58. Utility company work experience (which type of utility)	1	2	3	4	5
59. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
60. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 7: Firsthand

Firsthand questions assess aspects of the applicant that he or she would be able to evaluate based on his or her direct, objective knowledge. Firsthand questions do not assess an applicant's estimation of how others would evaluate them. Please indicate the extent to which the following questions are firsthand.

61. At least 18 years of age	1	2	3	4	5
62. High school diploma or GED may be required	1	2	3	4	5
63. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
64. Possession of a Class A driver's license	1	2	3	4	5
65. Ability to work in various weather conditions may be required	1	2	3	4	5
66. Number of years of construction experience	1	2	3	4	5
67. Previous experience working for a utility company	1	2	3	4	5
68. Utility company work experience (which type of utility)	1	2	3	4	5
69. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
70. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 8: Discrete

Discrete questions assess a single, unique behavior or simple count of unique instances of events, whereas summative questions require the summation or estimation of behavior over time. Please indicate the extent to which the following questions are discrete.

71. At least 18 years of age	1	2	3	4	5
72. High school diploma or GED may be required	1	2	3	4	5
73. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
74. Possession of a Class A driver's license	1	2	3	4	5
75. Ability to work in various weather conditions may be required	1	2	3	4	5
76. Number of years of construction experience	1	2	3	4	5
77. Previous experience working for a utility company	1	2	3	4	5
78. Utility company work experience (which type of utility)	1	2	3	4	5
79. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
80. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 9: Controllable

Controllable questions assess behaviors or actions that applicants chose to perform or not perform out of their own volition. Please indicate the extent to which the following questions are controllable.

81. At least 18 years of age	1	2	3	4	5
82. High school diploma or GED may be required	1	2	3	4	5
83. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
84. Possession of a Class A driver's license	1	2	3	4	5
85. Ability to work in various weather conditions may be required	1	2	3	4	5
86. Number of years of construction experience	1	2	3	4	5
87. Previous experience working for a utility company	1	2	3	4	5
88. Utility company work experience (which type of utility)	1	2	3	4	5
89. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
90. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 10: Equally Accessible

Equally accessible questions assess past behavior and experiences that all applicants had an equal opportunity to engage in. Please indicate the extent to which the following questions are equally accessible.

91. At least 18 years of age	1	2	3	4	5
92. High school diploma or GED may be required	1	2	3	4	5
93. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
94. Possession of a Class A driver's license	1	2	3	4	5
95. Ability to work in various weather conditions may be required	1	2	3	4	5
96. Number of years of construction experience	1	2	3	4	5
97. Previous experience working for a utility company	1	2	3	4	5
98. Utility company work experience (which type of utility)	1	2	3	4	5
99. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
100. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 11: Noninvasive

Noninvasive questions assess aspects of candidates that are non-personal in nature. Noninvasive questions would not be considered an invasion of privacy from the applicant's perspective. Please indicate the extent to which the following questions are noninvasive.

101. At least 18 years of age	1	2	3	4	5
102. High school diploma or GED may be required	1	2	3	4	5
103. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
104. Possession of a Class A driver's license	1	2	3	4	5
105. Ability to work in various weather conditions may be required	1	2	3	4	5
106. Number of years of construction experience	1	2	3	4	5
107. Previous experience working for a utility company	1	2	3	4	5
108. Utility company work experience (which type of utility)	1	2	3	4	5
109. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
110. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

Dimension 12: Susceptibility to Faking

Questions that are susceptible to faking are questions to which candidates can decipher the "correct" or "best" answer, allowing them to consciously distort their responses to make themselves look better, or increase their chances of passing the screening. Please indicate the extent to which the following questions are susceptible to faking.

111. At least 18 years of age	1	2	3	4	5
112. High school diploma or GED may be required	1	2	3	4	5
113. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
114. Possession of a Class A driver's license	1	2	3	4	5
115. Ability to work in various weather conditions may be required	1	2	3	4	5
116. Number of years of construction experience	1	2	3	4	5
117. Previous experience working for a utility company	1	2	3	4	5
118. Utility company work experience (which type of utility)	1	2	3	4	5
119. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
120. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 13: Socially Desirable

Socially desirable questions may lead applicants to unconsciously alter their true response because they want to be viewed favorably by others. Please indicate the extent to which the following questions might elicit a socially desirable response.

121. At least 18 years of age	1	2	3	4	5
122. High school diploma or GED may be required	1	2	3	4	5
123. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
124. Possession of a Class A driver's license	1	2	3	4	5
125. Ability to work in various weather conditions may be required	1	2	3	4	5
126. Number of years of construction experience	1	2	3	4	5
127. Previous experience working for a utility company	1	2	3	4	5
128. Utility company work experience (which type of utility)	1	2	3	4	5
129. Ability to meet overtime requirement in emergency situations	1	2	3	4	5
130. Ability and willingness to travel and transfer for Title 300 construction work	1	2	3	4	5

THANK YOU FOR YOUR PARTICIPATION!!!

Prescreening Item Taxonomy Survey
ABOVE ENTRY-LEVEL PHYSICAL JOB

Please carefully read the description of an above entry-level physical job below as it appears to a job applicant. After reading the job description, you will be presented with a series of prescreening question classification definitions and the questions applicants must answer when they are screening for this position. Beginning with Dimension 1, first review the definition of the dimension, and then read and rate each of the prescreening questions that candidates are required to answer based on the extent to which each individual prescreening question reflects the classification dimension. Please rate all the screening questions on a dimension before proceeding to the next dimension.

Job Description/Posting: GAS SERVICE REPRESENTATIVE

Responsibilities: Duties include interacting with residential and business customers while troubleshooting and re-lighting various gas appliances, investigating reported gas leaks, and responding to emergencies such as fires. Gas Service Representatives turn on and turn off gas and electric meters, install and change meters, make repairs, and build gas meter sets using tools such as pipe wrenches. The work environment may include working in confined spaces. Shift work and weekends may be required.

Requirements: MUST RESIDE WITHIN 30 MINUTES OF JOB LOCATION. Must possess excellent safety and customer focus skills, as well as mechanical skills (construction, plumbing, or basic electrical), with knowledge of proper use of tools and materials. Must possess [State] Class C Driver's License. Must pass [Company's written pre-employment] tests.

For each prescreening question classification dimension below, please read the definition of the question dimension to be assessed. For each question, please indicate the extent to which the question reflects the dimension, using the scale below. Note: please rate all items on one dimension before proceeding to the next dimension.

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 1: Verifiability

Verifiable questions tap aspects of applicant experience that can be externally confirmed by an independent source. Please indicate the extent to which the following questions are verifiable.

1. At least 18 years of age	1	2	3	4	5
2. High school diploma or GED may be required	1	2	3	4	5
3. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
4. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
5. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
6. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
7. Describe what good customer service means to [candidate]	1	2	3	4	5
8. Rating of overall communication skills	1	2	3	4	5

Dimension 2: Job-Relatedness

Job-related questions assess applicant experience and knowledge, skills, or abilities that are considered essential to successful performance on the job. Please indicate the extent to which the following questions are job-related.

9. At least 18 years of age	1	2	3	4	5
10. High school diploma or GED may be required	1	2	3	4	5
11. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
12. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
13. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
14. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
15. Describe what good customer service means to [candidate]	1	2	3	4	5
16. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 3: Objectivity

Objective questions ask applicants to recall factual information about their past experiences. Objective events would be those that could be corroborated by an external observer. Please indicate the extent to which the following questions are objective in nature.

17. At least 18 years of age	1	2	3	4	5
18. High school diploma or GED may be required	1	2	3	4	5
19. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
20. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
21. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
22. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
23. Describe what good customer service means to [candidate]	1	2	3	4	5
24. Rating of overall communication skills	1	2	3	4	5

Dimension 4: Minimum Qualifications/Knockout

Minimum qualifications questions assess skills, abilities, experience, or other attributes an applicant must possess in order to be considered further for a position; otherwise, the candidate is no longer considered. Please indicate the extent to which the following questions assess minimum qualifications.

25. At least 18 years of age	1	2	3	4	5
26. High school diploma or GED may be required	1	2	3	4	5
27. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
28. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
29. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
30. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
31. Describe what good customer service means to [candidate]	1	2	3	4	5
32. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 5: Historical

Historical questions assess behaviors that have taken place or continue to take place. Historical questions do not assess behavioral intentions. Please indicate the extent to which the following questions are historical.

33. At least 18 years of age	1	2	3	4	5
34. High school diploma or GED may be required	1	2	3	4	5
35. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
36. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
37. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
38. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
39. Describe what good customer service means to [candidate]	1	2	3	4	5
40. Rating of overall communication skills	1	2	3	4	5

Dimension 6: External

External questions assess past behavior and experiences that occur in specific, real-life situations. External questions do not assess unexpressed thoughts, attitudes, opinions, or reactions to events. Please indicate the extent to which the following questions are external.

41. At least 18 years of age	1	2	3	4	5
42. High school diploma or GED may be required	1	2	3	4	5
43. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
44. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
45. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
46. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
47. Describe what good customer service means to [candidate]	1	2	3	4	5
48. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 7: Firsthand

Firsthand questions assess aspects of the applicant that he or she would be able to evaluate based on his or her direct, objective knowledge. Firsthand questions do not assess an applicant's estimation of how others would evaluate them. Please indicate the extent to which the following questions are firsthand.

49. At least 18 years of age	1	2	3	4	5
50. High school diploma or GED may be required	1	2	3	4	5
51. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
52. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
53. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
54. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
55. Describe what good customer service means to [candidate]	1	2	3	4	5
56. Rating of overall communication skills	1	2	3	4	5

Dimension 8: Discrete

Discrete questions assess a single, unique behavior or simple count of unique instances of events, whereas summative questions require the summation or estimation of behavior over time. Please indicate the extent to which the following questions are discrete.

57. At least 18 years of age	1	2	3	4	5
58. High school diploma or GED may be required	1	2	3	4	5
59. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
60. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
61. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
62. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
63. Describe what good customer service means to [candidate]	1	2	3	4	5
64. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 9: Controllable

Controllable questions assess behaviors or actions that applicants chose to perform or not perform out of their own volition. Please indicate the extent to which the following questions are controllable.

65. At least 18 years of age	1	2	3	4	5
66. High school diploma or GED may be required	1	2	3	4	5
67. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
68. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
69. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
70. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
71. Describe what good customer service means to [candidate]	1	2	3	4	5
72. Rating of overall communication skills	1	2	3	4	5

Dimension 10: Equally Accessible

Equally accessible questions assess past behavior and experiences that all applicants had an equal opportunity to engage in. Please indicate the extent to which the following questions are equally accessible.

73. At least 18 years of age	1	2	3	4	5
74. High school diploma or GED may be required	1	2	3	4	5
75. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
76. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
77. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
78. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
79. Describe what good customer service means to [candidate]	1	2	3	4	5
80. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 11: Noninvasive

Noninvasive questions assess aspects of candidates that are non-personal in nature. Noninvasive questions would not be considered an invasion of privacy from the applicant's perspective. Please indicate the extent to which the following questions are noninvasive.

81. At least 18 years of age	1	2	3	4	5
82. High school diploma or GED may be required	1	2	3	4	5
83. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
84. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
85. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
86. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
87. Describe what good customer service means to [candidate]	1	2	3	4	5
88. Rating of overall communication skills	1	2	3	4	5

Dimension 12: Susceptibility to Faking

Questions that are susceptible to faking are questions to which candidates can decipher the "correct" or "best" answer, allowing them to consciously distort their responses to make themselves look better, or increase their chances of passing the screening. Please indicate the extent to which the following questions are susceptible to faking.

89. At least 18 years of age	1	2	3	4	5
90. High school diploma or GED may be required	1	2	3	4	5
91. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
92. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
93. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
94. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
95. Describe what good customer service means to [candidate]	1	2	3	4	5
96. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 13: Socially Desirable

Socially desirable questions may lead applicants to unconsciously alter their true response because they want to be viewed favorably by others. Please indicate the extent to which the following questions might elicit a socially desirable response.

97. At least 18 years of age	1	2	3	4	5
98. High school diploma or GED may be required	1	2	3	4	5
99. Possession of or ability to obtain a valid [State] driver's license may be required	1	2	3	4	5
100. Ability to arrive at work location in 30 minutes in an emergency situation	1	2	3	4	5
101. Ability to work flexible schedule, including overtime on weekends/holidays	1	2	3	4	5
102. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
103. Describe what good customer service means to [candidate]	1	2	3	4	5
104. Rating of overall communication skills	1	2	3	4	5

THANK YOU FOR YOUR PARTICIPATION!!!

Prescreening Item Taxonomy Survey
ENTRY-LEVEL CUSTOMER SERVICE JOB

Please carefully read the description of an entry-level customer service job below as it appears to a job applicant. After reading the job description, you will be presented with a series of prescreening question classification definitions and the questions applicants must answer when they are screening for this position. Beginning with Dimension 1, first review the definition of the dimension, and then read and rate each of the prescreening questions that candidates are required to answer based on the extent to which each individual prescreening question reflects the classification dimension. Please rate all the screening questions on a dimension before proceeding to the next dimension.

Job Description/Posting: CALL CENTER REPRESENTATIVE

A Call Center Representative responds to telephone inquiries and initiates necessary action or documentation in a polite, professional, and efficient manner. Serving the customer and performing the job in a safe manner is paramount. A successful candidate possesses the following skills and characteristics:

- Is prompt, courteous, and professional at all times.
- Communicates effectively on the telephone with customers
- Communicates and works well with peers and supervisors
- Has basic keyboarding skills
- Effectively operates various computer applications to resolve customer inquiries
- Operates telephone equipment
- Processes routine help tickets, billing adjustments requests, online feedback, customer correspondence, and service orders using various computer applications
- Identifies, analyzes, and successfully resolves customer inquiries over the telephone
- Is able to work overtime as required

Call Center Representatives are often a customer's first point of contact with [Company]. Therefore, regular attendance, professionalism, efficiency, attention to detail, and customer service skills are key to ensuring a positive interaction and impression with the public. In addition to the high impact, fast-paced environment, for those employees who learn the business and demonstrate the abilities listed above, advancement opportunities – including working on special projects, supervisory assignments, analyst, quality assurance, and training positions – are available.

Requirements: A high school diploma or equivalent and at least one year of customer service experience are required. Obtain a qualifying score on [Company's] written [Pre-employment Test]. Keyboard and telephone experience is desired. Bilingual skills are a plus.

Training & Scheduling: Four weeks of training and two weeks of mentoring. Overtime may be required during emergencies.

For each prescreening question classification dimension below, please read the definition of the question dimension to be assessed. For each question, please indicate the extent to which the question reflects the dimension, using the scale below. Note: please rate all items on one dimension before proceeding to the next dimension.

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 1: Verifiability

Verifiable questions tap aspects of applicant experience that can be externally confirmed by an independent source. Please indicate the extent to which the following questions are verifiable.

1. At least 18 years of age	1	2	3	4	5
2. High school diploma or GED may be required	1	2	3	4	5
3. Previous computer experience may be required	1	2	3	4	5
4. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
5. Years of customer service experience in a call center environment	1	2	3	4	5
6. Software programs/applications experience	1	2	3	4	5
7. Typing skills (words per minute)	1	2	3	4	5
8. Anticipated tenure in role	1	2	3	4	5
9. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
10. Years of customer service experience working with general public	1	2	3	4	5
11. Describe what good customer service means to [candidate]	1	2	3	4	5
12. Shift/work schedule availability	1	2	3	4	5
13. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 2: Job-Relatedness

Job-related questions assess applicant experience and knowledge, skills, or abilities that are considered essential to successful performance on the job. Please indicate the extent to which the following questions are job-related.

14. At least 18 years of age	1	2	3	4	5
15. High school diploma or GED may be required	1	2	3	4	5
16. Previous computer experience may be required	1	2	3	4	5
17. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
18. Years of customer service experience in a call center environment	1	2	3	4	5
19. Software programs/applications experience	1	2	3	4	5
20. Typing skills (words per minute)	1	2	3	4	5
21. Anticipated tenure in role	1	2	3	4	5
22. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
23. Years of customer service experience working with general public	1	2	3	4	5
24. Describe what good customer service means to [candidate]	1	2	3	4	5
25. Shift/work schedule availability	1	2	3	4	5
26. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 3: Objectivity

Objective questions ask applicants to recall factual information about their past experiences. Objective events would be those that could be corroborated by an external observer. Please indicate the extent to which the following questions are objective in nature.

27. At least 18 years of age	1	2	3	4	5
28. High school diploma or GED may be required	1	2	3	4	5
29. Previous computer experience may be required	1	2	3	4	5
30. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
31. Years of customer service experience in a call center environment	1	2	3	4	5
32. Software programs/applications experience	1	2	3	4	5
33. Typing skills (words per minute)	1	2	3	4	5
34. Anticipated tenure in role	1	2	3	4	5
35. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
36. Years of customer service experience working with general public	1	2	3	4	5
37. Describe what good customer service means to [candidate]	1	2	3	4	5
38. Shift/work schedule availability	1	2	3	4	5
39. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 4: Minimum Qualifications/Knockout
Minimum qualifications questions assess skills, abilities, experience, or other attributes an applicant must possess in order to be considered further for a position; otherwise, the candidate is no longer considered. Please indicate the extent to which the following questions assess minimum qualifications.

40. At least 18 years of age	1	2	3	4	5
41. High school diploma or GED may be required	1	2	3	4	5
42. Previous computer experience may be required	1	2	3	4	5
43. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
44. Years of customer service experience in a call center environment	1	2	3	4	5
45. Software programs/applications experience	1	2	3	4	5
46. Typing skills (words per minute)	1	2	3	4	5
47. Anticipated tenure in role	1	2	3	4	5
48. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
49. Years of customer service experience working with general public	1	2	3	4	5
50. Describe what good customer service means to [candidate]	1	2	3	4	5
51. Shift/work schedule availability	1	2	3	4	5
52. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 5: Historical

Historical questions assess behaviors that have taken place or continue to take place. Historical questions do not assess behavioral intentions. Please indicate the extent to which the following questions are historical.

53. At least 18 years of age	1	2	3	4	5
54. High school diploma or GED may be required	1	2	3	4	5
55. Previous computer experience may be required	1	2	3	4	5
56. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
57. Years of customer service experience in a call center environment	1	2	3	4	5
58. Software programs/applications experience	1	2	3	4	5
59. Typing skills (words per minute)	1	2	3	4	5
60. Anticipated tenure in role	1	2	3	4	5
61. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
62. Years of customer service experience working with general public	1	2	3	4	5
63. Describe what good customer service means to [candidate]	1	2	3	4	5
64. Shift/work schedule availability	1	2	3	4	5
65. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 6: External

External questions assess past behavior and experiences that occur in specific, real-life situations. External questions do not assess unexpressed thoughts, attitudes, opinions, or reactions to events. Please indicate the extent to which the following questions are external.

66. At least 18 years of age	1	2	3	4	5
67. High school diploma or GED may be required	1	2	3	4	5
68. Previous computer experience may be required	1	2	3	4	5
69. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
70. Years of customer service experience in a call center environment	1	2	3	4	5
71. Software programs/applications experience	1	2	3	4	5
72. Typing skills (words per minute)	1	2	3	4	5
73. Anticipated tenure in role	1	2	3	4	5
74. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
75. Years of customer service experience working with general public	1	2	3	4	5
76. Describe what good customer service means to [candidate]	1	2	3	4	5
77. Shift/work schedule availability	1	2	3	4	5
78. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 7: Firsthand

Firsthand questions assess aspects of the applicant that he or she would be able to evaluate based on his or her direct, objective knowledge. Firsthand questions do not assess an applicant's estimation of how others would evaluate them. Please indicate the extent to which the following questions are firsthand.

79. At least 18 years of age	1	2	3	4	5
80. High school diploma or GED may be required	1	2	3	4	5
81. Previous computer experience may be required	1	2	3	4	5
82. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
83. Years of customer service experience in a call center environment	1	2	3	4	5
84. Software programs/applications experience	1	2	3	4	5
85. Typing skills (words per minute)	1	2	3	4	5
86. Anticipated tenure in role	1	2	3	4	5
87. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
88. Years of customer service experience working with general public	1	2	3	4	5
89. Describe what good customer service means to [candidate]	1	2	3	4	5
90. Shift/work schedule availability	1	2	3	4	5
91. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 8: Discrete

Discrete questions assess a single, unique behavior or simple count of unique instances of events, whereas summative questions require the summation or estimation of behavior over time. Please indicate the extent to which the following questions are discrete.

92. At least 18 years of age	1	2	3	4	5
93. High school diploma or GED may be required	1	2	3	4	5
94. Previous computer experience may be required	1	2	3	4	5
95. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
96. Years of customer service experience in a call center environment	1	2	3	4	5
97. Software programs/applications experience	1	2	3	4	5
98. Typing skills (words per minute)	1	2	3	4	5
99. Anticipated tenure in role	1	2	3	4	5
100. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
101. Years of customer service experience working with general public	1	2	3	4	5
102. Describe what good customer service means to [candidate]	1	2	3	4	5
103. Shift/work schedule availability	1	2	3	4	5
104. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 9: Controllable

Controllable questions assess behaviors or actions that applicants chose to perform or not perform out of their own volition. Please indicate the extent to which the following questions are controllable.

105. At least 18 years of age	1	2	3	4	5
106. High school diploma or GED may be required	1	2	3	4	5
107. Previous computer experience may be required	1	2	3	4	5
108. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
109. Years of customer service experience in a call center environment	1	2	3	4	5
110. Software programs/applications experience	1	2	3	4	5
111. Typing skills (words per minute)	1	2	3	4	5
112. Anticipated tenure in role	1	2	3	4	5
113. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
114. Years of customer service experience working with general public	1	2	3	4	5
115. Describe what good customer service means to [candidate]	1	2	3	4	5
116. Shift/work schedule availability	1	2	3	4	5
117. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 10: Equally Accessible

Equally accessible questions assess past behavior and experiences that all applicants had an equal opportunity to engage in. Please indicate the extent to which the following questions are equally accessible.

118. At least 18 years of age	1	2	3	4	5
119. High school diploma or GED may be required	1	2	3	4	5
120. Previous computer experience may be required	1	2	3	4	5
121. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
122. Years of customer service experience in a call center environment	1	2	3	4	5
123. Software programs/applications experience	1	2	3	4	5
124. Typing skills (words per minute)	1	2	3	4	5
125. Anticipated tenure in role	1	2	3	4	5
126. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
127. Years of customer service experience working with general public	1	2	3	4	5
128. Describe what good customer service means to [candidate]	1	2	3	4	5
129. Shift/work schedule availability	1	2	3	4	5
130. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 11: Noninvasive

Noninvasive questions assess aspects of candidates that are non-personal in nature. Noninvasive questions would not be considered an invasion of privacy from the applicant's perspective. Please indicate the extent to which the following questions are noninvasive.

131. At least 18 years of age	1	2	3	4	5
132. High school diploma or GED may be required	1	2	3	4	5
133. Previous computer experience may be required	1	2	3	4	5
134. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
135. Years of customer service experience in a call center environment	1	2	3	4	5
136. Software programs/applications experience	1	2	3	4	5
137. Typing skills (words per minute)	1	2	3	4	5
138. Anticipated tenure in role	1	2	3	4	5
139. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
140. Years of customer service experience working with general public	1	2	3	4	5
141. Describe what good customer service means to [candidate]	1	2	3	4	5
142. Shift/work schedule availability	1	2	3	4	5
143. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 12: Susceptibility to Faking

Questions that are susceptible to faking are questions to which candidates can decipher the "correct" or "best" answer, allowing them to consciously distort their responses to make themselves look better, or increase their chances of passing the screening. Please indicate the extent to which the following questions are susceptible to faking.

144. At least 18 years of age	1	2	3	4	5
145. High school diploma or GED may be required	1	2	3	4	5
146. Previous computer experience may be required	1	2	3	4	5
147. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
148. Years of customer service experience in a call center environment	1	2	3	4	5
149. Software programs/applications experience	1	2	3	4	5
150. Typing skills (words per minute)	1	2	3	4	5
151. Anticipated tenure in role	1	2	3	4	5
152. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
153. Years of customer service experience working with general public	1	2	3	4	5
154. Describe what good customer service means to [candidate]	1	2	3	4	5
155. Shift/work schedule availability	1	2	3	4	5
156. Rating of overall communication skills	1	2	3	4	5

1	2	3	4	5
Not at all	Small extent	Moderate extent	Great extent	Completely

Dimension 13: Socially Desirable

Socially desirable questions may lead applicants to unconsciously alter their true response because they want to be viewed favorably by others. Please indicate the extent to which the following questions might elicit a socially desirable response.

157. At least 18 years of age	1	2	3	4	5
158. High school diploma or GED may be required	1	2	3	4	5
159. Previous computer experience may be required	1	2	3	4	5
160. Ability to meet overtime requirement as needed on weekdays/weekends	1	2	3	4	5
161. Years of customer service experience in a call center environment	1	2	3	4	5
162. Software programs/applications experience	1	2	3	4	5
163. Typing skills (words per minute)	1	2	3	4	5
164. Anticipated tenure in role	1	2	3	4	5
165. Provide reasons [candidate] would like to join the department/company team	1	2	3	4	5
166. Years of customer service experience working with general public	1	2	3	4	5
167. Describe what good customer service means to [candidate]	1	2	3	4	5
168. Shift/work schedule availability	1	2	3	4	5
169. Rating of overall communication skills	1	2	3	4	5

THANK YOU FOR YOUR PARTICIPATION!!!

APPENDIX B: BUSINESS ACCOUNT REPRESENTATIVE JOB DESCRIPTION

JOB DESCRIPTION:
Business Account Representative (Public Utility)

Responsibilities

Customer Service:
Delight customers by effectively applying interpersonal skills and consultative sales approaches. Possess and apply strong sales skills to serve small- and mid-size business customers, and meet customer satisfaction and Integrated Demand Side Management (IDSM) goals. Effectively position [Company] products and services as win-win solutions to customer business needs. Position [Company] as the preferred supplier of energy products and services. Provide a customer liaison to all [Company] customer contact processes (such as service planning, planned shutdowns, billing, credit, etc.). Collaborate with other Account Managers to provide services to local business customers. Understand all [Company] processes and procedures that touch the customer at a level sufficient enough to allow rapid and effective response to customer needs. Establish effective internal relationships to help clear barriers to customer satisfaction. Facilitate solutions and successfully advocate process improvements to better meet customer needs. In some cases, conduct energy efficiency surveys and prepare detailed reports of improvement recommendations, current and proposed usage, estimated savings, and payback periods. Educate customers on company and energy industry issues such as electric emergency procedures, rotating outage policy, regulatory developments and electric and gas safety. Interact with customers to address uneconomic bypass issues.

Technical Support:
Serve as a sales-appropriate technical expert for energy efficiency, demand response, and self-generation programs. Perform energy efficiency surveys of customer facilities and prepare basic financial analyses of potential energy efficiency investments. When needed, secure more advanced technical expertise (such as consultants, other department experts, peers). Act as a technical resource for customer rate and tariff concerns.

Special Projects:
Participate in special teams to enhance the business customer experience and attain [Company] IDSM goals. Support the implementation of team action plans to achieve company goals.

Safety and Compliance:
At all times, work safely and in compliance with legal, regulatory, and ethical standards. Improve customer opinion of [Company] as an advocate of public and employee safety.

Essential/Required:
- Ability and willingness to apply consultative sales skills; strong interpersonal and communication skills
- Technical skills to learn, understand, explain, and recognize opportunities with energy end use systems and IDSM measures/techniques (including but not limited to CEE, demand response, and self-generation); basic financial analyses, including payback and return-on-investment
- BA/BS (technical or business) or equivalent relevant work experience
- Willing and able to conduct 25% day travel and work 50-60 hour weeks when needed
- Must have access to a reliable vehicle and possess a solid driving record with the ability to meet [Company] insurance requirements

Marginal/Desired:
- Knowledge of energy industry issues; gas and electric transmission and distribution systems (including generation and self-generation); [Company] rates, tariffs, and contracts; [Company] products and services (such as IDSM, core gas and electric, transmission, direct access); [Company] customer contact processes (such as service, planning, planned shutdowns, billing, credit, etc.)
- Familiarity with technical tools available to address business customer needs (such as rate analysis tools, business energy tools, online billing history)
- Advanced degree, PE license, or other relevant certifications
- Experience applying portfolio-based consultative sales skills

APPENDIX C: CRITICAL RESUME ELEMENTS SUMMARY

Resume Evaluation Dimensions Guide

The following resume evaluation definitions were provided by an SME knowledgeable in resume evaluation in general, and for the BAR position in particular. Refer to these guidelines as you evaluate the candidates' resumes.

Objective: It is ok if the candidate does not have one, but if they do, is it related to the position? If it is related to the position, how is it related? Do you infer they tailored their objective to this specific position?

Summary of Skills: Is there a summary of skills provided, at either the top or bottom of the resume? (Summaries are important because they allow the reader to quickly scan for relevant skills required for the position.) If there is a summary, are the skills it highlights detailed in the body of the resume, within the actual descriptions of the jobs they previously held or currently hold? Are the skills clearly related and relevant to the position they are applying for (BAR)?

Current Employment Status: Is the candidate currently working? If not, how recently were they working? If they have not worked in over 6 months, is there a reason provided? Does the reason seem valid?

Work Experience: Is the work experience relevant to the position? (The job title or responsibilities should, to a large degree, synch up with the job for which the candidate is applying.) *Note: refer back to the requirements and qualifications as outlined in the BAR job description.* If the experience is not closely or obviously related, can you determine if most or some of the skills the candidate used/gained are reasonably transferable to the position for which the candidate is applying? If there is not a clear match, has the candidate shown—in any other part of the resume—they possess the skills required for the position, even if the candidate learned the skills in other environments, e.g., volunteering, training, or academia? Also, has their employment been steady? Have they spent 2-3 years at each company? This can be important and in some cases indicates how committed/stable they are.

Advancement: Does the resume indicate that the candidate has been promoted in their positions? Does the resume show that they have bettered themselves and won higher-level positions?

Education or Compensatory Experience/Training: Most positions imply that a four-year degree is required, but if the position is not inherently technical in nature, e.g., engineering or scientific, the educational requirement may also be met by the candidate possessing "compensatory experience," i.e., possessing many years of experience in a relevant area related to the position. (The BAR position is an example of a position that indicates a degree is required, but it is not an unequivocal requirement as this is not a highly technical position.) If the candidate has a degree, they should receive automatic credit for that.

If the candidate does not have a degree, but has relevant educational experiences or training, the candidate should also receive credit for that. If the candidate's total years of experience exceed the requirements, and appear to fulfill the "compensatory experience" mandate, they should receive credit for that as well.

Continued Education: If the candidate does possess a degree, but also has a graduate degree, they should receive credit for that as well.

Volunteerism: Is the candidate a volunteer in their community? Often, this indicates that they are conscientious, involved, and diligent. Hiring managers are always appreciative of volunteers.

Format: Is there organized chronology, proper spelling, grammar, and punctuation? Is there consistency with the use of italics, underlining, and capitalization? How would you rate the "readability" factor, i.e., the ability to scan the document and extrapolate critical data with ease?

Overall Rating: Keeping in mind the overall requirements and qualifications necessary for this position, based on the job description, as well as the overall skills and qualifications as presented in the candidate's resume, how would you rate the candidate's overall qualifications and fit with this position?

APPENDIX D: INITIAL INVITATION TO GRADUATE STUDENTS TO PARTICIPATE IN STUDY 2

Dear Graduate Student Colleagues:

I will be conducting Phase 2 of my data collection, involving a resume evaluation exercise, the week of [DATE], and I invite you to participate. If you participated in Phase 1 of the study, this does not preclude you from participation in Phase 2, nor are you required to participate in Phase 2. Likewise, if you did not participate in Phase 1, and would like to participate in Phase 2, you may still do so without having participated in Phase 1.

As mentioned in my initial communication last winter, I am conducting my dissertation research on the use of various preliminary job applicant evaluation procedures. Specifically, I am investigating the use of online applicant prescreening and resume evaluation in order to better understand how organizations can best leverage these tools to identify the top candidates.

Your involvement in the study, should you decide to participate, will occur in two parts. During the week of [DATE], you will complete a resume evaluation exercise. More details about this process will be provided to you at the data collection. Toward the end of the month, you will be participating in another resume evaluation exercise. In each case, you will be presented with a set of actual applicant resumes for a specific position (supporting materials such as a job description will be provided). You will be asked to evaluate the resumes based on the guidelines provided. The second data collection will proceed in a similar manner, with the primary difference being that you will receive the materials via mail, and will be provided a postage paid return envelope to mail your completed survey back to me. Each evaluation exercise should take approximately one hour, though some people may complete it in less time.

Should you decide to participate, please contact me via email. I am hoping to assemble small groups for data collection if possible, but I will also be available for single-person sessions, based on scheduling availability. In your email, please also indicate your best day(s) and time(s) to participate. I will notify participants of their scheduled session. Data collection will occur between [DATE], and [DATE].

You will receive further information on your scheduled date and time. If you have any questions, please contact me either via telephone or email.

Sincerely,

Matthew E. Paronto

APPENDIX E: SAMPLE UNSTRUCTURED RESUME EVALUATION SURVEY

August 2007

Dear Participant:

My name is Matthew Paronto. I am a graduate student in the Psychology Department at Portland State University conducting my doctoral dissertation study on preliminary job applicant evaluation procedures. You may not receive any direct benefit from participating in this study. However, this research may help others in the future, as the information you provide may lead to recommendations on organizational best practices regarding the use of these preliminary applicant evaluation procedures.

You are being asked to take part in this study so that we can start to better understand how organizations can best leverage the use of applicant selection procedures. This is extremely important for organizations both from a legal as well as from a financial standpoint, as hiring employees who are not a good fit with an organization due to poor screening procedures results in delays in getting the human capital needed for an organization to be successful.

If you decide to participate, you will be presented today with prototypical applicant resumes for a specific job and will be asked to evaluate them using the job description and evaluation guidelines provided to make an assessment of the quality of the applicant's resume. You will be presented with a similar exercise in approximately two weeks via mail. More details on the procedures are outlined in the instructions that accompany the study materials. Each evaluation exercise may take up to one hour, though some people may complete it in less time.

If for any reason you feel uncomfortable, you may stop participating at any time. Also, you may skip any questions you feel uncomfortable responding to. However, I assure you that your responses will be kept confidential, and there will be no way in which to match responses to the individuals who provide them, since your name will not be on the survey.

Participation is entirely voluntary. Your decision to participate or not will not affect your relationship with the researcher or with Portland State University in any way. Please keep a copy of this letter for your records.

If you have any concerns or questions about your participation in this study or your rights as a research participant, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, 503.725.4288. If you have any questions about the study itself, contact me at 503.381.2338 or Dr. Donald Truxillo of the Psychology Department at 503.725.3969.

Sincerely,
Matthew E. Paronto
Department of Psychology
Portland State University

RESUME EVALUATION EXERCISE INSTRUCTIONS

Please begin by reading the description of the Business Account Representative (BAR) position as it appears to an applicant applying for this job. This description outlines the requirements and qualifications needed for this position. Be sure to review and familiarize yourself with the requirements of this position before proceeding to the resume evaluation exercise.

Next, please review and familiarize yourself with the key dimensions of a resume as outlined by a Subject Matter Expert (SME) knowledgeable about resume evaluation, particularly for this position. You will be using this guide to aid you in evaluating the candidates' resumes.

On the following pages you will find 20 candidate resumes for the BAR position. For each resume, evaluate the candidate's resume, *keeping in mind the qualifications and requirements for the BAR position contained in the job description as well as the SME guidelines to resume evaluation* (both of which you may refer back to at any time). Place a check mark in the box corresponding to your evaluation of the candidate's resume, located at the top of the first page of the candidate's resume. The rating scale for evaluating each resume appears below.

1 DOES NOT MEET REQUIREMENTS: The candidate's resume does not meet the minimum requirements for the position.

2 MEETS REQUIREMENTS: The candidate's resume meets the minimum requirements for the position.

3 EXCEEDS REQUIREMENTS: The candidate's resume exceeds the minimum requirements for the position.

Please complete the evaluation of the resumes in the order in which they are presented. Should you decide to go back to an earlier resume to adjust a rating, be sure to clearly mark your intended rating (you may want to complete the exercise in pencil).

Candidate Profile 21

❒ 1	❒ 2	❒ 3
Does not meet requirements	Meets requirements	Exceeds requirements

SKILLS

- Supervised the ordering and distribution of fresh baked goods for a major wholesale bakery route distribution system to retail outlets and food service enterprises.
- Managed inventory levels for retail grocery stores including ordering and merchandising.
- Strong organization and planning skills with attention to detail and follow-through.
- Proficient on IBM PC with application software Word, Excel, Outlook, Lotus Notes, Internet Explorer, and Kronos
- Interest and abilities tested and confirmed through the following professionally administered testing programs: Myers-Briggs Type Indicator; Strong Interest Inventory; Career Occupational Preference System. Results available upon request

EXPERIENCE
[Company]
District Sales Manager
From: 07/1990 To: current
Wholesale Experience: Responsible for the maintenance of existing accounts; solicitation of new business; establishing and maintaining efficient route structures; inventory control; facility and vehicle maintenance; recruiting, training, scheduling, directing, and supervising subordinates; and accounts receivables, the maintenance of the Excel database for product produced at ten company and contract commercial bakeries. I am a member of the core training and integration team for both Route Sales Representatives and Supervisors.

Various [Companies]
Asst. Store Mgr, Dept Mgr
From: 06/1969 To: 07/1990
Retail Experience: Responsible for recruiting, training, scheduling, directing, and controlling labor and labor costs; inventory control including ordering, purchasing, and warehouse management; setting and controlling retail pricing and profit margins; schematic and display merchandising strategies; direct dealings with food brokers, salespeople, and customer relations

EDUCATION
[University], Bachelor's Degree, Biblical Studies, 05/2004

APPENDIX F: SAMPLE STRUCTURED RESUME EVALUATION SURVEY

August 2007

Dear Participant:

My name is Matthew Paronto. I am a graduate student in the Psychology Department at Portland State University conducting my doctoral dissertation study on preliminary job applicant evaluation procedures. You may not receive any direct benefit from participating in this study. However, this research may help others in the future, as the information you provide may lead to recommendations on organizational best practices regarding the use of these preliminary applicant evaluation procedures.

You are being asked to take part in this study so that we can start to better understand how organizations can best leverage the use of applicant selection procedures. This is extremely important for organizations both from a legal as well as from a financial standpoint, as hiring employees who are not a good fit with an organization due to poor screening procedures results in delays in getting the human capital needed for an organization to be successful.

If you decide to participate, you will be presented today with prototypical applicant resumes for a specific job and will be asked to evaluate them using the job description and evaluation guidelines provided to make an assessment of the quality of the applicant's resume. You will be presented with a similar exercise in approximately two weeks via mail. More details on the procedures are outlined in the instructions that accompany the study materials. Each evaluation exercise may take up to one hour, though some people may complete it in less time.

If for any reason you feel uncomfortable, you may stop participating at any time. Also, you may skip any questions you feel uncomfortable responding to. However, I assure you that your responses will be kept confidential, and there will be no way in which to match responses to the individuals who provide them, since your name will not be on the survey.

Participation is entirely voluntary. Your decision to participate or not will not affect your relationship with the researcher or with Portland State University in any way. Please keep a copy of this letter for your records.

If you have any concerns or questions about your participation in this study or your rights as a research participant, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, 503.725.4288. If you have any questions about the study itself, contact me at 503.381.2338 or Dr. Donald Truxillo of the Psychology Department at 503.725.3969.

Sincerely,

Matthew E. Paronto
Department of Psychology
Portland State University

RESUME EVALUATION EXERCISE INSTRUCTIONS

Please begin by reading the description of the Business Account Representative (BAR) position as it appears to an applicant applying for this job. This description outlines the requirements and qualifications needed for this position. Be sure to review and familiarize yourself with the requirements of this position before proceeding to the resume evaluation exercise.

Next, please review and familiarize yourself with the key dimensions of a resume as outlined by a Subject Matter Expert (SME) knowledgeable about resume evaluation, particularly for this position. You will be using this guide to aid you in evaluating the candidates' resumes.

On the following pages you will find 20 candidate resumes for the BAR position. For each resume, evaluate the candidate along each of the key resume dimensions on the appropriate rating sheet (which appears on the page following the candidate's resume), *keeping in mind the qualifications and requirements for the BAR position contained in the job description, as well as the SME guidelines to resume evaluation* (both of which you may refer back to at any time). The rating scale for each resume dimension appears below.

1 DOES NOT MEET REQUIREMENTS: The candidate's resume does not meet the minimum requirements on the dimension in question.

2 MEETS REQUIREMENTS: The candidate's resume meets the minimum requirements on the dimension in question.

3 EXCEEDS REQUIREMENTS: The candidate's resume exceeds the minimum requirements on the dimension in question.

In addition to making dimension ratings, please provide an overall rating of the candidate, which appears at the bottom of each candidate's rating sheet, using the following scale:

1 DOES NOT MEET REQUIREMENTS: The candidate's resume does not meet the minimum requirements for the position.

2 MEETS REQUIREMENTS: The candidate's resume meets the minimum requirements for the position.

3 EXCEEDS REQUIREMENTS: The candidate's resume exceeds the minimum requirements for the position.

Please complete the evaluation of the resumes in the order in which they are presented. Should you decide to go back to an earlier resume to adjust a rating, be sure to clearly mark your intended rating (you may want to complete the exercise in pencil).

Candidate Profile 36

Relevant Experience

[Company] 1/06 to Present Consumer Affairs

Research, analyze evaluate, develop and recommend alternatives on regulatory issues and policies. Independently receive, investigate, and resolve consumer complaints filed against utility companies. Provide testimony, apply code and consult with commission staff.

[Company] 1/00 to 1/06 Account Manager

Managed installed base. Maintained client relationships. Designed and sold technologically advanced voice &data services and equipment. Developed convergent applications addressing operational performance, improving business processes and delivering ROI. Provided portfolio of VOIP, SIP, Presence Management, Collaboration and Messaging solutions. Marketed managed services including: provisioning, facilities management, professional services and customized development. Consistently exceeded monthly and yearly sales objectives.

[Company] 1/97 to 12/99 Business Development

Developed business opportunities for a wide range of voice, data and IP Technology companies. Created, maintained and updated customized database pipelines. Served as a liaison between client, providers, manufacturers and developers. Represented, positioned and formed strategic alliances for B2B clients. Formulated abstracts of Fortune 1000 companies. Consulted with IT professionals including: CTO s, CIO s, COO s, CSO s and CEO s while initiating sales calls.

[Company] 2/88 to 12/96 Communications Analyst-Account Executive

Analyzed communication needs. Marketed hybrid and PBX equipment to mid-large size businesses. Demonstrated products, presented pricing, consummated sales and coordinated installations. Maintained and exceeded monthly quotas and annual sales objectives.

[Company] 10/84-1/88 Systems Engineer

Acted as a liaison between client, hardware and software engineers. Streamlined and developed production processes and procedures. Evaluated existing practices. Prepared efficiency recommendations. Conducted time studies.

[Company] 1/79-9/84 Service Representative

Processed moves, adds and changes. Collected delinquent accounts. Investigated disputed charges. Adjusted customer bills. Established new accounts. Sold yellow page advertising.

Education

DBA-Doctoral Student, [University] 5/06 to Present

MBA-Technology Management, [University]-1/04.

BS-Busines Management, [University]-4/81.

RESUME EVALUATION: CANDIDATE 36

Objective	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Summary of Skills	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Current Employment Status	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Work Experience	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Advancement	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Education or Compensatory Experience/Training	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Continued Education	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Volunteerism	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
Format	❒ 1 Does not meet requirements	❒ 2 Meets requirements	❒ 3 Exceeds requirements
OVERALL CANDIDATE RATING	❒ 1 Does not meet requirements for the position	❒ 2 Meets requirements for the position	❒ 3 Exceeds requirements for the position

APPENDIX G: RECRUITER RESUME INTERVIEW PROTOCOL

1. Which of the positions you recruit for would you consider to be a high-volume position?

2. Is there an up-to-date job description (e.g., based on a job posting) for this position? Where can this description be obtained/located?

3. How do you source for this position?

4. When you have collected all the résumés for this position, how do you sort through them to arrive at the résumés you want to review further?

5. When reviewing the résumés for this position, what qualifications are you looking for? (List every qualification, experience, etc., that you look for.)

6. Select a few résumés and walk through the evaluation process of each (high, medium, and low quality). What about each candidate's résumé makes them a high/medium/low-quality candidate?