PREDICTING OBJECTIVE MEASURES OF PERFORMANCE

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

_____
Kristina Eva Chirico

Certificate of Approval:

_____   _____
Adrian L. Thomas   Philip M. Lewis, Chair
Assistant Professor   Professor
Psychology   Psychology

_____   _____
John G. Veres III   Virginia E. O'Leary
Adjunct Professor   Professor
Psychology   Psychology

_____
Stephen L. McFarland
Acting Dean
Graduate School

PREDICTING OBJECTIVE MEASURES OF PERFORMANCE

Kristina Eva Chirico

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama

December 16, 2005

PREDICTING OBJECTIVE MEASURES OF PERFORMANCE

Kristina Eva Chirico

                            Signature of Author

                            December 16, 2005
                            Date of Graduation

DISSERTATION ABSTRACT

PREDICTING OBJECTIVE MEASURES OF PERFORMANCE

Kristina Eva Chirico

Doctor of Philosophy, December 16, 2005
(Master of Science, May 11, 2002)
(B.A., Agnes Scott College, 1998)

69 Typed Pages

Directed by Philip M. Lewis

In today's competitive job market, many organizations use various selection procedures
in order to hire the best possible employees. Most selection tools such as structured
interviews, mental ability tests, and personality inventories have been shown to predict
employee performance in terms of subjective measures (i.e., supervisory ratings).
However, organizations are more interested in predicting objective measures of
performance (e.g., sales). The purpose of this study was to determine whether biodata,
situational judgment, and role-play could be effective in predicting objective measures of
performance. Four objective measures (i.e., sales amount, number of orders, debt, and
the number of active staff members) were collected from 189 District Sales Managers
employed by an international company selling beauty and related products. The results

indicated that none of the objective measures are related to the predictors. Several

plausible explanations of these results are discussed.

ACKNOWLEDGMENTS

Style manual used: APA Publication Manual 5th Edition

Computer software used: Microsoft ®Word 2002, SPSS 10.0

TABLE OF CONTENTS

LIST OF TABLES

LITERATURE REVIEW

In today's competitive job market, organizations need to select the best possible employees. Most organizations use selection tools for this purpose. In order to evaluate the effectiveness of these tools, employers have to validate their selection tools against performance measures. However, not all performance measures are equivalent. Performance measures can be categorized as either objective or subjective. Measures such as sales volume and number of accidents or grievances are more "objective" while performance ratings, typically given by supervisors or peers, are more "subjective" in nature. It is important to keep in mind that the distinction is not strictly dichotomous. Most objective and subjective measures fall on a continuum.

There are numerous types of selection tools including structured interviews, mental ability tests, and personality and biodata inventories. Research has shown that all of these tools frequently predict employee performance when measured by supervisory ratings (e.g., Hunter & Hunter, 1984). Hunter and Hunter (1984) report the average validity of cognitive ability tests of .53 in comparison with .37 for biodata measures and .14 for interviews. Supervisory ratings typically include ratings on performance dimensions such as leadership, decision-making, and communication skills. However, one should keep in mind that these ratings are subjective in nature. Some selection tools, that predict supervisory ratings may not predict what is important to organizations, namely, objective

1

performance measures such as sales, number of orders, or profits. Prior research does not

provide conclusive evidence regarding which selection tools best predict these types of

"objective" performance measures. The primary purpose of the present study is to

evaluate three selection methods in terms of their ability to predict objective performance

measures. Specifically, the present study compares a biodata scale, situational judgment

test (SJT), and a role-play in terms of their potential to predict four objective

performance measures.

Performance Criterion Issues

Most validity studies of selection tools have focused on the predictors.

Researchers and practitioners alike accept the notion that if the predictor is correlated to a

relevant criterion, such as job performance or turnover, the predictor is valid. However,

researchers are increasingly recognizing that there are many criterion issues that need

further study. For example, Guion (1991) notes that performance measurement, which is

often used as a criterion in validity studies, poses several problems. One problem is that

performance is multi-dimensional in nature. Performance dimensions such as task

performance, output quantity and quality, organizational citizenship, turnover, and

accidents are all part of the performance domain. However, almost 85% of validity

studies ignore the multi-dimensional nature of performance and use only one overall

rating of performance as a global criterion (Lent, Aurbach, & Levin, 1971a). Likewise,

Fried (1991) and Williams and Livingstone (1994) made no distinction between objective

and subjective performance measures in their meta-analysis. Although objective and

subjective measures are related, they should not be used interchangeably (Bommer et al.,

1995).

In most validity studies performance is usually assessed by supervisory ratings of performance dimensions. In almost 60% of the validity studies reviewed, supervisory ratings were used as the sole performance criterion (Lent, Aurbach, & Levin 1971b). Bernardin and Beatty (1984) found in a survey of human resource managers that over 90% of the respondents used supervisory ratings as their primary source of performance ratings. Peer ratings were the second most widely used source of ratings. More recently, Viswesvaran (2002) found in his meta-analysis that almost 60% of studies use supervisory ratings as a performance criterion.

One obvious problem with such ratings is that personal and contextual biases can seriously impair their accuracy (Borman, 1979; DeNisi, Cafferty, & Meglino, 1984). For example, managers may be reluctant to give low ratings to poor performers. Therefore, managers may inflate their performance ratings, committing leniency errors. According to Bretz, Milkovich, and Read (1992) 60-70% of employees in most organizations are rated in the top level, which suggests that leniency error is common in supervisory ratings. In another instance of supervisory bias, halo error is said to occur when a supervisor assigns similar ratings across different dimensions of performance based on a general impression of the ratee (Pulakos, 1997).

Little attention has been paid in empirical attempts to validate selection tools to the objective criteria that are critical to most organizations (Thayer, 1992). Most selection instruments predict employee performance in terms of ratings on performance dimensions. However, in those few instances where both objective and subjective

3

criteria have been used, supervisory ratings produce a lower validity coefficient in comparison to more objective criterion such as job sample measures (Nathan & Alexander, 1988). This finding is most likely due to the fact that supervisory ratings are more likely to be affected by ratings errors than work sample tests. Nathan and Alexander (1988) compared validity coefficients from seven tests of clerical abilities (e.g., cognitive ability, memory, perceptual speed, motor ability) for five criteria. The criteria included two subjective measures (supervisory ratings and rankings) and three objective criteria (work samples, production quality and quantity). The work sample was the criterion most highly predicted by each type of test. The average validity coefficient for the work sample across all seven tests was .54, compared with .34 for supervisory ratings and .46 for supervisory rankings. Production quantity yielded an average coefficient of .37 and the validity coefficient for production quality was the lowest among criteria, averaging only .17. These correlations were corrected for range restriction and test unreliability.

Although the use of subjective criteria in research is wide spread, it is evident that the use of subjective criteria can be problematic. Since these ratings rely on human judgment, they are prone to various kinds of ratings errors. The most recent trend in Human Resource practice is to evaluate employees on the degree to which they possess certain competencies. However, Bernardin, Hagan, Kane, and Villanova (1998) point out that a measurement of one's competencies is not necessarily a measure of performance. The problems associated with the subjective performance measures have led practitioners and researchers to search for more objective ways to measure performance.

Objective Performance Measures

The use of objective performance measures has its own problems. Performance can be measured in seemingly simple terms such as manufacturing defects, quantity of output, and sales volume. But these measures can be contaminated by external factors that the employee cannot control. For instance, poor economic conditions can affect product sales, and faulty equipment can decrease the quality and quantity of output. This criterion issue becomes even more problematic as job complexity increases. Complex jobs typically have complex performance criteria.

Current study

In the present study, four objective performance criteria essential for success in a District Sales Manager position were defined. These "objective" criteria were sales amount, number of orders, amount of debt, and number of active salespeople for the period of one year. The purpose of the study was to evaluate three types of selection tools (i.e., biodata scale, situational judgment test, and role-play) in terms of their ability to predict several objective criterion measures. The next three sections will focus on description, advantages, criticisms, and the predictive power of these selection tools including theories of why these tools are predictive of performance.

Biographical Data

Despite the widespread use of biodata measures in research (and some use in practice), there is no accepted definition of what constitutes biographical data (biodata). All biodata scales contain items that measure various aspects of past behaviors, usually those behaviors that are believed to predict future behavior. Biodata items focus on past

experiences as well as background information. Mael (1991) suggests that the historical nature of biodata items is their single most important attribute. Biodata items should measure discrete and unique events that are relevant to the job and reflect external events that are objective and potentially verifiable. Gunter, Furnham, and Drakeley, (1993) point out that biodata items often cover personal circumstances such as habits, attitudes, health, human relations, childhood or teen experiences, recreation, hobbies, education, work, opinions, preferences, and personal attributes. Although biodata items are usually objectively verifiable and factual, they can also be focused on subjective attitudes and feelings.

Numerous studies have demonstrated that biodata measures can be valid predictors of many performance criteria, such as training success, performance ratings, wages, leadership performance, employee theft, adjustment, satisfaction, team performance, and safety performance (Hough & Paullin, 1994; Hunter & Hunter, 1984; Reilly & Chao, 1982; Stricker & Rock, 1998). As with other measures, validity coefficients for biodata measures differ across jobs and types of criterion. On average, corrected correlations range from .30 to .40 (Hunter & Hunter, 1984). Since the prediction of performance is the focus of the present proposal, theories that explain why biodata is predictive and studies showing the usefulness of biodata will be discussed further.

Why does biodata predict performance. Although the criterion-related validity of biodata has been well demonstrated, researchers are still searching for an explanation for why they work. The most accepted rationale for using biodata measures rests on the

6

notion that "life history information is a good predictor of the future job behavior of individuals" (Fine & Cronshaw, 1994, p.41). However, researchers are concerned what life history biodata actually measure. This concern stems from the lack of theory for understanding how individual characteristics and experiences lead to criterion performance.

One of the first steps taken in attempts to close this gap in theory was the work of Owens and Schoenfeldt (1979). They proposed a theoretical framework for biodata – the Developmental-Integrative model. Owens and Schoenfeldt (1979) suggested that participants with similar life experiences will behave in similar way as adults and can be subgrouped on this basis. They also suggest that different subgroups will respond to the same situations very differently. Subgrouping was shown to be effective in predicting occupational choice (Brush & Owens, 1979).

Another effort to theory building was carried out by Stokes Mumford, and Owens (1989). They presented an interactional model, called the Ecology Model, in which individual characteristics and the situations to which one gets exposed result in individual differences in life experiences. Individual characteristics influence not only people's reaction and behaviors in various situations but also determine which situation the individual will enter. Since biodata items are more interpretable when based on a theory, the Ecology Model is often used as a framework for biodata item generation. The items generated describe situations that are relevant to behaviors often occurring in the workplace.

Advantages of using biographical data. There are several advantages to using

7

biodata items. First, researchers agree that biodata have little or no adverse impact on protected employee groups in contrast to general mental ability tests (Mitchell, 1994; Stokes & Cooper, 2001). For example, Pulakos and Schmidt (1996) report a mean difference between Caucasians and African-Americans to be -.05. Second, the administrative and scoring cost for biodata items is quite low (Stokes & Cooper, 2001). Third, biodata are more resistant to faking, because most items are nonintuitive and there is no favorable or unfavorable answer. Furthermore, some biodata items are verifiable; therefore, applicants are less likely to exaggerate their responses (Mitchell, 1994). Finally, some studies found that biodata predictors can account for incremental variance in criteria beyond that accounted for by general mental ability and personality measures (e.g., Allworth & Hesketh, 1999; McManus & Kelly, 1999; Mount et al., 2000). Mount et al. constructed biodata scales that reflected the five factor model personality factors as well as general mental ability. All of the biodata scales correlated significantly with one or more personality factors. They concluded that biodata overlap with both, personality and general mental ability. Therefore, it might not be unreasonable to argue that various performance criteria might be better measured by a biodata scale because biodata include "wider aspects of personality and motivation than behavior alone"(Gunter, Furnham, & Drakeley, 1993, p. 2). Similarly, Stokes (1999) asserts that although biodata differs from other measures such as personality, interests, values, and abilities, it can capture constructs in all four domains.

   Criticism of Biodata. The failure to establish construct and content validity is one of the most common criticisms of biodata (Stokes & Cooper, 2001). There is significant

concern among practitioners and researchers over what is actually being measured by biodata.

Since the I/O literature seems to emphasize application of biodata mostly for selection purposes, most studies focus on criterion validity (Mumford, 1999). While the focus on criterion validity is useful, it does little to advance our knowledge and understanding of the relationship between the biodata and the criteria that are predicted. To do so, it is necessary to collect evidence that will allow us to draw causal inferences from the test scores. According to Mumford (1999) even high validity coefficients "do not allow us to draw the inference that the attribute being measured is, in fact, the cause of performance" (p. 118).

Another common criticism is that biodata is often not generalizable across context. Generalizability refers to the extent to which research findings and conclusions from a study conducted on a sample population can be applied to different persons, times, and settings (Cook & Campbell, 1979). In response to this criticism, Rothstein, Schmidt, Erwin, Owens, and Sparks (1990) showed that it is possible to develop biodata items that are not organizationally specific. Additionally, they found that external validity was not moderated by gender, race, age, education, tenure, or previous experience. Rothstein et al. (1990) pointed out that large sample size and utilization of multiple organizations can contribute to the generalizability of the results.

In contrast, Carlson, Scullen, Schmidt, Rothstein, and Erwin (1999), who built on Rothstein et al.'s (1990) study, suggested that multiorganizational sample and keying is not required for generalizability. Carlson et al. (1999) argue that there are four factors

9

that influence generalizability. The first factor is a sound reason to believe that the validity of the instrument will generalize to other populations. For example, their scale was designed to assess general managerial ability, not unique skills. The second factor is selection of valid criterion measures. The third factor is establishing validity on the item level as opposed to the scale level. Finally, adequate sample size is needed in order for a study to be generalizable. This is consistent with the previous study. Just for illustration, Rothstein et al. used over 10,000 participants and Carlson et al. used over 7,000 participants. Interestingly, empirical keying was used in both studies, which is in direct contradiction with previous assertions that this type of keying will impede generalizability. Perhaps future studies could compare the generalizability of biodata using both scaling methods.

Response distortion or faking has been a criticism of many non-cognitive predictors, including biodata. Although the extent to which faking influences criterion-related validity is not known, there is some evidence that faking influences validity. For example, Pannone (1984) found that faking on rationally developed scales introduced error variance that reduced validity coefficients. In addition, Douglas, McDaniel, and Snell (1996) found that faking had an impact on construct validity, criterion validity, and external validity. Douglas et al. (1996) argue that internal consistency becomes artificially high and construct validity diminishes when applicants fake answers consistently across items. They also found that faking reduces criterion validity. When applicants fake, their scores are not predictive of their job performance. This finding is consistent with Graham, McDaniel, Douglas, and Snell (2002). Lastly, Douglas et al.

(1996) indicate that the unexplained variance in generalizability studies can be "attributable to differences across studies in the proportions of applicants who are faking" (p. 130). Validity coefficients will be higher if a smaller percentage of applicants is faking. Conversely, if a larger percentage of applicants fake, validity coefficients will be lower.

Since faking can have an effect on predictive validity, it is imperative that practitioners and researchers try to limit the amount of faking on biodata measures. Lautenschlager (1994) provides several useful guidelines on how to minimize response distortion on biodata scales. He recommends including items or scales that would help us in detecting distortion. For example, Pannone (1984) included fake questions and Stokes and Cooper (2001) used a social desirability scale and impression management scale to uncover faking. It might also be useful to repeat some items to check on the reliability of responding. Obviously, this would be useful only in measures of considerable length. Another possibility is to warn the applicants that their responses will be verified. Kluger and Colella (1993) found that such warnings reduce the inclination to fake. However, warning probably is only credible when the biodata scale is composed mostly of factual items that can be verified.

Interestingly, some researchers argue that due to the non-intuitive nature of biodata (i.e., the connection between item content and the purpose of the test is not apparent to the applicant), they are less prone to faking in comparison to other non-cognitive measures, such as personality inventories (e.g., Mitchell, 1994). However, if items are written so that they are highly job-relevant (and thus face valid), the desirable

answer becomes apparent to the applicant. Items that are more job-relevant, and less historical, objective, discrete, verifiable, and external are more likely to be faked (Becker & Colquitt, 1992).

One potential problem arises with the use of non-intuitive items. If items are non-intuitive, applicants will be less likely to understand the relationship between the biodata measure and their job success. Elkins and Phillips (2000) found that the perceived job-relatedness of biodata influences applicants' perceptions of fairness. Practitioners wishing to use biodata must decide whether to use face valid items that are more prone to faking or non-intuitive items that may raise issues of fairness for some applicants.

Evidence of the predictive power of biodata. There is strong evidence that biodata can predict job performance on subjective measures. For example, Reilly and Chao (1982) in their meta-analysis reported that the best three predictors of entry-level performance (as measured by supervisory ratings) for various jobs were cognitive ability tests (average validity coefficient .53), work samples (average validity coefficient .44), and biodata (average validity coefficient .37). Based on four studies, the average validity coefficient between biodata and supervisory ratings for sales jobs was .42. The correction for range restriction was performed only for cognitive tests and not the other predictors. It is important to note that this correction most likely inflated the validity coefficient for the cognitive tests, which could lead to unfair comparison with other predictors.

Another study that utilized subjective performance measures (e.g., supervisory ratings) was carried out by Schmitt, Gooding, Noe, and Kirsch (1984). They found that

the best predictors of job performance across six different occupational groups were assessment centers (validity coefficient .43), work samples (validity coefficient .32) and biodata (validity coefficient .32). Interestingly, the validity coefficient for cognitive tests was only .22. These investigators corrected for sampling error but not range restriction.

Stokes, Toth, Searcy, Stroupe, and Carter (1999) constructed a biodata scale containing ten subscales in order to predict overall performance of retail and wholesale salespeople. Performance was measured as a composite of supervisory ratings on four dimensions: positive workers behavior, job knowledge, worker productivity, and sales ability. The biodata scales that assessed leadership (.17), dependability (.20) and ability to multitask (.17) were most predictive of performance.

A small number of studies, which have used mostly entry-level employees as participants have demonstrated that biodata can predict objective performance measures. Schmitt, Gooding, Noe and Kirsch (1984) stated that biodata was the only predictor that reached a modest validity coefficient against productivity (.20). Schmitt et al. (1984) did not indicate how productivity was measured, but it was presumably an "objective" measure. Biodata has also been found to predict objective performance in terms of sales volume. For example, Vinchur, Schippmann, Switzer and Roth (1998) conducted a meta-analysis in which they reported a validity coefficient between biodata and sales volume to be .17 for various sales positions. The biodata measures included in their study were quite heterogeneous and included items pertaining to age, job experience, grades, and club membership. Vinchur et al. (1998) also found that sales volume was predicted by a sales ability measure (.21, corrected for range restriction). The sales ability measures

included items pertaining to knowledge of selling techniques. Based on four studies, Reilly and Chao (1982) reported an average validity coefficient between biodata and productivity in sales jobs of .50. Reilly and Chao (1982) did not indicate how productivity was measured, but it was probably at least somewhat "objective." It is plausible that productivity was measured on the basis of some kind of production records. Even if productivity was assessed in terms of supervisory ratings, these ratings were most likely based on more objective record of performance. Dalessio and Silverhart (1994) reported that biodata made a marginal contribution to the prediction of an objective measure, in this case a monthly sales commission of life insurance agents. Although the contribution of biodata to the prediction was not statistically significant ($p<.08$), Dalesio and Silverhart (1994) state that a one point increase in the biodata score resulted in an increase of about \$215 in monthly commissions.

Approaches to biodata development. There are four approaches to biodata scale development: subgrouping, factorial, empirical and rational (also referred to as construct-oriented). When using the subgrouping technique, the patterns of life history are identified. The "patterns can be used to understand the developmental history of an individual and to predict future behavior based on his or her subgroup's pattern of past experiences and behavior" (Hein & Wesley, 1994, p. 172). The subgrouping approach is appropriate under three conditions. First, it is the best approach if one wants to predict multiple criteria. Second, a large sample size of at least three hundred is needed to assure that the resulting subgroups are adequately large. A large sample size is also needed if cross-validation is performed. Lastly, it is best used in situations when a non-linear

14

relationship is anticipated between the criteria and the predictor. If different patterns of abilities are capable of contributing to successful performance so that more than one type of person will be an effective performer, the subgrouping technique is appropriate.

In the empirical keying approach "biodata items are selected and weighted based on their empirical relationship to the criterion" (Mount et al., 2000, p. 300). Responses to the biodata items that have a strong relationship to the criterion receive a higher weight than those responses that have a weaker relationship with the criterion. There are two major criticisms of this approach. First, the generalizability of a scale developed in this manner is limited. The scale typically predicts a specific criterion for a specific job. Second, empirical keying does not further our understanding of the predictor-criterion relationship, since the items may not appear related to the criterion (Hogan, 1994).

As with other measures, biodata scales often include many items. The use of a factor-analytic approach stems from the desire to reduce the number of items. Factor analysis is a multivariate statistical procedure that reduces biodata items into limited number of clusters. It is a method of attaining biodata dimensions from item responses. This approach is often combined with the rational approach. A movement toward the rational approach in biodata development is evident in the psychological literature. When utilizing this approach, it is necessary to first identify and define the constructs underlying performance. The constructs can be defined on the basis of a job analysis or a review of the literature (Stokes & Cooper, 2001). Items are then developed to represent constructs considered critical for successful performance. According to Stokes and Copper validity is not sacrificed when a rational approach to item development is used.

Conclusions about the effectiveness of rational, empirical-keying, and factor-analytic biodata scale development strategies are inconsistent. Stokes and Searcy (1999) constructed scales using rational, empirically keying, and factor-analytic approaches and compared their ability to predict sales performance and overall performance. They concluded that rational scales were as predictive as empirical keying scales. Factor-analytic and rational scales have predicted several customer service criteria better than empirical keying (Schoenfeldt, 1999). A meta-analysis found similar levels of criterion-related cross-validities across the three biodata development strategies (Hough & Paullin, 1994). In the data set utilized in the present research, the assessors opted for the rational approach. This choice could be justified on the following grounds: 1. empirical keying is criticized as atheoretical; 2. subgrouping is not appropriate unless the sample size is very large and 3. the need for reducing the number of items was not anticipated.

From the evidence cited above, it seems that biodata can be predictive of subjective as well as objective measures. Since the validity coefficient reported by Reilly and Chao (1982), was larger when objective measure of performance was used, it is expected that biodata will be predictive of objective measures of performance used in the present study.

Situational Judgment Tests (Low-Fidelity Simulations)

Situational judgment tests belong to a broader category of work sample tests. Work sample tests can be ordered along a continuum where at one end there are situational judgment tests or low-fidelity simulations and on the other end are high-fidelity simulations, such as flight simulators or role-plays (Wood & Payne, 1998). In

low-fidelity simulations the candidates respond to a hypothetical situation they may encounter on the job. In contrast, high-fidelity simulations are more realistic in their approximation of the job situation, and the candidate's response is highly similar to what it would be on the job.

Situational judgment tests (SJTs) have been around for many years in several forms. Some of the formats of situational tests have been situational interviews (Latham, Saari, Pursell & Campion, 1980), assessment centers (Thorton & Byham, 1982), video-based situational tests (Chan & Schmitt, 1997), and paper-and-pencil situational judgment tests (SJTs). One of the first written situational judgment tests appeared in the 1920s (Moss, 1926). The George Washington Social Intelligence test contained a subtest called Judgment in Social Situations. This multiple-choice test included items that required the participants to respond to hypothetical work and social situations. Since then, many more SJTs assessing supervisory potential have been developed. The next section of this document presents theories about why SJTs work and includes several examples of SJTs.

Why SJTs predict performance. Most of the opinions about why SJTs work fall into one of two categories. First, some theorists suggest that SJTs are related to other constructs (i.e., cognitive ability, job experience) that are themselves predictors of performance. Others believe that situational judgment is a unique construct.

Several studies suggest that SJTs are related to cognitive ability. Cardall (1942) constructed the Practical Judgment Test, which contains multiple-choice items that describe common everyday business as well as social situations. The participants in this

study were asked to indicate which would be the best action to take in each of the given situations. Despite Cardall's conclusion that judgment was a unique factor, independent of other factors such as intelligence, other researchers noted that the Practical Judgment Test significantly correlates with tests of general intelligence (e.g., Carrington, 1949).

A SJT called How Supervise? was designed to measure supervisors' knowledge of and perceptiveness regarding interpersonal relations (File & Remmers, 1948, 1971). The test items concern difficult situations that a supervisor faces on daily basis; study participants were asked to indicate if they were agreed, disagreed, or were uncertain about a particular statement. File and Remmers (1971) report that studies done using the How Supervise? SJT showed a significant relationship between performance ratings and How Supervise? scores. They also report studies that found a relationship between cognitive abilities tests and the scores on the How Supervise? SJT. For example, Millard (1952) concluded that How Supervise? measures intelligence after he found that the correlations between this test and a test of general mental ability is between .62 and .71. Others reported this correlation to be around .45 (McDaniels et al., 2001; Weekly & Jones, 1999).

Borman et al. (1993) found performance on SJTs to be related to job experience. Chan and Schmitt (1997) supported this by arguing that finding a solution or handling a problem requires various skills and abilities. Therefore, judgment is multidimensional in nature.

Other researchers believe that situational judgment is a unique construct. In their view, SJTs measure tacit knowledge - a construct often described as "street smarts."

Sternberg (1997) argues that tacit knowledge is a crucial component of overall

managerial intelligence. Two other components that are needed for success in

managerial positions are analytical intelligence as measured by traditional cognitive

ability tests and creative intelligence. Sternberg, Wagner, Williams, Horvath (1995)

claim that tacit knowledge is an ability to solve real-world problems that is independent

of cognitive ability. They argue that while tacit knowledge develops based on

experience, the two constructs are separate because not all people are capable of

acquiring tacit knowledge from their experience. Wagner and Sternberg (1991)

published a measure of tacit knowledge called the Tacit Knowledge Inventory for

Managers (TKIM). The purpose of the measure is to identify individuals who could be

successful as managers or executives. The scenarios included in the TKIM are more

detailed than the scenarios included in previously mentioned tests.

Advantages of SJTs. SJTs offer several advantages in comparison to selection

methods such as assessment centers, structured interviews, and cognitive ability tests.

The first and most obvious advantage is the low cost of SJTs. SJTs are also easy to

develop, administer, and score.

Another positive aspect of SJTs is that, unlike cognitive ability tests, SJTs are not

problematic when it comes to adverse impact. Motowidlo and Tippins (1993) found

gender differences in SJT results to be less than a third of a standard deviation. The

differences in mean scores between racial groups are also smaller in comparison to

cognitive tests. Motowidlo et al (1990) found standardized mean differences between

Caucasians and African Americans to be .14 and .29 in two different samples. A study

by Pulakos and Schmitt (1996) showed that the reported standardized mean difference between Caucasians and African Americans was .41 and .02 between Hispanic Americans and Caucasians. This finding is consistent with Clevenger, Pereira, Wiechmann, Schmitt, and Harvey (2001) who report a mean difference between Caucasians and African Americans of .37 and .01 between Hispanic Americans and Caucasians.

The items in SJTs commonly present applicants with problem work situations. The applicant is required to choose the one response that he/she would be the most likely to make from several alternatives. Therefore, it can be argued that, unlike cognitive tests which typically assess maximum performance, SJTs measure likely job performance, which is another advantage of SJTs.

Criticism of SJTs. The only possible problem with SJTs cited in the literature is that they may not always accurately reflect actual work conditions (Motowidlo et al., 1990). High-fidelity simulations are typically more representative of the actual work conditions and demands.

Evidence of predictive power of SJTs. Several research studies have shown that SJTs can predict performance, at least in terms of subjective measures. Motowidlo, Dunnette, and Carter (1990) concluded that SJTs are a valuable tool in predicting job performance. Their study measured performance in terms of supervisory ratings on 10 performance dimensions. Motowidlo et al. (1990) referred to the SJT they developed as a "low-fidelity simulation" because the candidates who participated were presented with hypothetical situations that only approximated actual job stimuli. In this study,

applicants for entry-level management jobs were presented with descriptions of work

situations and were instructed to choose from among several possible responses for each

situation. The participants chose both the response that they would be most likely to

make and also the response that they would be least likely to make. The correlations

between the test and several job performance dimensions ranged between .28 and .37.

Motowidlo et al. (1990) reported that the SJT is not related to cognitive ability.

However, this assertion has to be viewed cautiously since the participants were selected

based on their aptitude scores and academic achievement, making restriction of range a

likely problem.

In a follow up study, Motowidlo and Tippins (1993) found that their test had a

predictive validity of .25 and concurrent validity of .20. Again, supervisory ratings were

used as a criterion, which is very common. McDaniel, Morgeson, Bruhn-Finnegan, and

Campion (2001) included 39 different SJTs within the total sample of more than 10,000

in their meta-analysis, most of which used supervisory ratings as a criterion. The

reported validity coefficient was .34, not corrected for range restriction. This level of

validity is comparable to the validity coefficients of assessment centers, biodata

measures, and structured interviews (e.g., Schmitt & Hunter, 1998).

A conclusion from a study carried out in the Center for Creative Leadership is

that situational judgment, as assessed by a tacit knowledge measure is the best predictor

of managerial performance. Managerial performance was judged by observers who

provided ratings of eight dimensions of performance (i.e., influencing others, task

orientation, verbal effectiveness) (Wagner &, Sternberg, 1990).

The only evidence that SJTs can be useful in predicting objective measures comes from another study by Wagner and Sternberg (1985). They found a correlation of .48 between the level of tacit knowledge and increases in merit salary. They also reported a correlation of .56 between the level of tacit knowledge and the amount of business branch managers generated for a bank. It is important to keep in mind that the studies conducted by Sternberg and colleagues included participants such as Yale undergraduate students and business managers; this is likely to have imposed a range restriction on measures of cognitive ability. Thus, the high correlation of tacit knowledge and performance measures might be a result of high level of cognitive ability among the participants.

It can be concluded that SJTs are better at predicting objective criteria than they are of subjective criteria. Although, there is a greater number of studies that successfully used SJT for the prediction of subjective measures, the validity coefficients are larger when objective measures were used.

Role-Plays (High-fidelity Simulation)

High-fidelity simulations do not represent an exact method or procedure. They can vary in complexity, number of participants, type of interaction (i.e., subordinate, customer, peer), or bandwidth, which is "the degree to which the entire job performance domain is represented by the tasks" in the simulation (Callinan & Robertson, 2000, p. 256). High-fidelity simulations such as role-plays, leaderless group discussions or in-basket exercises are typically a part of assessment centers. As a stand-alone method, role-plays are not frequently discussed in the selection literature.

Why role-plays predict performance. High-fidelity simulations such as role-plays

22

approximate actual work conditions more closely than low-fidelity simulations. Therefore, it can be assumed that high-fidelity simulations are better predictors of job performance (Motowidlo et al., 1990). High-fidelity simulations such as work sample tests have been found to be the single best predictor of job performance in comparison with 18 other selection procedures (Schmidt & Hunter, 1998). Schmidt and Hunter (1998) reported average validity for the work sample tests to be .54. The average validity coefficient for cognitive tests was reported to be .51. Both correlations were corrected for error in the criterion measure. The majority of the studies included in this meta-analysis measured job performance by subjective measures, namely supervisory ratings.

Advantages of role-plays. In addition to the high predictive potential discussed above, simulations have three more advantages. First, they are highly job-relevant and therefore are perceived as fair. Court cases involving simulations are rare in comparison to other selection methods. Furthermore, their use has been defended in six out of seven reported cases (Terpstra, Mohamed, & Kethley, 1999). The lack of cases challenging the use of simulations may be due in part to the second advantage, which is the fact that simulations have a low incidence of adverse impact on members of minority groups. When the scores of African-American candidates were compared to the scores of Caucasian applicants, the average effect was -.38 standard deviations (Schmitt, Clause, & Pulakos, 1996). Lastly, simulations can serve as realistic job previews, which can lead to self-selection by candidates.

Criticism of role-plays. The only criticism of high-fidelity simulation is that they are often complex and thus expensive to develop (Motowidlo et al., 1990). Motowidlo et

23

al. (1999) points out that the increase in fidelity is sometimes not worth the additional cost over other assessment methods.

Evidence of predictive power of role-plays. Prior research provides some evidence that role-plays can predict job performance in terms of objective as well as subjective measures. Squires, Torkel, Smither, and Ingate (1990) found that a participant's score on a 30-minute telephone role-play predicted the percentage of sales quota the participant would reach on the job; the percentage of sales quota reached served as an objective criterion. The sole objective of the role-play was for the telemarketing representative to sell a service contract for appliances. The validity coefficient for the percentage of sales quota reached was .31. The role-play also predicted supervisory ratings on three dimensions of performance: sales results (.33), sales skills (.38), and customer service (.39). The ratings of the supervisors might have been influenced by their knowledge of the participants' actual sales compared to quota.

O'Connell et al. (2002) found that a short role-play developed to assess the sales skills of retail salespeople predicted sales performance (.23). Sales performance was measured using the actual sales of each salesperson over a period of six months. The role-play also demonstrated incremental validity. It explained 4% more variance after biodata was entered into the regression.

Both of the studies described above used simple role-plays since the jobs were not very complex. However, role-plays might be very useful even when assessing candidates for more complex positions. Although role-plays are complex and expensive to develop, they can be of help in predicting performance in many situations. The use of simulations

in research and practice will enhance our understanding of complex managerial skills. Unlike paper-and-pencil tests, simulations provide the opportunity to measure more complex skills such as communication, building relationships, and leadership (Cleveland & Thornton, 1990). These skills are often referred to as competencies in performance appraisals and selection procedures for managerial positions. While role-plays might be suitable in the assessment of managerial skills, the studies described above used lower level employees as participants.

Although there are only a few studies, it is evident that role-play can be predictive of subjective and objective performance measures. The validity coefficients in the study conducted by Squires et al. (1990) were only slightly higher for the subjective criteria; therefore, it seems reasonable to argue that role-plays will be predictive of objective performance measures used in the current study.

Rationale for Skills and Competencies Used in Current Study

The predictors used in the current study were designed to capture skills and competencies critical for successful performance in the District Sales Manager position. It is evident from the review above that biodata, SJTs, and role-plays can predict a candidate's ability to perform on many job criteria, including objective performance criteria. Therefore, it is not unreasonable to argue that biodata, SJTs, and role-plays that tap into various performance dimensions deemed critical for success on the job can predict performance on objective performance criteria. Sales managers must possess sales ability as well as other skills relevant to all managerial positions. People skills and business skills are paramount for effective performance of all managers. It is without

question that people skills are at the center of every manager's job (Morand, 2001).

Boyatzis (1982) arrived at an integrated model of skills clusters needed by leaders at all

organizational levels. Boyatzis states that a skill cluster that is most important to middle-

level management is people skills. Few would disagree that, with competitive pressures,

accelerated change, and challenging economic conditions, managers at all levels need

sophisticated business management skills. Wood and Payne (1998) identified 12

competencies that are most commonly used in employee selection and development in

the UK. Among those 12 competencies, planning and organizing, analytical thinking,

result orientation, and business awareness can be categorized as business skills. Others,

such as leadership, developing others, building relationships, and communication, fall

into the people skills category.

The biodata scale and SJT in the present study were designed to capture selling

ability, people skills, and business skills. The role-play is designed to capture ten

competencies similar to those identified by Wood and Payne (1998). The following

competencies were assessed in the role-play: building relationships, leadership,

communications, prospecting/recruiting, results orientation, selling,

training/development, planning, analytical skills, and business management/judgment.

Purpose of the Study

This study contributes to the understanding of objective performance criteria in a

number of ways. First, it may be valuable to know if any of the four objective criteria

used in this study would yield acceptable validity coefficients. Organizations that

conduct validation studies are usually able to choose from a number of criteria. Since

26

organizations must follow technical and legal standards when conducting validation

studies, the degree to which job-related criteria correlate with valid predictors of

performance could affect the choice of criteria used in validation studies.

Second, various objective measures could be predictive of a different set of

performance dimensions. Investigating the predictive power of certain competencies and

skills could lead to a better understanding of employee performance. For example, some

competencies such as recruiting and training might be more predictive of how well one

performs in quantity of new recruits while result orientation might be predictive of the

amount of sales.

Third, the literature review shows that biodata, SJTs, and role-plays can predict

subjective performance. However, with the exception of SJTs, it has not been

demonstrated that these selection tools can predict managerial performance in terms of

objective criteria. Most studies provide some evidence that these tools can be predictive

of objective criteria in low-level positions.

Finally, it is not known which one, if any, of the selection tools under discussion

is more efficient in predicting objective criteria in the Sales Manager job title. This study

attempts to compare biodata scale, SJT, and a role-play in terms of their power to predict

objective performance measures of District Sales Managers.

Method

Participants

The sample consisted of 189 District Sales Managers from an international

company engaging in direct selling of beauty and related products. The District Sales

Managers in the company are mostly women representing diverse age and ethnic groups throughout the United States. The District Sales Managers' primary responsibilities include growing sales, which is accomplished by recruiting, motivating, and developing new sales representatives. Detailed position descriptions are included in Appendix A.

Development of Predictor Measures

Job analyses. A job analysis of a District Sales Manager's position was conducted to identify and verify the tasks performed by District Sales Managers, identify and verify competencies required for effective performance as a District Sales Manager, and develop items and simulations that accurately reflect the demands and situations encountered on the job. Information for the job analysis was collected from the following sources: 1. Documents pertaining to the job; 2. Site visits and interviews with District Sales Managers, Division Sales Managers, senior managers, and Human Resource Managers; 3. Job analysis questionnaires completed by subject matter experts (SMEs); and 4. SME meetings to review the collected information. Incumbents who have been in the District Sales Manager position for at least two years served as SMEs. The SMEs were asked to review task lists and evaluate job relatedness. Each SME provided ratings of the importance of each task to overall success for the District Sales Managers. The SMEs were also asked to select and edit biodata and situational judgment items from a drafted item list. They assigned weights for the biodata and SJT items. The scores for each of the biodata and situational judgment items ranged from one to four. The response judged as the most preferred one by the SMEs was scored at four points and the response judged to be the least preferred was scored at one.

Based on the SMEs evaluation of the tasks, a set of competencies was developed. Then, the SMEs rated the extent to which each competency was essential to effective performance in each call scenario. Due to time constraints and job complexity, only those competencies necessary for minimum performance on the job were included.

The SMEs identified a set of 10 competencies and 22 underlying skill dimensions (Appendix B). The SMEs assigned points to each competency reflecting relative importance of that competency. The total number of points assigned was to equal exactly 100. Then, the average assigned weight was calculated for each competency. For computational simplicity, the assigned weights were transformed such that mean weight was one. That is, if a behavior was neither more nor less important than the other behaviors, the weighted behavior score would remain unchanged. It was determined that the building relationship dimension is the most critical one and therefore was weighted at 1.4 while the remaining competencies were weighted equally at 1.0.

Evaluation Guidelines. Inferences about the candidate's skills were based on the behaviors elicited in the call scenarios. Therefore, detailed guidelines were prepared in order to standardize the evaluative inferences drawn from the behaviors. A list of 550 behaviors was classified into three skills levels: less than adequate, adequate, and more than adequate. Then, the SMEs listened to taped samples of each call scenario and then reviewed the behaviors associated with that call. They were asked to reclassify any behaviors that they felt were misclassified into the wrong skill level. A total of 546 behaviors were retained. These behavioral examples provided a concrete set of evaluation guidelines for assessors. For every call, there were one to seven behavioral

29

examples for each level of each of the 22 skill dimensions. By comparing behaviors demonstrated during a call scenario to the sample behaviors, the assessors were able to determine to what skill dimension and competency that behavior relates and to infer what level of effectiveness that behavior represents.

Description of predictor measures. This study uses three independent variables generated on the basis of the job analysis described above. The first independent variable consisted of 19 biodata items. The items were developed to tap business, people, and selling skills. The items pertained to various experiences such as selling, budgeting, and handling inventory, and interacting with people. An example of a biodata item is: "Which statement best describes your business experience?"

The second independent variable consisted of 11 situational judgment items. In the SJT, participants were asked to select the scenario representative of the best approach to a specific situation from the four choices given. The items were concerned with candidate's judgments in various situations such as dealing with people from different cultural background. The items were developed to tap business, people, and selling skills. An example of a situational judgment item is: "You are speaking to a group of Sales Representative candidates. One candidate is very interested but must speak to her family before deciding. Which of the following describes the best approach?" The biodata and SJT were administered at the same time. More examples of biodata and situational judgment items are included in Appendix C.

The third independent variable is a telephone role-play designed to assess 10 competencies critical for effective performance of the job of District Sales Manager. A

total of five call scenarios (four evaluated calls) were developed to adequately measure

each competency and present a representative set of situations encountered by District

Sales Managers. Table 1 shows which competencies were assessed during the each call

scenario.

   Assessor Training. All 10 assessors had advanced graduate training in

psychology or related social sciences. The assessors reviewed the competency model

and familiarized themselves with the information in the candidate's materials. Next, the

assessors reviewed the written role-play instructions for each simulation call and then

listened to a taped sample of that call. Assessors played various roles with each other

and the trainers to gain practice in playing the roles. Feedback was given on how the

roles were being performed. The assessors also practiced to ensure that the roles were

being played in a standardized fashion across assessors. The assessors reviewed the

behavioral dimension rated on each call and read through the evaluation guidelines for

each of these dimensions. Then, they listened to taped calls composed specifically for

training purposes. Assessors focused on observing, recording, and evaluating candidate

behavior relevant to each designated dimension and competency. Assessors rated the

taped calls and the ratings given by the assessors were compared with master ratings that

had been determined to be accurate for each simulation call by the Assessment Manager,

Training Manager, and the developers of the role-play. The ratings of the call samples

were discussed with the Training Manager. Behaviors exhibited in each call were

reviewed until consensus was reached with regard to how each behavior should be rated.

This process ensured that the assessors learned to apply similar standards across

candidates and that each assessor was using the standards similarly.

Candidate evaluation. Between two and three assessors independently evaluated each candidate. While role-playing, the assessors recorded behavioral observation and then utilized the evaluation guidelines in order to determine a rating for each skill dimension assessed on the call. Ratings were provided on a 5-point scale (1 = Less than Adequate, 3 = Adequate, 5 = More than Adequate). When calls were completed, the ratings for all skill dimensions within each competency were averaged within each of the rated calls. Next, an overall competency rating was calculated by averaging the ratings for the competencies across calls. Finally, competency ratings were averaged to arrive at a composite evaluation.

Quality control of the assessment process was ensured by conducting assessor rating reviews and by monitoring the calls. After each assessment was completed, ratings were reviewed by Quality Assurance Senior Assessors. Ratings were compared to assessor notes on each call. When discrepancies were found, the Quality Assurance Senior Assessor reviewed the ratings with the assessor and agreement was reached on the final rating. The management staff also monitored calls randomly to assure the quality and consistency of the role-plays. The ratings assigned on the monitored calls were also examined to ensure that the ratings were assigned in accordance with the evaluation guidelines.

Description of Criterion Measures

The measures used in this study capture objective performance data, which included sales amount, number of orders, debt, and active sales staff in the district (i.e.,

the number of sales representatives recruited minus the number of sales representatives terminated). Furthermore, each criterion measure is captured in terms of target figures, actual figures, and the percentage difference. The company has a thorough plan for setting target figures for each District Sales Manager. The target figures are based on qualitative and quantitative information. The quantitative part is derived from the company's source of sales model, which is based on sophisticated financial analysis. The average growth derived from this model for each district in 2003 was 6.3%. However, each Division Sales Manager (a position one level above the District Sales Manager) was able to override this percentage based on market opportunity within the district as defined by him or her. Thus, the range of growth was approximately 2-11% in 2003. The actual figure reflects the accomplishment of each District Sales Manager within a given year. Debt refers to the amount of money that was not collected for the orders in the district. In addition, target and actual sales to debt ratios were used in the analysis. The sales to debt ratio equals current year debt divided by current year sales, multiplied by 100. A lower ratio is more desirable.

Procedure

The participants were provided with a paper copy of the biodata scale and the SJT. They completed the inventory at their convenience. They recorded their choices on a provided answer sheet, which was to be returned via mail or fax. The participants were also required to return the inventory booklet via mail. In the instruction part of the inventory, they were asked to provide information that was as accurate as possible. The instructions also notified the participants that their answers regarding work experience

33

and history would be verified in the subsequent steps of the selection process.

After the biodata scale and the SJT were returned, the participants were asked to schedule a two-hour session to complete the role-play. They participated in role-play at their homes via their telephones. The participants scheduled a time that allowed them to complete the exercise without distractions. They were to prepare a place where they were able to spread out papers, write, and talk on the telephone. Scratch paper, calculator, pens/pencils and candidate's instructions were to be at hand.

The instructions were sent via mail. The instructions described, in general terms, the role the candidate would be playing and the fictitious company that candidate would work in during the role-play. It also presented a schedule of the program's activities, stating when calls and preparations time would occur. Finally, it provided an overview of all of the remaining support materials enclosed in another envelope that was not to be opened prior to the telephone assessment. A list of the support materials made available to the participants is found in Appendix D.

The candidates received a series of five phone calls. The first phone call was from a supervisor (not evaluated). The second phone call was from a current customer. The third phone call was from a current sales representative. The fourth phone call was a conference call with two current sales representatives. Finally, the last call was again from their supervisor. An example of the call scenarios is presented in Appendix E.

<div align="center">Results</div>

The data set was examined for normality of distributions, homogeneity of variance, and outliers. Internal consistency of the measures was assessed. Reasonable

<div align="center">34</div>

internal consistency of the measures was expected, as both have been developed to assess specific constructs. Coefficient alpha for the role-play is acceptable at .87. The biodata yielded an alpha coefficient of .62 and SJT yielded an alpha coefficient of .17. Lower internal consistency coefficients are usually acceptable in biodata scales since they are heterogeneous in nature (Mumford, Costanza, Connelly, & Johnson, 1996). However, the internal consistency of the SJT is not at an acceptable level. A closer look at the biodata total inter-item correlations shows that most of the items are at least somewhat correlated. The correlations range from -.03 to .53 most of which correlate at around .30. The inter-item correlations for the SJT reveal that the SJ items are not strongly correlated. In fact, most of the correlations are around zero and the highest correlation is only .15. The total inter-items correlations for the role-play range from .40 to .70 and only three of them are below .50. This suggests that both biodata and especially the situational judgment items are measuring multiple constructs.

All variables were transformed into z-scores. A total of 38 (0.7%) data points were removed from the analysis as they were found to be outliers. Outliers were defined as having a z-score of at least plus or minus three (Barnett & Lewis, 1984). An outlier was removed only for the pertinent part of the analysis (casewise removal). Descriptive statistics (mean and standard deviation) for the selection tests and for the performance measures were obtained. The means, standard deviations (reported in raw numbers), and intercorrelations of the variables examined in the study are presented in Table 2. The table shows that most target and actual figures for each objective measure are strongly correlated (r = .94 for sales, r = .70 for debt, r = .67 for sales to debt ratio, r = .93 for

orders, $\underline{r}$ = .90 for staff). The company's planning system (described above) is carefully devised; therefore, this result is not surprising.

The number of active staff is correlated with sales ($\underline{r}$ = .67), number of orders ($\underline{r}$ = .87), and debt ($\underline{r}$ = .25). This is to be expected as sales, orders, and debt tend to be higher with a larger sales force.

A composite variable of managerial effectiveness was created by adding actual figures of sales, orders, and staff. Actual debt was subtracted from the sum. As anticipated, this variable is correlated with most of the criterion measures. There are several negative correlations between the role-play and the criterion measures. Specifically, the role-play is negatively correlated with actual orders ($\underline{r}$ = -.16), actual staff ($\underline{r}$ = -.18), target staff ($\underline{r}$ = -.15), and overall managerial effectiveness ($\underline{r}$ = -.17). The negative correlations above were not anticipated. The discussion section explores the possible reasons for these results. None of the objective measures are related to the biodata items and situational judgment items. There is a modest but statistically significant correlation between the SJT and the role-play ($\underline{r}$ = .16). Given that they are both work sample tests requiring the candidates to make a decision about work situations, some relationship is to be expected. The correlation between role-play and biodata did not reach statistical significance ($\underline{r}$ = .12). There is also no significant relationship between the biodata scale and SJT ($\underline{r}$ = -.10), which supports the notion that they are empirically distinct.

Table 3 depicts intercorrelations among the role-play competencies and biodata scale and SJT. There are no statistically significant correlations between the biodata

scale and the role-play competencies. However, the SJT correlates with building

relationships ($\underline{r}$ = .21) and communication ($\underline{r}$ = .16). The biodata scale and SJT were

designed to measure sales, business, and people skills and a certain amount of overlap

with the role-play competencies was expected. Building relationships and

communications dimensions are related to people skills as measured by the SJT.

Correlations between the role-play competencies and the criterion measures can

be found in Table 4. As the table shows, 13 correlations are statistically significant.

However, all of them are in the opposite direction than was expected. Note that positive

correlations with debt and ratio figures are not desirable. There is a total of 150

correlations between the criterion measure and the role-play competencies. With a two-

tailed significance test at the 5% level, one would expect to obtain 15 correlations by

chance.

<div align="center">Discussion</div>

Overall, the results of this study did not support the notion that biodata scale, SJT,

or role-play simulations are effective predictors of objective measures of performance.

Several statistically significant correlations were found. However, all were counter to

what was expected. The reasons for these results are not entirely clear.

There are several possible explanations for the null findings with the objective

measures. First, Cook (1998) suggests that the sales criterion is less reliable than

subjective criteria because it is very complex and might be affected by factors other than

the manager's ability. Although the sales potential of the district has been taken into

consideration, the work of a District Sales Manager's is interdependent. The sales,

<div align="center">37</div>

orders, and debt are regulated by individual sales representatives and thus are outside of

the District Sales Manager's direct control. For this reason, sales, orders, and debt within

regions are not good measures of the District Sales Manager's individual productivity.

Most of the studies cited above have used criteria that more directly measured the

participants' performance such as increase in merit salary (Wagner & Sternberg, 1985) or

percentage of sales quota reached (Squires et al., 1990). Moreover, some of the District

Sales Managers were hired during the year in which the data collection occurred. Thus,

they were only able to influence the outcomes for several months. Their sales, orders,

debt, and number of active staff were adjusted (e.g., prorated) to reflect the entire year.

Second, the objective measures used in this study might be dynamic criteria. A

dynamic criterion is one that changes over time. Ghiselli and Haire (1960) used tests to

predict the amount of fares collected by taxi drivers over an 18-week period. They found

that the tests predicted fares collected in the first three weeks but not the last three weeks.

The District Sales Managers in the current study had a relatively short tenure (about one

year). It is possible that the measures would be more predictive of long-term results in

the district rather than short-term results. A distinction between short and long tenure is

not usually made in studies that use biodata, SJTs, and role-plays as predictors.

Third, the reliability coefficients of the biodata scale and situational judgment test

are very low (.62 and .17, respectively). Without high reliability of the measures, it is

impossible to reach high validity. The reliability in this study was measured by

coefficient alpha that assesses internal consistency of the scale. High internal

consistency reliability shows high similarity of content, that is, homogeneity of

measurement by all parts of a measure. This means that respondents' answers to one part or an item on a measure will be similar to their answers on all other parts or items of the measure. If the coefficient alpha is low, as in this study, it suggests that the scale is not homogeneous, which means that items on a measure are assessing more than one attribute. This assertion is consistent with Oswald, Schmitt, Kim, Ramsey, and Gillespie (2004).

Gatewood and Feild (2001) recommend a reliability coefficient of no less than .85 as the minimum for most selection tools. However, Mumford et al. (1990) suggest that the acceptable level of reliability for biodata is closer to .70 because biodata scales are typically designed to measure multiple constructs, same as SJTs. The complexity of situational judgment items can also affect reliability (Oswald et al., 2004). Oswald et al. (2004) reported internal validity as low as .22 for their SJTs. Despite the low reliabilities of the SJTs they were still predictive of student performance. The more typical range internal reliability coefficients for SJTs is from .56 to .73 (Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Weekly & Davis, 1999).

Other factors that could be affecting the magnitude of the reliability coefficients in this study are the length of the scale and item difficulty. In general, as the length of a scale increases, reliability coefficients also increase. The biodata scale used in this study contained 19 items, and the situational judgment scale contained 11 items, which is shorter than is typically seen in the literature. For example, Stokes and Searcy (1999) and Stokes et al. (1999) constructed biodata scales with more than 200 biodata items organized into several subscales. The SJTs also tend to be longer than the one used in the

present study, with the total of items ranging between 33 and 50 items (Clevenger et al., 2001, Motowidlo et al., 1990; Weekly & Jones, 1999).

Any differences that contribute to the variability in the individual scores will increase the reliability coefficient. For instance, if the items are very difficult or very easy, the differences among the participant's score will be diminished. The scores on the test will be very similar for everyone. As it can be seen in Table 2, the standard deviation for the scores on the biodata scale as well as the SJT is 3.15 and the means are 64.5.and 36.5, respectively. This means that 95% of the scores fall between 58 and 71 points on the biodata scale and 30-43 on the SJT. The minimum possible score on the biodata scale is 19 and the maximum is 76. For the SJT the range is from 11 to 44. It can be argued that the spread of the distribution is narrow, indicating that the item difficulty is low. It might be desirable to increase the difficulty of the items thus enhancing the reliability of both of the tools.

The role-play has a higher reliability because it is not influenced by the same factors as the biodata scale and SJT. More specifically, the inter-item correlations are much higher. The role-play is a longer and most likely more thorough assessment than the biodata scale and SJT. It also seems to be more difficult than the other tools. The standard deviation is 7.92 and the mean is 71.26. Therefore, 95% of the scores fall approximately between 55 and 87 points. This is a much larger range in comparison to the biodata scale and SJT. Despite the higher reliability coefficient, the role-play did not reach acceptable validity.

It is possible that the all of the predictors used in this study are better suited for

predicting subjective measures of performance since they were designed to capture competencies such as leadership and people skills, which might not be directly related to the objective measures. This conclusion is consistent with findings from other studies discussed earlier in this paper.

A number of studies (e.g., Schmitt et al., 1984) showed that biodata predict subjective measures. With the exception of the validity coefficient reported by Reilly and Chao (1982), the studies using subjective measures yielded larger validity coefficients in comparison to studies predicting objective measures. Only one study (Wagner & Sternberg, 1985) showed some evidence that SJT can predict objective measures of performance. There are more studies that successfully used SJT for the prediction of subjective measures. The same is true for role-play. There is very limited evidence to support the idea that role-plays can predict objective measures. The validity coefficients in the study conducted by Squires et al. (1990), for example, were higher for the subjective criteria.

It seems reasonable to argue that all of these selection measures are more predictive of subjective criteria. In the light of the null findings in the current study, a post-hoc analysis was conducted to corroborate the assertion. Four subjective measures (i.e., supervisory ratings) were available for the District Sales Managers. The managers received ratings from their supervisors on four dimensions: emotional intelligence, passionate driver, global builder, and talent nurturer. Within each of the four dimensions, the supervisors rated several competencies that underlie the particular dimension. (Appendix F). Behavioral examples of poor, adequate, and superior performance within

41

each competency were available to the supervisors.

The ratings ranged from one to three. Those District Sales Managers who exceeded expectations received a rating of 1, those who met expectations received a rating of 2, and those who fell below expectations received a rating of 3. The overall score was calculated. The results of the post-hoc analysis are presented in Table 5. Again, high validity coefficients for the biodata scale and SJT were not expected, due to their low reliability. Several validity coefficients reached statistical significance. Biodata predicts the scores on Global Builder ($r = -.15$) and Talent Nurturer ($r = -.16$) dimensions. Situational judgment was found to be predictive of Global Builder ($r = -.15$). The role-play is predictive of the ratings on Global Builder ($r = -.21$), Talent Nurturer ($r = -.18$), and the overall rating ($r = -.16$). Scores on the role-play competencies were also predictive of some of the supervisory ratings. In general, the competencies are predictive of the performance dimension that contained a rating on that competency. For example, the supervisors rated three competencies under the dimension of Talent Nurturer. All three role-play competencies are predictive of the rating of Talent Nurturer dimension. The post-hoc analysis supports the notion that biodata scales, SJTs, and role-playing are more effective in predicting subjective measures.

This study is not without limitations. First, one of the most apparent limitations is the use of predictors with low reliability. Secondly, the criteria used in this study may have been contaminated. Third, all District Sales Managers in this study had a short tenure. They might have lacked the experience to produce the desired bottom line results. Another limitation is that the measures used in this study were designed to

42

capture competencies necessary for the job of District Sales Managers, which might not

be directly related to the objective measures. Finally, the objective measures of

performance may have not been the best measures of individual performance, but rather a

measure of collective performance of many sales representatives in each of the districts.

Future studies should utilize objective measures that are directly under the

manager's control and tests designed to be more reflective of these measures.

Longitudinal design should be used in future studies since it is feasible to expect that the

selection tools would predict long-term rather than short-term results as predicted by

objective measures.

References

Allworth, E., & Hesketh, B. (1999). Construct-oriented biodata: Capturing change-related and contextually relevant future performance. *International Journal of Selection & Assessment, 7*, 97-111.

Barnett, V., & Lewis, T. (1984). *Outliers in Statistical Data*, John Wiley & Sons, New York.

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.

Bernardin, H. J. & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work.* Boston: Kent.

Bernardin, H. J., Hagan, C. M., Kane, J. S., & Villanova, P. (1998). Effective performance management: A focus on precision, customers, and situational constrains. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 3-48). San Francisco: Jossey-Bass Publishers.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & McKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587-605.

Borman, W. C. (1979). Format and training effects on rating accuracy and rating

errors. *Journal of Applied Psychology, 64*, 410-421.

Borman, W. C., Hanson, M. A., Opper, S. H., Pulakos, E. D., & White, L.A. (1993). Role of early supervisory experience in supervisor performance. *Journal of Applied Psychology, 78*, 443-449.

Bretz, R. D. Jr., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management, 18*, 321-352.

Boyatzis, R. E. (1982). *The Competent Manager: A Model for Effective Performance*, New York, NY: John Wiley & Sons.

Brush, D. H., & Owens, W. A. (1979). Implementation and evaluation of an assessment classification model for manpower utilization. *Personnel Psychology, 32,* 369-383.

Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*, 248-260.

Cardall, A. J. (1942). *Preliminary manual for the Test of Practical Judgment*. Chicago: Science Research.

Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology, 52*, 731-756.

Carrington, D. H. (1949). Note of the Cardall Practical Judgment Test. *Journal of Applied Psychology, 33*, 38-45.

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

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, S. V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417.

Cook, M. (1998). *Personnel selection: Adding value through people*. Chichester, England: Wiley & Sons, Ltd.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Miffin Company.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich, Inc.

Dalessio, A. T. & Silverhart, T. A. (1994). Combining biodata test and interview information: Predicting decisions and performance criteria. *Personnel Psychology, 47*, 303-315.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 33*, 360-396.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). The validity of non-cognitive measures decays when applicants fake. *Academy of Management Proceedings*, 127-131.

Elkins, T. J., & Philips, J. S. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85*, 479-484.

File, Q.W., & Remmers, H. H. (1948). *How Supervise? Manual (Revised).* New York: Psychological Corporation.

File, Q.W., & Remmers, H. H. (1971). *How Supervise? Manual (Revised).* New York: Psychological Corporation.

Fine, S.A., & Cronshaw S. (1994) In G. S., Stokes, M. D., Mumford, & W. A., Owens (Eds.). *Biodata handbook: Theory, research and use of biographical information in selection and performance prediction.* (pp. 39-64). Palo Alto, CA: Cpp Books.

Fried Y., (1991). Meta-analytic comparison of the job diagnostic survey and job characteristic inventory as correlates of work satisfaction and performance. *Journal of Applied Psychology, 76*, 690-697.

Ghiselli, E. E., & Haire, M. (1960). The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology, 13*, 225-231.

Graham, K. E., McDaniel, M. A., Douglas, E. F., Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology, 16*, 573-592.

Guion, R. M. (1991). Personnel assessment, selection, and placement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., vol. 2, pp. 327-397). Palo Alto, CA: Consulting Psychologists Press.

Gunter, B., Furnham, A., & Drakeley, R. (1993). Biodata: Biographical indicators of business performance. London: Routledge.

Hein, M., & Wesley, S. (1994). Scaling biodata through subgrouping. In G.S. Stokes, M.D. Mumford, & W.A. Owens (Eds.), *Biodata Handbook: Theory, Research, and Use of Biographical Information in Selection and Performance Prediction* (pp. 171-196). Palo Alto, CA: Consulting Psychologists Press.

Hogan, J. B. (1994). Empirical keying of background data measures. In G. S., Stokes, M. D., Mumford, & W. A., Owens (Eds.). *Biodata handbook: Theory, research and use of biographical information in selection and performance prediction.* (pp. 69-107). Palo Alto, CA: Cpp Books.

Hough, L. M., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G.S. Stokes, and M.D. Mumford (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction.* (pp. 109-145). Palo Alto, CA: CPP Books.

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

Kluger, A.N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology, 46*, 763-781.

Latham, G. P., Saari, L. M., Pursell, E., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.

Lautenschlager, G.J. (1994). Accuracy and faking of background data. In G.S., Stokes, M.D., Mumford, & W.A., Owens (Eds.). *Biodata handbook: Theory, research and use of biographical information in selection and performance prediction* (pp. 391-419). Palo Alto, CA: Consulting Psychologists Press.

Lent R.H., Aurbach, H.A., & Levin, L.S. (1971a). Research design and validity assessment. *Personnel Psychology, 24*, 247-274.

Lent R.H., Aurbach, H.A., & Levin, L.S. (1971b). Predictors, criteria and significant results. *Personnel Psychology, 24*, 519-533.

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

McDaniel M. A., Morgenson, F. P, Bruhn-Finnegan, E., Campion M.A, & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of literature. *Journal of Applied Psychology, 86*, 730-740.

McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the Life Insurance Company. *Personnel Psychology, 52*, 137-148.

Milllard (1952). Is How Supervise? an intelligence test? *Journal of Applied Psychology, 36*, 221-224.

Mitchell, T. W. (1994). The utility of biodata. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.). *Biodata handbook: Theory, research, and use of biographical*

information in selection and performance prediction. (pp. 485-516). Palo Alto, CA: CPP Books.

Morand, D. A. (2001). The emotional intelligence of managers: Assessing construct validity of a non-verbal measure of "people skills." *Journal of Business and Psychology, 16*, 21-33.

Moss, F. A. (1926). Do you know how to get along with people? Why some people get ahead in the world while others do not. *Scientific American, 135*, 26-27.

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

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

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

Mumford, M. D. (1999). Construct validity and background data: Issues, abuses, and future directions. *Human Resource Management Review, 9*, 117-146.

Mumford, M.D., Costanza, D.P., Connelly, M.S., & Johnson, J.F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology, 49,* 3611-398.

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test

validation: A meta-analytic investigation. *Personnel Psychology, 41*, 517-534.

O'Connell, M. S., Hattrup, K., Doverspike, D., & Cober, A. (2002). The validity of "mini" simulations for Mexican retail salespeople. *Journal of Business and Psychology, 16,* 593-599.

Owens, W.A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology Monograph, 65*, 569-607.

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

Pannone, R. D. (1994). Blue collar selection. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 261-273). Palo Alto, CA: Consulting Psychologists Press.

Pulakos, E. D. (1997). Ratings in job performance. In D.L. Whetzel and G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 291-318). Palo Alto, CA: Davies- Black Publishing.

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

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. (1990). Biographical data in employment selection: Can validities be made

generalizable? *Journal of Applied Psychology, 75,* 175-184.

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

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C.L. Cooper and I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*. Chichester: Wiley.

Schoenfeldt L. F. (1999). From dust bowl empiricism to rational constructs in biographical data. *Human Resources Management Review*, 9, 147-167.

Schoenfeldt, L. F., & Mendoza, G. C. (1994). Developing and using factorially derived biographical scales. In G.S. Stokes, M.D. Mumford, & W.A. Owens (Eds.), *Biodata Handbook: Theory, Research, and Use of Biographical Information in Selection and Performance Prediction* (pp. 147-170). Palo Alto, CA: Consulting Psychologists Press.

Squires, P., Torkel, S. J., Smither, J. W., & Ingate, M. R. (1991). Validity and generalizability of a role-play test to select telemarketing representatives. *Journal of Occupational Psychology, 64,* 37-47.

Sternberg, R. J. (1997). Managerial Intelligence. *Journal of Management, 23*,

475-493.

Sternberg, R. J., Wagner, R. K., Williams, & W. M., Horvath, J. A. (1995). Testing common sense. *American Psychologist. 50*, 912-927.

Stokes, G. S. (1999). Introduction to special issues: The next one hundred years of biodata. *Human Resource Management Review, 9*, 111-116.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (1989). Life history prototypes in the study of human individuality. *Journal of Personality, 57*, 509-545.

Stokes, G. S., & Cooper, L. (2001). Content/construct approaches in life history form development for selection. *International Journal of Selection and Assessment, 9,* 138-151.

Stokes G. S., & Searcy C. A. (1999). Specification of scales in biodata form development: rational vs. empirical and global vs. specific. *International Journal of Selections and Assessment, 7*, 72-85.

Stokes, G. S., Toth, C. S., Searcy, C. A., Stroupe, J. P. & Carter, G. W. (1999). Construct/rational biodata dimensions to predict salesperson performance: Report on the U.S. Department of Labor sales study. *Human Resource Management Review, 9,* 185-218.

Stricker, L. J., & Rock, D. A. (1998). Assessing leadership potential with a biographical measure of personality traits. *International Journal of Selection and Assessment, 6,* 164-183.

Terpstra, D.E., Mohamed, A., & Kethley, R. B. (1999). An analysis of federal court cases involving nine selection devises. *International Journal Selection and*

Assessment, 7, 26-34.

Thayer, P. W. (1992). Construct validation: Do we understand criteria? *Human Performance, 5,* 97-108.

Thorton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance.* New York: Academic Press.

Vinchur, A. J., Shippmann, J. S., Switzer, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83*, 586-597.

Viswesvaran, C. (2002). Absenteeism and measures of job performance: A meta-analysis. *International Journal of Selection & Assessment, 10*, 12-17.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 52*, 1236-1247.

Wagner, R. K., & Sternberg, R. J. (1990). Street smarts. In K. E. Clark & M. B. Clark (Eds.), *Measures of Leadership* (pp. 493-504). West Orange: NJ: Leadership Library of America.

Wagner, R. K., & Sternberg, R. J. (1991). Tacit Knowledge Inventory for Managers: User manual. San Antonio, TX: Psychological Corporation.

Weekly, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.

Williams, C. R., & Livingstone, L. P. (1994). Another look at the relationship between performance and voluntary turnover. *Academy of Management Journal, 37,*

267-298.

Wood R., & Payne, T. (1998). *Competency-based recruitment and selection.*

West Sussex, England: John Wiley & Sons.

Appendices

<u>Appendix A</u>

<u>District Sales Manager Position Description</u>

Summary of position

The ultimate responsibility of a District Sales Manager is to grow and sustain profitable sales by meeting sales plan. District Sales Managers are key to implementing field strategy and are critical to the achievement of direct selling excellence. They implement strategies by recruiting, motivating and training Representatives and in so doing achieve sales objectives. They help the Representative achieve personal and career goals.

ROLES AND RESPONSIBILITIES

I. TO GROW SALES PROFITABLY THROUGH ACHIEVEMENT OF TARGET KEY INDICATOR PERFORMANCE.

- Develop plans and systematic courses of action to accomplish identified goals and objectives in the areas of sales, additions, removal rate, staff count, order count, activity, average order, Presidents Club growth, number of customers served.
  - Create opportunities to incent Representatives to participate in greater earning opportunities.
  - Utilize technology to identify business opportunities
  - Organize and present product opportunities through sales meetings
  - Use administrative systems and sales reports to monitor performance level relative to target objectives.
  - Take corrective action on both a campaign and longer-term basis.
- Grow the customer and Representative staff through active prospecting and recruiting. Conducts contacts with Representatives with an emphasis on selling the opportunity and generating leads.
- Apply product knowledge, needs assessment, and presentation skills to influence customers to buy or sell the company products
- Analyze business and market penetration reports to determine priority customer and Representative growth opportunities
- Apply industry and financial knowledge to monitor and control performance of resources and budgetary allowances
  - Monitor Representative business to minimize bad debt and returns
- Establish relationships with Representatives. Identify needs for training and motivation to improve Representative retention.
- Identify and develop a strong flexible support staff to support the needs of the district.

II.   SUPPORT AND IMPLEMENT KEY BUSINESS STRATEGIES THAT LEADS TO INCREASED EARNING OPPORTUNITIES FOR REPRESENTATIVES

- Provide approved training to newly recruited Representatives that will help them establish a customer base and maximize their sales and earnings opportunity.
  - Enthusiastically promote and manage the New Representative Development process.
- Motivate, train and develop established Representatives to manage and grow their businesses and achieve personal goals.
  - Coach and develop Representatives to ensure the understanding of the leadership opportunity
  - Promote and develop the Beauty Advisor program to Representative
  - Promote and create interest in becoming an E-Representative

- Take initiative to understand and participate in all strategic initiatives

EXPECTATIONS

- Develop specific plans and actions to achieve net sales plan, staff growth plan, order plan and Presidents Club growth.
  - Conduct a minimum of 4-8 contacts with Representatives each day or the number required to meet the district sales plan.
  - Conduct a minimum of 6-8 contacts with prospects each day or the number required to meet the district sales plan.
  - Obtain a minimum of 2-10 leads per contact or the number required to meet the district sales plan.
- Utilize demographic data to achieve market penetration plan.
- Utilize total source of additions to achieve target additions including 4 District Sales Manager appointments per campaign
- Achieve targeted number of Beauty Advisors per district (10/district)
- Achieve leadership growth and level advancement (minimum of 3 unit leaders per district)
- Achieve New Representative productivity standards (LOA 1 - $1,000, LOA2-4 - $2500, LOA 1-6 - $5000)
- Delegate non-sales generating activities.
- Manage time effectively; prioritize duties.
- Conduct a planning day every two weeks to analyze progress and modify plans appropriately.
- Spend a minimum of 6 hours a day in the field (with the exception of planning day) with additional time devoted as required to developing strategies or fulfilling district administration requirements.
- Maintain a positive attitude; promote teamwork in the District and Division.

PHYSICAL REQUIREMENTS

- Must be able to lift 35 lbs. on a frequent basis and a maximum of 50-55 lbs. on an occasional basis.
- Must be able to perform stooping, bending and lifting from low positions (floor level or slightly above an automobile trunk). Requires ability to set up, break down and load displays of large quantities of products.
- May require going up and down stairs and steps while carrying materials.
- May require extended periods of standing.
- May require driving for extended periods of time.

OTHER REQUIREMENTS

- Evening and weekend work may be required.
- High School diploma or equivalency.
- Valid drivers' license required.
- Must successfully complete a pre-placement physical, drug test and background check.

Building Relationships
> Establishes rapport and conveys personal interest in each person
> Takes ownership of problems affecting others

Leadership
> Provides clear direction/sets clear goals and expectations

Communication
> Speaks clearly and concisely
> Uses appropriate language
> Project enthusiastic tone

Prospecting/Recruiting
> Continuously looks for potential Sales Representatives
> Effectively promotes the benefits of being a Sales Representative

Results Orientation
> Uses time efficiently
> Always is accountable for own decisions

Selling
> Persists to overcome the objections of others
> Gains commitment from others to act in a certain manner

Training & Developing
> Identifies the strengths and weaknesses of all employees
> Suggests ways to improve performance
> Provides feedback to others
> Provides recognition for positive performance

Planning
> Organizes work activities and prioritizes them accordingly
> Ensures priorities are completed in a timely fashion

Analytical Skills
> Probes to obtain needed information
> Reviews data/information and identifies key points with complex information

Business Management & Judgment
> Understands the business and how it operates
> Considers the consequences of decision alternatives

Examples of Biodata and Situational Judgment Items

Which statement best describes your experience with coaching or teaching adults in a work setting?

    A. I have informally coached co-workers on the job.
    B. I have no experience coaching or teaching adults in a work setting.
    C. I coached and taught sales representatives as part of my job responsibilities.
    D. I coaches and taught when a new worker joined our work group.

Which statement best describes your experience in planning and managing budgets?

    A. I have had decision-making experience for a budget of less than $10,000.
    B. I have never had a decision-making responsibility for a budget of nay type.
    C. I have had decision-making experience for a budget of more than $10,000 and less than $25,000.
    D. I have had decision-making experience for a budget greater than $25,000.

Which statement best describes your educational background?

    A. I have had some college business courses, such as accounting, sales or marketing.
    B. I have an Associates Degree or the equivalent in Business or related field.
    C. I have a high school education.
    D. I have a Bachelors Degree.

You are conducting a planning session and setting goals for your next sales campaign. Which of the following describes the best approach?

    A. Identify the poorest performing Sales Representatives and set specific and challenging goals to improve their performance.
    B. B. Set specific and challenging goals for Sales Representatives who, in your opinion, have demonstrated the greatest potential.
    C. Set general and modest goals to motivate new Sales Representatives.
    D. Set specific and challenging goals for your best Sales Representatives.

You are speaking to a group of Sales Representative candidates. One candidate is very interested but must speak to her family before deciding. Which of the following describes the best approach?

    A. Ask her about the objections her family may have and coach her about what to say.

B. Ask for a commitment date when she will decide and provide her with the company brochure.
C. Gain her agreement to meet her family so that you can present the opportunity to them personally.
D. Obtain her telephone number and ask if you can call in a week to learn her decision.

You scheduled a coaching session with a new Sales Representative who has been with the company for two weeks. At the session, you want to establish the right tone and set expectations. Which of the following describes the best approach?

A. Tell her you are not her "boss" but someone there to help her succeed and that you are confident that she will succeed by following the sales process.
B. Tell her you are confident she has the skills to succeed and that you monitor the performance of all new Sales Representatives carefully during their first six weeks.
C. Tell her that you are tough but fair and that if she follows the company's sales process she will succeed.
D. Find positive things to say about her work during the first two weeks and explain that your role is to provide specific sales ideas to help her succeed.

List of Support Materials Available to the Candidate

1.   **Background about Albee Products**. Provides the Vision and a brief summary of Albee Products, the company, its business strategy, and products.

2.   **Your Role and Goals as the Sales Manager**. Describes the skills and expectations of a Sales Manager at Albee Products as well as the goals you are expected to achieve.

3.   **Guidelines and Overview of Terms**. You should follow these guidelines carefully when dealing with the situations you will face during this Program. You are to use these terms, and follow them, when answering questions and resolving other issues.

4.   **Benefits of being an Albee Products Direct Sales Representative**. Provides details on the many benefits of working as an Albee Products Direct Sales Representative.

5.   **Keys to Recruiting Individuals to Become Albee Products Direct Sales Representatives**. Provides several of the Keys to Recruiting Albee Products Direct Sales Representatives..

6.   **Organization Chart**. An organization chart showing your District.

7.   **Weekly Planner**. Your Weekly Planner is included for your use and reference. Listed are your appointments and activities for the current week.

8.   **Voice Mail and E-Mail Transcriptions**. The transcriptions from several Voice Mails and E-Mails you have recently received.

9.   **Direct Sales Representative Sales Results - Canterbury District**. Sales from selected Albee Products Direct Sales Representatives from your District for three Sales Cycles.

10.   **District Overall Performance Against Goals**. – Canterbury District. Overall performance for the last three sales cycles against your three performance goals.

Role-Play Scenarios

1. The District Sales Manager receives a telephone call from his/her Division Sales Manager (supervisor). The Division Manager is excited about a new promotion, but is somewhat "pressed" for time. The Division Manager provides the candidate with specific information about new promotion that will begin this month. He/she also informs the candidate that he/she will be calling back later to further discuss some issues regarding the candidate's recent sales results. (During this call the candidate in NOT evaluated, and this information should be used by the candidate throughout future calls.)

2. A current customer calls to talk about becoming a Sales Representative. However, this customer is hesitant about becoming a Sales Representative. But, information about this customer is provided in the candidate's materials, and the candidate should talk to this customer and begin to "recruit" him/her to become a Sales Representative.

Subjective Performance Dimensions

Emotional Intelligence –understands and successfully manages one's self and one's relationship with others. Strives for self-awareness and self-regulation, lives the company's values, demonstrates empathy and influence, and is obsessive about Representatives and customers.

      Building Relationships
      Personal Effectiveness
      Interpersonal Skills
      Communication
      Customer Focus

Passionate Driver – is accountable for results governed by business ethics, embraces change, courage to make tough calls, raises the bar, acts decisively using facts and discipline

      Results Orientation
      Business Management and Judgment
      Selling
      Analytical Skills

Global Builder – breaks down barriers, strategic thinker and action oriented implementer, manages matrix relationships, diversity of perspectives; works across geographical boundaries and works closely with operation partners.

      Diversity Management
      Planning
      Technical Skills

Talent Nurturer – assures that the right people in the right jobs at the right time, develops next generation of leaders, promotes those who deliver results and develops talent, differentiates performance.

      Prospecting/Recruiting
      Training & Developing
      Leadership

Table 1

Competencies Evaluated During Call Scenarios

	Call Scenarios			
Competency	Discussion with potential Sales Representative	Discussion with a Sales Representative	Conference call with two Sales Representatives	Discussion with a supervisor
1. Building relationships	X	X	X	No
2. Leadership	X	X	No	X
3. Communication	X	X	X	X
4. Prospecting/Recruiting	X	X	X	No
5. Result Orientation	X	X	X	X
6. Selling	X	X	X	No
7. Training/Developing	No	X	X	No
8. Planning	No	X	X	X
9. Analytical Skills	X	No	X	X
10. Business Management/Judgment	X	X	X	X

Note. Only competencies that were considered essential by more than 60% of the SMEs for a scenario
were evaluated in that scenario.

Table 2

Descriptive Statistics and Intercorrelations for Variables

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Sales - actual	1,315,248.00	341,058.14	-																	
2. Sales - target	1,350,354.00	369,202.92	.94**	-																
3. Sales - % difference	4.23	6.95	-.05	.13	-															
4. Debt - actual	28,474.96	13,592.97	.45**	.45**	.37**	-														
5. Debt - target	25,872.98	12,600.06	.50**	.61**	.28**	.70**	-													
6. Ratio - target	1.96	.76	.07	.11	.27**	.60**	.80**	-												
7. Ratio - actual	2.09	.88	.08	.14	.42**	.86**	.60**	.67**	-											
8. Debt - % difference	-13.02	39.54	.11	.11	-.08	-.31*¹	.29**	.25**	-.34**	-										
9. Orders - actual	7,095.32	1,781.46	.85**	.80**	-.07	.24**	.24**	-.21**	-.12	.04	-									
10. Orders - target	7,203.76	1,889.85	.77**	.83**	.15*	.27**	.31**	-.15*	-.06	.05	.93**	-								
11. Orders - % difference	1.55	7.89	-.08	.09	.74**	.17*	.15*	.14	.21**	-.02	-.06	.24**	-							
12. Staff - actual	358.00	80.47	.72**	.67**	-.04	.25**	.26**	-.09	-.08	.00	.87**	.80**	-.05	-						
13. Staff - target	349.29	81.66	.65**	.70**	.20**	.33**	.32**	-.03	.05	-.04	.81**	.88**	.29**	.90**	-					
14. Staff - % difference	-2.66	9.10	-.05	.10	.63**	.16*	.11	.08	.21**	-.11	-.04	.19*	.81**	-.12	.26**	-				
15. Managerial Effectiveness	-0.04	2.41	.79**	.71**	-.21**	-.04	.09	-.34**	-.39**	.16*	.93**	.84**	-.13	.86**	.75**	-.14	-			
16. Biodata	64.50	3.15	.00	.01	.10	.03	-.03	-.03	-.03	-.06	.06	.08	.10	.07	.06	.00	.04	-		
17. Situational Judgment	36.52	3.15	-.04	-.08	-.04	-.02	-.01	.00	.08	-.06	.06	-.07	-.10	-.04	-.05	-.06	.00	-.10	-	
18. Role-play	71.26	7.92	-.06	-.09	-.01	.04	.02	.10	.09	-.03	-.16*	-.12	.05	-.18*	-.16*	.04	-.17*	.12	.16*	-

Note. N = 155-187.

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

66

Table 3

Descriptive Statistics and Intercorrelations for Role-Play Dimensions

Variable	$\underline{M}$	$\underline{SD}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\underline{6}$	$\underline{7}$	$\underline{8}$	$\underline{9}$	$\underline{10}$	$\underline{11}$	$\underline{12}$
1. Building Relationships	73.74	8.65	-											
2. Leadership	70.46	11.57	.42**	-										
3. Communication	79.35	9.56	.42**	.51**	-									
4. Prospecting/Recruiting	63.89	9.70	.32**	.32**	.26**	-								
5. Result Orientation	70.81	9.07	.59**	.48**	.45**	.49**	-							
6. Selling	69.09	15.74	.46**	.34**	.23**	.33**	.43**	-						
7. Training/Developing	62.83	12.16	.53**	.42**	.37**	.34**	.48**	.33**	-					
8. Planning	70.48	12.74	.38**	.51**	.47**	.32**	.39**	.36**	.40**	-				
9. Analytical Skills	52.03	13.17	.33**	.38**	.29**	.22**	.25**	.29**	.49**	.47**	-			
10. Business Management/Judgment	69.88	9.42	.49**	.52**	.48**	.38**	.46**	.32**	.58**	.58**	.46**	-		
11. Biodata	64.50	3.15	.07	.13	.11	.03	.14	.12	.02	.04	.03	.15	-	
12. Situational Judgment	36.52	3.15	.21**	.13	.16*	.05	.14	.14	.10	.12	-.02	.08	-.10	-

Note. N = 155-187.

$*$ $\underline{p}$ < .05, $**\underline{p}$< .01

Table 4

Correlations for Role-Play Dimensions and Criterion Measures

	Role - Play Dimensions									
Criterion Measures	Building Relationships	Leadership	Communication	Prospecting/ Recruiting	Result Orientation	Selling	Training/ Developing	Planning	Analytical Skills	Business Mngt/Judgment
1. Sales - actual	-.13	.-03	.08	.01	.02	-.08	-.03	-03	.00	-.06
2. Sales - target	-.11	.01	.01	.01	-.01	-.09	-.05	-.10	-.02	-.12
3. Sales - % difference	.09	.01	-.11	.13	.02	.07	.01	-.13	-.09	-.08
4. Debt - actual	-.02	.00	.10	.19*	.11	.00	.01	-.02	.03	-.02
5. Debt - target	-.06	.04	.03	.13	.11	.04	-.02	-.09	.00	.03
6. Ratio - target	.04	.05	.04	.17*	.13	.12	.04	.03	.00	..03
7. Ratio - actual	.03	-.05	.04	.20*	.07	.04	.05	.00	.03	.07
8. Debt - % difference	-.08	.02	-.06	-.08	.03	.04	.00	-.04	-.03	-.08
9. Orders - actual	-.14	-.14	-.01	-.08	-.01	-.13	-.08	-.15*	-.03	-.13
10. Orders - target	.-11	.-11	.-04	-.06	.00	-.10	-.03	-.16*	-.01	-.12
11. Orders - % difference	.08	.02	-.10	.03	-.03	.06	.03	-.10	.02	-.03
12. Staff - actual	-.11	-.15*	-.08	-.08	.03	-.15*	-.13	-.17*	-.02	-.18*
13. Staff - target	-.10	-.13	-.11	-.05	.03	-.11	-.07	-.19*	-.01	-.17*
14. Staff - % difference	.02	.02	-.06	.02	-.02	.08	.11	-.10	.03	-.03
15. Managerial Effectiveness	-.17*	-.11	-.09	-.13	.00	-.13	-.11	-.13	.01	-.17*

Note. N = 155-187.

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

Table 5

Correlations for Subjective Measures

Variable	Emotional Intelligence	Passionate Driver	Global Builder	Talent Nurturer	Total
Biodata	-.09	.03	-.15*	-.16*	-.11
Situational Judgment	.-10	.-09	-.16*	.03	-.10
Overall Role-Play	-.06	.05	-.21**	-.18*	-.16*
Building Relationships	-.14	.06	-.18*	-.09	-.15*
Leadership	-.06	.09	-10	-.18*	-.14
Communication	-.15*	-.13	.06	-.16*	-.11
Prospecting/Recruiting	-.11	-.11	-.07	-.19*	-.14
Result Orientation	.06	-.17*	-.02	-.18*	-.13
Selling	.02	-.16*	-.01	.05	-.08
Training/Developing	.07	-.15*	-.22**	-.18*	-.16*
Planning	.03	.09	-.14	.12	-.05
Analytical Skills	.10	-.16*	-.23**	.03	-.18*
Business Management/Judgment	-.10	-.20**	.06	.09	-.09

Note. N = 127-186.

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