THE ROLE OF JOINT VERSUS SEPARATE EVALUATION IN SELECTION PROCEDURE DECISIONS

Sarah R. Kirkendall

A Thesis
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

December 2012

Committee:
Scott Highhouse, Advisor
Margaret Brooks
Mary Hare
ABSTRACT

Scott Highhouse, Advisor

Practitioners struggle to convey statistical information concerning employee selection procedures to managers. Joint vs. separate evaluation has been used to study preference reversals in the field of medical decision making, as well as other domains. A similar design was used to test the preference for the popular interview and the less popular, yet more predictive test of general mental ability. Although a complete preference reversal was not found, people did in fact prefer the test of GMA to the interview when information about the two hiring procedures were presented together rather than when presented separately.
This thesis is dedicated to my family –

Thank you for your continual love & support.
ACKNOWLEDGEMENTS

First, I would like to acknowledge my advisor, Dr. Scott Highhouse, for his guidance and support through the thesis process. I would also like to express gratitude to my committee members, Dr. Margaret Brooks and Dr. Mary Hare, for their contributions and insight into my research design and manuscript. Special thanks go to Paula Watson, Sue Thorton, and the HSRB office for their help in allowing this project to run smoothly. Finally, I would like to thank my family and friends who have provided continual encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>GMA versus the Unstructured Interview</td>
<td>2</td>
</tr>
<tr>
<td>Preference for the Unstructured Interview</td>
<td>3</td>
</tr>
<tr>
<td>Resistance Toward Standardized Tests</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER I. EVALUATION</td>
<td>8</td>
</tr>
<tr>
<td>Joint vs. Separate Evaluation</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER II. NUMERACY</td>
<td>13</td>
</tr>
<tr>
<td>CHAPTER III. METHOD</td>
<td>14</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td>Procedure</td>
<td>14</td>
</tr>
<tr>
<td>CHAPTER IV. ANALYSES</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER V. RESULTS</td>
<td>18</td>
</tr>
<tr>
<td>CHAPTER VI. DISCUSSION</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER VII. CONCLUSION</td>
<td>23</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>27</td>
</tr>
<tr>
<td>APPENDIX A. THE NUMERACY SCALE</td>
<td>33</td>
</tr>
<tr>
<td>APPENDIX B. CONSENT LETTER</td>
<td>35</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES/TABLES

<table>
<thead>
<tr>
<th>Figure/Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Summary of Analysis of Variance: Difference between Hiring Procedure, Evaluation Mode, and Numeracy</td>
<td>24</td>
</tr>
<tr>
<td>2 Mean Ratings of Hiring Procedure (Interview and test of GMA) for High and Low Numeracy</td>
<td>25</td>
</tr>
<tr>
<td>3 Mean Ratings for Hiring Procedures in Joint and Separate Conditions</td>
<td>26</td>
</tr>
</tbody>
</table>
INTRODUCTION

One goal of employers is to select the best applicants for each job in the most efficient manner. A goal of many Industrial/Organizational (I-O) psychologists is to create assessment methods that would assist employers in choosing the best employees. There seems to be a disconnect, however, between research and practice. When attempting to gather information about a candidate from an initial application blank, it is easy to evaluate an applicant’s past work experience and educational background from such forms. However, when deciding on subsequent assessments in the hiring process, employers are underutilizing selection devices that have been validated (Terpstra & Rozell, 1993).

Schmidt and Hunter (1998) published a meta-analysis which reviewed findings on eighteen different selection methods, including measures of General Mental Ability (GMA), integrity tests, personality tests, interviews (structured and unstructured), work samples tests, etc. They concluded that the greatest amount of variance in job performance could be accounted for by using a cognitive ability test. Nevertheless, research suggests that employers do not perceive these tests to be useful. One possible reason that employers are not using such methods is that they are not familiar with the findings concerning different selection procedures (Rynes, Colbert, & Brown, 2002). Another explanation could be that they are not able to evaluate the statistical terminology used to present research findings and therefore are neglecting valuable information when deciding which selection procedure to implement (Cosmides & Tooby, 1996). If academic researchers could express their findings in a manner that could be more easily interpreted, employers may be willing to reevaluate their hiring practices.

The purpose of this study was to evaluate whether presenting validity information about specific selection procedures in a comparative (versus alone) manner would lead decision makers
to value the methods differently. This study focused on cognitive ability tests, which have been found to be strong predictors of future job performance and compared them with the popular, but less predictive unstructured interview (Hunter, 1980; Hunter & Hunter, 1984).

### GMA versus the Unstructured Interview

Measures of GMA have been found to account for more variance (corrected $r$'s ranging from .26 to .58) in future job performance than any other single selection procedure (Hunter, 1980; Hunter & Hunter, 1984). Not only does GMA predict job performance, it also accounts for variance in areas such as, career advancement achieved by an individual and performance in training programs (Schmidt & Hunter, 2004). The more complex a job, the better GMA predicts job performance. Schmidt and Hunter (2004) argued that people with higher GMA are able to acquire the knowledge necessary to perform a job faster than those with lower GMA, which leads to overall higher levels of job performance. However, managers are reluctant to use measures of cognitive ability, as well as other standardized paper-and-pencil tests (e.g. Diab, Yueh Pui, Yankelevich, & Highhouse, 2009; Terpstra & Rozell, 1993, 1997). The most commonly used selection method is still the unstructured interview (Buckely, Norris, & Wiese, 2000) with validity estimates ranging from .14 (Hunter & Hunter, 1984) to .26 (Wiesner & Cronshaw, 1988) for predicting future job performance.

Graves and Karen (1996) called for HR practitioners to reexamine current hiring practices and evaluate the predictive validity of unstructured interviews for outcomes such as productivity, turnover, satisfaction, etc. According to the authors, employers must be informed of consequences of poor hiring decisions. Collecting data regarding the effectiveness of hiring practices such as unstructured interviews is the first step. Graves and Karen noted that there are many opportunities for unstructured interviews to be compromised by managers’ own
preferences for and beliefs about applicants’ characteristics. If an interviewer seems unprepared, as might happen during an unstructured interview, the organization may be perceived to be unprofessional. Personal prejudices may also be transparent during unstructured interviews. For example, a study by Rynes, Bretz, and Gerhart (1991) found that 50 percent of women reported experiencing negative events (ranging from an interviewer commenting on the woman’s appearance to being asked to meet in a hotel room) during an unstructured interview due to their gender. Kuncel (2008) suggested that using objective measures, such as a cognitive ability test, to narrow the pool of applicants would decrease the negative effects of using a subjective method of evaluation when making the final hiring decision. Despite research findings that suggest the unstructured interview is an ineffective tool for predicting future job performance, they remain well liked and commonly used among managers.

Preference for the Unstructured Interview

A survey conducted in the 1950’s found that 99 percent of organizations used an interview to evaluate applicants (Ulrich & Trumbo, 1965). Still today, the interview is the most popular selection tool (Buckley, et al., 2000). Employers place great value on human judgment and prefer subjective methods of evaluation to objective paper-and-pencil tests (Highhouse, 2008). Brooks, Guidroz, and Chabrabarti (2009) found that selection policies that utilize a holistic method are preferred over procedures that utilize mechanical methods.

Medical decision-making research has provided empirical evidence demonstrating peoples’ preference for human judgment over mechanical prediction (e.g. Arkes, Shafter, & Medow 2007; 2008). In Arkes et al.’s studies, participants preferred physicians who did not use decision aids when forming a diagnosis to those who consulted a decision aid. They also felt that physicians were perceived as less professional and less capable when decision aids were used.
When a favorable diagnosis was delivered, participants rated those physicians who used a computer-based diagnostic support system less favorably than those physicians who used no aid. On the same note, physicians were rated less negatively when a diagnosis was unfavorable, and no decision aid was consulted. This provides further evidence that, in general, people prefer doctors who do not use decision aids when forming a diagnosis.

Decision makers (employers) value their intuition, however, research has shown that relying on intuition alone is not good practice (Kleinmuntz, 1990). Managers routinely make quick decisions when faced with complex problems and it is often the case that they rely on their intuition to do so (Patton, 2003). However, Dawes (2002) warns against the use of intuition alone when making decisions. There have been several reasons offered as to why intuition may fail. First, research suggests that people are not good at evaluating their performance on past decisions. Self-checks and feedback systems are necessary, in order to evaluate past decisions that were made based on intuition (Patton, 2003). Second, if a poor-performing employee is hired, the manager is likely to blame external factors, rather than his or her intuition (Hastie & Dawes, 2001; Louie, 1999; Pezzo & Pezzo, 2007). Lastly, it is suggested that, when using intuition, people dismiss unique characteristics of situations and instead over simplify them in order to make a quick decision (Bonabeau, 2003).

Decision makers (i.e. interviewers) also suffer from errors and biases (Pezzo & Pezzo, 2007). Past successes are attributed to their ability to select a high performer and failures are attributed to external forces that the manager had no control over. If an employee is unsuccessful, managers are likely to label the employee as a phony, blame trainers, or other members of the company who may have played a role in hiring the employee, rather than find fault in their own hiring practices. The confirmation bias suggests that managers recall instances
of successful hires, while ignoring examples of poor hiring decisions. They look to confirm their believed expertise. Because there is often no objective scoring of an unstructured interview, and therefore no ability to evaluate the procedure; the manager is able to selectively recall and reinterpret what was discussed and, in turn, be more confident than warranted in his or her abilities to predict future job performance. Feedback is necessary when attempting to explain the usefulness of selection procedures (Phillips & Gully, 2008). Fisher (2008) suggests that feedback would be a useful tool for showing managers the errors made when using unstructured interviews. A simple tally measure could be used to keep track of successful and unsuccessful hires; however, this is not a common practice.

Perhaps because predictive validity is not a factor which managers commonly use to evaluate the value of selection procedures, they are more likely to discount evidence that suggests other hiring practices are more “valid.” Research has shown that interviewers can evaluate the personality traits of Extraversion, Agreeableness, and Openness to Experience. However, the trait of Conscientiousness is the most predictive of future job performance, yet the most difficult to measure with an interview (Barrick, Patton, & Haugland, 2000). Although research findings are not supportive of the use of unstructured interviews, managers are reluctant to use more predictive methods, such as a structured interview (Van der Zee, Bakker, & Bakker, 2002).

There have been a number of reasons offered as to why employers resist using paper-and-pencil tests and other valid methods of applicant evaluation. For instance, managers prefer holistic methods of evaluating employees because there is an overall preference for making decisions based on human judgment and they believe that experience, along with intuition will result in a better hiring decision than a paper-and-pencil test (Highhouse, 2008). Mechanical or
standardized methods of evaluating candidates are viewed as being unprofessional, impersonal, and less comprehensive than less standardized, holistic methods (Diab, et al., 2009).

Resistance Toward Standardized Tests

A number of studies have explored the gap between research and practice in the area of employee selection (e.g. Rynes, Colbert, & Brown, 2002; Terpstra & Rozell, 1993, 1997). For example, Terpstra and Rozell (1993) evaluated five staffing practices that, although predictive of employee performance, were traditionally underutilized by employers. The practices under investigation were: (1) studies used to determine recruiting sources which result in the hire of high-performing employees, (2) studies which evaluate the validity of selection procedures, (3) structured/standardized interviews, (4) measuring cognitive ability/aptitude, and (5) the practice of using biographical information blanks (BIBs) within weighted application blanks (WABs). The authors found that the application of such practices was positively correlated with higher annual profit, profit growth, and, in some industries, overall performance. However, on average, less than two of the above mentioned five staffing practices were actually being used within the sample of organizations. When asked to rank specific hiring practices, managers listed work samples, references/recommendations, unstructured interviews, structured interviews, and assessments centers as the top five selection methods. Specific aptitude tests, personality tests, general cognitive tests, and biographical information blanks (all traditionally paper-and-pencil measures) were the least favored selection methods (Terpstra & Rozell, 1997). This suggests that, although research has shown numerous advantages to using predictive staffing procedures, few organizations are implementing such practices.

Lievens, Highhouse, and De Corte (2005) found that employers valued the information gathered via a holistic method, such as an unstructured interview, more than the same
information obtained by paper-and-pencil test. Retail supervisors were asked to rate applicants based on their personality and GMA. When supervisors were told that GMA scores were obtained via an unstructured interview, rather than a paper-and-pencil test, they held the scores to be more important. Similarly, supervisors placed more value on scores obtained by an unstructured interview to those obtained by a paper-and-pencil test when evaluating personality traits (specifically extraversion).

Not only are employers unlikely to use paper-and-pencil tests, they are also unfamiliar with research findings on specific hiring practices. Rynes, Colbert, and Brown (2002) presented HR professionals with research findings (some true and some fabricated) and asked whether or not they believed the findings to be true or false. In regard to staffing procedures, there were major discrepancies between research findings and what HR professionals hold to be true. Less than half of the HR professionals agreed with statements that were consistent with research findings. Seventy-two percent of HR professionals believed that conscientiousness was a better predictor of future job performance than intelligence. Also, 46% of respondents felt that when hiring for a low-level job, screening for intelligence disadvantages companies. This suggests that HR professionals are either misinformed or not at all familiar with research findings. Perhaps if data were presented in a manner, which was easier to understand the usefulness and predictive power of objective measures, such as an intelligence test; they would be more likely to implement more effective hiring procedures.
CHAPTER I. EVALUATION

It is important that managers be able to understand and interpret data that is presented to them concerning specific selection methods. A study by Hazer and Highhouse (1997) found that the better managers understood information they were presented with, the less likely they were to be distracted by inconsequential details surrounding alternatives. The authors measured managers’ reactions to utility analysis, a method for quantifying the usefulness of specific selection methods in terms of dollars and cents. The authors were interested in the ability of managers to comprehend the material and the perceived usefulness of the data. It was found that managers preferred the simplest utility analysis technique to more complex techniques. Specifically, when data is difficult to comprehend, simple presentation methods may be most useful.

Presentation of information and phrasing of questions can also affect the ability of people to make correct judgments. Previous research has found that when presented with a medical diagnosis problem stating information in the form of a frequency (1 in 100) and providing questions that caused people to carefully think about information presented to them, people were more likely to correctly use Bayesian reasoning and consider base rates (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). In fact, reasoning was actually impaired when information was presented in terms of percentages (1%) and frequencies rather than only presenting information as frequencies (Cosmides & Tooby, 1996). This suggests that it is very important to consider not only what data is presented, but also how it is presented.

Joint vs. Separate Evaluation

Evaluation mode refers to how alternatives are presented to decision makers. In the judgment and decision making literature evaluation mode is often referred to as having two
levels: joint evaluation and separate evaluation. In a joint evaluation mode alternatives are presented to decision makers simultaneously, whereas in the separate evaluation mode alternatives are presented independently from other alternatives.

Hsee (1996a; 1996b; Hsee, Lowenstein, Blount, & Bazerman, 1999) argued that the way in which information is presented is very important when it comes to choosing between alternatives. By simply viewing alternatives jointly (rather than alone) preferences can be reversed; what was favored in a separate evaluation may no longer be favored when alternatives are viewed jointly (Hsee, 1996a; 1996b; Hsee et al. 1999; Bazerman, Lowenstein, & White, 1992). Hsee (1996a; 1996b) goes on to explain that certain information is difficult to evaluate without a referent and therefore that information may not be considered when making a decision. However, the difficult to evaluate information may be very important, and not until the options are evaluated jointly will the decision maker be able to weigh it in his or her decision.

For example, Hsee demonstrated how joint versus separate presentations could result in preference reversals. Subjects were provided with information about Dictionary A and Dictionary B and asked which they would be most likely to buy (Hsee, 1996b).

Dictionary A was published in 1993, has 10,000 entries and has no defects.

Dictionary B was published in 1993, has 20,000 entries and has a torn cover, but otherwise is like new.

In the case of the dictionaries, the number of entries is harder to evaluate independently than the condition of the book (i.e. the number of entries is lower in evaluability). A person is not able to decide whether or not 10,000 or 20,000 entries is an important factor if the alternatives are presented separately. Without the ability to compare the number of entries side-by-side, decision makers are unable to recognize the importance of the information. Therefore, the information
concerning the number of entries is ignored and Dictionary A is chosen significantly more times than Dictionary B because it is easy to visualize a dictionary free of defects. However, when the two options are evaluated simultaneously, preferences are reversed and Dictionary B is chosen significantly more times than Dictionary A. This is because it is much easier to evaluate the number of entries comparatively and see that, although Dictionary B has a torn cover, it is worth choosing because it has twice as many entries as Dictionary A (Hsee, 1996b). Zikmund-Fisher, Fagerlin, and Ubel (2004), found that the phenomenon can also be seen when making medical care decisions, such as, deciding among doctors.

Medical decision-making studies have also found preference reversals. For example, Zikmund-Fisher et al. (2004) also use the example of choosing a fertility clinic; the first clinic has a 33% success rate and is 15 minutes away and the second clinic has a 40% success rate, but is 45 minutes away. In this case the success rate is low in evaluability when viewed in isolation; therefore the short drive time makes the first clinic a better choice. However, when viewed simultaneously, the success rate (33% vs. 40%) is easier to evaluate and becomes more important than drive time.

The above examples stress the importance of presentation method. When options are presented separately, the information which is difficult to evaluate without comparison (e.g. number of entries/success rate) plays less of a role than when presented jointly (Hsee, et al., 1999). Evaluability research has shown that by allowing people to view options jointly, so that they are able to compare attributes, the once difficult to understand information is no longer ignored and can be used to reach a final decision. This research adds value not only to the field of judgment and decision making, but also the field of employee selection. Although the research community is confident in the ability of standardized tests to predict high performing employees,
the lay community is still resistant. This resistance could be a result of a lack of information, or a lack of evaluability. If a manager views information about only one selection procedure, he/she may find that some of the data is difficult to understand and therefore ignore it. However, if a manager were presented with information about multiple selection procedures simultaneously, he/she would be able to compare multiple criteria against each other and difficult to understand data could be considered.

This study examined whether people are more positively disposed toward GMA tests when presented alongside an unstructured interview rather than independently. According to the Evaluability Hypothesis (Hsee, 1996b), information that is difficult to evaluate without comparison may be underutilized when making a decision. Just as in the Dictionary example, the number of entries (10,000 vs. 20,000) was difficult to evaluate independently, so to are the validity coefficients (.20 vs. .51) in regard to the selection procedures.

An overall preference for the unstructured interview, as it is the most common method of evaluating applicants was expected. However, when both an unstructured interview and GMA test are presented jointly, the discrepancy between the two selection procedures will be reversed completely. Although statistical information about selection methods is difficult to understand, when data is presented jointly, decision makers will show a preference for the method that possesses greater predictive validity.

Hypothesis 1: There will be an interaction between hiring procedure (interview vs. test of GMA) and evaluation mode (joint vs. separate).

By modeling previous studies that have found preference reversals, I expected to find similar results among selection methods when unstructured interviews and cognitive ability tests are presented jointly rather than separately. This is because the difficult to evaluate information
(validity) will be compared in the joint evaluation mode and allow for the decision maker to make a distinction between the two alternatives.

A control condition was added to test the overall preference for an unstructured interview that previous research (e.g. Buckley et al., 2000) has found. In this condition participants were asked to rate their preferences for an unstructured interview and a test of GMA with no corresponding validity information.

Hypothesis 2: Decision-makers will show an overall preference for the more commonly used unstructured interview over the less popular test of GMA.
CHAPTER II. NUMERACY

Another variable of interest is numeracy. Numeracy, or numerical ability, is the ability to process numbers and basic probability. Previous studies have found that people high in numeracy are better at making judgments concerning probabilities (Hamm, Bard, & Scheid, 2003) and risk information that is presented in numeric form (Gurmankin, Baron, & Armstrong, 2004) than those low in numeracy. Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert (2006) found that individuals high in numeracy were less likely to fall prey to framing effects and use numeric principles when making decisions. A survey in 2002 found that nearly half of the American population was innumerate, lacking the basic mathematical skills to use numbers (Kirsch, Jungeblu, Jenkins, & Kolstad, 2002).

Given that so many people are unable to appropriately use numeric information, it is even more important to examine the effect of how such information is presented to decision makers. If simply presenting data differently can reduce the discrepancy between individuals who are high and low in numeracy, people could make more informed financial, medical, consumer, as well as other decisions. We also examined the differences between individuals high and low in numeracy and their ability to use statistical information when deciding between selection procedures. It was expected that decision makers who are high in numeracy would be more likely to use validity information when rating selection procedures. They would also place more value on the validity information than those low in numeracy. Thus, effects of presentation mode should be stronger among the low-numeracy participants.

Hypothesis 3: There will be a 3-way interaction between hiring procedure (interview vs. test of GMA) and evaluation mode (joint vs. separate) and numeracy (high vs. low).
CHAPTER III. METHOD

Participants

Participants were drawn from a pool of undergraduate psychology students at a university in the Midwestern United States (n = 270). The university population is 54% female, 81% Caucasian, and average age of 21 years. Of the 270 participants, 51 evaluated the test of GMA alone and 53 participants evaluated the interview alone. The presentation of the test of GMA and the interview was counterbalanced in the joint conditions. Fifty-four people were presented the test of GMA first and another 54 people were initially presented the interview. Participation was completely voluntary. Participants were recruited within their psychology classes. Permission to collect data during a regular class meeting was granted by instructors. Instructors awarded credit for participation at their discretion.

Procedure

This study used a 2 (evaluation mode: joint vs. separate) x 2 (hiring procedure: test of GMA vs. interview) factorial design. Participants were randomly assigned to a condition and asked to imagine themselves as a hiring manager for a company. Specifically, they were instructed:

Imagine that you are a manager in charge of hiring new employees for a company. Your organization is looking for an employee hiring procedure in which the best applicants are identified.

All conditions began with a brief explanation about the participant’s role in choosing a selection procedure. In the Separate Condition, they were instructed, “Please provide your reaction to the proposed hiring method.” In the Joint Condition, they were instructed, “Please provide your
reaction to the proposed hiring methods. (Please read both before responding to the scales.)”

They were then presented with either one or two selection procedures.

Hiring Procedure 1: Applicants will undergo a traditional job interview that is tailored to each applicant. Studies have shown that these interviews correlate .20 with later performance on the job.

Hiring Procedure 2: Applicants will undergo a standardized test of general intelligence that presents the same questions to each applicant. Studies have shown that these tests correlate .51 with later performance on the job.

Validity for GMA was presented in the form of a correlation coefficient and was taken from Schmidt and Hunter’s (1998) review of selection methods. Job performance in occupations of medium complexity are correlated with GMA $r=.51$. Validity for unstructured interviews was taken from Huffcutt, and Arthur’s (1994) review of employment interviews for entry-level positions ($r=.20$). Both Schmidt and Hunter (1998) and Huffcutt and Arthur (1994) corrected for range restriction and measurement error. Participants were presented with both of the selection procedures in the joint conditions and only one of the selection procedures in the separate conditions. The presentation of the hiring procedures was counterbalanced in the joint conditions. Control conditions were also added to test the general preference for the interview over the test of GMA. In these conditions, the same information was presented to the participants, minus correlation coefficients (e.g., “Hiring Procedure 1: Applicants will undergo a traditional job interview that is tailored to each applicant”).

After reading about the hiring procedure(s), participants were asked to respond to three items assessing the effectiveness of a hiring procedure:

How good of a choice is Hiring Procedure 1(2) for hiring future employees?
Hiring Procedure 1(2) is very appealing to me.

Hiring Procedure 1(2) is an effective way to hire new employees.

For the first item participants responded using a 5-point Likert-type scale from 1 (extremely bad choice) to 5 (extremely good choice). Participants were asked to rate their agreement on a 5-point Likert-type scale from 1 (strongly disagree) to 5 (strongly agree) on the last two items.

In order to assess individual differences in numeric ability, the Numeracy Scale (Lipkus, Samsa, & Rimer, 2001) was administered to participants (Appendix A). The measure consists of three general numeracy items and seven risk numeracy items. An example of a general numeracy item is, “Imagine that we role a fair, six-sided die 1,000 times. Out of 1,000 roles, how many times do you think the die would come up even (2, 4, or 6)?” Examples of risk numeracy items include: a) Which of the following numbers represents the biggest risk of getting a disease? ____ 1 in 100, ___ 1 in 1000, ___ 1 in 10, and b) If Person A’s risk of getting a disease is 1% in ten years, and person B’s risk is double that of A’s, what is B’s risk? The scale has a coefficient alpha = .78. Brooks and Yueh Pui (2010) found that numeracy is positively related to GMA, yet a distinct construct. The authors also found that numeracy was positively related with self-reported rational decision style ($r = .23$) and age ($r = .16$), and negatively correlated with the personality facets of extraversion ($r = -.22$) and neuroticism ($r = -.19$).
CHAPTER IV. ANALYSES

A 2 (evaluation mode: joint vs. separate) x 2 (hiring procedure: test of GMA vs. interview) between-subjects ANOVA was used to test the first hypothesis. Only the ratings for the primary selection procedure in the joint conditions were used when running these analyses. For example, when participants were presented with Hiring Procedures 1 and 2 simultaneously, only ratings for Hiring Procedure 1 were compared with ratings for Hiring procedure 1 in the separate condition. Conversely, when participants were presented with Hiring Procedure 2 and Hiring Procedure 1 simultaneously, only ratings for Hiring Procedure 2 were compared with ratings for Hiring Procedure 2 in the separate condition. The control condition, in which no validity information was presented, was used to evaluate the overall preference for an unstructured interview versus a general mental ability test using an independent samples t-test. Finally, a median split was used to categorize participants as high or low in numeric ability. One hundred thirty-six people were categorized as “high numeric” and 76 as “low numeric.”
CHAPTER V. RESULTS

Independent variables were evaluation mode (joint vs. separate) and hiring procedure (test of GMA vs. interview) and the dependent variables were the three-item scale that measured preference. The three items evaluating the interview had a coefficient alpha = .85 and the three items evaluating the test of GMA had a coefficient alpha = .87.

Hypotheses 1 predicted an interaction between evaluation mode and hiring procedure. This was tested using a 2 x 2 between-subjects ANOVA (Table 1). The ANOVA revealed an interaction approaching significance between hiring procedure and evaluation mode, $F (1, 211) = 3.71, p = .056$ (Figure 1).

Figure 1 shows the interaction between evaluation mode and hiring procedure. As you can see, people preferred the interview ($M=2.96; SD=.84$) to the test of GMA ($M=2.62; SD=.78$) when they were presented separately ($t=-2.11, p<.04;\text{ Cohen's }d=.42$). In the joint presentation condition the mean for the test of GMA ($M= 3.44; SD=.77$) was slightly higher than the interview mean ($M=3.36; SD=.80$) but, this was not statistically significant ($t=.57; p=.57;\text{ Cohen’s }d=-.10$). The data suggest that the general preference for the interview does not sustain when interview validity is presented concurrently with the GMA test validity.

Hypothesis 2, decision-makers will show an overall preference for the unstructured interview over the test of GMA absent validity information, was tested using an independent samples t-test for equality of means. This revealed that the test of GMA ($M = 2.97; SD = .83$) was viewed less favorably than the interview in the control condition ($M = 3.51; SD = .95$); $t (56) = -2.32, p = .024, d = -0.61$. This is consistent with Hypothesis 3. People preferred the interview to the test of GMA when no correlation coefficients were presented.
Hypothesis 3 stated that the difference between joint and separate conditions in evaluation of GMA tests would be greater for the low-numeracy participants. A 2 (evaluation mode: joint vs. separate) x 2 (hiring procedure: test of GMA vs. interview) x 2 (numeracy; low vs. high) between-subjects ANOVA was conducted (Table 1). No significant interactions were found between numeric ability and evaluation mode or hiring procedure. These findings are inconsistent with Hypothesis 4. Mean ratings for people high and low in numeracy can be found in Table 2.
CHAPTER VI. DISCUSSION

Past research has found that not all employee selection procedures are equal when predicting future performance (Schmidt & Hunter, 1998; Terpstra & Rozell, 1993). Schmidt and Hunter (1998) reviewed 18 selection procedures, including the interview and test of GMA. The authors found that the single best predictor of future job performance is a test of general mental ability. However, the interview remains the most commonly used method of selection among managers (Buckley et al., 2000). This gap between research and practice may be attributed to the methods of communication used when presenting research findings concerning selection methods to managers. One reason for this gap may be that managers are not familiar with or do not understand the statistical information concerning various methods of employee selection (Cosmides & Tooby, 1996; Rynes et al., 2002). It is important to find a way to present data which is easy to understand yet consistent with research findings.

This study used different presentation methods of statistical information to measure perceived effectiveness of the commonly used interview and the less popular, yet superior test of GMA. It was hypothesized that in separate conditions, when only information about the interview or the test of GMA was presented, participants would prefer the interview. This hypothesis was supported. We expected however that, when information concerning the interview and test of GMA were presented jointly, participants would prefer the test of GMA. Although, we did not find a preference reversal, we did observe an interaction between selection procedure and hiring mode (see Figure 1). Although both the interview and the test of GMA were perceived as more useful when presented jointly, the test of GMA was perceived substantially more favorably when presented jointly (M=3.44; SD=0.77) then when presented in isolation (M=2.62; SD=0.78; \(t=5.46, p<.001\); Cohen’s \(d=1.06\)). The difference between
interview favorability when presented alone (M=2.96; SD=0.84) versus when presented jointly (M=3.36; SD=0.80) was less substantial (t=2.54, p<.05; Cohen’s d=0.49). This suggests that when the two hiring procedures were presented alongside one another, people were able to compare the statistical information more easily. Even if they could not understand what a correlation coefficient means or how to interpret it, they could see that the test of GMA was relatively better than the interview. Simply presenting information about two hiring procedures side-by-side may help convey a message that would have been lost if only one procedure was presented.

This can be explained by the general evaluability theory (Hsee & Zhang, 2010), which suggests that people are “value sensitive”; they form decisions based on the importance of certain attributes that are quantifiable, for example, cost, time, or size. A preference is based on three factors: evaluation mode (joint vs. separate), the amount of knowledge a person already has about the attributes, and whether the attributes are able to be evaluated without comparison. Although this study was not focused on measuring the second factor, it can be assumed that most participants have been exposed to correlation coefficients, yet are unable to evaluate the information without a referent.

One possible explanation for the lack of a complete preference reversal in my study could be the overwhelming preference for the interview (Buckely et al., 2000). Employee selection research has shown both applicants (Lievens, DeCorte, & Brysse, 2005) and managers (Terpstra & Rozell, 1997) prefer the interview to any other type of hiring procedure. Even the correlation coefficient comparison of $r=.51$ (test of GMA) vs. $r=.20$ (interview) was not enough to overcome the strong preference for the interview. Hypothesis 3 tested the overall preference for the interview that has been found in previous research. It was no surprise that when the interview
and test of GMA were presented jointly minus any statistical information, the interview was
preferred to the test of GMA.

Not only is the interview preferred, it is possible that it is also more familiar. Research
has shown that familiarity leads to attraction (e.g. Moreland & Zajonc, 1982). The participants in
this study were upper-level psychology students, most of who have probably applied for a job at
some point in their lives (most likely entry-level positions). If this assumption is correct, it may
be that most of the participants have had more experience with the interview process, than a
cognitive ability test. Simply because the interview is more familiar to participants it is also more
attractive. It is difficult to overcome such strong preconceptions about hiring procedures;
however, this study gives promise to researchers who are looking for ways to communicate
findings to nonacademic audiences. One limitation of this study is the use of an undergraduate
student sample that is most likely unfamiliar with making hiring decisions and the implications
for such decisions. A sample of managers would be more informative, as they deal with hiring
decisions regularly. The next step in this line of research is to sample managers to see if our
findings hold. This exercise may be even more salient to managers than it was for
undergraduates and result in stronger findings.
CHAPTER VII. CONCLUSION

I found no significant interaction among individual differences in numeracy, hiring procedure, and evaluation mode. This did not support the hypothesis 4, that individuals high in numeracy would rate the test of GMA differently than those low in numeracy. A potential explanation for the lack of significant differences between high and low numeric people could be due to performance on the Numeracy Scale. Lipkus et al. (2001) found that 15-21 percent of participants answered the General Numeracy items correctly, where 24 percent of people in this study correctly answered the same questions. The researchers also found that 29-34 percent of participants answered the Risk Numeracy items correctly, however only 12 percent of people in this study answered all of the Risk Numeracy items correctly. It is uncertain as to why the sample did not perform according to norms.

This study suggests that the format of presentation is very important when communicating statistical information to lay people. More informed decisions could be made when information is presented in a comparative manner rather than independently. As researchers, we need to find a means of communication that is aimed at our target audience, CEO’s, managers, employees, etc. If presenting comparative data helps decision makers make better-informed choices, then there is little reason not to make this a regular practice. This could be a simple fix to a large communication problem.
### Table 1

Summary of Analysis of Variance: Difference between Hiring Procedure, Evaluation Mode, and Numeracy

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hiring Procedure</td>
<td>1</td>
<td>0.68</td>
<td>1.07</td>
<td>.00</td>
</tr>
<tr>
<td>Evaluation Mode</td>
<td>1</td>
<td>17.07</td>
<td>26.77</td>
<td>.01**</td>
</tr>
<tr>
<td>Numeracy</td>
<td>1</td>
<td>0.89</td>
<td>1.40</td>
<td>.00</td>
</tr>
<tr>
<td>Hiring Procedure*</td>
<td>1</td>
<td>2.87</td>
<td>4.50</td>
<td>.00*</td>
</tr>
<tr>
<td>Evaluation Mode*</td>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
<td>.00</td>
</tr>
<tr>
<td>Numeracy</td>
<td>1</td>
<td>.25</td>
<td>0.39</td>
<td>.00</td>
</tr>
<tr>
<td>Hiring Procedure*</td>
<td>1</td>
<td>.66</td>
<td>1.04</td>
<td>.00</td>
</tr>
<tr>
<td>Evaluation Mode*</td>
<td>1</td>
<td>.64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.*

* $p < .05$

** $p < .01$
Table 2
Mean Ratings of Hiring Procedure (Interview and test of GMA) for High and Low Numeracy

<table>
<thead>
<tr>
<th></th>
<th>Joint Evaluation</th>
<th></th>
<th>Separate Evaluation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Numeracy</td>
<td>High Numeracy</td>
<td>Low Numeracy</td>
<td>High Numeracy</td>
</tr>
<tr>
<td>Test of GMA</td>
<td>2.97</td>
<td>3.25</td>
<td>2.97</td>
<td>3.07</td>
</tr>
<tr>
<td>Interview</td>
<td>3.56</td>
<td>3.34</td>
<td>3.34</td>
<td>3.27</td>
</tr>
</tbody>
</table>
Figure 1.
Mean Ratings for Hiring Procedures in Joint and Separate Conditions
REFERENCES


on trust and comfort with hypothetical physician risk communication. *Medical Decision

patient’s prostate cancer screening outcome probability judgments*. Paper presented at
the annual meeting of Society for Judgment and Decision Making, Vancouver, British
Columbia, Canada.

Sage Publications, Inc.

Hazer, J. T. & Highhouse, S. (1997). Factors influencing managers reactions to utility analysis:
Effects of SD$_{y}$ method, information frame, and focal intervention. *Journal of Applied
Psychology, 82, 1*, 104-112.

*Industrial Organizational Psychology, 1*, 333-342.


and preference reversals in health care decisions. *Medical Decision Making, 24, 2, 142-148.*
APPENDIX A.

THE NUMERACY SCALE

General Numeracy Items

1. Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 roles, how many times do you think the die would come up even (2, 4, or 6)?

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each by a single ticket to BIG BUCKS?

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?

Risk Numeracy Items

1. Which of the following numbers represents the biggest risk of getting a disease? ___ 1 in 100, ___ 1 in 1000, ___ 1 in 10

2. Which of the following numbers represents the biggest risk of getting a disease? ___ 1%, ___ 10%, ___ 5%

3. If Person A’s risk of getting a disease is 1% in ten years, and person B’s risk is double that of A’s, what is B’s risk?

4. If Person A’s chance of getting a disease is 1 in 100 in ten years, and person B’s risk is double that of A’s, what is B’s risk?

5. If the chance of getting a disease is 10%, how many people would be expected to get the disease: A: Out of 100? B: Out of 1000?

6. If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.
7. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?
October 6, 2010

TO: Sarah Kirkendall
Psychology

FROM: Hillary Harms, Ph.D.
HSRB Administrator

RE: HSRB Project No.: H11T040GX2

TITLE: The Role of Joint vs. Separate Evaluation in Selection Procedure Decisions

You have met the conditions for approval for your project involving human subjects. As of October 5, 2010, your project has been granted final approval by the Human Subjects Review Board (HSRB). This approval expires on September 9, 2011. You may proceed with subject recruitment and data collection.

The final approved version of the consent document(s) is attached. Consistent with federal OHRP guidance to IRBs, the consent document(s) bearing the HSRB approval/expiration date stamp is the only valid version and you must use copies of the date-stamped document(s) in obtaining consent from research subjects.

You are responsible to conduct the study as approved by the HSRB and to use only approved forms. If you seek to make any changes in your project activities or procedures (including increases in the number of participants), please send a request for modifications immediately to the HSRB via this office. Please notify me, in writing (or email: hsr@bgsu.edu) upon completion of your project.

Good luck with your work. Let me know if this office or the HSRB can be of assistance as your project proceeds.

Comments/ Modifications:
Stamped consent form is coming to you via campus mail.

c: Scott Highhouse

Research Category: EXEMPT #2