THE APPLICATION OF AUTOMATED SCORING TECHNIQUES TO PREDICT PERSONALITY AND PERFORMANCE FROM JOB INTERVIEWS

Christopher Frost

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

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan

May 2013

Accepted by the Faculty of the College of Graduate Studies,

Central Michigan University, in partial fulfillment of the

requirements for the doctorate of philosophy degree

Thesis Committee:

Neil Christiansen, Ph.D.	Committee Chair
Matthew Prewett, Ph.D.	Faculty Member
Randall B. Hayes, Ph.D.	Faculty Member
May 23, 2013	Date of Defense
Roger Coles, Ed.D.	Dean College of Graduate Studies
July 1, 2013	Approved by the College of Graduate Studies

ACKNOWLEDGEMENTS

I would like to first thank all of those individuals who transcribed interviews to assist me in completing my dissertation. This was certainly a long and tedious process and therefore I am very grateful for all the assistance I received from research assistants, friends, and family members. I would also like to acknowledge HireVue for allowing me to view their online interviews. Without this rich data source I would not have been able to complete this project. Furthermore, I need to thank the graduate students at Central Michigan University who completed interview ratings. Finally, I would like to thank my advisor Dr. Neil Christiansen. Neil has served as a mentor and friend during my time in graduate school and today.

ABSTRACT

THE APPLICATION OF AUTOMATED SCORING TECHNIQUES TO PREDICT PERSONALITY AND PERFORMANCE FROM JOB INTERVIEWS

by Christopher T. Frost

The current study attempts to utilize an automated scoring methodology to predict applicant personality and interview performance. This study is unique in two ways. First, it is one of the first attempts to apply automated techniques to job interviews. Second, it will utilize two automated software programs, BETSY and LIWC, to generate automated scores. Individuals applying for a pharmaceutical sales position through participating in a structured online interview served as participants. Study 1 found that using BETSY and LIWC together was relatively effective at reproducing self-report personality scores. Study 2 examined the ability automated scores based on interview ratings to predict actual hiring decisions. Unfortunately, the results of this study were not nearly as encouraging as the Study 1 findings. Overall, while results may not support the contention that automated scores are reproducing self-report scores, there are multiple reasons to be encouraged by the findings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The employment interview is one of the most popular selection methods used in organizations today (Podsakoff, Whiting, Podsakoff, Mishra, 2011). Research indicates that reviewing resumes and application blanks are the only two selection methods used more frequently than the employee interview (Dipboye, 1997; Dipboye & Jackson, 1999). In addition to their popularity, organizations believe the interview is the most important aspect of the selection process (Eder & Buckle, 1988). This belief is unquestionably justified as the job interview has proven to be one of the most valid predictors of job performance (McDaniel, Whetzel, Schmidt, & Maurer, 1994, Huffcut & Arthur, 1994).

While interviews have proven to be valid predictions of job performance (Schmidt & Hunter, 1998), they are not without disadvantages. Specifically, one of the most obvious disadvantages to job interviews is the amount of time required to complete the interview process. To properly conduct an interview, time must not only be set aside to conduct the interview itself, but interviewers must be trained in order to appropriately conduct and score the interview. This training is a vital step in the interview process as adequately training interview raters tends to result in more reliable ratings (Schmitt & Chan, 1998). Overall, the total time commitment required for the interview process is so extensive that it is estimated that the time spent conducting an interview for a new job opening is equal to two full days of staff time (Gatewood & Field, 2001). In addition to the significant time requirements involved with conducting job interviews, improper scoring processes present another disadvantage to the job interview process. That is, when human raters are used to provide interview performance ratings, there are a number of biasing factors that may lead in inaccurate ratings (Huffcut, 2011). Considering these two disadvantages, a method that both reduces time requirements along with reducing rater bias and rater errors in interview ratings is certainly worthwhile to pursue.

One technique that may drastically reduce the time requirements associated with the interview process is application of an automated scoring method. By using an automated scoring process to evaluate interviews, the amount of time required to develop an appropriate scoring scheme as well as the time required to train raters are both significantly reduced. An automated scoring process would eliminate the necessity of this training process while still maintaining the reliability that results from conducting rater training. Furthermore, the application of an automated scoring system to job interviews maintains additional advantages such as increased objectivity, reducing rater biases, a reduction in applicant faking effects, as well as a more comprehensive examination of the information derived from the job interview responses.

CHAPTER II

LITERATURE REVIEW

Personality and the Job Interview

Among the numerous constructs assessed during the job interview, one of the most frequently assessed constructs is an applicant's personality (Huffcut, Conway, Roth & Stone, 2001). Additionally, personality is considered to be an important component of the job interview itself (Fletcher, 1987; Keenan, 1982). Although not as popular as the job interview itself, the examination of applicant's personality traits has increased a great deal in recent years (Goldberg, 1993; Hogan, 1991; Hough & Ones, 2001). Research findings displaying relationships between personality and job relevant criteria have no doubt contributed to this increase in popularity (Barrick & Mount, 1991; Zhao & Seibert, 2006). Although there has been considerable debate concerning the usefulness of personality in the selection setting (Tett & Christiansen, 2007; Morgeson, Campion, Dipboye, Hollenbeck, Murphy, Schmitt, 2007), a considerable amount of research has shown that personality can be a useful aspect to consider in the selection domain (Barrick & Mount, 1991). A variety of meta-analyses have also demonstrated the usefulness of personality testing in predicting a variety of job outcomes (Tett & Christiansen, 2007). In sum, personality appears to be a vital element of the job interview and the selection process in general. Fortunately, applying an automated scoring process to job interviews can be used not only to provide an interview performance score; it can also be used to predict personality based on job interview responses.

Applying Automated Scoring to the Job Interview

Automated scoring processes have been shown to be relatively accurate in predicting a variety of outcomes based on written or spoken passages (Cohn, Mehl, & Pennebaker, 2004, Fast & Funder, 2008; Hirst & Peterson, 2009; Pressman & Cohen, 2007; Yarkoni, 2010). However, this process has seldom been applied to spoken language and never applied in the job interview context. The current research aims to fill this void and apply an automated scoring process to the job interview in order to predict job interview performance ratings as well as an applicant's personality characteristics.

Study 1 will examine the ability of an automated scoring process to predict job applicant's personality characteristics. Furthermore, Study 1 will also examine the word and word pairs that are used more often by those with specific trait elevations. The purpose of Study 2 will be to examine the ability of the automated scoring processes to predict expert ratings of job applicant's interview performance as well as actual hiring decisions. Similar to Study 1, Study 2 will also examine the words and words pairs that are used more often by those job applications who received increased interview ratings as well as those applicants that were given a second interview. Finally, and perhaps most importantly, the present research will also aim to develop an automated scoring protocol that is designed specifically to predict meaningful job interview outcomes by analyzing interview response content. This final objective is of primary importance as it considers the context of the job interview.

For the purposes of this study, two Automated Essay Scoring (AES) programs will be used. The Bayesian Essay Test Scoring System (BETSY) will be used to identify the words and word pairs that differentiate between those applicants who received high interview ratings and those applicants who received decreased ratings as well as differentiating between those applicants who have certain trait elevations and those who are low in that particular trait. In addition to BETSY, the Linguistic Inquiry and Word Count (LIWC) software will also be used to predict self-report personality as well as interview performance ratings.

Automated Essay Scoring (AES)

AES is defined as the computer technology that evaluates and scores the written prose (Shermis & Barrera, 2002; Shermis & Burstein, 2003; Shermis, Raymat, & Barrera, 2003). Typical AES software programs such as BETSY and LIWC simply involve analyzing word counts and word pairs. Essentially, based on the amount of times that a particular word appears in a text passage, the AES process will determine how to score this text passage. Up to this point in time, this AES software has primarily been applied to the education setting, specifically to examine writing passages such as student essays. This methodology has proven to be particularly useful (Attali, 2004; Burstein & Chodorow, 1999) as scores/grades generated by the AES software are very similar to the scores generated by human raters or graders. This is a logical relationship as the ability of students to recognize and reference key words or concepts in their writing will result in increased scores. It is often the case that instructors may simply look for key words or references to key concepts when grading student essays. Therefore, one would expect that this simple word count analysis would be quite useful in this context.

However, when applied outside of the education setting, this word count method might seem overly simplistic. That is, will simple word count and word pair analysis from interview responses result in a meaningful index of a job applicant's personality and performance? Fortunately, research shows that using simple word count methods to analyze written responses has proven valuable in a variety of domains (Kufner & Back, 2010). This success is likely due in no small part to the intensive processes used to develop the text analytic software programs.

The Development of Text Analytic Software

Original Development. The first development of a text analysis technique dates back to the 1950's when a content analysis method was used to identify Freudian themes in writing samples (Gottschalk & Gleser, 1969; Gottschalk, Gleser, Daniels, & Block, 1958). This method simply consisted patients talking for five minutes in a "stream of Conscientiousness" while the audio was recorded. Judges then listened to the recorded audio and determined to what extent certain themes such as depression and death were evident. This "Gottschalk" method was also used in an attempt to diagnose patients with possible brain damage or mental disorders (Tauscik & Pennebaker, 2010). Attempts have been made to convert this Gottschalk method into a computer program, but these attempts have had little success (Gottschalk & Bechtel, 1993).

General Inquire. The first general purpose text analytic software that utilized computerized software was developed by Philip Stone (Rosenberg & Tucker, 1978). The program was labeled the General Inquirer and has proven to be relatively successful in identifying mental disorders, predicting personality traits, and evaluating speeches quality (Martendale, 1990). While this program has experienced some success, the primary disadvantage of the General Inquirer is that it utilizes complex algorithms to evaluate the text. Therefore, the user and the participant or patient would not be fully aware of *why* a particular score was given based on a particular piece of writing of spoken language (Tauscik & Pennebaker, 2010).

Weintraub Method. While most of the text analytic research up to this point in time that examined spoken language had primarily focused on the diagnoses of patients in a clinical setting, Walter Weintraub attempted to generalize this process and began to analyze words that are used in everyday language. In what was no doubt an extremely labor intensive process, Weintraub analyzed simple pronouns (e.g., I, me, my) that were used in political speeches by hand counting the number of times each of these words were used. Based on this analysis, it was revealed that these self-referencing words were linked to certain depression characteristics as well as other self-focused variables. Considering these results, it was proposed that by simply analyzing the number of pronouns someone uses, one can identify their psychological state. Unfortunately, and perhaps due to the labor intensive data collecting process, this method was largely ignored at the time (Tauscik & Pennebaker, 2010). Despite this general lack of attention from researchers at the time of the method's development, future research has revealed findings that lend credence to the results found by Weintraub. That is, research has found that those with a self-focus tend to magnify negative emotion and blame themselves more frequently than those who do not have a high self-focus. Numerous studies have shown a meaningful relationship between depression and self reports of self-focused attention (Ingram, Lumry, Cruet & Sieber; Smith & Greenberg, 1981).

Current. Overall, text analytic research has been taking place for over 50 years and a great deal of success has been had. However, processes proposed in previous research have not been without significant disadvantages such as an extremely time intensive or complex scoring process. Today, there are text analytic software programs available that aim to measure personality characteristics while maintaining both a high level of both simplicity and reliability. One of the most popular text analytic software programs is the LIWC (Pennebaker, Francis, & Booth, 2001). Another recent AES software program which has displayed success in scoring written responses is the BETSY software.

Overview and Development of LIWC Software

One of the most popular AES programs is the LIWC software, developed by Pennebaker, Francis, and Booth (2001). The stimulus for the development of the LIWC software occurred when researchers attempted to determine a participant's emotional style based on written essays. This research consisted of judges that rated the emotionality of essays. It was found that the time required for judges to read each of the essays was quite time intensive. Furthermore, the agreement between the judges was not as high as the researchers would have liked. Therefore, a more objective and efficient process was clearly needed.

The LIWC software contains 64 dictionaries/dimensions, most of which are designed to evaluate particular personality dimensions such as achievement, negative emotion, and anger. Each dictionary contains multiple words, usually 200 or more, that are linked to the dimension being assessed (the linking process will be discussed in more detail later). For example, one of the dimensions/dictionaries included in the LIWC software is anger. Some of the words that are included in the anger dictionary include; mad, blame, annoy, rude, and fight. Therefore, if a text passage contains the words mad or blame at a high frequency, this passage will then be scored high in the anger dimension. Additionally, it is important to note that the LIWC software examines the percentage of words that belong to a particular dictionary relative to all of the words used in a particular passage. Therefore, if one text passage simply contains more text than another, this will not result in this passage being scored higher on each of the dimensions simply because more words are used.

The LIWC software scoring method is entirely based on the dictionaries that it employs. Therefore, determining what words are linked to each dimension is obviously a critical process. This process of placing words into particular dictionaries took place in four stages. First, words that could potentially be included in a particular dictionary were identified by examining dictionaries, thesauruses, and questionnaires. A comprehensive list of each of the words was developed by research assistants. Second, groups of three judges independently rated whether each word candidate was appropriate for the overall category. Third, all categories of word lists were updated using the following three criteria; a word remained in a particular dictionary only if two of the three judges had included it in their list, a word was deleted from the dimension if two of the three judges said it should be eliminated, and a word was added to the list if two of the three judges said it should be included. The fourth step in this process involved repeating this process with a separate group of three judges. The separate group of judge's agreement with the first group ranged from 93% to 100%. For additional information pertaining to the development of the LIWC software, see Tausczik & Pennebaker (2010).

Perhaps the most useful feature of the LIWC software is the ability for users to generate their own dictionary. Users can generate their own personality traits and input words which are indicative of that particular personality trait. For example, consider a researcher that is interested in using the LIWC software to predict the Five Factor Model (FFM) of personality based on writing samples. One method this researcher could use is to display evidence of convergent validity by simply examining the default LIWC dimensions which should conceptually be related to each of the FFM dimensions. For example, if the negative emotion dimension contained in the LIWC default dictionary was closely correlated with Neuroticism, this would indicate that the LIWC is accurately identifying personality traits. However, a more reliable way to assess the FFM would be by developing an independent dictionary for each of the personality traits included in the FFM. In developing this dictionary, the user would simply input words in the dictionary which they believe are relevant to the trait which they are attempting to assess. For example, when creating a dictionary which measures Extraversion, a researcher may wish to include words such as talk, friends, party, gathering, etc. The more often these words are used, the higher Extraversion score this passage will receive.

While this idea of developing five independent dictionaries to examine each of the FFM personality traits is appealing, it is not without disadvantages. The primary disadvantage is that the user must determine which words should be included in each dictionary. As discussed previously, the initial process which was used to develop the default LIWC dictionaries took place in four steps and involved a great deal of time and resources. Additionally, the user would have to make sure to include enough words in each of the FFM dictionaries in order for each characteristic to be reliably assessed. Finally, the researcher should give strong consideration to context when developing dictionaries. For example, the words that differentiate between personality traits in written essays may not be the same words that differentiate between job applicants when analyzing their interview responses. Therefore, when attempting to apply the LIWC software to the analysis of job interview responses, the researcher should consider the context of the situation as this may have an impact on the types of words that are used.

Overall, it is clear that there are many challenges a user faces when attempting to develop a user-generated LIWC dictionary. Fortunately, another AES program, BETSY, has the capability to at least mitigate, if not completely eliminate, the challenges discussed here. This BETSY software has the capability to identify words which should be included in the LIWC user-generated dictionaries. Therefore, using the BETSY and LIWC programs together may be able to generate a job interview specific scoring process.

Bayesian Essay Test Scoring System

In addition to the LIWC software, a second type of software that can be used to predict personality as well as interview performance ratings is the Bayesian Essay Test Scoring System (BETSY). While this software is typically used in an educational setting to grade student essays and is rarely applied to the assessment of personality traits or interview performance, the software does have this capability. The primary difference between BETSY and other software programs, such as LIWC, is that the BETSY software must be trained to analyze text passages appropriately. At the start of this training process, text passages must be arranged into groups based on a particular variable, in this case personality traits or interview performance ratings.

In order to more easily describe the training process BETSY uses, consider a situation in which a researcher intends to utilize the BETSY program to predict individual's Extraversion levels based on written text. First, this researcher must obtain some index (self-report, observe ratings, etc.) of each participant's Extraversion level. Next, the researcher would place the participants into groups based on this Extraversion index. The researcher must then separate the participants into distinct groups based on their Extraversion levels. For example, the researcher may choose to identify three separate groups, high Extraversion, average Extraversion, and low Extraversion. The participants who scored in the top 33% in terms of Extraversion would be placed in the high Extraversion group, the participants who scored in the middle 33% would be placed in the average Extraversion group, and the participants who scored in the bottom 33% would be placed in the low Extraversion group. The researcher then inputs this information into the BETSY software. The BETSY software then examines each of these groups and identifies the words that are used must more frequently by the high Extraversion group than the low Extraversion group. For example, we might expect that the participants in the high Extraversion group would use words such as *talk, spoke, friends,* or *party* more often than those who are in the low Extraversion group. If this is in fact the case, then BETSY will recognize these words as important to predicting an individual's Extraversion level. This entire process is referred to as the training process. The BETSY software now recognizes important words which differentiate between individuals who are high and low in Extraversion. Although only three categories are used in this example, the BETSY software is capable of analyzing up to five separate categories.

When this training process is complete, *new* text passages can then be uploaded into the BETSY software and BETSY will use the words identified in the training process to predict which category each file should be placed into. Based on the previous example, if a text passage contains the word *friends* or *party* a great deal, BETSY will recognize these important differentiating words and predict that this text passage should belong to the high Extraversion group. BETSY's final output reports the probability that each text passage should belong to each of the groups predefined by the researchers. In this example, BETSY would yield three probabilities, the probability that a text passage belongs to the high Extraversion group, the probability that a text passage belongs to the average Extraversion group, and the probability that a text passage belongs to the low Extraversion group.

The fact that BETSY is trained based on relevant text passages in the primary advantage of this program. This training process ensures that BETSY is scoring passages based on relevant words which are relevant to the specific context. Another advantage of the BETSY software is that the words which are used to differentiate between different groups are revealed in the output. That is, each word is displayed and the amount of time the word is used by participants in each group is shown.

However, while this training process represents the primary advantage of the BETSY software, some of the most pertinent disadvantages to BETSY are also a function of this training process. Unfortunately, the BESTY software requires a large amount of text passages to be adequately trained. In many cases, the researcher may not have a great deal of participants at their disposal. For example, using 100 or more participants only for the purposes of adequately training the BETSY software may not reasonable as the researcher must obtain as much meaningful information from each participant. For these reasons, the BETSY software has primarily been employed as a way to score student essays and rarely applied to assistant in identifying personality traits from text passages. However, utilizing the BETSY software in combination with other AES programs can eliminate many of these disadvantages.

Developing Interview Specific Scoring Process

Both the LIWC and BETSY software programs have unique advantages and disadvantages. The LIWC software allows users to quickly input a text passage and have that text passage scored on a variety of personality traits. However, these default personality traits included in the LIWC software are somewhat generic and also are based on a specific context which may not generalize to other contexts such as the job interview. Meanwhile, the BETSY software allows users to examine any personality trait they desire and generates personality scores based on the relevant context. However, this training process requires a great deal of text passages, something which is not always available to the user.

While each of these programs has unique advantages and disadvantages, utilizing a process which employs both of these programs may be very advantageous. This process would involve first obtaining ratings on a particular variable for each participant, then dividing these participants into groups based on their scores on this particular variable and inputting this into BETSY. The BETSY software would then be used to reveal the words which differentiated between those high and low on the variable. These words which differentiated between individuals high and low on a particular variable would then be used to develop a user-generated LIWC dictionary.

When reviewing this process, a pertinent question may arise. *Why not simple continue to utilize BETSY alone after the relevant words are identified? Is that not the next step that BETSY takes after the training process?* There are two reasons why BETSY will not be used independently in this research. First, *all* of the words which BETSY identifies as potential differentiators will not be inputted into the LIWC user-generated dictionary. In the current study, for a word to be included in the LIWC dictionary, it must be conceptually relevant to the personality trait which is being examined. This analysis serves the function of a filter as words that are meaningless but still display a difference between those high and low on a particular personality trait will not be included in the LIWC dictionary. Using this human analysis also reduces the amount of text passages that need to train the BETSY software, as this is one of BETSY's primary disadvantages. Using a relatively small number of text passages in the training process, using BETSY alone would likely not yield accurate results due to random error. That is, there may be multiple words which are meaningless to the construct at hand but yet still show a large difference between the amount of times they were used by those high in a trait and the amount of times they were used by those low on a particular trait. Typically, BETSY controls for this by simply requiring a large amount of text passages. Obviously, the large the amount of text passages used, the less random error will occur. However, this random error can also be controlled for by carefully analyzing the differentiating words identified by BETSY before simply putting them into the LIWC software. This method allows for identification of meaningful words and an accurate scoring scheme based on much fewer text passages.

The second reason why BETSY will not be used independently in this process is that the BETSY software only examines the absolute frequency of the differentiating words. However, the LIWC software analyzes the words in terms of the amount of times the differentiating words are used relative to all of the words used in the entire text passage. Therefore, the score for each personality dimension in the LIWC software is displayed as a percentage (words in dictionary used/total words in the document). This is relative index is particularly advantageous with respect to this research as it is certainly possible that some applicants will talk more than others during the job interview. If BETSY is used independently when applied to job interviews, an interviewee who talks more during the interview will likely be rated high across all personality and performance dimensions.

Overall, creating user-generated dictionaries and using the BETSY software and LIWC software in combination has many advantages with respect to creating a more efficient and accurate automated scoring process in general. However, there are additional advantages realized that are specific to the job interview process with respect to both interview performance as well as the identification of relevant personality traits.

Advantages of Interview Specific Scoring Process

While the advantages of creating a new and specific scoring process (ie. user generated dictionaries) will likely have certain benefits in nearly any context, these advantages are likely realized more so in the job interview context than in any other context in which an automated scoring process may be applied. More specifically, there are at least three primary advantages to using a new interview specific automated scoring process in order to predict meaningful outcomes from job interview response content. In reality, it may be argued that attempting to predict meaningful interview outcomes without applying this new scoring process is a futile effort due to the numerous disadvantages of previous scoring processes discussed about.

The first advantage of using a job interview specific scoring methodology is that it allows users to identify narrow and relevant personality traits. This function allows for the researcher to examine traits which are particularly relevant to the job in question; or job performance in general. As previously discussed, and displayed in Appendix F, the personality traits included in the default LIWC software are somewhat generic and are unlikely to display meaningful relationships with job performance. In support of this contention, previous research has proposed that it is important to consider a situational specificity perspective when attempting to identify personality traits that will predict job performance (Tett & Christiansen, 2007). That is, not all personality traits should be examined for a particular job, as only traits considered relevant for the job will be useful in the prediction of job performance.

However, even across contexts, some personality traits have proven to be more valid predictors of job performance than others. For example, some personality traits such as achievement orientation and Conscientiousness have displayed meaningful relationships with job performance while other traits such as Openness have displayed relatively weak relationships with performance criteria (Barrick & Mount, 1991). Additionally, it is recommended that the job interview be developed to assess primarily the knowledge, skills, and abilities that are required for the particular job (Gatewood & Field, 1999).

Researchers and practitioners that have a clear understanding of a particular job based on a job analysis or previous experience should be able to accurately identify traits relevant to performing the job at a high level. Based on this information, a scoring protocol and user-generated dictionaries can be developed which specifically examine these relevant traits rather than assessing a wide range of somewhat generic traits. For example, perhaps an organization may wish to only assess traits such as achievement orientation that have been shown to predict relatively well across all jobs, along with a few personality traits that have been identified as being particularly relevant to the job.

A second important advantage of creating user-generated dictionaries is that developing a user-generated dictionary maintains the advantage of being able to predict constructs other than personality traits. While the LIWC software default scoring procedures are intended only to measure personality traits, this process can be improved upon as the LIWC has the capability to examine other constructs beyond personality traits. As mentioned previously when discussing the training process that BETSY uses, if text passages can be grouped together based on a particular variable then this variable can be predicted based on text analysis. Obviously, the ability of the automated scoring process to predict constructs beyond personality (ie. performance) has clear implications for the selection process.

A third advantage of user-generated dictionaries is that it allows the researcher to develop a dictionary that is particularly relevant to the job interview context. This advantage is particularity meaningful when one considers the fact that in some situations individuals will speak in a much different manner than they would in other situations. For example, during a job interview there may be certain words which are not used nearly as frequently as they would be in a typical conversation. Additionally, it also may be possible, if not probable, that particular words are used *more* often in interviews than in typical conversations. For example, if an individual frequently uses words such as accomplishment and succeed in casual conversation, this may indicate that they are a very conscientious and achievement oriented. However, during a job interview, an applicant will attempt to appear to display the characteristics of an ideal employee (Fletcher, 1990). While attempting to accomplish this, applicants in general will be much more likely to use words such as accomplishment, motivated, and/or succeed. For example, during the job interview, in order to make a favorable impression, even an unmotivated applicant may say "*I am extremely motivated to succeed when attempting to accomplish goals* *and will not give up until the job is done right*" or "*I consider my success at my previous job one of my greatest accomplishments*". Therefore, these words may not serve the function of differentiating among job applicants in terms of achievement orientation or Conscientiousness given that all applicants will likely use these words at a higher frequency than they would in a typical conversation.

Based on the fact that the LIWC software was developed based on casual conversations, it is important to examine if previous research based on casual conversations generalize to the interview process. In addition to the attempts applicants may make to appear more favorable, there are numerous other reasons why words spoken during a typical or casual conversation may not be the same as words spoken during a job interview. For example, in a formal interview context applicants may feel more obligated to speak about specific experiences and will perhaps be more cautious about revealing certain other information (Keats, 2009). Participants may also feel additional anxiety during the interview given that they are in a face to face situation in which they are being evaluated. The well known hawthorn effect demonstrates that people perform and act differently when they are aware that they are being observed (Adair, 1984). This anxiety is likely to increase when one considers the obvious evaluation present during the job interview.

Furthermore, Keats (2010) makes the contention that the situation plays a key role in determining the word choices used by an individual. The context of a particular situation can direct and focus the types of issues, ideas, or questions that arise. In more formal interview contexts applicants may feel more obligated to participate, be led in particular directions in speaking about experiences, or be more cautious in disclosures. Given the variety of situational factors that are present during the job interview, it is easy to imagine why an applicant's interview responses may differ from other less formal conversations they may participate in.

Despite the contentions made here, some research tends to indicate that the context of a particular situation does *not* influence the words which are used (Pennebaker & King, 1999). However, this research has limited implications with respect to predicting personality and interview performance based on the job interview context. The research conducted by Pennebaker & King (1999) consisted of participants writing short essays in the form of diaries in which they would describe their day. Due to the fact that the types of words used did not fluctuate from day to day, it was proposed that regardless of the situation the participant was in, the words used were relatively stable. This research did not consider a context nearly as structured as a job interview. The differences between contexts from one day of the week to another day of the week are much less obvious than the differences between a casual conversation and a job interview. Therefore, the contention made here remains. That is, in a job interview setting, the words which are used will be different than the words used in a casual conversations context and therefore an independent scoring system must be used in order to assess an individual's personality and job performance ratings based on the job interview.

Overall, while both the LIWC software and BETSY software have shown encouraging results in their intended domains, using these two software programs in combination may yield a process that has obvious and significant practical applications to the selection domain. Specifically, using these software programs in combination may assist in developing a context specific scoring scheme, predict for constructs outside of personality traits, and also allow for the prediction of narrow and job relevant personality traits.

Previous Text Analysis Research

Text analysis methods such as word count, word pairs, and linguistic style have been used in a variety of domains. Some of these domains include but are not limited to: counseling

(Weintraub, 1989), qualitative research (Keats, 2009), teaching techniques, predicting personality (Holleran & Mehl, 2008), and identifying clinical symptoms such as depression in patients (Rodriguez, Holleran, Mehl, 2010). Analyzing the words and writing style that an individual uses has proven to be useful in many of these domains by diagnosing clinical symptoms, identifying if a student understands course material, and organizing large amounts of qualitative data.

Until recently, researchers did not have much success when trying to correlate word use and personality traits. However, recently a variety of studies have shown that personality can be reliably assessed by examining word counts (Fast & Funder, 2008). Previous research involving text analysis is typically used to examine written text such as diaries or other writing samples. Novels, diaries, pager messages, and blogs have each been linked to the author's emotional, social, and cognitive styles (Cohn, Mehl, & Pennebaker, 2004, Fast & Funder, 2008; Hirst & Peterson, 2009; Pressman & Cohen, 2007; Yarkoni, 2010). In terms of the Industrial Organizational literature, more pertinent examples include work-narratives along with other forms of text displaying the ability predict important job outcomes as well as other variables (Hirsh & Peterson, 2009; Abe, 2009; Niederhoffer, Pennebaker, 2002). Due to the fact that text analysis is a relatively new research area in terms of predicting personality, a summary of the recent research which has examined text analysis and personality is provided below. The strong majority of word analysis studies have used the Linguistic Inquiry and Word Count (LIWC) software developed by Pennebaker et al. (2001).

Written Text Analysis and Personality

Hirsh and Peterson (2009) examined self-narratives and their ability to predict the big five personality traits. In this study, undergraduates participated in an automated writing program which required them to recall instances which occurred in the past as well as plans for the future. A simple word count procedure was used and the LIWC text analysis software was used. Each of the Big Five personality traits was strongly correlated with word use patterns which were theoretically appropriate for that trait. For example, dimensions which are generated by the LIWC software such as achievement and work were closely associated with Conscientiousness ($r = .22$ and $.21$ respectively). The LIWC dimensions such as sad and negative emotion were closely associated with Neuroticism ($r = .29$ and $.26$ respectively).

Kufner and Back (2010) analyzed written short stories and examined the correlations between the big five personality traits and the big five traits identified using the LIWC software. Personality was measured based on a self-report personality report as well as peer judgments. In their short stories, participants were required to use the words “plane crash” “parlormaid” “fireworks” “Middle Ages” and “supermarket”. These words were chosen in an attempt to induce a certain amount of creativity from the participants as the words had little relationship to each other. The study showed that all three of these methods to measure personality (word count, peer judgments, and self report) were each closely correlated with each other.

Given the relatively novel nature of text analysis and predicting personality traits, most of the current research typically only examines the Big Five personality traits. Lee and Cohn (2009) went beyond the Big Five traits and attempted to assess coping strategies by analyzing writing samples. In this study participants wrote about a stressful college experience and then completed three coping measures. Again, the LIWC software was used to analyze these writing samples. Results showed the amount of negative emotion words that a participant used was negatively related to problem focused coping. Additionally, insight words were highly negatively related to emotion focused coping. This research shows that more narrow personality traits such as coping can be assessed using text analysis methods.

Two research studies have been conducted which examine depressed writing samples and LIWC to indentify depressed individuals. The first study conducted by Rude, Gortner, and Pennebaker (2004) used written essays by currently depressed, formerly depressed, and never depressed college students. This research was consistent with previous models of depression, specifically the self-focus model proposed by Pyszcynski and Greenberg (1987). This model essentially proposes that individuals who are depressed will focus on themselves more often when interpreting an event rather than on others around them or on the situation in general. Lending credence to this contention, the results showed that depressed college students used more negative words and also used the word “I” more often than never-depressed college students. There was little difference between the never depressed college students and formerly depressed college students.

A second study that examined depression and word analysis was conducted by Stirman & Pennebaker (2001). This study examined the poetry writing of suicidal poets and non-suicidal poets. Similar to the previous study, and also confirming the results of the self-focus model, this study also found that suicidal poets tended to use more “I” words, focusing on themselves. While non-suicidal poets used more collective and “we” words.

Finally, an interesting but perhaps less relevant study by Kozbelt and Burger Pianko (2007) examined the words students in song writing. The words were used to attempt to predict a variety of personality traits, although the primary focus of the study was on creativity. Results based on the word counts were mixed and majority of the correlations of trivial. This is perhaps less surprising as student's who are attempting to write a successful and appealing song will be less likely to use personal emotions and feelings than those who are writing in a diary or a blog.

Text Analysis and Performance

Abe (2009) examined college student's weekly journals in an attempt to identify words which predicted supervisor's ratings of performance in a practicum course. The study revealed that positive emotional words as well as insight words such as *free, perfect, aware,* and *know,* were positively related with each of the dimensions of supervisor's ratings. Meanwhile, "we" words were more closely associated with supervisor ratings of interpersonal relations. A third and final interesting longitudinal finding from this research is that as the practicum course progressed, students began to use more "we" such as (*help, said,* and *relationship)* words and less anxiety words such as *afraid, aversion* and *fear.* This indicates that students were likely becoming more comfortable in the practicum course as it progressed and therefore did not experience or express as much anxiety. More "we" words were used as the class progressed because students likely formed friendships or at least improved working relationships with other classmates and expressed these relationships in the weekly journal.

Text Analysis and Cross Cultural Differences

In a study which examined 250 students written text from introductory courses, Donahue (2008) compared the differences between French and United States students using text analysis. The results of the study showed that factors such as the type of the assignment given, the student's level in school, and the texts read leading up to the assignment, had a stronger impact on the text used rather than the national setting. By using text analysis to examine these cross cultural differences, this research contends that there may be a general innate "discourse of academic learning" trait that all students apply rather than relying on cues from their natural cultural norms for guidance on how to write.

Spoken Text Analysis and Personality

While the vast majority of text analysis research has examined text which was originally in text format such as diaries and journals, a few research studies have converted spoken language into text and then analyzed this text. One of these studies conducted by Fast and Funder (2008) involved participants completing a 1 hour life history interview. This study identified a variety of words which were relevant to certain personality traits. The primary goal of this research was to identify words that although were used at a low frequency, may be useful for identifying certain personality constructs. The authors contended that even though words may be used at a very low frequency they may still differentiate between individuals with high or low levels of a particular personality trait.

One of the only other examples of research which applies text analytic software to spoken language in an attempt to identify personality analyzed presidential speeches. This study conducted by Slatcher, Chung, and Pennebaker (2007), analyzed television interviews, press conferences and campaign debates for four presidential and vice presidential candidates; John Kerry, John Edwards, George W. Bush, and Dick Cheney. This study examined the political candidates cognitive complexity, femininity, depression, presidentiality, and honesty. Results showed that there were meaningful differences on each of these constructs across candidates. Cheney's speech displayed the most cognitive complexity, presidentiality, and was the least feminine. John Kerry's language use was most strongly related with depression. Bush's language was most similar to that of an older person and was relatively dishonest compared to the other candidates. The findings of this research seem to generally support the common public opinion of each candidate.

Overall, there have been a variety of studies which have demonstrated the usefulness of the LIWC software in identifying personality traits. However, a great deal of this research has focused on written passages such as essays and diaries, while few studies have focused on converting spoken language into text and then analyzing the results. Furthermore, of the few studies that have examined word counts in spoken language, none of these studies have examined spoken language within the job interview context. The job interview context proposes unique circumstances in which certain restraints are present which may cause the words used in a typical job interview to be much different than the words spoken in a less formal context.

Scoring Methodologies

The current study will consider three separate scoring methodologies. These three methodologies will be utilized to examine the effectiveness of both software programs when used independently as well as when they are utilized in combination. The first scoring method, expert dictionary scoring, will produce only LIWC default dimensions. The expert dictionary scores will be based only on the frequency of words in a particular passage. The identification of the words included in each dictionary is discussed above.

The second scoring methodology, Bayesian scoring, will utilize the BETSY software independently. The BETSY software will be trained based on transcribed interview content and then will score additional interviews based on transcribed spoken content. BETSY will generate a probability that each transcribed sample belongs to each particular group.

Finally, the third scoring methodology will utilize both BETSY and LIWC software programs. These custom scores will include differentiating words, identified by BETSY, which are uploaded into the LIWC software and scored. Custom dictionaries will be created within LIWC for each of the personality traits pertinent to hypothesis here. Table 1 below summarizes the scoring methodologies that will be utilized in the current study.

Table 1. *Scoring Methodologies*

Scoring Method	Software	Scoring Process
1.Expert Dictionaries	LIWC	Scores generated from LIWC default dictionaries using only LIWC default words
2.Bayesian	BETSY	Categories created to train BETSY, scores are generated based on training
3.Custom	LIWC and BETSY	Words identified by BETSY are examined and uploaded into LIWC

CHAPTER III

STUDY 1 LITERATURE REVIEW

Interview-Based Assessment of Personality

Personality is one of the most frequently assessed constructs in the job interview, even more so than constructs such as knowledge, skills, and mental ability (Huffcut, Conway, Roth & Stone). Traditionally, these ratings of personality are made by the interviewer or other expert rater's who observe the interview. Unfortunately, despite the regularity that personality is assessed during the interview, there is not a great deal of research which examines this process. However, although it is not as strong as one might hope, there is some evidence which supports the contention that human raters are able to accurately identify personality characteristics based on interview responses.

Meta-analytic results that compared self ratings and observer ratings revealed that, among the Big Five personality traits, Extraversion displayed the greatest observed correlation, .45, between self and observer ratings (Connolly, Kavanagh, & Viswesvaran, 2007). The additional four facets of the Big Five traits each displayed a correlation between .30 and .38. In addition to the big five personality traits, research suggests that lay judges are fairly accurate in assessing other's integrity levels based on a brief ten minute interaction (Townsend, Bacigalupi, Blackman, 2006). While these correlations are certainly meaningful and represent a high degree of overlap between self and observer ratings, there is certainly unique variance between these two perspectives.

One method that has been shown to lead to an increase in the accuracy of a rater's personality ratings is to develop an interview guide designed to assess personality traits. When a specific interview guide is developed that is designed to assess specific personality traits, interviewer scores tend to correlate very highly with self-report personality measures (Trull, et al. 1998). Trull et al (1998) developed interview questions based on NEO items. For example, the Neuroticism item “I often worry about things that might go wrong” was transformed into the following interview questions/protocol: “Do you worry a lot about the future or things that might go wrong? [If yes] “What kinds of things do you worry about? [After response] “What proportion of the time?” Using this protocol, results showed that correlations among interviewer personality ratings and self-report scores were extremely high. The smallest observed correlation among the Big Five was Openness which revealed a correlation of .65, while Extraversion revealed the highest correlation, .84.

While the results from this study are encouraging, and further the belief that personality can be revealed through analyzing interview responses, these types of results are unlikely be revealed in a selection setting. The primary reason for this is that a job interview will seldom consist of questions that are directly designed to assess a particular personality trait, or at least not to the extent that the scale developed by Trull (1998) was designed to do so. Therefore, providing ratings of personality based on typical job interview questions is much more challenging than providing ratings for interview responses to questions that are designed to assess particular traits. Lending support to this argument is a study conducted by Barrick, Patton, and Haugland (2000). This study consisted of more typical job interview questions that were not specifically designed to assess personality traits. The results of this study revealed much more modest correlations between self report ratings and interviewer ratings, .27.

Furthermore, with respect to the accuracy of personality judgments made based on interview responses, research has shown that these judgments are more accurate when based on unstructured interviews than structured interviews (Blackman, 2002; Townsend, Bacigalupi, Blackman, 2006). However, research has shown that the structured interview maintains increased reliability and validity over the unstructured interview (Hunter & Hunter, 1984). It would appear based on these research findings that practitioners are faced with an unavoidable trade off. One may choose to utilize the structured interview and maintain increased validity and reliability while also being subject to less accurate assessment of applicant's personality traits. If the unstructured interview is used, an applicant's personality may be accurately assessed; however the overall interview may suffer decreased reliability and validity. While this trade off may appear to be inherent to the job interview process, the application of an automated scoring technique may allow for accurate assessment of applicant's personality traits while the reliability and validity of the job interview is also maintained.

This finding that judge's are relatively accurate when predicting interviewee's personality based on job interviews may be somewhat surprising, especially when considering that that job interviews are typically designed to measure specific job-relevant constructs (Huffcut, et al. 2001) and not intended to reveal personality traits. Considering this, it may be useful to consider what cognitive processes judges use when examining applicants personality traits as well as the observable characteristics that are most closely attended to. The processes used by judges are seldom examined in research and are perhaps worthy of further consideration given judge's relative accuracy in identifying applicant's personality. Due to the fact that the content of the interview responses will likely be indicative of job relevant constructs and not personality, perhaps judges are not only attending to only the content of an applicant's responses. There may be a variety of other elements, such as non-verbal behaviors, that judges used to develop personality ratings. DeGroot (2009) attempted to shed light on this process and found that judge's ratings of personality traits were in fact heavily influenced by nonverbal cues. This study showed that when judges watched interviews with no audio, their personality ratings correlated very highly with those individuals who watched the interview with audio.

Improving Judge's Assessment of Personality

Overall, while researchers have certainly enjoyed some success with respect to identifying personality based on interview responses, this process could certainly be improved. This is the case for four reasons. First, interviewer ratings of personality correlated very highly with self-report personality only when questions were used which were designed to examine specific personality traits. This is not realistic to a typical job interview process as job interviews are typically designed to measure specific job relevant constructs (Huffcut, et al, 2001).

Second, it appears that interviewers may be much more effective when attempting to identify broad personality traits, such as Extraversion. Predicting more narrow/specific traits seems to be more difficult to accurately assess. Considering that each job has unique responsibilities which may require these somewhat narrow personality traits, assessing these more narrow traits which are relevant to specific job requirements will likely be beneficial. For example, a customer service job will likely involve a strong interpersonal component. For this job, the personality trait of Agreeableness and extraverted may be highly related to performance. However, for a job such as a computer programmer these traits will likely show a trivial relationship with job performance and other important job outcomes. Therefore, when using a personality measure in the selection setting it is vital to consider what the specific job is and consider which personality traits are relevant to the job. In one of the most popular meta-analysis concerning the big five personality traits, Barrick and Mount (1991) found that Conscientiousness showed a consistent relationship with performance across all jobs. However, other personality traits were only related to performance for specific occupational groups. For example, Extraversion was a valid predictor for jobs concerning social interaction, managers, and sales.

This reasoning is that a certain personality trait may be a valid predictor for some jobs but not others is referred to as situational specificity (Muprhy, 2000). Supporting this view is the fact that research has shown that there is considerable variability in the validity of tests across different occupations (Gutenberg, Arvey, Osburn and Jenneret, 1983, Hunter and Hunter, 1984). This situational specificity that is likely to have a large impact on the validities of certain personality traits.

The LIWC software allows users to create their own dimensions and develop dictionaries which contain words which are linked to that dimension. Therefore, ideally organizations would be able to first identify what personality traits are relevant for a particular job and use or develop a dictionary which is developed specifically to examine only traits which are relevant for that job and only words which have proven to be related to that personality trait in the job interview context.

Third, the use of automated essay scoring may allow for organizations to enjoy the increased validity and reliability which is present in the structured interview, while also maintaining the ability to accurately predict personality traits. Additionally, research has revealed that those making personality judgments of interviewee's personality traits give strong weight to nonverbal cues. While this has proven to be relatively effective, it is possible that there is information contained in the content of the interview responses that the judges are not attending to the extent that they should be. Using an automated process to predict personality may be very beneficial in complimenting a more subjective evaluation of personality traits.

Overall, it is clear that the use of an automated scoring process to evaluate interviews has numerous advantages. This advantages apply both to the identification of personality traits and yielding an accurate interview performance score. To develop an automated scoring process that is specific to the job interview, the current research will utilize Automated Essay Scoring (AES) software to analyze employee's interview responses. Specifically, the Bayesian Essay Test Scoring System (BETSY) and Linguistic Inquiry and Word Count (LIWC) software will be used. While the LIWC software has proven to predict personality traits in previous research (Holleran & Mehl, 2008), no current research has applied this software to the job interview context specifically. The vast majority of previous research involving the LIWC software has focused on using the LIWC software to analyze written text such as diaries or essays (Stirman & Pennebaker, 2001). Therefore, this research will analyze the ability of this software to generalize to the job interview context as well as creating an automated scoring system that is specific to the job interview.

Predicting Personality from Word Counts

As mentioned previously, the method of simply analyzing word counts to predict personality from interviews may seem relatively simple. One might argue that only analyzing the words which a job applicant uses during the interview will not provide enough meaningful information to predict personality traits. However, while these concerns are certainly legitimate, research does in fact lend support to the contention that word counts derived from job interviews are linked to personality traits. Specifically, there are three reasons why we can expect personality and language to be closely linked based on word counts alone (Fast & Funder, 2008). First, the interview contains a great deal of information (ie. words). In fact, the amount of words used in an interview is so vast that it is virtually impossible for the interviewer or interview raters to attend to each spoken word. Combining this vast amount of information with the fact that automated software can analyze every single word, increases the confidence one might have when considering the ability of word count analysis to predict personality based on job interview responses. Second, research has shown that word use is very closely related to personality (Fast & Funder, 2008). Third, spoken and written samples tend to be closely related to each other. Analysis of word counts within written samples has already displayed meaningful results in terms of predicting personality traits.

Structured interviews generally consist of two types of interviews, the situational interview and the past behavioral interview (Campion, Palmer, & Campion, 1997). The situational interview requires an applicant to respond to hypothetical situations in terms of the actions they would take in that situation (Latham, Saari, Pursell, & Campion, 1980). The past behavioral interview requires applicants to indicate recall how they have responded to certain situations in the past (Latham & Sue-Chan, 1996). Responses to these questions generally consist of detailed stories where a relatively large amount of words must be spoken in order to develop a coherent dialogue. Although human ratings of these responses have been highly correlated with job performance (McDaniel, Whetzel, Schmidt, & Maurer, 1994), as discussed previously, it is impossible for an individual human rater to assess every single word or phrase which was uttered during the interview process (Fast & Funder, 2008). This would be a task that is simply too cognitively complex for one rater to complete. Pennebaker et al. (2003) contends that when individuals are participating in a conversation, the task of paying attending to the flow of the conversation and what is actually being said is so complex and requires so many cognitive resources that the ability to recognize if a particular word was used or not becomes virtually impossible. Further, Chung and Pennebaker (2007) propose that simple word counts provide subtle information which can be more reliable than even the most conscientious rater.

When one considers the limitations of human raters to attend to each word spoken during the interview, the benefits of text analytic software becomes obvious. By using text analytic software to analyze the interview, the software will be able to analyze each word spoken during the entire interview. This process would allow raters to capitalize even further on the detailed information that the interview provides. Typically there is a vast amount of information which results from an employee interview, using an automated process affords researchers, organizations and practitioners the ability take advantage of each piece of it. Overall, while on the surface, word count analysis may seem too simplistic to generate a meaningful prediction of personality. This belief may only be maintained due to the fact that the human eyes and ears are simply incapable of recognizing the vast amount of information that is available for the automated scoring process to utilize.

The second reason to believe that simple word counts can be used to extract personality from the job interview is that the idea of using word counts scores is congruent with the very definition of personality (Fast & Funder, 2008). For example, if someone is generally narcissistic they will tend to focus on themselves. Therefore one would expect such an individual to use words such as *I*, *me*, or *myself* more often than a person who is low in narcissism. As will be discussed later, research has shown that human raters are relatively accurate when attempting to predict personality based on interview responses (Connolly, Kavanagh, & Viswesvaran, 2007). Human raters, although it may not be a conscious process, are likely to analyze the words that a person uses when attempting to determine an individual’s personality during an interview.

A final reason why word counts may be used to accurately predict personality traits based on interview responses is that typically written samples and spoken samples tend to be closely related (Pennebaker & King, 1999). Recent research has shown that written text such as diaries or essays has been able to reliably predict personality (Cohn, Mehl, & Pennebaker, 2004). Therefore, if spoken text and written text are in fact closely related, we would expect that word counts could also predict personality based on spoken language relatively well. Overall, based on the amount of information provided in the job interview, the ability of the automated scoring process to analyze every single spoken word, the relationship between word use and personality, and the relative success of previous research, we can be at least reasonably confident that using word count analysis can be used to reliably assess personality based on interview responses.

This research will assess each of the Big Five personality traits utilizing a self-report methodology. Similar to the method used by Hirsh & Peterson (2009), where written samples were used, this research will attempt to demonstrate that LIWC dimensions that are conceptually related to specific self-report personality traits will be positively related to each other. Although previous research has demonstrated meaningful relationships among these constructs, it is relatively unknown how well these dimensions will relate to each other in the job interview context. However, given the pervious success of research using the expert dictionaries included in the LIWC software, it is expected that Expert Dictionary scores and Self-Report scores will yield meaningful relationships. Therefore the following hypotheses are proposed.

Hypothesis 1: Self-Report personality scores will be positively related to expert dictionary personality dimensions that have a strong theoretical and rational link

While it is expected that the Expert Dictionary Personality dimensions will relate to Self-Report personality traits that are conceptually related, the primary purpose of this research is to examine the ability of a Custom Automated personality scores to predict Self-Report personality in the job interview context. That is, it is expected that the Custom Personality scores generated by using the BETSY and LIWC software programs will display positive relationships with Self-Report personality scores. Therefore, the following hypothesis is proposed.

Hypothesis 2: Custom Automated Personality Scores will be positively correlated with Self-Report Personality Scores.

As discussed here, it is expected that the Custom Personality Scores will be more effective in predicting Self-Report Personality Scores than the Expert Dictionary Scores or Bayesian scores alone. Bayesian scores will be generated using the BETSY software. While the BETSY software will primarily be used to identify differentiating words, the BETSY software is also capable of scoring samples. Before scoring can take place, BETSY must first be trained based on files that are placed into groups based on a particular criterion. The resulting Bayesian Personality Scores will be compared to the Custom Personality Scores. While the Bayesian Scores will evaluate all words that appear in the transcriptions, the Custom Scores will be based only on words that demonstrate differentiating power and maintain a strong theoretical basis. Based on the fact that the custom scores will be based only on words that differentiate and maintain a strong theoretical basis, the following hypothesis is proposed.

Hypothesis 3: Customized Automated Scores will explain a significant amount of incremental variance in Self-Report Personality Scores, over and above what is predicted by Bayesian Scores alone.

CHAPTER IV

STUDY 1 METHOD

Participants

Job applicants who completed an on-line interview for a pharmaceutical company served as participants for this study. A total of 189 applicants completed an online personality questionnaire and therefore fulfilled the requirements to serve as participants in this study. Applicants were applying for a pharmaceutical sales position by completing the online interview. Most participants had previous job experience in the sales industry. Applicants from a variety of locations completed the online interview.

Each of the job applicants that had completed an online interview prior to July 13th, 2012 was sent an e-mail with a link to an online survey. By completing the survey, the applicants were entered into a drawing in which four participants would be randomly selected and awarded 50 dollars. Applicants were informed explicitly that the results of the survey would have no impact whatsoever on chances of obtaining the job.

Procedure

Job Interview

The job interview itself consists of ten separate questions. Applicants are given three minutes to answer each question. To initiate the interview, applicants simply log on to the website, upload their resume, and enter the interview. The applicant is not permitted to access the interview unless they have a functional audio and video recording device that is currently recording. When the applicant enters the online interview the question appears for the applicant on the screen along with a text box that displays the amount of time they have left to complete the question. If the applicant completes the question before the three minutes have expired, they are instructed to simply click to the next question. The interview consists of ten questions; applicants are given three minutes to provide a response to each of these questions. The ten questions are located in Appendix D. A total of five research assistants transcribed each of the interviews.

Automated Scoring

Expert Dictionary Personality Scores. To generate Expert Dictionary Personality Scores, the transcribed interview responses were uploaded into the LIWC software. Each candidate's response will then be automatically scored on each of the LIWC default dimensions. Only the words included in the default LIWC dictionaries will be utilized to score each candidate's responses. Words to be included in each dictionary were identified by experts during the LIWC's developmental process.

Bayesian Personality Scores. After all the data were gathered, the participants were rank ordered on each personality trait examined. For each personality trait, three separate groups were created. One group consisted of the highest 10 scores, another group consisted of applicants with the lowest 10 scores, and the final group will consist of the 10 applicants who scored closest to the median score. After the three separate files were created for each personality trait examined, these groups were then uploaded into the BETSY software. The BETSY software was then trained based on three groups of ten interviews.

The remaining interviews were then loaded into the BETSY software. BETSY then scores each interview based on the transcribed content. This scoring is based on Bayesian theory as provides three separate probability statistics for each transcript are generated. More specifically, the probability that each interview transcript belongs to each of the three categories created is generated. The probability is based on the similarity between the words used in each transcript and the transcripts included in each group.

Finally, these three probability scores were combined to form one Bayesian score for each personality trait. To create this Bayesian score, the probability that a transcript belonged to the high group was multiplied by 3, the probability that it belonged to the median/medium group was multiplied by 2, and the probability it was in the low group was multiplied by 1. An aggregate of these three products was utilized as the Bayesian Personality score.

Custom Personality Scores. In addition to creating the Bayesian Personality scores, the BETSY software also identified the words that occurred with much greater frequency among one group as compared to the other groups. These words were then examined by the researcher. Words that occurred most frequently in one particular group and also had a strong theoretical basis were added to the custom LIWC dictionary. This custom LIWC dictionary was utilized to generate the Custom Personality Scores for each personality trait.

In addition to analyzing words that are spoken more frequently by those with trait elevations, the current research will also examine words that are spoken more frequently by those with low trait scores. Unfortunately, the LIWC software that is utilized here is only capable of increasing an individual's score when a particular word appears. That is, only words that differentiate and are used more often by those in the higher group can be included in the custom LIWC dictionary. Therefore, two customized dictionaries were created for each personality trait. This allowed for the generation of custom personality scores that examined words used less often by those with trait elevations. More specifically, one customized dictionary consisted only of words that were used more frequently by those with elevated trait levels (using the standard

LIWC methodology). A second customized dictionary contained words that were used more frequently by those applicants with low levels of a particular trait. Each of the scores generated by this second dictionary was multiplied by -1 for each trait. These transformed automated scores were then added to the scores generated by the original dictionary which contained words that were used more frequently by those with trait elevations. This final aggregate score was utilized as the final Custom Personality Score for each personality trait.

Measures

Five Factor Model of Personality

In order to examine the personality traits that are included in the Five Factor Model of Personality, a revised version of the 50-item IPIP was used (Donnellan, Oswald, Baird, Lucas, 2006). Participants were presented with 35 statements and asked to indicate the extent to which each statements describes them on a 1 (does not describe me at all) to 5 (completely describes me) scale. A total of seven items were used to assess each of the personality factors included in the FFM, resulting in a total of 35 items. Due to time restrictions, the entire 50-item IPIP could not be used. Items were selected by the researchers based on redundancy and appropriateness of each item. The abridged IPIP scale that was used in this research can be found in Appendix A.

Additional Personality Traits

The Thinkwise Selection Development Index (TSDI) was utilized to assess additional personality traits. The TSDI contains 32 competencies that are each measured using eight items. Based on the interview questions, the job applied for, and the interview responses, four of the traits included in the TSDI were identified as being particularly relevant for the job and also revealed in the job interview (Kingwood, 2008). These traits included Action Orientation (I never procrastinate on difficult tasks), Customer Focus (I can easily convince others to purchase products or services), Conflict Management (I am comfortable making quick decisions), and Interpersonal Savvy (I am easy to get to know). Items were selected by researchers to eliminate redundancy as well as minimize the time needed to complete the survey. The items from the TSDI that were used in this research can be found in Appendix B.

Affectivity

To measure affectivity, an abridged version of the Positive and Negative Affect Schedule (PANAS) was used (Clark, Watson, & Tellegen, 1988). Participants indicated how well 12 words described how they generally felt on a 1 to 5 scale. Six of the words are generally reflective of positive affectivity, such as excited and proud. The other six words, such as distressed, and nervous, are designed to measure negative affectivity. The entire PANAS scale used in this study can be found in Appendix C.

Materials

To analyze the content of interview responses, two software programs will be used. The first is the Linguistic Inquiry and Word Count (Pennebaker, Francis, Booth, 2007). This software was originally developed with the intention to analyzing the emotional content of an individual's writing (Pennebaker, Matthias, Mehl & Niederhoffer, 2008). However, recently the use of the LIWC software has been used to analyze a variety of texts, including personal narratives, press conferences, and transcripts from typical conversations (Pennebaker & Graybeal, 2001). This software is by far the most frequently used software in terms of analyzing text and has shown to be highly correlated with ratings made by trained judges (Hirsh, Peterson, 2009). LIWC uses a word count method in which it searches for over 2300 words or stems. Each of these words is linked to a particular personality trait or other dimensions, the entire list of dictionaries can be found in Appendix F.

The second software program that will be used is the Besyian Essay Test Scoring System (BETSY). The method of using the Besyian theorem has shown to be promising (McCallum, Rosenfeld, Mitchell, Ng, 1998; Rudner, Liang, 2002). The BETSY program has primarily been used to analyze student essays and is used to put essay into specific categories. For the purposes of this research, the BETSY program was primarily utilized to identify words that are used more often by those high in a personality trait than those who are in low in the same personality trait. These words were then used to develop a dictionary scoring system which can be used in the LIWC program.

CHAPTER V

STUDY 1 RESULTS

Displayed in Table 2 are the descriptive statistics and correlations among the Self-Report Personality scores. In general, reliabilities were acceptable and correlations among the personality traits were as expected. With the exception of Neuroticism and Customer Focus, all reliabilities were .70 or higher.

Table 2. *Self-Report Personality Correlations*

	M	S.D.	1	2	3	4	5	6	7	8	9	10	11
1.Extra	31.14	3.43	.78										
2.Open	27.90	3.47	.48**	.72									
3.Agree	29.32	3.49	.50**	.50**	.72								
4.Cons	30.34	3.71	.27**	.26**	.38**	.79							
5.Neuro	14.30	3.93	-.46**	-.36**	-.29**	-.38**	.69						
6.Act Ori	33.18	3.92	.57**	.33**	.32**	.49**	-.54**	.70					
7.Con	33.43	4.14	.54**	.35**	.40**	.48**	-.65**	.75**	.75				
8.Cust	31.57	3.21	.41**	.38**	.54**	.29**	-.23**	.41**	.51**	.56			
9.In	36.00	3.59	.69**	.41**	.59**	.39**	-.40**	.60**	.56**	.66**	.78		
10.P. A.	26.89	2.74	.63**	.42**	.39**	.39**	-.43**	.59**	.63**	.47**	.70**	.80	
11.N.A.	8.58	2.89	-.36**	-.31**	-.18*	-.38**	.53**	-.61**	-.54**	-.31**	-.40**	-.39**	.78

Note. *N*=189. Extra = Extraversion, Open = Openness, Agree = Agreeableness, Cons = Consciousness, Neuro = Neuroticism, Act Ori = Action Orientation, Con Man = Conflict Mangement, Cust Foc = Customer Focus, In Savvy = Interpersonal Savvy, P.A. = Positive Affectivity, N.A. = Negative Affectivity. Reliability for each scale is displayed on the diagonal axis. * = $p<.05$ ** = $p<.01$

The first hypothesis examined relationships among Expert Dictionary Personality Scores and Self-Report Personality Scores. Displayed in Table 3 are the relationships among Expert Dictionary Scores and Self-Report Personality Scores. Overall, hypothesis 1 was not supported. More specifically, the Self-Report Extraversion and the Friends Expert Dictionary Scores were the only two constructs that displayed a meaningful relationship in the expected direction.

Overall, by examining Table 3, it is apparent that utilizing only Expert Dictionary Scores to predict Self-Report Personality is ineffective in the job interview context. Examining the results as a whole, of the 84 correlations displayed in Table 3 only five of these correlations reached statistical significance. Although not all of these relationships are theoretically appealing, this further demonstrates the lack of relationship observed between Expert Dictionary Personality Scores and Self-Report Personality. Furthermore, it is worth noting that there are numerous Expert Dictionary Dimensions that are not displayed in Table 3. However, many of these personality scores are not pertinent to this study and are not conceptually related to the personality traits examined here. Therefore, there were no hypothesized relationships between these additional dimensions and self-report personality traits. Overall, the data provides virtually no support for hypothesis 1.

Table 3. *Expert Dictionary Personality and Self-Report Personality Correlations*

Expert Dictionary Dimensions	Related Self-Report Personality Trait	Self-Report Personality						
		E	O	A	C	N	PA	NA
Negative Emotion	Neuroticism and NA	.10	-.04	.10	-.04	.05	.10	.03
Anxiety	Neuroticism	-.04	-.08	-.02	-.07	.07	-.01	.19*
Anger	Neuroticism	.13	-.09	.07	-.04	.16	-.05	.10
Sad	Neuroticism	.12	.01	.01	-.10	.04	.14	-.05
Cognitive	Conscientiousness	-.13	.01	-.06	-.05	-.01	.08	.01
Insight	Conscientiousness	-.07	.03	-.10	-.05	-.02	.04	.05
Work	Conscientiousness	-.02	.05	.00	.03	.07	-.02	-.08
Achieve	Conscientiousness	-.05	.11	-.02	.08	-.02	.03	-.01
Tentativeness	Openness	.05	.12	.06	.00	-.18*	.08	.05
Social	Extraversion	-.11	-.03	-.15	-.02	.05	-.03	.16
Friend	Extraversion	.17*	-.14	-.07	-.12	.16	.19*	-.17*
Positive Emotion	Positive Affectivity	-.12	.02	-.01	.02	.02	-.04	.06

Note. *N*=189, E= Extraversion, O=Openness, A=Agreeableness, C=Conscientiousness, N=Neuroticism, PA=Positive Affectivity, NA=Negative Affectivity,* = *p*<.05 ** = *p*<.01

The first step in creating Custom Personality Scores is utilizing the BETSY software to identify key differentiating words. In order to identify these words, the interview transcripts were separated into three groups based on their standing on each self-report personality trait. Displayed in Table 4 below are the words identified by BETSY that differentiated between groups. More specifically, these words were used most often by one particular group (high, medium, low) of applicants. Furthermore, these are the words that are utilized to create the Custom Dictionaries that are utilized to create Custom Personality Scores.

Table 4. *Differentiating Words among Self-Report Personality*

Personality trait	Differentiating words
1.Extraversion	phone, motivated, career, accomplishments, develop, confident, sell, talked, chatted, customer, he, nervous*, patient*, tough*, improve*
2.Conscientiousness	progress, struggle, better, overcome, patient, win, all, best, money, solve, schedule, sad*
3.Openness	schedule, talk, control, working, stress, chatted, plan, he, people, conversation, sad*, Chilcott*
4.Agreeableness	develop, career, learn, office, solved, phone, he, nurse persistence*, persistent*, money*, no*, sign*, paid*
5.Neuroticism	sad, quit, employee, Chilcott, Warner, I, nervous, talked*, customer*, solved*, rank*, progress*, good*, great*
6.Action Orientation	progress, customer, he, number, win, developed, provide, better, solved, control, persistent, schedule, sell, career, motivated, persistence
7.Customer Focus	struggle, he, learn, sell, number, skill, ask, sales, I*, quit*, no*, employee*
8.Conflict Management	product, progress, customer, better, win, struggle, career, employee*, quit*, I*, patients*, sad*, no*
9.Interpersonal Savvy	sell, growing, phone, learn, trust, better, customer, he, solved, quit*, no*
10.Positive Affect	sell, motivated, better, solved, control, he, overcome, customer, customers, sale, talk, employee*
11.Negative Affect	employee, give, tell, tough, wouldn’t, sell*, accomplishments*, growing*

Note. *N*=30. Applicants were divided into three groups, each group consisting of ten interview transcripts. All words displayed in the table were used at least twice as often by those in one group as compared to another. *= Words used less frequently by those with elevated personality scores.

As indicated in Table 4, the BETSY software identified multiple words that were used more frequently by those applicants who displayed relatively elevated or lower levels of a particular trait. For example, an applicant low in extraversion was more likely to use the word nervous than an applicant who was highly extraverted. It is worth noting that there were multiple words in addition to those displayed in Table 4 that differentiated among applicants. However, these words did not have a strong theoretical basis as to why they would reliably differentiate between candidates' personality characteristics. The words shown here are also utilized to create the customized LIWC dictionaries. These dictionaries were used to create Custom Personality Scores.

In general, there was a relatively high degree of overlap among the personality traits in terms of which words effectively differentiated. This is somewhat expected for two reasons. First, the personality traits themselves are highly correlated. Regardless of the reason for these relationships, the self-report scores were utilized to identify differentiating words here. Second, the majority of the personality traits examined here involve an interpersonal component. For example, elevated levels of extraversion, openness, and interpersonal savvy would each likely increase the probability of interpersonal interaction. Therefore, words such as *he*, *nurse*, *people* and *customer* effectively differentiated for multiple personality traits. More specifically, candidates who displayed high levels of these interpersonal personality traits would be more likely to discuss previous interpersonal interactions during the job interview.

Hypothesis 2 examined the relationship between Custom Personality Scores and Self-Report Personality Scores. Displayed in Table 5 below are the relationships among Custom Personality Scores and Self-Report Personality Scores. Based on a preponderance of evidence, as displayed in Table 5 hypothesis 2 is supported. Of the 11 personality traits examined here, eight displayed significant relationships between the Custom Scores and Self-Report Scores. Agreeableness and Extraversion displayed the strongest relationships between Custom Scores and Self-Report Scores, yielding correlations of .39 and .34 respectively.

Table 5. *Self-Report Personality and Custom Personality Correlations*

	Self-Report Personality										
Custom Scores	Ex	Open	Agree	Con	Neu	Act	ConM	CusF	IntSav	PA	NA
1.Extraversion	.34*	.21*	.17*	-.06	-.19*	.32**	.27**	.26**	.34**	.26**	-.16
2.Openness	.06	.28**	.05	.00	.02	-.14	-.05	-.05	.09	.04	-.02
3.Agreeableness	.13	.14	.39**	.08	-.01	-.02	-.01	.15	.07	.02	.00
4.Conscientiousness	.07	.08	.02	.23**	-.10	.18*	.19*	.06	.19*	.20*	-.02
5.Neuroticism	-.01	.02	-.02	-.08	.15	-.06	-.18*	-.11	-.04	.07	.00
6.Action Orientation	.11	-.05	.01	-.03	-.08	.25**	.15	.08	.10	.09	-.13
7.ConflictM	-.04	-.02	-.11	.03	-.11	.12	.24**	.09	.08	.18*	-.04
8.Customer Focus	-.09	.01	.02	.02	-.05	.02	.09	.15	.02	.09	-.04
9.Interpersonal Savvy	.08	.04	.09	-.06	-.07	.07	-.02	.01	.20*	.05	.04
1.Positive Affect	.03	-.01	.00	.04	-.01	.15	.11	.06	.13	.15	-.01
11.Negative Affect	-.01	.01	-.08	-.06	.02	-.07	-.02	-.03	.01	.03	.25**

Note. N = 189, Ex=Extraversion, Open=Openness, Agree=Agreeableness, Con=Conscientiousness, Neu=Neuroticism, Act=Action Orientation, ConM=Conflict Management, CusF=Customer Focus, IntSav=Interpersonal Savvy, PA=Positive Affectivity, NA=Negative Affectivity, * = p<.05 ** = p<.01

While the current research will primarily utilize the BETSY software to identify key differentiating words, BETSY’s scoring capabilities will also be utilized. BETSY scoring will be used to generate Bayesian Personality Scores. The purpose of this analysis is to demonstrate that utilizing both software programs in combination and creating Custom Personality Scores provides additional predictive power over utilizing Bayesian scores. Displayed in Table 6 are the results of this analysis. It is worth noting that the BESTY software recommends that 100 or more text samples be used in each training group. However, the current research utilized only ten text samples in each group due to sample size limitations. Despite this limited sample, multiple significant relationships between Bayesian Personality scores and Self-Report Personality Scores were observed. More specifically, four (Extraversion, Action Orientation, Customer Focus, and Conflict Management) of the eleven relationships between an applicant’s self-report personality and Bayesian personality scores were significant at the .01 level. These correlations ranged from .22 to .31. While relationships of this magnitude are certainly noteworthy, they are certainly not strong enough to indicate that that the Bayesian personality scores can be used to simply replace the self-report personality scores. Furthermore, there were numerous relationships between Self-Report Scores and Bayesian scores that were not significant. However, these results remain encouraging, particularly when one considers that BETSY recommends that 100 or more text samples are utilized for each group. Furthermore, it may be the case that personality traits such as extraversion and agreeableness are easier to assess based only on words spoken during a job interview.

Table 6. *Bayesian Personality and Self-Report Personality Correlations*

	Self-Report Personality										
Bayesian	Ex	Con	Open	Agree	Neu	Act	CusF	ConM	IntSav	PA	NA
1.Extra	.25**	-.08	.06	.06	-.13	.23**	.11	.11	.16	.19*	-.06
2.Cons	-.13	.02	-.16	-.07	.02	-.05	.01	-.08	-.12	-.09	.10
3.Open	-.11	-.08	.03	-.03	.23**	-.12	.02	-.14	-.19*	-.17	.06
4.Agree	.09	.02	.21*	.18*	-.08	.11	.17*	.08	.04	-.02	-.07
5.Neu	.09	.04	.12	.20*	-.04	.03	.07	.04	.01	-.05	.10
6.Action	.10	.07	.16*	.09	-.16	.22**	.18*	.23**	.05	.13	-.11
7.Cust Focus	.21*	.10	.24**	.15	-.17*	.27**	.30**	.26**	.26**	.22*	-.11
8.Con Man	.18*	-.01	-.05	.01	-.14	.28**	.01	.31**	.12	.20*	-.11
9.Interp Sv	.16	.07	.06	.11	-.09	.23**	.25**	.24**	.13	.19*	-.09
10.P.A.	.22*	.04	.19*	.12	-.07	.22*	.22*	.14	.16	.21*	-.13
11.N.A.	.07	.02	.05	.12	.00	.09	.12	.09	.16	.16	.06

Note. *N*=159 * Extra=Extraversion, Open=Openness, Agree=Agreeableness, Cons=Conscientiousness, Neu=Neuroticism, Act=Action Orientation, Con Man=Conflict Management, Cust Focus=Customer Focus, Intper Sv=Interpersonal Savvy, P.A.=Positive Affectivity, N.A.=Negative Affectivity, *p*<.05, ** = *p*<.01

Hypothesis 3 proposed that, despite the limited number of words utilized, the Custom Personality Scores would add incremental variance over the Bayesian Personality scores in the prediction of the Self-Report Personality Scores. Table 7 displays the results of the regression analysis utilized to examine this hypothesis. Overall, of the 11 personality traits examined here, eight of the Custom Personality Scores added unique variance in the prediction of the Self-Report Scores. Therefore, there is sufficient evidence here to support hypothesis 3. It is particularity noteworthy that Custom Scores provided incremental variance even for even those personality traits that displayed strong relationships between Bayesian scores and Self-Report Scores.

Table 7. *Incremental Variance Provided by Custom Personality*

Personality	Beta	B	Total R^2	$R^2\Delta$	Sig
Extraversion					
Bayesian	.25	1.21	.06	--	.00
Custom	.30	1.87	.15	.09	.00
Conscientiousness					
Bayesian	.02	.10	.00	--	.91
Custom	.23	4.65	.05	.05	.01
Openness					
Bayesian	.07	.41	.00	--	.66
Custom	.29	3.8	.09	.09	.00
Agreeableness					
Bayesian	.18	1.02	.03	--	.03
Custom	.39	3.69	.18	.15	.00
Neuroticism					
Bayesian	-.03	-.23	.00	--	.66
Custom	.15	.53	.02	.02	.08
Action Orientation					
Bayesian	.19	1.12	.05	--	.01
Custom	.22	7.32	.11	.06	.01
Customer Focus					
Bayesian	.29	1.37	.09	--	.00
Custom	.13	1.61	.11	.02	.14
Conflict Management					
Bayesian	.27	1.61	.09	--	.00
Custom	.20	2.41	.13	.04	.01
Interpersonal Savvy					
Bayesian	.26	1.78	.10	--	.00
Custom	.22	2.01	.12	.04	.01
Positive Affectivity					
Bayesian	.20	.74	.04	--	.02
Custom	.14	2.45	.06	.02	.11
Negative Affectivity					
Bayesian	.03	.12	.00	--	.51
Custom	.25	4.88	.07	.07	.01

Note. *N*=189.

CHAPTER VI

STUDY 1 DISCUSSION

The purpose of Study 1 was to examine the effectiveness of automated scoring to reproduce Self-Report Personality Scores. The automated scores were customized through utilizing two separate automated scoring programs, each of which has shown encouraging results when utilized independently. However, due to the uniqueness of the job interview context, it is expected that utilizing both programs to generate job interview specific scores will be useful. Overall results indicate that, the relationships among the automated scores and self-report personality scores may not support the contention that automated scores are completely reproducing the self-report scores. However, there are still multiple reasons to be encouraged by the results observed here.

First, the relationship between Bayesian and Self-Report Scores was somewhat surprising and encouraging. Despite the fact that only 10 interview transcripts were placed into each group, there were multiple meaningful relationships observed between Self-Report and Bayesian scores. These relationships will likely only be magnified if a larger sample is utilized to train the BETSY software.

A second and perhaps more important reason to be encouraged by these results is this research clearly highlights the importance of context. That is, this research lends support to the contention that words that are utilized in the job interview are different than words typically utilized in every day conversation or free writing samples. These relatively low stress environments were utilized to develop the LIWC software. Therefore, an automated scoring technique that is unique to the job interview context needs to be developed if one wishes to generate automated scored based on interview content. Given the complete lack of meaningful relationships between Expert Dictionary and Self-Report Personality Scores, along with the incremental variance explained by the Custom Personality Scores in the prediction of Self-Report Scores; it is clear that a customized scoring method is needed in this context.

Furthermore, these results are also encouraging given the simplicity of the methodology utilized as well as the novelty of the research in general. More specifically, only word use is examined to generate automated scores. Additionally, this is the first attempt to apply automated scoring techniques to the job interview context.

Limitations and Future Directions

Any research that utilizes a self-report methodology is obviously subject to the limitations inherent to this methodology. Additionally, participants in this study may have attempted to complete the personality survey in a way that would make them appear to be a desirable candidate. This is more likely given that the email sent to the applicants came from the organization making the hiring decisions. However, applicants were informed that their responses would have no impact on hiring decisions. This likely at the very least mitigated the chances that an applicant would distort their responses.

A second limitation of this study is the relatively simplistic methodology that is utilized. That is, only word usage is examined and therefore the manner in which the words are used is not considered. For example, if an applicant uses words such as ‘shy’ or ‘afraid’, typically one would assume that the use of these would indicate that this applicant is introverted. However, if an applicant indicated “I am never shy around customers, I am never afraid to speak my mind”, this is obviously not indicative of an introverted individual. Overall it is still likely that words such as these will be in general, related to introversion. However, the fact that this relatively simple word count methodology does not take into account the manner in which a word remains a weakness of this Study. Future research may consider the possibility of examining the manner in which words are utilized. As software programs become more efficient, it may be possible to only include words in automated scoring when utilized in a certain manner or context.

As mentioned above, despite a relatively small sample size this study still yielded encouraging results. Future research may benefit from utilizing a larger sample in order to identify differentiating words. This may allow researchers to examine the validity of this methodology across different samples. More specifically, a research methodology might consist of utilizing 200 or more text samples simply to train the BETSY software and identify key words. The next step of this research would involve applying these key words to an independent sample. Additionally, it may be interesting to allow BETSY to score this sample independently. Based on the relative success of BETSY's independent scoring capabilities observed in this study, this may yield encouraging results.

Finally, examining multiple personality measurement techniques may also yield further interesting results. Previous research has demonstrated that personality is one of the most frequently assessed constructed in the employee interview. Therefore, collecting personality ratings from interviewers in which they rate the personality of the interviewee may yield exciting insights. This would allow for the examination of words used by applicants that cause them to be evaluated as high or low on a particular personality trait. Examining the degree of overlap between these words and words that differentiate based on self-report would be very interesting.

CHAPTER VII

STUDY 2 INTRODUCTION

While Study 2 will use a similar methodology to Study 1, Study 2 will attempt to predict actual hiring decisions while Study 1 attempted to predict applicant's personality traits. As previously discussed, by using both the LIWC and BETSY software programs to create custom scores, it allows users the ability to predict any variable for which they can derive a meaningful measurement of. Obviously, an applicant's performance during the interview is one of the most meaningful outcomes that can be derived from the employee interview. There are at least four potential benefits to using an automated scoring process to predict interview performance ratings.

Benefits of Automated Process in Predicting Performance

First, an automated scoring process provides an entirely objective evaluation to a process that is traditionally a relatively subjective process. Second, an automated scoring process is able to capitalize on *all* of the interview response content. Third, it increases the utility and efficiency of the interview by extracting other meaningful information from an interview that would otherwise have to be evaluated separately, as well as obviously reducing the time requirements that are inherent to the job interview process. Finally, by using interview responses, as opposed to self-report responses, the issue of applicant faking is likely to be greatly reduced. It is important to keep in mind that the current state of the literature does not allow for each of these potential benefits to be realized immediately. A great deal of additional research needs to be conducted in order to fully realize each of these possible benefits.

Objectivity

Although human ratings of interview responses have proven to be sufficient to yield high validity for applicant interviews (Posthuma, Morgeson, & Campion, 2002), it is likely that increasing the objectivity of these ratings may yield further increases in validity. Unlike objectively scored selection methods such as cognitive ability tests, interview ratings are dependent upon human interview ratings. Therefore, characteristics that are not necessarily related to performance in the job may be considered (Huffcut, 2011). Research has highlighted numerous factors which may result in inaccurate or biased evaluations of individuals (Gatewood & Field, 1999; Schmitt & Chan, 2001; Huffcut, 2011). These factors include but are not limited to halo error, severity bias, similarity to the applicant, applicant attractiveness, and leniency bias. Additionally, Schmitt (1976) reported some evidence for the presence of primacy and recency effects in the interview evaluation process. Overall, the process of assigning a specific performance score to an interview is a very complex process and multiple factors are likely to play a role. Two of the diagrams displayed here highlight many of the factors that may play a role. Obviously, an automated process would eliminate the biases involved with human judgment and therefore would be very useful in this context.

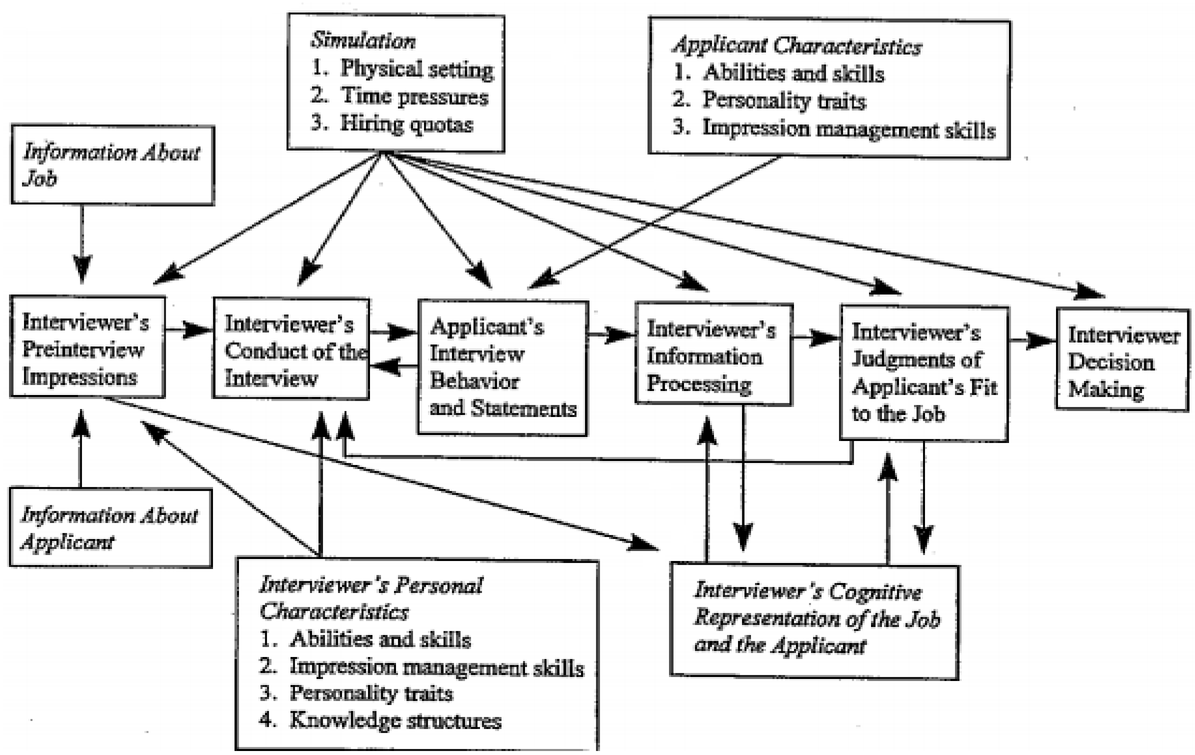

Figure 1. *Representation of Various Influences on Interview Decisions (Schmitt & Chen, 1998)*

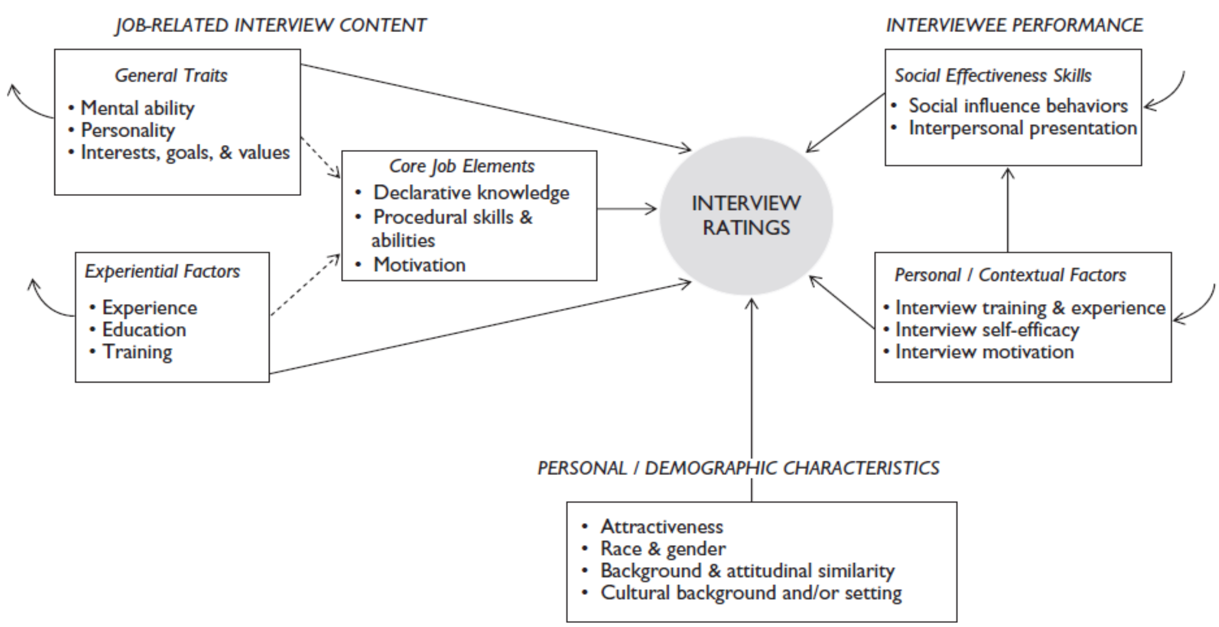

Figure 2. *Model of the major sources of constructs in employment interview ratings.* Curved arrows indicate that general traits and experimental factors can influence both core job elements and interviewee performance (Huffcut, 2011)

As demonstrated in the figures above, personal characteristics can have a strong impact on interview ratings (Burnett & Motowidlo, 1998). More specifically, perceived similarity to the applicant as well as interpersonal attractiveness have each been shown to be positively related to subjective qualifications (Graves & Powell, 1995). Perceived similarity to the applicant in general has also been shown to lead directly to increases in interview performance ratings (Schmitt, Pulakos, Nason, Whitney, 1996). It is important to consider that similarity to an applicant may be based on a variety of factors. For example, similarity to an applicant may be based on gender, attitudes or background similarity (Cable & Judge, 1997; Posthuma et al., 2002; Raza & Carpenter, 1987). For example, several studies have revealed that minority interviewers had a tendency to give higher ratings to job applicants of the same race (Lin, Dobbins & Farth, 1992; McFarland et al., 2004).

When similarity to the interviewer is not considered, demographic characteristics such as race, gender, and age still have been shown to influence interview ratings. Huffcut and Roth (1998) found an overall effect size of .25 when examining 31 studies that examined White-Black race differences in interview ratings. Additionally, this same study revealed an effect size of .26 across 15 studies that examined White-Hispanic differences.

In terms of gender differences, Huffcut (2001) found that males typically receive much higher interview ratings with respect to job related physical skills and work history/experience. Females on the other hand received much higher ratings in terms of general physical attributes and moral standards. Additionally, research has revealed gender differences that are present when evaluating non-verbal behaviors. That is, interview raters do not appear to give the same consideration to non-verbal cues displayed by women as they do men (Burnett, 1993). Although the findings have been somewhat mixed, age also appears to play a role in interview performance ratings (Avolio & Barrett, 1987). Overall, when examining demographic characteristics, it appears that many of the stereotypes that pertain to particular minorities or genders may be present during the interview process and therefore may bias the interview ratings.

In addition to personal characteristics, several other factors have been shown to influence interview ratings. For example, conformation bias has been found to influence interview ratings (Harris, 1989). This confirmation bias is particular likely to play a role when recruitment of a particular job applicant is present. Based on this recruitment, the job applicant will be perceived to be a very qualified job applicant. Therefore, this applicant may receive an unjustified increase in interview ratings, thus confirming the interviewer's original perceptions of the applicant.

Displayed above in Figure 2 is a model proposed by Huffcut (2011) that considers many sources which may contribute to interview ratings. Two of the proposed general sources, job-related interview content and interviewee performance, are meaningful to evaluate in the interview and thus should be evaluated. In addition to these sources, a third general source that contributes to interview ratings, personal factors, consists of constructs such attractiveness, similarity to interviewer and cultural background. These elements ideally should not be considered in the interview ratings as they are not related directly to dimensions of the job. Unfortunately, Huffcut (2011) found that across 187 studies, personal and demographic characteristics displayed a *stronger* relationship with interview ratings than interviewee performance sources.

While recent reviews seemed to reveal differences in interview ratings based on certain demographic effects, recent research has revealed somewhat contradictory results. A study conducted by McCarthy, Iddekenge, and Campion (2010) found that the effects of interviewee similarity to the interviewer on demographic characteristics did *not* have any impact on interview ratings. This study revealed a very impressive sample size (n= 19,931) and the interviews examined were each highly structured interviews. The researchers took great care to make sure that each of the interviews incorporated 15 components of a structured interview that are outlined by Campion et al (1997).

These 15 components include (1) interviews were based on a comprehensive job analysis (2) within the situational and past-behavioral interviews, the same questions were asked of each candidate, and within the experience-based interview, similar questions were asked of each candidate; (3) the use of prompts and follow-up questions was limited; (4) three different questioning techniques were employed (i.e., experienced-based, situational, past-behavioral); (5) each interview allowed sufficient time for interviewers to ask several questions; (6) ancillary information was controlled; (7) candidates were encouraged to ask questions after the structured phase of the interview process was complete 8) interviewers evaluated each dimension using behaviorally anchored rating scales 9) descriptive scale anchors were derived from KSAO definitions, previously developed interviews and responses from previous candidates 10) interviews were trained on the importance of note taking during the interview process 11) a panel of two interviewers evaluated each candidate 12) the same set of interviewers conducted the interviews for each applicant 13) the interviewers did not discuss candidates between interviews 14) all interviewers were extensively trained to ensure proficiency in conducting and scoring the interview and statistical procedures (unit weighting) were used to combine ratings within each interview.

It is somewhat clear, that if great care is taken in making sure that each of these 15 steps is carried out to the fullest extent, there will be minimal differences based on demographic characteristics. However, most structured interviews conducted within organizations will not involve each of these 15 steps being carefully carried out to the fullest extent. Additionally, several previous studies, as previously discussed, have in fact revealed that demographic characteristics do play a role in interview ratings. Overall, based on the results of these previous studies, it is clear using an automated scoring process that eliminates the biasing effects of factors such as applicant similarity, gender, age, and race will certainly be beneficial to the interviewing process.

Capitalize on All of Interview Content

The employee interview is typically the first and only opportunity that an employer has to meet face to face with a job applicant during the selection process. For this reason, several inferences are made with respect to a specific candidate based on a single interview (Podsakoff, Whiting, Podsakoff, Mishra, 2011). Given that in most hiring processes only one interview is conducted, it is of vital importance to obtain as much valid information as possible from this interview. The amount of time needed to conduct an interview also increases the importance of gathering as much information as possible from the interview. Unfortunately, due to cognitive limitations of human interviewers, it is likely that a great deal of information contained in the interview is not utilized. Even if an interviewer that is responsible for providing interview performance ratings is entirely focused for the duration of the interview, it is impossible for this person to evaluate every single word and phrase uttered during the interview (Chung & Pennebaker, 2007). To evaluate each specific aspect of the entire interview would prove to be a task that is too cognitively complex for any individual assessor to complete successfully (Fast & Funder, 2008).

In addition to the interviewer not being capable of attending to each spoken work during the interview, Fast and Funder (2008) contend that there may be psychological information contained in the interview which is not obvious to the naked eye. That is, even if the interviewer was capable of attending to each portion of the interview responses, they may still not be able to recognize why each piece of information is relevant. Furthermore, Pennebaker et al. (2003) suggested that during a social interaction people are generally focused on comprehending the meaning of what the speaker is saying rather than simply identifying the specific words and sentences that are used. If an interviewer focuses on the flow of each sentence, it is virtually impossible to attend to each time a specific word or subject is used. Therefore, an automated scoring process provides advantages in that each word spoken during the entire interview will be evaluated. Therefore, all of the information gathered from the interview will be utilized to the fullest extent possible.

A relevant and analogous example of interviewer rater's inaccuracies and cognitive limitations is found in the assessment center literature. Research conducted on assessment centers has shown that when raters are required to rate fewer dimensions the ratings tended to be more accurate (Woehr & Arthur, 2003). This effect can also be applied to interview ratings. That is, when an assessor is required to rate numerous dimensions of an interview they will likely experience a cognitive overload and therefore will likely to be subject to an increased probability of committing halo error and/or inaccurate ratings. Therefore, the use of automated software can be very beneficial as it reduces the cognitive load placed on the interviewer or other raters. The amount of constructs or dimensions that the interviewer is required to examine can be reduced as the automated scoring system will assist with the evaluation of additional constructs that would have previously been assessed by the interviewer.

Increased Efficiency

The process of evaluating job candidates can prove to be a very expensive and lengthy process (Gatewood & Field, 1999). Therefore, it is advantageous to obtain as much relevant information as possible from a job interview. By using all of the information which results from the interview, it may be possible to eliminate other elements of the selection process that are designed to assess other relevant constructs that the interview may not be intended to examine. Through the use of automated scoring of interview responses, one may be able to objectively evaluate an employee's personality, cognitive ability, leadership style, coping style, and perhaps most importantly interview performance. This would potentially entirely replace the process of administering a personality measure, cognitive ability test, and the interview ratings that must be conducted by the interviewer to score the interviewee's performance.

Although less likely, the automated scoring of interview responses could also be used as a screening tool. If an individual uses certain words or phrases they may be eliminated from the hiring process. This application is admittedly less likely as interview assessors are very unlikely to not notice an applicant using a phrase which could potentially screen them out. It also is unlikely that an applicant would willingly reveal such information about themselves during a job interview (Lievashina & Campion, 2007).

For ideal efficiency, two other processes may be applied. First, during the selection process there are traditionally numerous amounts of resumes and job applicants which must be considered. Job resumes and application blanks are the only selection tool used more frequently than the job interview (Dipboye, 1997; Dipboye & Jackson, 1999). Those involved in the selection process may benefit greatly from using text-analysis software in order to quickly organize the sometimes overwhelming amounts of text (Ford, 2005).

Second, if voice recognition software is used to transcribe interview responses into text, the efficiency of automated scoring would be dramatically improved. For example, rather than having multiple raters watch or listen to a full interview and then score the interview, the automated scoring would almost instantaneously score the interview responses. This is something that should certainly be considered for future research.

Reducing Faking

Although research has shown that personality is a valid predictor of important job outcomes such as job satisfaction (Judge, Heller, Mount, 2002) and job performance (Barrick & Mount, 1991), personality reports are typically based on a self-report personality measure. Due to this strong reliance on self-report to examine personality; one of the drawbacks of applying personality tests to the selection domain is applicant faking. In fact, concern with applicant faking remains the most frequently referenced criticism of personality assessment in the selection domain (Ellingson, Sackett, & Hough, 1999).

When engaging in the selection process, applicants will generally want to appear as a worthy job candidate. Therefore, applicants will generally want to appear conscientious, hard working, and intelligent (Zerbe & Paulhus, 1987). It is important to note that faking during an interview differs slightly from faking which may take place during traditional self-report personality measures. While completing a selection measure, the job candidate will simply try to emphasize desirable characteristics, while in a job interview the job candidate is more likely to present themselves as having the knowledge and abilities which are relevant to the job (Levashina & Campion, 2007).

In order to fake a self-report personality test, an applicant must simply provide a paper and pencil response that is inaccurate. Therefore, faking a self-report personality test is not a demanding cognitive process. However, it is likely much more difficult to fake a job interview response. This is particularity true for interview questions that require applicants to recall previous experiences (Levashina & Campion, 2007). Therefore, if interviews are used to assess personality constructs, the faking concern is at least mitigated.

Interviews questions are typically based on explanation of past behaviors as well as behavioral intentions based on hypothetical situations (Campion, Palmer, & Campion, 1997). In order to fake a response to a past behavioral question, an applicant would be required to mentally create a situation which did not actually occur while also creating an artificial personal reaction to this situation. This is without a doubt a more cognitively complex task than simply providing an inaccurate response to a paper and pencil measure. In fact, research has shown that job interview questions which require a response to a past behavior question are more resistant to faking than situational questions (Levashina & Campion, 2007). This is at least in part due to the increased cognitive complexity involved in generating an artificial previous experience.

Despite the increased cognitive complexity and general resistance to faking, using automated scoring of interviews to infer personality is unlikely to entirely eliminate the effect of faking. While being interviewed, the same motives that drive an applicant to fake a personality test are still present. The applicant will want to appear as intelligent and conscientious as possible. This desire may lead the applicant to distort or even completely manufacture certain situations and/or behaviors (Levin & Zickar, 2002). However, at the very least, using automated scoring to infer personality from interview responses will make it more difficult for an applicant to fake personality. Overall, while using automated scoring to evaluate interview responses and infer personality may not completely eliminate the threat of applicants faking personality, it certainly at the very least should reduce the chances of applicant faking. Given the overall importance and practicality of identifying those individuals who are faking or are more likely to fake and the fact that it may still be possible to fake a job interview, it is important to at least attempt to determine who is faking an interview response. This is particularly important when considering that even small differences in scores may have substantial impact on selection decisions (Christiansen, Goffin, Johnston, & Rothstein, 1994).

Although much more job interview relevant research is needed, research has shown that it may be possible to actually detect individuals who are faking by using automated scoring of interview responses. A study conducted by Newman, Penneback, Berry, and Richards (2003) was able to identify liars at a rate of 67%. In this study, liars showed lower cognitive complexity, used fewer self-references, and used more negative emotion words. Due to the fact that people who are lying or faking an interview response must develop a fictitious story rather than drawing on past experiences, their responses may differ from someone who is discussing an experience which actually occurred. Further research has shown that motion words and fewer third person pronouns were significant predictors of lying or deception by participants when they were instructed to tell a lie (Bond & Lee, 2005). This result also appears to be generalizable to the virtual world as a similar pattern has emerged when examining the content of instant messages (Hancock, Curry, Goorha, & Woodworth, 2008).

Overall, although the assessment of interview performance ratings has yielded impressive validity with respect to job performance, there are certainly disadvantages. In addition to these disadvantages involved with assessing interview performance, there are also limitations when interviewers or other raters attempt to assess personality based on interview responses. Similar to human ratings of interview performance, research has shown that rater's are relatively effective in making judgments about an applicant's based on interview responses. However, there are multiple ways in which the use of an automated scoring process could enhance the prediction of personality traits based on job interview responses.

Applying Automated Scoring to Interview Process

One important note needs to be made with respect to the application of the automated scoring process to job interviews. While there are certainly numerous potential benefits to applying automated scoring to interview responses, it is not recommended that the results obtained through automated scoring be used independently when evaluating interviewees. That is, the subjective/human component should still be included in the evaluation. While computer based scoring techniques are more efficient and objective, human coders are better able to evaluate the subjective component of the interview process (Fast & Funder, 2008). The subjective component to evaluating interviews is likely an important process as it is likely that both verbal and nonverbal information may reveal important information which is relevant to the job (Posthuma, Moregson, Campion, 2002). Furthermore, research has shown that as many as seven nonverbal cues may contribute to interview performance ratings and job performance ratings (Burnett, 1993). These nonverbal cues include smiling, hand movements, back/side lean, body orientation, physical attractiveness, dress, and vocal attractiveness. It is also worth noting that nonverbal cues can also assist interviewers judgments of personality traits, therefore cues from interviews should not be considered a bias as they provide information that must be attended to (DeGroot & Gooty, 2009). Therefore, particularity for occupations which involve a customer service component, applicants who are able to display effective non-verbal behaviors during conversations are likely to be better performers than those who do not display appropriate nonverbal behaviors.

Up to this point in time, interviews have shown to be a very valid predictor of job performance and the evaluation process of these interviews has been primarily a subjective process. Therefore, it is clear that this subjective evaluation is a meaningful process that should not be ignored. Research has shown that the first impression that a job candidate makes during an interview is related to higher interview ratings as well as more internship offers (Barrick, Swider, & Stewart, 2010). This first impression that is formed should not be viewed entirely as error or bias, but simply part of the overall evaluation. This first impression represents something that would not be evaluated when using automated scoring. With respect to customer service jobs in particular, this first impression is likely something that should be given a certain amount of weight in the process of scoring the interview.

Based on the evidence presented above, it is advocated here that the use of automated scoring be used as a complement to the human subjective scoring methods. Using this method, it may be possible to improve the validity of interviews. Implementing an entirely objective scoring process may find additional information which could add to the already valid information that is found by using the subjective scoring method. In addition to examining the ability of automated scoring to predict personality traits, this research will also examine the ability of the automated scoring to accurately predict interview ratings.

Hypothesis 4: Custom Expert Interview Ratings will be positively correlated with hiring decisions.

In addition to simply examining the relationship between expert interview performance ratings and automated performance scores, it is interesting to examine the words and words pairs that lead to higher interview ratings. As previously discussed, one of the primary critiques of selection interviews is that it is not clear exactly what constructs are being assessed. Therefore, by examining the words which are used by interviewee's that receive high ratings, it may be possible to gain insights into what constructs are being used by rater's when scoring the interview.

Research Question 1: What words and word pairs are used more frequently by those applicants who receive high interview performance ratings?

CHAPTER XIII

STUDY 2 METHOD

Participants

As in Study 1, participants for this portion of the study consisted of job applicants who have completed an online interview for a pharmaceutical company. However, participants in this portion of the study will consist of those job applicants that did not complete the online personality questionnaire. The pharmaceutical company provided hiring process decisions on 191 applicants that did not complete the online interview, these 191 applicants therefore will serve as participants in this study. Furthermore, 62 additional randomly selected applicants will be rated by six research assistants. Hiring decisions for these additional 62 applicants was not obtained, only hiring decisions were provided for the 191 applicants.

Procedure

Hiring Decisions

The online interview represents the first step of the hiring process used by the pharmaceutical company. The second step is a phone interview; if an applicant performs well in the online interview they are given the opportunity to participate in the phone interview. Thirty-eight percent of the applicants in this study who complete the online interview are then given the opportunity to advance to the third step in the interview process and participate in a phone interview. Of the applicants who complete the phone interview, approximately 20% are offered an opportunity to advance to the fourth stage of the interview and are given the chance to participate in a face to face interview. Finally, the fifth and final step consists of a field ride with a current sales representative and a job offer.

To convert the hiring decisions into a numeric index, applicants were given 1 point for each stage they advanced to. Therefore, those applicants who only completed the online interview were given 1 point. Additionally, those applicants who were given a job offer received 5 points.

Custom Expert Ratings

In addition to the 191 applicants that served as participants in Study 2, 62 applicants were randomly selected and rated by expert interview raters. Two interview raters were assigned to rate each applicant. The interview raters provided ratings for four dimensions as well as an overall rating. Similar to the process utilized in Study 1, the 62 applicants rated were rank ordered on each dimension and then were then divided into three groups for each dimension. One group consisted of the top 21 applicants in terms of interview ratings, another group consisted of the next 20 applicants, and the final group consisted of the 21 applicants receiving the lowest scores. BETSY then identified the words that differentiated among these groups. Words that differentiated and also had a strong theoretical basis were included in custom dictionaries to create a custom expert interview score for each interview rating dimension. Because expert interview ratings were only obtained for 62 applicants, Bayesian expert rating scores will not be created here.

Measures

Expert Interview Ratings

Experts viewed the job applicant’s interview and provided ratings for each applicant. Ratings were provided for Achievement Orientation, Customer Focus, Conflict Management, and Interpersonal Effectiveness. Raters were required to indicate the extent to which each applicant displayed particular performance dimensions based on a 1 to 7 scale. In addition to these ratings, raters provided a general rating that is designed to assess the overall quality of the applicant's responses. To obtain this general rating the raters were asked to indicate their likelihood of hiring each applicant based on a 1 to 5 scale in which 5 = certainly would hire, 1 = would not hire under any circumstance. The full rating scale can be found in appendix E.

Prior to providing applicant ratings, raters participated in a training session. During this training session, raters were given an overview of the interview process. Raters were also given the list of questions utilized during the interview as well as a rating form which contained a detailed description of each of the competencies they would be rating. Along with a description of the competency, each rating form contained specific interview questions that would likely yield pertinent information regarding the given competency.

CHAPTER IX

STUDY 2 RESULTS

Study 2 examined the relationships between automated interview ratings actual hiring decisions. In order to generate the automated scores for each of the interview ratings dimensions, BETSY was again utilized to identify words that differentiated between applicants. Table 8 contains the words that differentiated among those receiving high and low ratings on each dimension. Furthermore, these words were also utilized to create the custom dictionaries that were used to generate the automated scores for each dimension.

Table 8. *Differentiating Words among Interview Ratings*

Rating Dimension	Differentiating Words
Action Orientation	relationships*, gained*, speak*, no, scared*, phone, quit*, improving*, I, pharmaceutical, talking, person, love, Chilcott, kept, talked, candidate*, win
Problem Solving	struggle*, growing*, I, pharmaceutical, talking, person, experience, Chilcott, what, working, develop
Customer Focus	provide, help, trust, she, I, pharmaceutical, people, love, doctors, what, talked
Interpersonal Effectiveness	scared*, agree*, person, drugs, doctor, another, project, best, working, win
General	talking, people, person, Warner, Chilcott, compete, award*, proud*, project, working, responsible

Note. N= 62. * = Words utilized more frequently by those receiving low interview ratings

It is worth noting that an effort was made in order to minimize the overlap among words included in each rating dimension. The correlations among the interview ratings were substantial, as displayed in Table 9. Therefore, the words that differentiated among the interview ratings were highly similar across all dimensions. Given the lack of relationship among custom interview rating scores, as displayed in Table 10, it appears that this effort was successful.

Table 9. *Expert Ratings Inter-Correlations*

	Mean	S.D.	1	2	3	4	5
1. Action Orientation	5.29	.86	--				
2. Problem Solving	4.67	.94	.75**	--			
3. Customer Focus	4.89	.79	.57**	.70**	--		
4. Interpersonal Savvy	4.77	.98	.83**	.74**	.60**	--	
5. General	3.47	.79	.70**	.75**	.56**	.76**	--

Note. N= 62, * = p<.05, ** = p<.01

Displayed in Table 10 are the correlations among automated interview ratings as well as actual hiring decisions. Unfortunately, the automated ratings generated by using the words included in Table 8 did not predict actual hiring decisions. Hypothesis 4 was not supported as automated customized interview ratings of Action Orientation, Problem Solving, Customer Focus, Interpersonal Savvy, and general each displayed non meaningful relationships with actual hiring decisions.

Table 10. *Custom Expert Ratings and Hiring Decisions Correlations*

	1	2	3	4	5	6
1. Hiring Decisions	--					
2. Action Orientation	.09	--				
3. Problem Solving	-.06	-.06	--			
4. Customer Focus	-.04	.20**	.01	--		
5. Interpersonal Savvy	-.08	-.01	-.02	.12	--	
6. General	.01	-.12	-.04	-.03	-.05	--

Note. *N*= 191, * = p<.05, ** = p<.01

While Custom Interview Ratings did not display meaningful relationships with hiring decisions, this may be due in part to the fact that the hiring decision metric was highly skewed. That is, more than half of the applicants did not advance beyond the online interview process. Therefore a simple mean comparison of those applicants who advanced beyond the online interview and those applicants who did not was conducted. Table 11 displays the results of this analysis. Unfortunately, the results of this analysis were similar to those revealed in Table 10 and does not provide any additional support for hypothesis 4.

Table 11. *Custom Expert Ratings and Dichotomized Hiring Decision*

Custom Interview Rating Dimension	Online Interview Only		Phone Interview		Sig
	Mean	S.D.	Mean	S.D.	
Action Orientation	.79	1.23	.74	.98	.75
Problem Solving	.88	.34	.81	.30	.14
Customer Focus	.62	.37	.64	.38	.72
Interpersonal Savvy	.35	.28	.34	.28	.73
General	.05	.10	.07	.10	.49

Note. Applicants given online interview only (N = 72), phone interview (N=119).

In addition to the hypothesis proposed, the current research will also examine those words that differentiate among those applicants who advanced in the hiring processes and those who did not. Displayed in Table 12 below is a list of these words. Again, it is worth noting that BETSY identified numerous differentiating words that are not included in Table 12. These words are not included as they did not have a strong theoretical basis as to why they would differentiate among applicants. Results here seem to indicate that those individuals that discuss their interpersonal skills and personality traits are more likely to advance in the interview process than those applicants that discuss experiences or knowledge of the pharmaceutical arena.

Table 12. *Differentiating Words among Hiring Decisions*

Sell	Confidence	Drugs*
Customers	No	Patients*
Accomplishments	Thanks	Science*
Read	Agree	Nurses*
Client	Solved	Bad*
Motivated	Account	Schedule

Note. * = Words used more frequently by those applicants that did not advance in hiring process

Given the numerous hypothesis proposed in the current study, Table 13 summarizes each of the hypotheses examined here. It is clear from Table 13 that hypotheses examined in Study 1 were supported much more frequently than those in Study 2.

Table 13. *Hypothesis Summary*

Hypothesis Theme	Personality Trait	Supported?
1. Self-report personality traits will be related to similar expert dictionary dimensions	Neuroticism	No
	Conscientiousness	No
	Openness	No
	Extraversion	No
	Negative Affectivity	Partially
	Positive Affectivity	No
2. Custom personality scores will be positively correlated with self-report personality scores	Extraversion	Yes
	Agreeableness	Yes
	Neuroticism	No
	Openness	Yes
	Conscientiousness	Yes
	Conflict Management	Yes
	Interpersonal Savvy	No
	Action Orientation	Yes
	Customer Focus	No
	Positive Affectivity	Yes
	Negative Affectivity	No
3. Custom personality scores will predict self-report personality scores beyond what is predicted by utilizing Bayesian scores alone.	Extraversion	Yes
	Agreeableness	Yes
	Neuroticism	No
	Openness	Yes
	Conscientiousness	Yes
	Conflict Management	Yes
	Interpersonal Savvy	Yes
	Action Orientation	Yes
	Customer Focus	No
	Positive Affectivity	No
	Negative Affectivity	Yes
4. Custom expert interview ratings will be positively related to hiring decisions.	Action Orientation	No
	Problem Solving	No
	Customer Focus	No
	Interpersonal Savvy	No
	General	No

CHAPTER X

STUDY 2 DISCUSSION

While interviews are clearly a valid selection methodology, they are not without drawbacks. Two of the most obvious drawbacks to the interview process are the frequency of rater errors and the extensive time commitment that interviews require. Therefore, identifying key words that are indicative of increased interview performance has numerous practical applications. Study 2 attempted to identify key differentiating performance words based on expert ratings and demonstrate the predictive power of these words in terms of predicting actual hiring decisions.

While Study 1 results were encouraging in general, the results from Study 2 were not nearly as promising. However, despite the lack of support for the relationships proposed in Study 2, there are still interesting results to be observed. For example, by examining the words that were used more frequently by those applicants that advanced further in the hiring process, one can gain insight into the traits and experiences that evaluators valued. Based on an examination of these words, it appears that the evaluators at did not place a high value on experiences and knowledge of the pharmaceutical sales arena. For example, words such as drug, patients, and science were actually used more frequently by those candidates who were not given a second interview. However, words which suggest strong interpersonal skills were used more frequently by those candidates that advanced in the hiring process. That is, words such as confident, customers, agree, and thanks were all used more frequently by those candidates that were given a second interview. Furthermore, words that indicate a candidate is highly conscientious and organized were also used more frequently among those given a second interview; words such as motivated, solved, and scheduled.

Limitations and Future Directions

Unfortunately, it appears that halo error was quite prevalent in the expert interview ratings. The inter-correlations among each of the rating dimensions was substantial, at least .56 in all cases. Therefore, there was a great deal of overlap among the words that differentiated across each of the interview rating dimensions. As mentioned previously, this resulted in some differentiating words to not be included in some dictionaries in order to minimize overlap between the interview rating dimensions.

In addition to the pervasiveness of halo error within the interview ratings, the interview raters had little knowledge of the job itself. The raters were simply given the title of the role and job competencies. This combined with the interview rater's lack of experience in this job area in general likely contributed to the lack of meaningful relationships among automated interview ratings and actual hiring decisions.

Finally, the evaluation process that was utilized may have also played a role in the lack of support for the predicted relationships in Study 2. That is, the second step in the evaluation process, after the online interview, is a phone interview. This conversation likely plays a more significant role in the hiring process than the initial online interview. During the phone interview, the evaluator obviously has the capability to probe into areas of interest. This likely allows the evaluator to gain a more comprehensive view of the applicant. Therefore, the evaluator is unlikely to give strong weight to something an applicant said during the online interview after this phone interview has taken place. Therefore, because interview raters only had access to the online interview, this makes it less likely that interview ratings will relate closely to actually hiring decisions.

After the phone interview, the applicant moves on to an on-site visit. Needless to say, this face-to-face interaction is likely to permeate the evaluator's minds much more so than the initial online interview. Essentially, as the applicant moves along in the hiring process, the initial online interview becomes less and less important in their evaluation. It appears that the online interview here serves more of a screening function rather than a decision making tool. Overall, the actual hiring decisions are being made based on multiple interactions with each applicant. By contrast, the interview ratings are being made after only observing the online interview. Given this somewhat dramatic contrast, it is perhaps not surprising that automated interview ratings did not predict actual hiring decisions.

Unfortunately, this study was not able to examine the relationship between expert interview ratings and hiring decisions. Based on the results of this study, it appears likely that the relationship between expert ratings and hiring decisions may be rather weak. This analysis could not be conducted as hiring data was not obtained for those applicants rated by the experts. Furthermore, this research would have also benefitted from obtaining ratings that were used in the hiring process. For this study, each applicant was simply rated based on how far they advanced in the hiring process. Given that approximately half of the applicants did not make it beyond the online interview process, all of these applicants were given an identical score. This lack of variance virtually eliminates the possibility of conducting a meaningful correlational analysis. Furthermore, this also prevented the researchers from creating Bayesian performance scores or custom performance scores. More specifically, candidates could not be accurately divided into groups based on the hiring data provided.

A relatively recent technological advancement that may greatly benefit this research is voice recognition software. Due to the less than ideal sound quality of the online interviews utilized here, voice recognition software was unable to accurately recognize the spoken language. However, if further research is conducted, researchers and practitioners should ideally identify numerous words that differentiate effectively prior to creating automated scores. After these words are identified, the voice recognition software could be trained to recognize these specific words. Therefore, while the voice recognition software may not display impressive accuracy for all words that are spoken, the differentiating words that are useful in the scoring will be more likely to be recognized.

CHAPTER XI

GENERAL DISCUSSION

Although many of hypothesis of this study were not supported, there are at least two encouraging results that can be taken from this study. First, it is clearly beneficial to utilize the two software programs together and create custom scores to predict self-report personality. Custom automated scores generated utilizing both software programs displayed much more encouraging results than scores generated by utilizing only one program. More specifically, Bayesian scores generated by BETSY alone were not nearly as predictive of self-report personality as the custom scores. Additionally, despite the extensive development process, the LIWC expert dictionary dimensions did not display meaningful relationships with candidates' self-report personality scores. As discussed previously, this is likely due to the impact of the interview context. That is, due to the evaluative context, the words spoken during a job interview are likely to be different than those words written or spoken in a non-evaluative context. Therefore, when identifying a set of words that differentiate among candidates, the current study indicates that it is necessary to create a customized list of words that have displayed effective differentiating capabilities in the given context.

Overall, the relationships between the custom personality scores and the self-report personality scores, with the exception of Neuroticism, were relatively strong. However, relationships were generally not substantial enough to support the contention that the custom scores are effectively reproducing the self-report personality scores. Despite this, the results observed here remain encouraging for multiple reasons.

First, these results are particularity encouraging when considering the number of words included in each customized dictionary. By comparison, each LIWC default dictionary that proposes to measure a personality characteristic includes 200 or more words. The number of words included in each dictionary used in this study ranged from 8 to 16 words. Furthermore, these results are also encouraging given that only a small set of the total sample was uploaded into the BETSY software and utilized to identify differentiating words. With a larger sample uploaded into the BESTY software, it is expected that the words identified would be more reliable in differentiating power.

Finally, it is worth noting that most custom personality trait scores displayed meaningful relationships with the self-report personality scores while displaying much weaker correlations with other traits self-report personality scores. That is, these results display impressive discriminant and convergent validity. Overall, this certainly lends support to the accuracy of the custom personality scores. This is even more impressive when considering the relatively strong relationships among the self-report personality dimensions. Consider the fact that self-report Extraversion and self-report Agreeableness were closely related to each other (r = .48). However, the custom Openness score was strongly related to self-report Openness, while displaying virtually no relationship with self-report Extraversion. Finally, considering that this was one of the first attempts to apply automated scoring to the job interview context, expecting exceptionally encouraging results may be overly optimistic.

Overall, while a great deal of research must be conducted before this methodology can be utilized to drive any hiring decision, results of this study indicate that this methodology may have promising future applications. That is, this process may allow for an instant generation of an objective interview score after a job interview is completed. This will at the very least mitigate many of the biases involved in the job interview process. Furthermore, although it is not recommended that this methodology be utilized independently to drive selection decisions, the instant generation of an interview score would dramatically mitigate the extensive time demands associated with the job interview.

APPENDICES

APPENDIX A

ABRIDGED IPIP

Participant Instructions: Please indicate on a 1 to 5 scale how much each statement describes you. 1 = This statement does not describe me at all 5 = This statement completely describes me.

Item	Factor	Text
1	E	I am the life of the party.
2	A	I sympathize with others' feelings.
3	C	I get chores done right away.
4	N	I have frequent mood swings.
5	O	I have a vivid imagination.
6	E	I don't talk a lot. (R)
7	A	I am not interested in other people's problems. (R)
8	C	I often forget to put things back in their proper place. (R)
9	N	I am relaxed most of the time. (R)
10	O	I am not interested in abstract ideas. (R)
11	E	I talk to a lot of different people at parties.
12	A	I feel others' emotions
13	C	I like order.
14	N	I get upset easily.
15	O	I have difficulty understanding abstract ideas. (R)
16	E	I keep in the background. (R)
17	A	I am not really interested in others. (R)
18	C	I make a mess of things. (R)
19	N	I seldom feel blue. (R)
20	O	I do not have a good imagination. (R)
21	E	I am skilled in handling social situations
22	A	I make people feel at ease
23	C	I waste my time (R)
24	N	I panic easily
25	O	I enjoy hearing new ideas
26	E	I Feel comfortable around people
27	A	I believe that others have good intentions
28	C	I pay attention to details
29	N	I feel comfortable with myself (R)
30	O	I tend to vote for liberal political candidates
31	E	I know how to captivate people
32	A	I accept people as they are
33	C	I am always prepared
34	N	I rarely get irritated (R)
35	O	I carry the conversation to a higher level

Note. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness; (R) = Reverse Scored Item. Original 50-item IPIP-FFM available at http://ipip.ori.org/newQform50b5.htm.

APPENDIX B

THINKWISE PERSONALITY ITEMS

Please indicate the extent to which you agree or disagree with each statement based on a 1 to 5 scale. 1= Strongly disagree, 2= Disagree, 3= Neither agree nor disagree, 4= Agree, 5= Strongly Agree.

Item	Factor	Item Content
1	A	I never procrastinate on difficult tasks.
2	CM	I am comfortable making quick decisions.
3	CF	I can easily convince others to purchase products or services.
4	I	I am easy to get to know.
5	A	I often take it upon myself to start a project.
6	CM	I sometimes procrastinate when I need to make tough decisions.(Reverse)
7	CF	I possess a natural talent for influencing people.
8	I	I enjoy comforting those in need.
9	A	I sometimes need a push to get started on tasks.(Reverse)
10	CM	I sometimes second-guess decisions I've made.(Reverse)
11	CF	I always employ diplomacy and tact when delivering news to others.
12	I	Most people would describe me as a very caring person.
13	A	Others consider me to be a self-starter in my job.
14	CM	I'm known for making timely and effective decisions.
15	CF	I always say what I think.(Reverse)
16	I	I am comfortable meeting people for the first time.
17	A	I am sometimes unsure of myself. (Reverse)
18	CM	I rarely lose my composure at work.
19	CF	I often possess keen insight into others' feelings and perceptions.
20	I	I am usually shy around people I don't know.(Reverse)
21	A	I have taken frequent stands in the face of strong opposition.
22	CM	I remain focused and under control in times of conflict.
23	CF	I'm very concerned about how others will react to what I say.
24	I	I shy away from crowds of people.(Reverse)
25	A	I'm reluctant to speak my mind freely when there might be a negative reaction.(Reverse)
26	CM	When faced with a conflict, I always stay in control.
27	CF	I am always receptive to others' comments.
28	I	I am very good at sensing what others are feeling.
29	A	Sometimes I lack the confidence to make a tough decision.(Reverse)
30	CM	When faced with a crisis on the job, I always remain composed.
31	CF	I'm known for taking a personal interest in customers.
32	I	I seldom pay attention to others' reactions.(Reverse)

A = Achievement Orientation. CM = Conflict Management. CF = Customer Focus. I = Interpersonal Savvy

APPENDIX C

PANAS SHORT FORM

Participant Instructions: This scale consists of a number of words that describe different feelings and emotions. Read each word and indicate to what extent you generally feel this way, that is, how you feel on average. Use the following scale to describe how you usually are; 5 = Extremely, 4 = Quite a bit, 3 = Moderately, 2 = A little, 1 = Very slightly or not at all.

Number	Item	Scale
1	Alert	Positive Emotion
2	Excited	Positive Emotion
3	Distressed	Negative Emotion
4	Nervous	Negative Emotion
5	Enthusiastic	Positive Emotion
6	Determined	Positive Emotion
7	Fearful	Negative Emotion
8	Guilty	Negative Emotion
9	Active	Positive Emotion
10	Proud	Positive Emotion
11	Upset	Negative Emotion
12	Irritable	Negative Emotion

APPENDIX D

JOB INTERVIEW QUESTIONS

1. Why are you interested in a career in pharmaceutical sales?
2. What accomplishment or sale are you most proud of?
3. Give me an example of a time when your first sales strategy for a customer didn't work, and you had to develop a new approach. What did you do?
4. Give me an example of a time you lost business and describe how you got it back.
5. Some customers never seem satisfied. Give me an example of one of your more difficult customers and how you dealt with them.
6. Tell me about a time when a potential customer promised you a deal or an order and they didn't deliver, how did you handle the situation and what was the outcome
7. Have there been challenging times in your current role? How have you dealt with these and been able to be successful despite these challenges?
8. Outside of price, how would you sell against a similar product? (Mr./Ms. Candidate please feel free to simulate any part of the sales call here.)
9. Since you began your professional career in sales how have you evolved as a professional?
10. In these last few minutes give us your best "close".

APPENDIX E

INTERVIEW RATING FORM

Rater Instructions: Please rate the extent to which each applicant displayed each of these abilities on a 1 to 7 scale, 7 = to a great extent, 1 = not at all.

Action Orientation
Instructions: To rate this dimensions, it will be useful to consider questions 2, 7 and 9.

1. The applicant was comfortable during the interview and sure of themselves.
2. The applicant was very confident in their abilities.
3. When faced with a difficult situation, the applicant seems very willing to approach it.
4. The applicant was motivated to achieve.

Problem Solving
Instructions: To rate this dimensions, it will be useful to consider questions 3, 4, 6, and 7.

1. To what extent did the applicant display the ability to *analyze* information, *access* potential solutions, and *develop* an effective course of action?
2. To what extent did the applicant display the ability to *execute* a course of action which leads to an effective solution?
3. To what extent did the applicant display the ability to *generate* original and creative ideas for solving problems?
4. The applicant displayed the ability to effectively adapt to new situations.

Customer Focus
Instructions: To rate this dimensions, it will be useful to consider questions 3, 4, 5, 6, 8, 10.

1. The applicant displayed the ability to *influence* customers and other individuals.
2. The applicant places a high priority on customer satisfaction.
3. Even if a customer is upset, the applicant displayed the ability to confront the situation and deal with the situation appropriately.
4. The applicant truly enjoys working with customers and making the customer happy.

Interpersonal Effectiveness
Instructions: To rate these dimensions, it will be useful to consider questions 5, 8 and 10.

1. To what extent was the interviewee able to influence the views or actions of others through interpersonal interaction?
2. To what extent was the interviewee able to clearly express their thoughts and ideas during the interview?
3. To what extent was the interviewee enthusiastic during the interview?
4. The applicant displayed appropriate and effective non-verbal behavior during the interview.

General

Instructions: Relative to the other interviews you have watched; please provide an overall rating for the performance of the applicant during the interview.

5 – It was one of the best interviews I have watched. If I was in a position to hire applicants for the sales rep position, I would certainly hire this person.
4- I would probably hire this person but there were certainly some interviews which were better.
3 – The interview was average. Half of the participants performed better than this applicant and half of the applicants performed worse. I might hire this person if there were a lot of positions available.
2- Based on the interview, this was one of the least hirable applicants. I would not choose to hire this person.
1- This was one of the worst interviews I viewed. I would not hire this person under any circumstances.

APPENDIX F

DIMENSIONS IDENTIFIED BY DEFAULT LIWC DICTIONARIES

LIWC Dictionary Dimension	**Example words from Dimension**
All pronouns	I'd, I'm, he'd, her
Proper Pronoun	he'd, I've, myself, our
1st person singular	I, I'd, mine, myself
1st person plural (we)	lets, our, ours, us, we
Total 2nd person (you)	you, you'd, three, you're
SheHe	he, he'd, she, she'd
They	their, they, they'll
Improper Pronoun	anyone, anything, it, it's
Article	a, a lot, an, the
Verbs	accepted, arrive, hit, taken
Auxiliary verbs	be, did, didn't, do, does
Past	accepted, became, been
Present	admit, it's, look, cant
Future	could've, gonna, might
Adverbs	about, beyond, clearly, here
Prepositions	about, beside, by, for, except
Conjunctions	how, however, nor, or, plus
Negations (negate)	ain't, aren't, hasn't, isn't
Quant	all, couple, enough, equal
Numbers	billion, dozen, once, quarter
Swear	butt, dumb, heck, hell
Social	heard, share, reply, help
Family	grandma, husband, kin
Friends	girlfriend, mate, pal, roommate
Humans	citizen, girl, gentleman, ladies
Affect	affection, good, loves, annoy
Positive emotion	agrees, alright, good, profit, great
Negative emotion	anger, fired, bad, furious, sucks
Anxiety	distraught, fear, guilt, impatient
Anger	attack, critical, bitter, outraged
Sad	Depressed, dull, gloom, hopeless
Cognition	adjust, origin, define, overall
Insight	learn, lesson, became, rational
Cause	change, control, create, depending
Discrepancy	lack, impossible, needs, mistake
Tentativeness	appears, doubt, guess, depends
Certainty	commits, confident, correct, defined
Inhibition	blocking, contain, delay, deny
Inclusion	out, we, with, plus

Exclusion	vs, whether, without
Perception	caress, concert, cool, say
Seeing	eye, gaze, image, look,
Hearing	listen, loud, noise, ring, song
Feel/touch	fire, grab, grip, hand, hot
Biology	arm, elbow, face, slender
Body	belly, bone, pulse, chest
Health	nurse, pain, chills, clinic
Sexual	condom, foreplay, cuddle
Ingestion	brunch, candy, chew, coffee
Relative	driven, driving, prior, term, stops
Motion	attended, bring, carrying, catch
Space	bending, onto, open, corner
Time	already, new, old, busy, past
Work	hiring, homework, income, benefits
Achievement	outcome, beat, power, practice, pride
Leisure	poetry, play, blog, chillin, bath
Home	chore, dishwasher, dresser
Money	bet, poor, price, casino, cash
Religion	bible, pastor, god, faith
Death	Dead, grave, ghost, grieve
Assents	huh, yep, okay, yes

REFERENCES

Abe, J.A. (2009). Words that predict outstanding performance. *Journal of Research in Personality, 43,* 528-531.

Adair, G. (1984). "The Hawthorne effect: A reconsideration of the methodological artifact". *Journal of Applied Psychology, 69,* 334–345.

Attali, Y. (2004) *Exploring the feedback and revision features of Criterio*n. Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA.

Avolio, B.J. & Barrett, G.V. (1987) Effects of age stereotyping in a simulated interview. *Psychology and Aging, 2,* 56-63.

Barrick, M.R. & Mount, M.K. (1991). The Big Five Personality Dimensions and Job Performance: A meta-analysis. *Personnel Psychology, 44,* 1-26

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium:What do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9–30.

Barrick, M.R., Patton, G.K., Haugland, S.N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology, 53*, 925-951.

Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial Evaluations in the Interview: Relationships with Subsequent Interviewer Evaluations and Employment Offers. *Journal of Applied Psychology, 95, 1163-1172*.

Blackman, M.C. (2002) Personality judgment and the utility of the unstructured interview. *Basic and Applied Social Psychology, 24,* 241-250.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19,* 313-329.

Burnett, J.R., & Motowildo, S.J. (1998). Relations between different sources of information in the structured interview. *Personnel Psychology, 51,* 963-983.

Burnett, J.R. (1993). *Utilization and validity of nonverbal cues in the structured interview. Unpublished doctoral dissertation,* University of Florida, Gainesville.

Burstein, J. & Chodorow, M. (1999). Automated Essay Scoring for nonnative English speakers. Proceedings of the ACL99 Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing, College Park, MD.

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

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

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fielder (Ed.), *Frontiers in social psychology* (pp. 343–359). New York: Psychology Press.

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847–860

Connolly, J.J., Kavanagh, E.J. & Viswesvaran, C. (2007). The Convergent Validity between Self and Observer Ratings of Personality: A meta-analytic review. *International Journal of Selection and Assessment, 15*, 110-117.

DeGroot, T. & Gooty, J. (2009) Can nonverbal cues be used to make meaningful personality attributions in employment interevies? *Journal of business psychology, 24,* 179-192.

Dipboye, R. L. (1997). Structured interviews: Why do they work? Why are they underutilized? In N. Anderson & P. Herriott (Eds.), *International handbook of selection and assessment* (pp. 455–473). New York: Wiley.

Dipboye, R. L., & Jackson, S. L. (1999). Interviewer experience and expertise effects. In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 229–292). Thousand Oaks, CA: Sage.

Donahue, T. (2008). Cross-Cultural Analysis or Student Writing: Beyond Discourses of Difference. *Written Communication, 25,* 319-354.

Donnellan, B. M., Oswald, F.L., Baird, B.M., Lucas, R.E. (2006). The Mini-IPIP Scales: Tiny-Yet- Effective Measures of the Big Five Factors of Personality. *Psychological Assessment, 18*, 192-203.

Eder, R. W., & Buckley, M. R. (1988). The employment interview: An interactionist perspective. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 6, pp.75–108). Greenwich, CT: JAI Press.

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology, 94,* 334–346

Fletcher, C. (1987) Candidate personality as an influence on selection interview assessments. *Applied Psychology: An International Review, 36,* 157-162.

Ford, J.M. (2005). Review of Data and Text Mining: A Business Applicants Approach. *Personnel Psychology, 58*, 267-271.

Gatewood, R.D. & Field, H.S. (2001). Human Resource Selection (5th eds).

Goldberg, L.R. (1993) The structure of phenotypic personality traits. *American Psychologist, 48,* 26-34.

Gottschalk, L. A., & Bechtel, R. (1993). *Computerized content analysis of natural language or verbal texts*. Palo Alto, CA: Mind Garden.

Gottschalk, L. A., & Gleser, G. C. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley: University of California Press.

Gottschalk, L. A., Gleser, G. C., Daniels, R., & Block, S. (1958). The speech patterns of schizophrenic patients: a method of assessing relative degree of personal disorganization and social alienation. *Journal of Nervous and Mental Disease, 127,* 153-166.

Graves, L.M. & Powell, G.N. (1995). The effect of sex similarity on recruiters' evaluations of actual applicants: A test of the similarity-attraction paradigm. *Personnel Psychology, 48,* 85-98.

Groom, C.J. & Pennebaker, J.W. (2002). Words. *Journal of Research in Personality, 36,* 615-621.

Gutenberg, R.L., Avery, R.D., Osburn, H.G., & Jenneret, P.R. (1983). Moderating effects of decision-making/information processessing job dimensions on test validities. *Journal of Applied Psychology, 68,* 602-608.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45,* 1-23.

Harris, M.M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology, 42,* 691-726.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43,* 524–527.

Hogan, R. (1991) Personality and personality measurement. In: Dunnette, M.D. and Hough, L.M. (eds), *Handbook of industrial and organizational psychology* (2nd edn). Palo Alto, CA: Consulting Psychologists Press, pp. 873-919.

Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality, 42,* 747–754

Hough, L.M. & Ones, D.S. (2001). The structure, measurement, validity, and use of personality variables in industrial work and organizational psychology. In Anderson, N., Ones, D.S., Sinagil, H.K. and Viswesvaran, C. (eds), *Handbook of Industrial work and organizational psychology* (Vol. 1, Personnel Psychology). London: Sage, pp. 233-277.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79,* 184–190.

Huffcut, A.I., Conway, J.M., Roth, P.L., Stone, N.J. (2001). Identification and Meta-Analytic Assessment of Psychological Constructs Measured In Employee Interviews. *Journal of Applied Psychology, 86,* 897-913.

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

Huffcut, A.I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and assessment, 19*, 62-81.

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

Ingram, R.E., Lumry, A.B., Cruet, D., & Sieber, W. (1987) Attentional processes in depressive disorders. *Cognitive Therapy and Research, 11,* 351-361.

Janasik, N., Honkela, T., Bruun, H. (2009). Text Mining in Qualitative Research Application of an Unsupervised Learning Method. *Organizational Research Methods, 12,* 436-460.

Judge, T.A., Heller, D., & Mount M.K. (2002). Five-Factor Model of Personality and Job Satisfaction: A Meta-Analysis. *Journal of Applied Psychology, 87*,530-541.

Keenan, A. (1982). Candidate personality and performance in selection interview, *Personnel Review, 11*, 20-22.

Kingwood Group (2008). ThinkWise Selection and Development System Technical Report. Grand Rapids, Michigan.

Kozbelt, A. & Burger-Pianko, Z. (2007). Words, Music, and Other Measures: Predicting the Perertoire Popularity of 597 Schubert Lieder. *Psychology of Aesthetics, Creativity, and the Arts, 1*, 191-203.

Krajewski, H.T., Goffin, R.D., McCarthy, J.M., Rothstein, M.G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational and Organizational Psychology, 79,* 411-432.

Kufner, A.C., Back, M.D., Nestler, S., Egloff, B. (2010) Tell me a story and I will tell you who you are! Lens model analysis of personality and creative writing. *Journal of Research in Personality, 44,* 427-435.

Latham, G. P., Saari, L. M.,8 Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 569–573.

Latham, G. P., & Sue-Chan, C. (1996). A legally defensible interview for selecting the best. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 134–143). Westport, CT: Quorum Books/Greenwood Publishing

Lee, H.S. & Cohn, L.D. (2009) Assessing Coping Strategies by Analyzing Expressive Writing Samples. *Stress and Health, 26*, 250-260.

Levashina, J. & Campion, M.A. (2007). Measuring Faking in the Employment Interview: Development and Validation of an Interview Faking Behavior Scale. *Journal of Applied Pschology, 92,* 1638-1656.

Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies, and bullshit: Understanding faking and its effects on selection decisions using theory, field research, and simulation. In J. M. Brett & F. Drasgow (Eds.), *The psychology of work: Theoretically based empirical research* (pp. 253–276). Mahwah, NJ: Erlbaum

Lin, T.R. Dobbins, G.H., & Farth, J.L. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology, 77,* 363-371.

McCarthy, J.M., Iddekinge, C.H., Campion, M.A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology, 63,* 325-359.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599 – 616.

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

McFarland, L.A., Ryan, A.M., Sacco, J.M., & Kriska, S.D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of management, 30,* 435-452.

Mehl, M., Gosling, S., & Pennebaker, J. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90,* 862–877.

Mehl, M., & Pennebaker, J. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857–870

Morgeson, F.M., Campion, M.A., Depboye, R.L., Hollenbeck, J.R., Murphy, K., Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683-729.

Mueller-Hanson, R.A., Heggestad, E.D., Thornton, G.C. (2006) Individual differences in impression management: an exploration of the psychological processes underlying faking. *Psychological Science, 48,* 288-312.

Murphy, K.R. (2000). Impact of Assessment of Validity Generalization and Situational Specificity on the Science and Practice of Personnel Selection. *International Journal of Selection and Assessment, 8,* 194-206.

Newman, M.L., Pennebaker, J.W., Berry, D.S., Richards, J.M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin, 29,* 665-675.

Niederhoffer, K.G. & Pennebaker, J.W. (2002) Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology, 21*, 337-362.

Pennebaker, J.W., Francis, M.E., Booth, R.J. (2001). Linguistic Inquiry and Word Count (LIWC): LIWC, Mahwah, NJ, Erlbaum.

Pennebaker, J.W. & Graybeal, A. (2001). Patterns of natural language use: disclosure, personality, and social integration. *Current Directions in Psychological Science, 10,* 90-93.

Pennebaker, J.W., & King, L.A. (1999) Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296-1312.

Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547–577.

Podsakoff, N.P., Whiting, S.W., Podsakoff, P.M., Mishra, P. (2011) Effects of Organizational Citizenship Behaviors on Selection Decisions in Employment Interviews. *Journal of Applied Psychology, 96,* 310-326.

Posthuma, R.A., Morgeson, F.P., Campion, M.A. (2002) Beyond Employment Interview Validity: A Comprehensive Narrative Review of Recent Research and Trends Over Time. *Personnel Psychology, 55,* 1-55.

Pyszcynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of depression. *Psychological Bulletin, 102,* 122-138.

Raza, S.M. & Carpenter, B.N. (1987). A model of hiring decisions in real employment interviews. *Journal of Applied Psychology, 72,* 596-603.

Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of Personality, 78,* 575–598.

Rosenberg, S. D., & Tucker, G. J. (1978). Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry, 36,* 1331-1337.

Rude, S., Gortner, E.M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion, 18,* 1121-1133.

Schmitt, N., Pulakos, E.D., Nason, E., Whitney, D.J. (1996) Likability and similarity as potential sources of predictor-related criterion bias in validation research. *Organizational Behavior and Human Decision Processes, 68,* 272-286.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology, 29,* 79-101.

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.

Slatcher, R.B., Chung, C.K., Pennebaker, J.W., Stone, L.D. (2007) Winning words: Individual differences in linguistic style among U.S. presidential and vice presidential candidates. *Journal of Research in Personality, 41,* 63-75.

Smith, T.W., & Greenberg, J. (1981). Depression and self-focused attention. *Motivation and Emotion, 5,* 323-331.

Stirman, S.W., & Pennebaker, J.W. (2001). Word use in poetry of suicidal and non-suicidal poets. *Psychomatic Medicine, 63,* 517-522.

Tausczik, Y.R. & Pennebaker, J.W. (2010) The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29,* 524-54.

Tett, R.P. & Christiansen, N.D. (2007). Personality Tests at the Crossroads: A response to Moregeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967-993.

Townsend, R.J., Bacigalupi, S.C., Blackman, M.C. (2007). The accuracy of lay integrity assessments in simulated employment interviews, *Journal of Research in Personality, 41,* 540-557.

Trull, T.J., Widiger, T.A., Useda, D.J, Holcomb, J., Axelrod, S.R., Stern, B.L., Gershuny, B.S. (1998). A structured Interview for the Assessment of the Five-Factor Model of Personality. *Psychological Assessment, 10,* 229-240.

Watson, D., Clark, L.E., Tellegan, A. (1988) Development and Validating of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology, 6,* 1063-1070.

Weintraub, W. (1989). *Verbal behavior in everyday life*. New York: Springer.

Woehr, D.J., & Arthur, W. (2003). The Construct Related Validity of Assessment Center Ratings: A Review and Meta-Analysis of the Role of Methodological Factors. *Journal of Management, 29*, 231-258.

Yarkoni, T. (2010). Personality in 100,000 words: A large scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*, 363-373.

Zerbe, W.J., & Paulhus, D.L. (1987). Socially desirable responding in organizational behaviors: A reconception. *Academy of Management Review, 12*, 250-264.

Zhao, H. & Seibert, S.E. (2006) The Big Five personality dimensions and entrepreneurial status: a meta-analytic review. *Journal of Applied Psychology, 91*, 259-271.