8-2020

Resume Ratings: The Influence of Rater Individual Differences

Emily Ann Frye

Resume Ratings: The Influence of Rater Individual Differences

by

Emily Ann Frye

A thesis submitted to the College of Psychology and Liberal Arts of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Industrial/Organizational Psychology

Melbourne, Florida
August, 2020

We the undersigned committee hereby approve the attached thesis, "Resume Ratings: The Influence of Rater Individual Differences," by Emily Ann Frye.

Patrick Converse, Ph.D.
Professor
School of Psychology

Gary Burns, Ph.D.
Professor and I/O Chair
School of Psychology

Debbie Lelekis, Ph.D.
Associate Professor
School of Arts and Communication

Lisa Steelman, Ph.D.
Professor and Dean
College of Psychology and Liberal Arts

Abstract

Title: Resume Ratings: The Influence of Rater Individual Differences

Author: Emily Ann Frye

Advisor: Patrick Converse, Ph.D.

Resumes remain a popular selection tool in practice but are rarely researched. Moreover, little is known about how much variability is present across resume ratings and how much of that variability may be attributed to the resume rater's own individual differences. Therefore, the present study aims to address these issues by investigating the influence of resume raters' characteristics on resume hirability and personality ratings. More specifically, drawing from the lens model and related research, the present study examines the association between resume rater personality, dispositional intelligence, gender, experience, and cognitive ability and hirability and personality ratings. Using a cross-sectional survey design, the present study sampled 102 participants who have had to rate and/or evaluate resumes for their current or previous job(s). Participants completed target individual difference construct measures and participated in a resume rating activity using four student resumes. Results from the study indicated that there was variability in hirability ratings and personality rating accuracy across raters but the rater individual differences did not significantly relate to these outcomes with the exception of dispositional intelligence, which related to the accuracy of personality ratings. These findings suggest that more research is needed to investigate variability in resume ratings.

Table of Contents

List of Figures

List of Tables

Acknowledgement

First, I would like to express my deepest appreciation to my thesis advisor, Dr. Patrick Converse. Pat, I am extremely grateful to you for taking me on as an advisee late in the year and for the invaluable advice and guidance you gave me along the way to help push this thesis to its completion. I have appreciated your unwavering support and insightful suggestions, and I have enjoyed working with you over the past several months. I would also like to extend my deepest gratitude to Dr. Gary Burns whose knowledge and prior research in the topic area of this thesis was instrumental to the foundation and completion of it. Gary, I have appreciated your support and advice as well as your helpful suggestions. Additionally, I am very grateful to and appreciative of Dr. Debbie Lelekis for her time and support throughout this thesis.

I must also thank Florida Tech Career Services, specifically Maria Hidalgo and Dona Gaynor, whose support in the data collection for this thesis cannot be overestimated. Thank you both for working with me and making this thesis possible.

I would also like to extend a special thanks to Monica DeFelice-Robinson with the Office of Graduate Programs and Kristin Heifner with Evans Library at Florida Tech. I am extremely grateful to you both for your patience and time that you gave to answer my questions, meet with me, and provide all the guidance I needed for formatting this thesis.

Lastly, I would like to express my thanks to Kyle Crook for all of his love and support.

Chapter 1
Introduction

In the academic realm of selection research, resumes appear to be a largely lost and forgotten tool. Nevertheless, despite changes and trends in selection processes and procedures over the past several decades, resumes have arguably remained one of the most consistent selection tools in practice and are still frequently used today. In fact, it has been reported that a single corporate job opening will receive 250 resumes on average (Glassdoor, 2015). With this in mind, it is surprising that resumes are not more heavily researched and do not seem to be valued as much in the science world as they are in practice.

To illustrate this divide between science and practice in regard to resumes, a brief review of resume research is necessary. Resume research emerged in the early 1970s coming out of research on biographical data, more commonly known as biodata. Specifically, prior research has indicated that the first formal research definition of a resume was provided in a biographical data research article written by James Asher in 1972 (Arnulf, Tegner, & Larssen, 2010). In his article, Asher discussed biographical data items in the context of job applications. In expanding on these items, Asher demonstrated that biographical data that was historical and verifiable was commonly used in selection and also held strong validity. The historical and verifiable biographical data items that Asher referred to were items commonly found on a typical resume. However, despite this breakthrough for science that was followed by a spike in resume research from the 1970s until the early 2000s, practice had already been utilizing resumes for about 20 years prior to the 1970s (Hebberd, 2013). Furthermore, resumes have been thriving now more than ever in practice, surviving through popular business trends such as technological advances. Resume research, on the other hand, does not paint the same picture, as it has been on a decline since the early 2000s with only a few pieces of resume research coming out per year. With all of this

information in mind, it is necessary for the field to continue developing our understanding of the psychological mechanisms associated with resumes as well as their influences on selection decisions in order to close this gap.

Thus, the proposed study aims to add to the limited psychological understanding of resumes by examining the impact of rater individual differences on two types of resume ratings: personality ratings and hirability ratings. A recent review of resume research demonstrates that what we know about resumes now is geared more towards benefitting job seekers by showing empirical support for the inclusion or exclusion of various resume items (Risavy, 2017). In contrast, there is much less focus on those who screen resumes. More specifically, there is limited research on actual judgments made from resumes and of that limited amount there is little to no research on the resume judgment process from the recruiter/screener side. Thus, the present study will uniquely add to existing literature on resume research by first investigating variation in resume ratings across raters and then examining rater individual differences as a potential source of that variation. Using supporting research from the perceptions and judgments domain, the present study develops a better understanding of the potential underlying factors influencing resume evaluations. Results from this study theoretically advance resume research and inform practitioners about the potential underlying influences that impact their own resume ratings.

Chapter 2
Background

Judgments in Personnel Selection

Judgments are a commonplace in personnel selection; without them, decisions would not be made in regard to which applicants to hire for a job. Furthermore, it has been asserted that judgments are necessary to all personnel selection decisions because they represent the underlying cognitive process of those decisions (Guion, 2011). Thus, in order to understand judgments made in resume contexts, it is important to first understand what judgments entail in selection generally.

Judgments first and foremost imply prediction (Guion, 2011). More specifically, judgments in selection largely concern the attempt to predict a candidate's future behavior (i.e., job performance). Judgments help practitioners make decisions about applicants, and thus selection research largely pertains to answering the question: how can we make judgments regarding selection better? In order to make judgments, individuals (judges) need information along with the ability to interpret and integrate that information (Guion, 2011). In the selection context, resumes act as a form of this information and practitioners (i.e., recruiters or hiring managers) act as the judges. That is, when practitioners view an applicant's resume, they make judgments that they develop from the information contained within the resume.

Lens Model

One way to think about the judgment process regarding resumes is through the use of the lens model as a conceptual framework. Prior research has utilized the lens model to illustrate selection decision-making, and this model has even been used to illustrate the resume screening process as well (Kausel et al., 2016; Guion, 2011; Burns, 2004). The lens model, which was originally conceptualized by Egon Brunswick in 1952, views the general human judgment process in terms of the components necessary for judgmental achievement or accuracy (Kaufmann, Reips, & Wittmann, 2013; see Figure 1 for an illustration of the lens model). More specifically, the lens model posits the underlying process by which an individual attempts to judge an unobservable criterion in another individual (Nestler & Back, 2013). There are three main components of the judgment process within the lens model framework: a judgment made regarding a criterion, cues or predictors of the criterion, and the criterion itself (Kausel et al., 2016). The criterion represents a directly unobservable feature or characteristic, such as personality, within an individual. Because the criterion is not directly observable, the judge uses observable signals, referred to as cues, within the environment that relate to the unobservable characteristic. The extent to which the criterion relates to a cue is referred to as cue validity, and the strength of a cue's weight in the judgment is referred to as cue utilization. Accuracy of the judgment results when there are valid cues within an environment and when a judge effectively utilizes those cues (see Nestler & Back, 2013, for a discussion of these components).

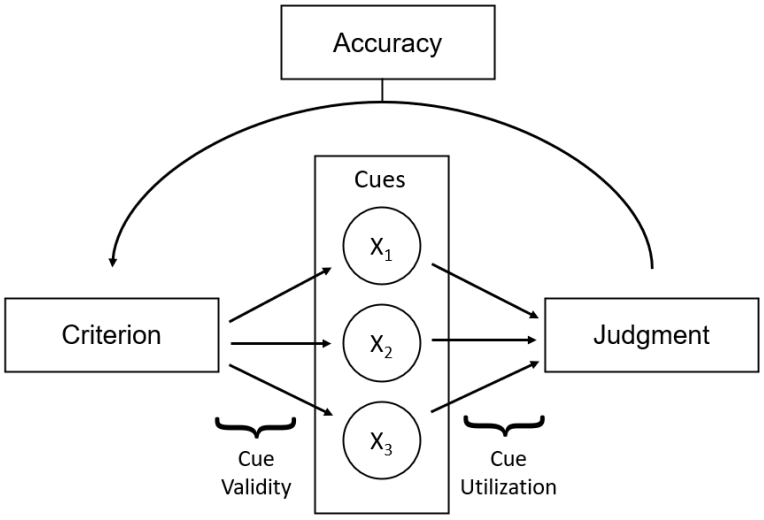

Figure 1 – An illustration of the lens model.

Relating the lens model back to resumes in a selection context, this model suggests that the information contained within a resume can be seen as the cues that signal unobservable characteristics within the job applicant. These cues are then viewed and utilized by the practitioner who acts as the judge. The practitioner then makes judgments based on those cues in order to reach a decision about the applicant. Thus, the lens model framework indicates that the resume acts as a "lens" through which judges attempt to see characteristics of the applicant (Burns, 2004). Figure 2 presents an illustration of the lens model applied to resumes in a selection context.

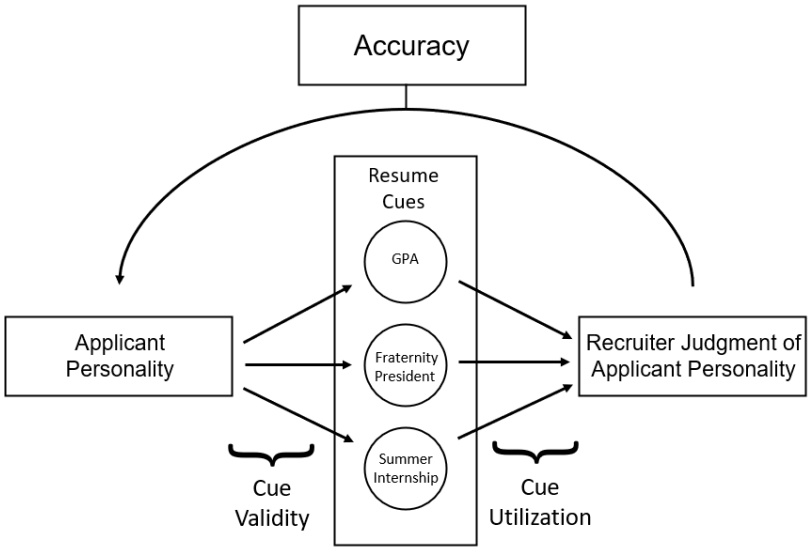

Figure 2 – An illustration of the lens model applied to resume contexts.

Resume Content and Inferences

As mentioned previously, information contained within a resume is mainly biographical data; that is, resume information tends to be historical and verifiable and represents past experiences or events (Arnulf, Tegner, & Larssen, 2010; Asher, 1972). Resumes are one of the most commonly used selection tools, and arguably one of the biggest reasons behind their use comes from the notion upon which biographical data rests: past behavior is indicative of future behavior (see Harvey-Cook & Taffler, 2000). Thus, applicants' past experiences will likely indicate their future experiences and successes or failures within those experiences as well. In fact, prior research has demonstrated that biographical data is reliable and demonstrates strong validity in selection contexts (Harvey-Cook & Taffler, 2000; Asher, 1972). Furthermore, prior resume research has shown that specific biographical data items within a resume context relate to future job performance and/or predictors of future job performance (Cole, Field, & Giles, 2003; Anderson & Shackleton, 1990; Brown & Campion, 1994; Schmidt & Hunter, 1998). With the wide array of evidence supporting reliability and validity for biodata, it is no wonder why resumes are so popular.

The specific biographical data items contained within a resume can vary greatly. However, a recent review of the resume literature (Risavy, 2017) synthesizes prior resume research based on the empirical findings for what information should and should not be included on a resume. This review suggested the following sections should be, and have commonly been, included on a resume: personal information; personal opening, job objective, career objective, and summary of qualifications; education; work experience; references; scholarships, awards, and honors; hobbies, interests, and extracurricular activities; and willingness to relocate and travel (Risavy, 2017). Though a detailed review of the specific sections and formatting contained within the resume is outside the scope of this paper, it is important to note that most of the prior research on resumes has examined the influence of specific sections on resume evaluations (e.g., Bright & Hutton, 2000; Hakel, Dobmeyer, & Dunnette, 1970; Knouse, 1994; Nemanick & Clark, 2002; Wilkin & Connelly, 2012; Ross

& Young, 2005). From this literature, two types of resume evaluations, or judgments, have emerged: hirability judgments and personality judgments.

Hirability Judgments

Hirability can be defined as the extent to which applicants are determined to be employable for the job at hand. In the resume screening context, resume screeners "act as an initial employment gatekeeper" (p. 5) by determining which applicants should be shortlisted and which applicants should be rejected (Cole, Feild, Giles, & Harris, 2009). Thus, within any resume screening context, the ultimate judgment to be made from the resume would regard the applicant's hirability. There is limited theoretical development underlying hirability or how hirability judgments are formed; however, attribution theory has been commonly applied to selection contexts. In fact, a review of attribution theory in personnel selection research posited that this theory provides the ability for organizational researchers to better understand how employers evaluate and determine hirability from the contributions of the applicant and the situational environment (Knouse, 1989). According to attribution theory, behavior can be explained by either internal/dispositional attributions, such as abilities or personality, or external/situational attributions, such as task difficulty (Knouse, 1989). In this light, resume screeners use the information contained within a resume to make judgments about the applicant by making attributions and evaluations regarding the applicant's characteristics such as abilities, personality, motivation, and job fit (Cole, Rubin, Feild, & Giles, 2007). Relating back to the beginning of the background section, these attributions and evaluations inform predictions; resume screeners make attributions and evaluations in an attempt to predict the applicant's future behavior on the job.

Three studies investigating biodata phenomenology in resumes indirectly supported attribution theory while eliciting the inferences that recruiters draw from resume information (Brown & Campion, 1994). The purpose of this investigation was to examine recruiters' perception and use of biodata in terms of attributes for making selection screening decisions. The attributes considered in the study included ability attributes, defined as basic human capacities (e.g., math), and nonability attributes, defined as human

qualities (e.g., motivation; Brown & Campion, 1994). Across three studies, results demonstrated that resume biodata was judged by recruiters as representing both ability and nonability attributes and recruiters were not only able to distinguish between the types of attributes, but they also reliably judged nonability attributes (Brown & Campion, 1994). This demonstrates that recruiters consider resume information to represent important attributes for hirability, and furthermore, recruiters utilize that same information to make their evaluations or judgments. Other research has also demonstrated recruiters' use of specific information from resumes in hirability judgments. Some studies have examined specific resume characteristics such as behavioral coursework, objective statements, resume determinateness, and even college sport participation as components that influence a resume raters' hirability judgments (Rynes, Lawson, & Ilies, 2003; Thoms, McMasters, Roberts, & Dombkowski, 1999; Oliphant & Alexander, 1982; Tanguay, Camp, Endres, & Torres, 2012).

Personality Judgments

Personality judgments are another type of judgment that can be made based on resumes, though research on this issue is still in initial development. Nonetheless, personality judgments in general have been a long-standing research avenue, and examining applicant attributes, such as personality, has been recognized as an important consideration by recruiters (Knouse, 1989).

Personality judgment research has been significantly influenced by the work of David Funder. Regarding personality judgment in his work, Funder wrote, "Judgments of personality are attempts to identify the psychological properties of people, such as personality traits, that help to explain what they have done in the past and to predict what they will do in the future" (Funder, 1995, p. 652). Reflecting back on selection judgments regarding the prediction of future behavior, evaluating an individual's personality in this frame of reference seems to fit well in selection research.

The personality judgment process has been conceptualized through a model developed by Funder called the realistic accuracy model (RAM), which has been asserted to be based on the lens model (see Figure 3; Funder, 1995; Letzring, 2008). In his research, Funder laid the groundwork for the underlying process behind making accurate personality judgments by proposing four stages that need to occur for accurate personality judgments (Funder, 1995, 1999). The four stages are: relevance, availability, detection, and utilization. The personality judgment process begins with an attribute of a target individual. This individual must convey information in some form that is both relevant to the specified attribute and available to the judge. Next, the judge must be able to detect the information from the target as well as utilize the relevant pieces of that information in order to make an accurate personality judgment (Funder, 1995, 1999).

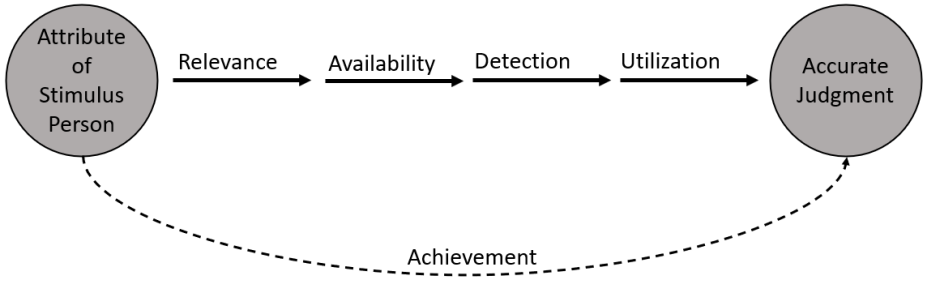

Figure 3 – An illustration of the Realistic Accuracy Model. Adapted from *Personality Judgment: A Realistic Approach to Person Perception* (p. 119) by D. C. Funder, 1999, Academic Press. Copyright 1999 by Academic Press.

Relating this information back to the resume process, the information that is elicited from the target would be the resume itself. As discussed previously, the resume contains biographical data representing an applicant's prior experiences and characteristics. A field study by Cole and colleagues (2003) examined the relationship between specific resume biodata (e.g., overall grade point average, supervisory experience, and membership in professional societies) and the personality of the applicant (i.e., the Big Five). The results

of this study demonstrated that resume biodata items do have links to an applicant's personality (Cole, Feild, & Giles, 2003). For example, overall grade point average and completing a summer internship position both positively correlated with conscientiousness, and holding an elected office position for a college club positively correlated with extraversion. Thus, the information contained in resumes does appear to be relevant to the applicant's personality. In sending the resume to an employer, the applicant is making his or her relevant information available to the judge, namely the employer. The employer must then be able to detect the information contained within the resume and utilize that information to make a judgment. Therefore, the resume screening process in and of itself allows not only for personality judgments to occur, but also for those judgments to be accurate if utilized correctly.

Chapter 3
Hypothesis Development

Variability in Judgment

Despite the usefulness of the judgment process for selection decisions, judgments are not always consistent or accurate, and different judges viewing the same stimuli can produce different or conflicting results (Guion, 2011). To demonstrate this point, the field of forensic psychology has frequently found variability across raters in the literature (e.g., Guarnera, Murrie, & Boccaccini, 2017; Murrie & Warren, 2005; Murrie et al., 2008). Ratings within these studies commonly involve forensic evaluations such as ratings of criminal defendants for legal sanity as well as incompetence to stand trial. Surprisingly, evidence in these studies shows that forensic evaluators can disagree in their ratings and that there is a wide variation in forensic evaluations. For example, a study conducted by Murrie and colleagues (2008) found that, across 60 clinicians, ratings of defendants' incompetence to stand trial ranged from 0% to 62% across a total of more than 7,000 evaluations. This provides evidence that some forensic evaluators are more likely to find defendants incompetent to stand trial than other evaluators (Murrie et al., 2008). Moreover, another forensic psychology study examining rater variability asserted, "we expect that clinicians are more or less interchangeable" (Harris, Boccaccini, & Murrie, 2015, p. 321). The article went on to provide evidence that recent field studies suggest that clinicians are not interchangeable and that variation in ratings across clinicians does exist. Within personnel selection research, no studies to date have focused specifically on the notion of variability across resume raters. Similar to forensic contexts, it appears that recruiters and other resume raters are often assumed to be interchangeable but little evidence directly speaks to this issue. Thus, it is imperative to investigate variability across raters in the resume context.

Though no study within resume research has solely focused on examining rating variability, previous resume studies have alluded to the existence of variability due to characteristics of the resume raters themselves. Within the framework of the lens model, this can be supported through the notion that judges may each have inherent traits or characteristics that cause them to utilize the same stimulus cues differently than other judges. As an example of other research that has alluded to rater characteristics contributing to rating variability, studies by Cole and colleagues (2004, 2007) controlled for personal characteristics of the recruiter, such as gender and experience, asserting that prior research shows that these characteristics can influence selection-related judgments (Hitt & Barr, 1989). Additionally, it has been asserted that when it comes to personality judgments, judgmental achievement or accuracy is an individual difference inherent to the judge; in other words, certain traits exhibited in an individual make them a "good judge" (Funder, 1999; Rogers & Biesanz, 2019; Letzring, 2008). Thus, these studies provide reason to believe that variability across resume ratings not only exists, but also may be attributed to the raters' traits.

Before examining specific rater characteristics that may be influencing factors in resume rating variability, it is first important to examine how much variability is present in resume ratings. The notion of variability in resume ratings can be understood differently in terms of hirability judgments versus personality judgments. For hirability judgments, examination of resume rating variability across different raters can focus on hirability ratings themselves but not on the accuracy of those ratings because applicants' "true scores" are difficult to determine. Therefore, the current study will first examine how much variability is present in hirability ratings:

> *Research Question 1*: How much variability is present across raters on resume hirability ratings?

On the other hand, for personality judgments, accuracy can be examined due to the fact that the applicant's personality can be directly measured (and thus compared against

personality judgments to determine accuracy). Therefore, the current study will also examine how much variability is present in personality judgment accuracy:

> *Research Question 2*: How much variability is present across raters on resume personality rating accuracy?

Individual Differences Influencing Hirability Judgments

As mentioned previously, there is reason to believe that variation exists in resume judgments of hirability, and some of that variability can likely be attributed to characteristics of the raters. An example of this can be found in a resume study by Camp and colleagues (2014). Within this study, the authors examined the influence of raters' prior sports experiences on resume ratings by drawing from the similarity-attraction effect, signaling theory, and self-categorization theory. The authors argued that a resume rater's own sports experience would impact the way the raters rated resumes of student athletes. Results supported this notion by showing that the resume rater's years of experience in sports participation positively influenced resume ratings (Camp et al., 2014). Though this study focused on the specific issue of sports experience, the results clearly show that resume raters' own characteristics can influence the way they rate resumes.

Rater Gender

One factor that may influence resume-based hirability ratings is rater gender. This can be demonstrated through the hawk-dove effect. The hawk-dove effect, or problem, asserts that some raters demonstrate a tendency to rate consistently leniently or stringently (McManus, Thompson, & Mollon, 2006). Raters who demonstrate the tendency to rate with higher standards and thereby rate more harshly are considered hawks and are referred to as being stringent. On the other hand, raters who demonstrate the tendency to rate with lower standards and thereby rate more favorably are considered doves and are referred to as being lenient (see McManus, Thompson, & Mollon, 2006, for further discussion of the hawk-dove effect).

Gender has been shown to relate to stringency and leniency through various studies. For example, a clinical study by McManus and colleagues (2006) found that males tended to be more likely than females to be stringent in ratings. Additionally, other studies have asserted that females tend to make more favorable evaluations and be more lenient than

males (Rose & Andiappan, 1978; London & Poplawski, 1976). With these findings, it can be inferred that gender effects resulting in stringency and leniency may carryover to resume contexts for resume ratings. In particular, males may be more stringent resume raters and therefore make less favorable judgments regarding applicants, while females may be more lenient raters and make more favorable judgments regarding applicants. Thus, the first proposed hypothesis is:

> *Hypothesis 1*: Male raters will make less favorable hirability judgments from resumes than female raters.

Rater Personality

Another factor that may influence resume-based hirability ratings is rater personality. Though there are many ways to define and conceptualize personality, one of the most common conceptualizations within a selection context is through the five-factor model. Commonly referred to as the Big Five, the five-factor model is a conceptualization of personality as a classification of traits into five dimensions: neuroticism, extraversion, openness, agreeableness, and conscientiousness (McCrae & Costa, 1989). Common conceptualizations and characterizations of each dimension, which have been adapted from McCrae and John (1992), are as follows: Neuroticism has been conceptualized as the tendency of an individual to experience distress and is often characterized by nervous tension, frustration, and guilt. Extraversion has been interpreted as an interpersonal aspect related to sociability and is characterized by affiliation, optimism, and energy. Openness has been interpreted as a broader form of intellect and is characterized by creativity, wide interests, originality, and curiosity. Agreeableness has been defined as humane aspects that relate to warmth and compliance, such as altruism, compliance, submission, and trust. Conscientiousness has been interpreted as one's tendency to organize and direct their own behavior and is characterized by thoroughness, neatness, organization, diligence, and possessing an achievement orientation.

A rater's personality may influence ratings because personality may influence the way that a rater sees and utilizes available information. Furthermore, prior research has supported this notion by asserting that personality is a contributing factor that influences ratings or evaluations (e.g., Guarnera, Murrie, & Boccaccini, 2017; Finn, Cantillon, & Flaherty, 2014; Sabzwari, Pinjani, & Nanji, 2018). One study by Miller and colleagues (2011) explicitly examined the link between a rater's personality and rating outcomes in the forensic psychology field. In this study, the authors sought to investigate the influence of raters' personality on scoring tendencies for the Psychopathy Checklist-Revised (PCL-R). They found that personality of the rater not only accounted for score variance, but that specific traits and facets of personality were significant contributors. More specifically, facets of conscientiousness and agreeableness related to scoring tendencies such that conscientious raters tended to score offenders higher for psychopathy and agreeable raters tended to score offenders lower for psychopathy (Miller et al., 2011). Additionally, a study by Bernardin and colleagues (2000) found that both conscientiousness and agreeableness influenced rating tendencies in a manner similar to the Miller and colleagues study. Thus, highly conscientious raters, who characteristically complete tasks more thoroughly, may pay more attention to the details within a resume and thereby rate the applicant more strictly or stringently. On the other hand, highly agreeable raters, who are characteristically friendly and possess an optimistic view of human nature, may be less strict in their ratings because they may not want to reject applicants. Therefore, the following hypotheses are proposed:

> *Hypothesis 2*: Raters high in conscientiousness will make less favorable hirability judgments from resumes than raters low in conscientiousness.
>
> *Hypothesis 3*: Raters high in agreeableness will make more favorable hirability judgments from resumes than raters low in agreeableness.

Aside from agreeableness and conscientiousness having an influence on ratings, neuroticism has also been suggested to have an influence as well (Finn, Cantillon, & Flaherty, 2014). In particular, one study examining the impact of neuroticism on the

evaluation of job candidates found that highly neurotic raters tended to make more negative judgments of hirability (Unsal & Caliskur, 2004). Given that those who are high on neuroticism have a tendency to experience negative emotions or emotional instability, it may be that resume raters who are highly neurotic may rate applicants more negatively or harshly from their resumes due to this general negative tendency. Thus, the following hypothesis is proposed:

> *Hypothesis 4*: Raters high in neuroticism will make less favorable hirability judgments from resumes than raters low in neuroticism.

Individual Differences Influencing Personality Judgments

Like hirability judgments, personality judgments are also posited to be influenced by characteristics of the judge/rater. Unlike hirability judgments, however, personality judgments can be compared against actual personality levels (obtained through self-reports) to examine accuracy. Thus, the accuracy of the personality judgment and the individual differences that influence accuracy will be discussed here.

As asserted by Funder (1995, 1999), the accuracy of personality judgments can be conceptualized through the four-stage RAM in which all stages (i.e., availability, relevance, detection, and utilization) must be achieved. In addition to this model, Funder also asserted that certain conditions could enable better judgmental accuracy. One such condition was the concept of the "good judge" (Funder, 1999). The notion behind the good judge is that certain individuals are able to make accurate judgments due to their own, inherent traits (Christiansen et al., 2005). In other words, these individuals' traits enable them to properly utilize information about a target's personality in order to make an accurate personality judgment. Personality and selection studies have both arguably been concerned with identifying the good judge, though results were not always promising (Funder, 1995). However, research still often investigates the good judge to this day, and some studies have been able to identify that the good judge does exist and thus that there are meaningful differences across individuals in personality judgmental accuracy (Rogers & Biesanz, 2019; Letzring, 2008).

Rater Cognitive Ability

One factor that may influence the accuracy of resume-based personality ratings is rater cognitive ability. Cognitive ability, or intelligence, can be generally referred to as an individual's general mental capability as it relates to reasoning, planning, solving problems, thinking abstractly, comprehending complex ideas, and learning (Gottfredson, 1997). It has been asserted by Christiansen and colleagues (2005) that "making judgments

is an extremely demanding cognitive process" (p. 125). For example, making personality judgments from resumes involves identifying cues present within the resume, such as undergraduate GPA and extracurricular activities, and utilizing those cues by understanding the significance of or weight they hold for specific judgments. These two cognitive processing components relate directly to two stages of the RAM: detection and utilization. If a rater is able to detect the appropriate cues in a resume and utilize them in such a way to make an accurate personality judgment, then the rater will achieve judgmental accuracy. Raters with a higher level of cognitive ability may be more likely to achieve judgmental accuracy because their cognitive capability to detect and utilize cues may be stronger and may enable them to cognitively process more effectively. Supportive of this, prior research on rating accuracy has found that rater intelligence does relate to rating accuracy (e.g., Borman, 1979; Smither & Reilly, 1987; Hauenstein & Alexander, 1991). Even within his own research on RAM, Funder (1995) asserted that the judge's intelligence would influence the accuracy of personality judgments. Furthermore, the study by Christiansen and colleagues (2005) "revisited" the notion of the good judge and found that individuals higher on general mental ability tended to make more accurate personality judgments. Thus, based on these results, the following hypothesis is proposed:

> *Hypothesis 5*: Raters high in cognitive ability will make more accurate personality judgments inferred from resumes than raters low in cognitive ability.

Rater Dispositional Intelligence

Another factor that may influence the accuracy of resume-based personality ratings is rater dispositional intelligence. First proposed and validated in the study by Christiansen and colleagues (2005), dispositional intelligence represents an individual difference construct that is characterized by an individual's knowledge about personality and its links to behavior. Furthermore, the components of dispositional intelligence include knowledge of the link between traits and behaviors, understanding of the relevance of situation and trait interaction, and proficiency of trait concepts. Christiansen and colleagues not only proposed this construct, but also developed and validated a measure and tested it in relation

to accuracy of personality judgments. The results of the study demonstrated that dispositional intelligence could be reliably measured, and that it positively related to judgmental accuracy. Additionally, a couple of other studies have used the measure since and found that the dispositional intelligence trait positively related to judgmental accuracy (i.e., Powell & Bourdage, 2016; Merbedone, 2012). Though these studies were not conducted in a resume setting, the findings suggest that an individual's level of knowledge about personality should positively relate to their ability to identify and correctly utilize cues found within a resume in order to make accurate personality judgments. Thus, the study hypothesizes:

> *Hypothesis 6*: Raters high in dispositional intelligence will make more accurate personality judgments inferred from resumes than raters low in dispositional intelligence.

Rater Personality

Another factor that may influence the accuracy of resume-based personality ratings is rater personality. Similar to the argument presented previously for rater personality influencing the way a rater makes evaluations, rater personality can also be posited to influence the accuracy of personality ratings. Prior research has supported this notion. Christiansen and colleagues (2005) as well as the study by Merbedone (2012) found that both rater openness to experience and rater conscientiousness positively related to making accurate personality judgments. It may be that those high in conscientiousness may pay more attention to detail and are able to utilize information presented in a resume better in order to make more accurate judgments. Openness to experience, on the other hand, has been found to correlate with social intelligence and relate to intellect which can enable judgmental accuracy (Shafer 1999; Christiansen et al., 2005). Therefore, those with high openness to experience may be able to more easily pick up and utilize resume cues. Thus, the following hypotheses are proposed:

Hypothesis 7: Raters high conscientiousness will make more accurate personality judgments inferred from resumes than raters low in conscientiousness.

Hypothesis 8: Raters high in openness to experience will make more accurate personality judgments inferred from resumes than raters low in openness to experience.

Rater Experience

A final factor that may influence the accuracy of resume-based personality ratings is rater experience. While no studies have explicitly examined the impact of rater experience on resume ratings directly, studies in other areas suggest that a negative relationship exists between rater experience and the accuracy of ratings. For example, a study by Kennealy and colleagues (2017) found that a significant portion of rating variance in a juvenile risk assessment rating (i.e., California – Youth Assessment and Screening Instrument) was attributable to rater experience and the accuracy of those ratings depended on experience such that those with less experience made more accurate juvenile risk ratings than those with more experience. Along similar lines, a meta-analysis of lens model studies across various domains also found the relationship between experience and rating accuracy to be negative, suggesting that experienced judges may rely on intuitive judgment more so than those with less experience (Karelaia & Hogarth, 2008). More specifically, this negative relationship between rater experience and accuracy may be due to the fact that those with more experience may be in job roles or positions where they are faced with making more complex decisions. In these contexts, individuals may need to rely more on their intuition to make those decisions. This can transfer to and has been commonly seen in hiring contexts where managers and practitioners have been found to have a preference to make decisions following their "gut" or intuition (Lodato, 2008). However, in relying on intuition, faults in judgment can occur (Guion, 2011). Therefore, those with more experience may tend to rely more on intuition to reach decisions which can then result in inaccurate judgments. Thus, the following hypothesis is proposed:

Hypothesis 9: Raters with more experience will make less accurate personality judgments inferred from resumes than raters with less experience.

Chapter 4
Method

Participants

Raters

Participants designated as raters were the main, targeted sample for this study and included adults who have had to rate and/or evaluate resumes for their current or previous job(s). One hundred and twenty-two rater participants were recruited through Amazon Mechanical Turk (MTurk) for the purposes of this study. However, 18 participants were excluded due to failed attention checks. Thus, the final sample for analysis was 104. Approximately 66% of the sample was female and 81% was White with the remaining 10% Asian, 4% Hispanic/Latino, 2% Mixed, and 1% representing Black or African American, American Indian or Alaskan Native, or Other. Age of the sample ranged from 22 to 69 years of age with a mean age of approximately 43 (*SD* = 12.17). Approximately 20% of the sample held the job of a recruiter or human resources representative/professional when they had rated resumes. The remainder of the sample either selected they were a hiring manager or other role when they had rated resumes. Lastly, the years of experience that participants had rating resumes ranged from 0 to 30 years with a mean of 7.3 years (*SD* = 7.5).

Ratees

Participants designated as ratees posed as job applicants by submitting their resumes for use in the resume rating activity that was to be completed by the rater participants. Based on prior research, a sample of four ratees was determined to be

sufficient for this study (see Thoms et al., 1999; Aspers & Derous, 2017; Bright & Hutton, 2000). Thus, ratee participants consisted of four senior undergraduate students whose age ranged between 22 to 30 years. Prior research has supported the use of college students for resume research because students' resumes tend to display different experiences (e.g., extracurricular activities and scholastic honors) aside from solely job-related ones, which requires the resume rater to focus on and make evaluations from resume content that does not relate solely to prior job experience (Thoms et al., 1999; Tanguay et al., 2012). These different experiences displayed within the resume are more likely to relate to aspects of the student's personality and other inherent traits (Cole, Field, & Giles, 2003). Each ratee had a unique school major (i.e., psychology, environmental science, business administration, and human factors), and two of the participants were male and the other two were female. These participants were recruited through the local university Career Services department via email. Ratees' compensation for participation included a report summarizing their resume's ratings that were given by the rater sample.

Procedure

This study began upon approval from the local university Institutional Review Board, and a cross-sectional and correlational survey research design was utilized to investigate the research questions and hypotheses. Therefore, an online survey was the data collection method for the rater sample. This survey was developed and conducted through an online survey platform called Qualtrics. The survey contained demographic questions as well as several measures of target constructs to examine the hypotheses (see Measures below). In addition to these measures, the survey also contained a resume rating activity that consisted of a job description accompanied by four resumes and two types of evaluations per resume. Each resume was one page and was presented on its own page with the ratees' names, addresses, and other contact information blacked out. The job description was obtained through the local university Career Services department and described a real, entry-level job for a management trainee that did not require a specific major. The job description was presented to ratees during recruitment, and ratees submitted their resume to be used for rater evaluation purposes. Thus, the rater sample was prompted to review the job description and then review each resume that was tailored to fit this job description. Raters then evaluated each resume in terms of hirability and personality. Utilizing a real, entry-level job description along with real student resumes improved ecological validity for this study.

Measures

Rater Measures

Demographics

Two targeted demographics were collected from the rater participants: gender and resume rating experience. Resume rating experience was collected through a

combination of experience measures utilized in previous studies (see Camp et al., 2014; Tanguay et al., 2012). Thus, rater participants were asked their number of years of experience screening resumes, type of job role possessed when screening resumes (e.g., recruiter, hiring manager), and approximate number of resumes screened. For the purpose of hypothesis testing, number of years screening resumes was utilized for analyses. Type of job role and number of resumes screened were used only for exploratory purposes and not hypothesis testing.

Cognitive Ability

Cognitive ability was assessed via the 5-item abbreviation of the International Cognitive Ability Resource (ICAR) 16-item test. ICAR is a public-domain assessment tool that facilitates cognitive ability assessments, and the ICAR 16-item sample test is one of the most common cognitive ability assessments developed by ICAR (ICAR, 2014). The 5-item version of this 16-item test was developed and validated in a study by Kirkegaard and Bjerrekaer (2016) for the purpose of creating a version of the 16-item test that did not take too long for participants to complete. This 5-item test includes four different item types for assessing cognitive ability: verbal reasoning, letter and number series, matrix reasoning, and three-dimensional rotation. A sample item is: “What number is one fifth of one fourth of one ninth of 900?”

Personality

Personality of the rater was assessed via a Big Five personality assessment. Specifically, this study assessed personality via a short-form of the 50-item International Personality Item Pool (IPIP) based on Goldberg’s (1992) markers of the Big Five (IPIP, n.d.). The Mini-IPIP scale is a 20-item assessment measuring the Big Five and was developed and validated across five studies by Donnellan and colleagues (2006). The questionnaire asks participants to rate the extent to which

they feel the items accurately describe themselves. A sample item is: "Am not interested in other people's problems."

Dispositional Intelligence

Dispositional intelligence was measured via the short-form of the dispositional intelligence scale developed and validated by Christiansen and colleagues (2005). Dispositional intelligence assesses an individual's knowledge of personality as well as knowledge of personality's relationship to various behaviors (Christiansen et al., 2005). A sample item is: "Which of the following situations are most relevant to the trait of sociability?"

Ratee Hirability Judgments

A four-item Likert scale used in previous resume studies by Cole and colleagues (2007; 2009) was used in the present study to assess ratee hirability. These four items were developed and utilized by Cole and colleagues (2007; 2009) from existing selection decision research (Cable & Judge, 1997; Kristof-Brown, 2000; Singer & Bruhns, 1991). A sample item is: "How likely is it that you would recommend the applicant be hired?"

Ratee Personality Judgments

Rater impressions of ratees' personality were assessed through the method developed and used by Cole and colleagues (2009). Drawing from prior research from Costa and McCrae (1992) as well as Goldberg (1992) on personality markers, Cole and colleagues conducted two pilot studies to develop and test the use of personality markers, or adjectives, that relate to the Big Five personality dimensions as a way for resume raters to rate applicants' personality. Thus, in their study, resume raters were asked, "To what extent do you agree or disagree that each adjective accurately describes this applicant?" Twenty-five adjectives were listed with five adjectives representing each of the Big Five personality dimensions.

Responses for adjectives within each dimension were averaged to create a composite adjective trait score, following the procedures described by Cole and colleagues (2009).

Ratee Measures

Personality

Ratee personality was assessed via the same adjective/trait method developed by Cole and colleagues (2009) that was described previously. The difference was the frame of reference: ratees were asked instead, "To what extent do agree or disagree that each adjective accurately describes you?" The same 25 adjectives and scale points as described by Cole and colleagues (2009) were used.

Chapter 5
Results

All of the results presented were developed using IBM SPSS Statistics. The research questions were analyzed using measures of variability such as standard deviation and variance while the hypotheses were analyzed using linear regression. First, before examining the research questions and hypotheses, the data were cleaned by checking for missing values as well as missed attention checks. There were two attention checks, one placed in the beginning and one near the end. If a participant missed one of the two attention checks, then he/she was removed from the sample for analysis. A total of 18 participants were removed due to missed attention checks. The data were then checked for outliers using z-scores. Two outliers were identified and removed. Thus, the remaining sample for analysis purposes was 102 participants. Descriptives, correlations, and reliabilities were computed for all variables and are presented in Table 1.

Table 1

Descriptives, Correlations, and Reliabilities for All Variables

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Gender	1.66	.48	—										
2. Experience (in years)	7.06	6.98	.04	—									
3. Cognitive Ability	2.47	1.38	.10	-.05	.56								
4. Dispositional Intelligence	21.79	6.85	.15	.08	.43**	.88							
5. Extraversion	2.46	1.00	-.07	.19	-.24*	-.16	.86						
6. Agreeableness	3.83	.89	.16	.07	-.04	-.07	.15	.88					
7. Openness to Experience	3.62	.96	-.17	.13	.08	.07	.26**	.28**	.79				
8. Neuroticism	2.58	.95	.05	-.25*	-.03	-.18	-.33**	-.34**	-.32**	.83			
9. Conscientiousness	3.88	.86	.02	.15	-.04	.06	.16	.16	.18	-.47**	.80		
10. Average Resume Hirability Rating	4.08	.66	-.02	-.03	-.13	-.32**	.03	.04	.06	.17	-.05	.95	
11. Average Resume Personality Accuracy	.39	.16	.04	.09	.15	.43**	-.20*	.17	-.05	.03	.04	.20*	.65

Note. *$p < .05$, **$p < .01$. Reliabilities are reported as Cronbach's alpha values. The alpha value for Average Resume Hirability Rating was computed by averaging the alpha values found for the four hirability items for each resume. The alpha value for Average Resume Personality Accuracy was computed by treating the alpha values of each resume's personality accuracy score as an item to input for the reliability analysis.

The research questions concerned exploring the variability found in resume hirability ratings as well as in the accuracy of resume personality ratings. The average resume hirability rating was calculated by creating a composite score for each resume rated by each participant. Then, an overall composite score was created by averaging the four resume ratings from each participant. The average accuracy of resume personality rating was calculated using a composite score as well; however, first, the inter-item correlation of each participant's personality rating for each of the four resumes and each of the resume holder's own personality rating was computed. The resulting correlation demonstrated accuracy, which for the purpose of this study is operationalized as self-other agreement. More specifically, for each rater, once the correlations were calculated for each resume holder, these correlations were then averaged across the four resume holders to create a composite accuracy score. With these two variables computed, variability was examined and results can be found in Table 2.

Table 2

Variability in Resume Ratings

	M	*SD*	*Variance*	*Min.*	*Max.*
Average Resume Hirability Rating	4.08	.66	.44	2.06	5.75
Average Accuracy of Resume Personality Ratings	.39	.16	.03	-0.14	0.67

Note. $N = 102$.

From this analysis, resume hirability ratings showed a standard deviation of 0.66 and variance of 0.44. The resume hirability rating was comprised of a six-point scale (1-6) with minimum rating of 2.06 and a maximum of 5.75 (see Figure 4). The 75th percentile of the average resume hirability ratings was 4.52 with median of 4.03, and the 25th percentile was 3.63.

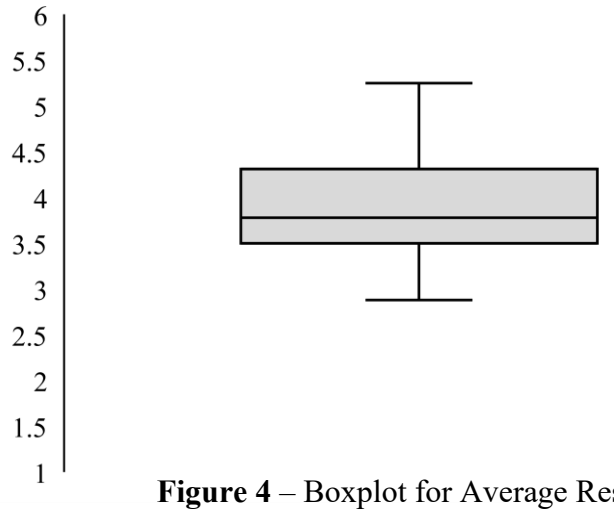

Figure 4 – Boxplot for Average Resume Hirability Ratings

The average accuracy of resume personality ratings showed a standard deviation of 0.16 and a variance of 0.03. Because the accuracy of resume personality ratings represented a correlation coefficient, the accuracy scores could range from -1.0 to 1.0. The minimum average accuracy of resume personality ratings was -0.14 and the maximum was 0.67 (see Figure 5). The 75th percentile was 0.52 with median of 0.41, and the 25th percentile was 0.28. These findings appear to be consistent with the notion that there is variability across raters in terms of hirability ratings and accuracy of personality ratings based on resumes.

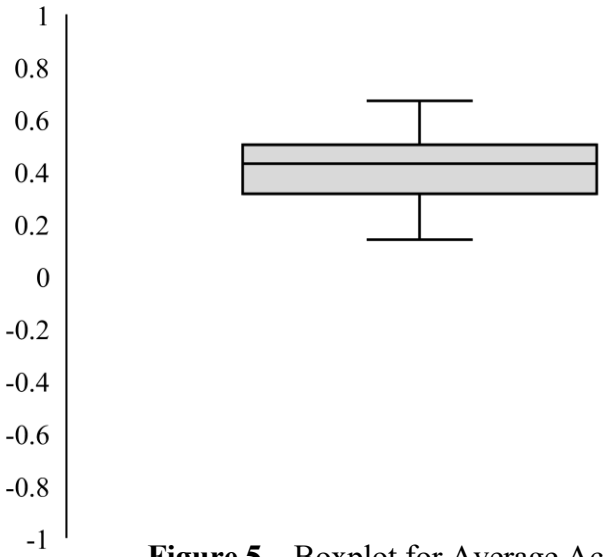

Figure 5 – Boxplot for Average Accuracy of Resume Personality Ratings

Linear regression was then used to analyze all the hypotheses. The hypotheses focusing on gender, conscientiousness, agreeableness, and neuroticism as predictors of resume hirability ratings were examined first and results can be found in Table 3. The first hypothesis posited that males would rate resumes less favorably than females. Results from the analysis showed that gender did not significantly relate to resume hirability ratings, thus failing to support Hypothesis 1. The second hypothesis posited that individuals higher in conscientiousness would rate resumes less favorably than those lower in conscientiousness. Results from the analysis showed that conscientiousness did not significantly relate to resume hirability ratings, thus failing to support Hypothesis 2.

Table 3

Summary of Regression for Resume Hirability Ratings

	R	R^2	*SE* of the Estimate	$R^2\Delta$	*b*	*SE*	*t*
Model 1	.214	.046	0.66	.05			
Gender					-0.03	0.14	-0.21
Conscientiousness					0.01	0.09	0.13
Agreeableness					0.08	0.08	0.98
Neuroticism					0.10	0.08	1.15

Note. SE = Standard Error.

The third hypothesis posited that individuals higher in agreeableness would rate resumes more favorably than those lower in agreeableness. Results from this analysis showed that agreeableness did not significantly relate to resume hirability ratings, thus failing to support Hypothesis 3. The fourth hypothesis posited that individuals higher in neuroticism would rate resumes less favorably than those lower in neuroticism. Results from the analysis showed that neuroticism did not significantly relate to resume hirability ratings, thus failing to support Hypothesis 4.

Next, the hypotheses focusing on cognitive ability, dispositional intelligence, conscientiousness, openness to experience, and experience as predictors of the accuracy of resume personality ratings were examined, and results can be found in Table 4. The fifth hypothesis posited that individuals higher in cognitive ability would make more accurate personality ratings from resumes than those lower in cognitive ability. Results from the analysis show that cognitive ability did not significantly relate to the accuracy of resume personality ratings, thus failing to support Hypothesis 5. The sixth hypothesis posited that individuals higher in dispositional intelligence would make more accurate personality ratings than those

lower in dispositional intelligence. Results from the analysis showed that dispositional intelligence significantly predicted the accuracy of resume personality ratings, $b = 0.01$, $p < .01$, thus supporting Hypothesis 6.

Table 4

Summary of Regression for Resume Personality Accuracy

	R	R^2	*SE* of the Estimate	$R^2\Delta$	*b*	*SE*	*t*
Model 1	.388	.151	0.16	.15			
Cognitive Ability					0.001	0.01	0.12
DI					0.009	0.003	3.48**
Conscientiousness					0.009	0.02	0.45
Openness					-0.01	0.02	-0.71
Experience (in years)					0.002	0.002	1.03

Note. DI = Dispositional Intelligence, *SE* = Standard Error, $^{**}p < .01$.

The seventh hypothesis posited that individuals higher in conscientiousness would make more accurate personality ratings than those lower in conscientiousness. Results from the analysis show that conscientiousness did not significantly relate to the accuracy of resume personality ratings, thus failing to support Hypothesis 7. The eighth hypothesis posited that individuals higher in openness to experience would make more accurate personality ratings than those lower in openness to experience. Results from the analysis showed that openness to experience did not significantly relate to the accuracy of resume personality ratings, thus failing to support Hypothesis 8. The ninth hypothesis posited that individuals with more experience rating resumes would make less accurate personality ratings than those with less experience. Results from this analysis showed that experience did not

significantly relate to the accuracy of resume personality ratings, thus failing to support Hypothesis 9.

Chapter 6
Discussion

Resumes remain one of the most commonly used selection tools in practice, yet little is known about the underlying mechanisms of the resume rating process. Furthermore, we often assume that all resume raters (e.g., recruiters) will make approximately the same rating regarding a single resume, but no study to date has examined this. Thus, the present study aimed to examine the variability present in resume ratings as well as investigate if rater individual differences can be linked to that variability.

Overall, the variability in resume hirability ratings was modest, yet may be meaningful because the variability found implies that different decisions were made regarding a resume holder's hirability status across participants. For example, the standard deviation of the average resume hirability rating was 0.66 and the mean was 4.08. Given this, a rater who provides a rating that is one standard deviation above the mean (4.74) would give a hirability rating that is more than one point (on the 1-6 scale) higher than a rater who provides a rating that is one standard deviation below the mean (3.42). This is potentially significant because this type of difference in rating could produce different outcomes for candidates. Thus, the current findings suggest that not all resume raters tend to make similar hirability judgments from resumes, even though this is often assumed to be the case.

The accuracy scores for resume personality ratings also demonstrated modest, yet potentially meaningful variability. With the average accuracy of resume

personality ratings at 0.39, this suggests that most participants were rating resume holders' personality somewhat similarly to resume holders' own personality ratings. In other words, participants were able to detect and somewhat accurately rate a resume holder's personality from the resume. However, the standard deviation of accuracy was 0.16, meaning it was not unusual for accuracy to be as low as 0.23 (one standard deviation below the mean) or as high as 0.55 (one standard deviation above the mean). Consistent with this, the 25th percentile value (0.28) and the 75th percentile value (0.52) almost match one standard deviation above and below the mean. This suggests there was nontrivial variability in accuracy as well. Additionally, the minimum accuracy score found was -0.14 suggesting that at least some participants were unable to accurately detect personality from the resumes. Note, however, that none of the accuracy scores were substantially negative (i.e., no one scored lower than -0.14), indicating that none of the participants judged the personality of the resume holder to be the opposite of what it was. In sum, these findings support prior research examining personality ratings from resumes (see Cole, Feild, & Giles, 2003; Cole et al., 2009) and further add to the evidence that not only can resumes elicit applicant personality, but also that raters can often detect the applicant's personality.

The hypotheses specifically addressed rater individual differences as a potential source of variability in ratings. For resume hirability ratings, findings from this study showed that individual differences such as gender, conscientiousness, agreeableness, and neuroticism did not significantly relate to a tendency to rate resumes more favorably or less favorably. This differs from prior rating research where all of those individual differences have been found to influence rating tendency. It could be that resumes, though popularly used in practice, may provide limited information regarding a candidate thereby leading resume raters

to make inconsistent judgments from resumes. It could also be that the sample gathered for this study from MTurk was biased and not entirely representative of those individuals who have screened resumes for a current or prior job. For example, only 20.2% of the sample held a position as a recruiter or human resources representative/professional while the remaining 79.8% of the sample held the position of hiring manager or other, meaning that a majority of the sample was not a recruiter nor in human resources. Future research should investigate the influence of these individual differences in a more representative sample aligned with recruiters or human resources representatives.

The next set of hypotheses examined rater individual differences as a potential source of variability in the accuracy of resume personality ratings. Rater cognitive ability, conscientiousness, openness to experience, and experience rating resumes all did not significantly relate to personality rating accuracy. The only individual difference found to relate to the accuracy of resume personality ratings was dispositional intelligence. This finding is significant because prior research on dispositional intelligence has mostly found dispositional intelligence to influence personality rating accuracy from interpersonal contexts (i.e., interviews; see Christiansen et al., 2005). Additionally, other research, and even Funder's (1995; 1999) own work on personality judgments, has largely implied and supported the personality judgment process from interpersonal settings only. Therefore, the current finding adds support to the few prior studies that have shown that personality judgments can be made and made accurately without interpersonal interactions (see Merbedone, 2012, Nestler & Back, 2013, Küfner, et al., 2010).

One reason that findings from the present study were largely non-significant could be related to the aggregation of the four resume ratings into a composite. Aggregating across four resume ratings may have muddied the variation that would have been represented in each resume. For example, if one resume was viewed negatively by all participants, then including this in the aggregation of all the ratings could have influenced the results. To examine this further, post-hoc exploratory analyses were conducted for each hypothesis on each resume individually. Results from these analyses, however, showed the same findings: all individual differences examined in this study, with the exception of dispositional intelligence, did not significantly relate to resume rating tendencies.

Limitations

There are several limitations that should be noted for the present study. First, due to the use of a cross-sectional and correlational research design, causation cannot be determined for the dispositional intelligence and accuracy relationship. Second, the main study survey was relatively lengthy, and it took participants approximately 24 minutes to complete. Thus, it is possible that survey fatigue might have occurred, which could have caused a distortion in participant responses. However, attempts were taken to mitigate the survey length. Before the present study was launched, a pilot study was conducted using the long-form of the target individual difference constructs (i.e., personality, cognitive ability, and dispositional intelligence) and having eight resumes comprise the resume rating activity. Results and feedback from the pilot showed that survey length was a major constraint to participants, with the average time taken to complete the pilot survey around 55 minutes. Therefore, using relatively short scales to measure the targeted individual difference constructs as well as using only four resumes for the resume rating activity dramatically reduced the survey length and task complexity. On the other hand, however, reducing to four resumes may have limited the findings because four resumes may not have been enough to detect rating tendencies. Third, another limitation to this study could have been the way accuracy was operationalized, namely as self-other agreement. As indicated by Cronbach (1955), there are different ways to measure and interpret accuracy of social perception and caution should be taken when using only one method for measuring accuracy. It has been argued that accuracy should be partitioned and measured in different components; thus, only using one method to measure accuracy as in this study may limit the findings (Cronbach, 1955; Kenny & Albright, 1987). Finally, several limitations associated with the use of MTurk

workers should be noted as well. One primary concern of collecting research participants through MTurk is ensuring the right population is targeted. For example, the present study required the use of a screener; that is, the targeted sample needed for this study included individuals who have had to rate and/or evaluate resumes for their current or previous job(s). In MTurk, it is not possible to ensure that the participants are responding accurately to the screener. To mitigate this, however, the use of a separate screener and distractor question survey was used where participants took a brief, 6-item survey containing distractor questions (e.g., "Did you buy a house or car in the past two years?") along with the screener. Participants who passed the screener question were then invited to take the present study's survey. Another drawback of using MTurk workers is the lack of motivation and attention that they generally have when completing survey research studies (Goodman, Cryder, & Cheema, 2013). However, the use of two attention checks, one in the beginning of the survey and one near the end, in the present study may have mitigated this limitation. More specifically, any MTurk worker who did not pass both attention checks was eliminated from analyses.

Future Directions

In addition to addressing current study limitations, future research might also examine other issues. For example, the analysis of variability showed modest variation in resume hirability ratings. Future research might investigate this variability further or look for other factors, such as biases or similarity-attraction effects, that may account for that variability. Additionally, findings suggested that on average participants in this sample were somewhat accurate in their ratings of the personality of the resume holder. This finding supports future research investigating resume personality ratings because it shows that personalities can be detected accurately from resumes. Along similar lines, most research, including the present study, has examined resume personality ratings using student resumes. Future research should thus investigate resume personality ratings using non-student resumes. Another consideration for future research that pertains to the use of resumes is the notion of variability within the resume itself. More specifically, when it comes to job candidate profiles in consideration for job suitability or hirability, research has shown that raters tend to negatively rate candidate profiles exhibiting more variability (Fox, Bizman, & Oren, 1995). That is, candidate profiles that exude variable traits and characteristics of the candidate can lead raters to rate the candidate more negatively as the rater perceives variable candidate profiles as less stable. While the variability of each resume/ratee candidate within this study was not examined, future research should investigate and examine profile variability to account for the effects of that variability in job hirability ratings.

References

Anderson, N., & Shackleton, V. (1990). Staff selection decision making into the 1990s. *Management Decision*.

Arnulf, J. K., Tegner, L., & Larssen, Ø. (2010). Impression making by résumé layout: Its impact on the probability of being shortlisted. *European Journal of Work and Organizational Psychology*, *19*(2), 221-230.

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology*, 25, 251– 270.

Aspers, C. & Derous, E. (2017). Are they accurate? Recruiters' personality judgments in paper versus video resumes. *Computers in Human Behavior, 73*, 9-19.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85*(2), 232-236.

Borman, W. C. (1979). Individual differences correlates of accuracy in evaluating others' performance effectiveness. *Applied Psychological Measurement, 3*, 103-115.

Bright, J. E., & Hutton, S. (2000). The impact of competency statements on résumés for short-listing decisions. *International Journal of Selection and Assessment*, *8*(2), 41-53.

Brown, B. K., & Campion, M. A. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology*, *79*(6), 897.

Burns, G. N. (2004). *Personality Judgments Based on Résumés: Cue Utilization, Validity, and Accuracy of Observer Ratings* (Doctoral dissertation, Central Michigan University).

Cable, D. M. & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82*(4), 546-561.

Camp, R. R., Tanguay, D. M., Endres, M. L., & Torres, E. M. (2014). The impact of raters' level and type of sports experiences on inferences drawn from resumes. Journal of Managerial Issues, 26(4), 388-407.

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance, 18*(2), 123-149.

Cole, M. S., Feild, H. S., & Giles, W. F. (2003). What can we uncover about applicants based on their resumes? A field study. *Applied HRM Research*, *8*(2), 51-62.

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2004). Job type and recruiters' inferences of applicant personality drawn from resume biodata: Their relationships with hiring recommendations. *International Journal of Selection and Assessment, 12*(4), 363-367.

Cole, M. S., Rubin, R. S., Feild, H. S., & Giles, W. F. (2007). Recruiters' perceptions and use of applicant résumé information: Screening the recent graduate. Applied Psychology: An International Review, 56(2), 319-343.

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2009). Recruiters' inferences of applicant personality based on resume screening: Do paper people have a personality? Journal of Business and Psychology, 24(1), 5-18.

Costa, P. T., & McCrae, R. R. (1992). *The revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52*(3), 177-193.

Donnellan, M. B., 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*(2), 192-203.

Finn, Y., Cantillon, P., & Flaherty, G. (2014). Exploration of a possible relationship between examiner stringency and personality factors in clinical assessments: a pilot study. *BMC Medical Education, 14*(1), 1052.

Fox, S., Bizman, A., Hoffman, M., & Oren, L. (1995). The impact of variability in candidate profiles on rater confidence and judgements regarding stability and job suitability. *Journal of Occupational and Organizational Psychology, 68*, 13-23.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*(4), 652-670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Glassdoor. (2015, January 20). *50 HR and Recruiting Stats That Make You Think.* Glassdoor for Employers. https://www.glassdoor.com/employers/blog/50-hr-recruiting-stats-make-think/

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26-42.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making, 26*, 213-224.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13-23.

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science, 3*(2), 143-152.

Guion, R. M. (2011). Judgment In Assessment and Prediction. In *Assessment, Measurement, and Prediction for Personnel Decisions* (pp. 429-466). Routledge.

Hakel, M. D., Dobmeyer, T. W., & Dunnette, M. D. (1970). Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. *Journal of Applied Psychology*, *54*(1p1), 65.

Harris, P. B., Boccaccini, M. T., & Murrie, D. C. (2015). Rater differences in psychopathy measure scoring and predictive validity. *Law and Human Behavior, 39*(4), 321-331.

Harvey-Cook, J. E., & Taffler, R. J. (2000). Biodata in professional entry-level selection: Statistical scoring of common format applications. *Journal of occupational and organizational psychology*, *73*(1), 103-118.

Hauenstein, N. M., & Alexander, R. A. (1991). Rating ability in performance judgments: The joint influence of implicit theories and intelligence. *Organizational Behavior and Human Decision Processes, 50*(2), 300-323.

Hebberd, L. (2013, February 16). *The 500 year history of the resume*. Undercover Recruiter. https://theundercoverrecruiter.com/history-resume-info/

Hitt, M. A., & Barr, S. H. (1989). Managerial selection decision models: Examination of configural cue processing. *Journal of Applied Psychology, 74*(1), 53-61.

The International Cognitive Ability Resource Team. (2014). *International Cognitive Ability Resource*. https://icar-project.com

International Personality Inventory Pool. (n.d.) *Big-Five Factor Markers*. International Personality Inventory Pool. https://ipip.ori.org/newBigFive5broadKey.htm

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*(3), 404–426.

Kaufmann, E., Reips, U. D., & Wittmann, W. W. (2013). A critical meta-analysis of lens model studies in human judgment and decision-making. *PloS one*, *8*(12).

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27-44.

Kennealy, P. J., Skeem, J. L., & Hernandez, I. R. (2017). Does staff see what experts see? accuracy of front line staff in scoring juveniles' risk factors. *Psychological Assessment, 29*(1), 26-34.

Kenny, D. A. & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102*(3), 390-402.

Kirkegaard, E. O. & Bjerrekaer, J. D. (2016). ICAR5: Design and validation of a 5-item public domain cognitive ability test. *Open Differential Psychology.*

Knouse, S. B. (1989). The role of attribution theory in personnel employment selection: A review of the recent literature. Journal of General Psychology, 116(2), 183-196.

Knouse, S. B. (1994). Impressions of the resume: The effects of applicant education, experience, and impression management. *Journal of Business and Psychology*, *9*(1), 33-45.

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

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

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of research in personality*, *42*(4), 914-932.

Lodato, M. A. (2008). *Going with your gut: An investigation of why managers prefer intuitive employee selection* (Doctoral dissertation, Bowling Green State University). Retrieved from https://search-proquest-com.portal.lib.fit.edu/docview/621741890?accountid=27313

London, M., & Poplawski, J. R. (1976). Effects of information on stereotype development in performance appraisal and interview contexts. *Journal of Applied Psychology, 61*(2), 199-205.

McCrae, R. R., & Costa, P. T., Jr. (1989). More reasons to adopt the five-factor model. *American Psychologist, 44*(2), 451-452.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of Personality, 60(2), 175-215.

McManus, I. C., Thompson, M., & Mollon, J. (2006). Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP (UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education, 6*(1), 42.

Merbedone, M. A. (2012). *Personality perceptions based on social networking sites: An examination of the effect of information quantity and individual differences on judgmental accuracy.* (Doctoral dissertation, Florida Institute of Technology).

Miller, A. K., Rufino, K. A., Boccaccini, M. T., Jackson, R. L., & Murrie, D. C. (2011). On individual differences in person perception: Raters' personality traits relate to their psychopathy checklist-revised scoring tendencies. *Assessment, 18*(2), 253-260.

Murrie, D. C., & Warren, J. I. (2005). Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Professional Psychology: Research and Practice, 36*(5), 519-524.

Murrie, D. C., Boccaccini, M. T., Zapf, P. A., Warren, J. I., & Henderson, C. E. (2008). Clinician variation in findings of competence to stand trial. *Psychology, Public Policy, and Law, 14*(3), 177-193.

Nemanick, Jr, R. C., & Clark, E. M. (2002). The differential effects of extracurricular activities on attributions in resume evaluation. *International Journal of Selection and Assessment*, *10*(3), 206-217.

Nestler, S., & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science, 22*(5), 374-379.

Oliphant, V. N., & Alexander, E. R. (1982). Reactions to resumes as a function of resume determinateness, applicant characteristics, and sex of raters. Personnel Psychology, 35(4), 829-842.

Powell, D. M., & Bourdage, J. S. (2016). The detection of personality traits in employment interviews: Can "good judges" be trained? *Personality and Individual Differences, 94*, 194-199.

Risavy, S. D. (2017). The resume research literature: where have we been and where should we go next?. *Journal of Educational and Developmental Psychology*, *7*(1), 169.

Rogers, K. H., & Biesanz, J. C. (2019). Reassessing the good judge of personality. *Journal of personality and social psychology*, *117*(1), 186.

Rose, G. L., & Andiappan, P. (1978). Sex effects on managerial hiring decisions. *Academy of Management Journal, 21*(1), 104-112.

Ross, C. M., & Young, S. J. (2005). Resume preferences: Is it really "business as usual"?. *Journal of Career Development*, *32*(2), 153-164.

Rynes, S. L., Trank, C. Q., Lawson, A. M., & Ilies, R. (2003). Behavioral coursework in business education: Growing evidence of a legitimacy crisis. Academy of Management Learning & Education, 2(3), 269-283.

Sabzwari, S. R., Pinjani, S., & Nanji, K. (2018). Effect of faculty personality, rating styles, and learner traits on student assessment in medical education: A mixed-method study from the aga khan university, karachi. *Education for Health: Change in Learning & Practice, 31*(2), 103-108.

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*(2), 262.

Shafer, A. B. (1999). Relation of the Big Five and factor v subcomponents to social intelligence. *European Journal of Personality, 13*, 225–240.

Singer, M. S., & Bruhns, C. (1991). Relative effect of applicant work experience and academic qualification on selection interview decisions: A study of between-sample generalizability. *Journal of Applied Psychology, 76*, 550–559.

Smither, J. W., & Reilly, R. R. (1987). True intercorrelations among job components, time delays in rating, and rater intelligence as determinants of accuracy in performance ratings. *Organizational Behavior and Human Decision Processes, 40*, 369-391.

Tanguay, D. M., Camp, R. R., Endres, M. L., & Torres, E. (2012). The impact of sports participation and gender on inferences drawn from resumes. *Journal of Managerial Issues, 24*(2), 191-206.

Thoms, P., McMasters, R., Roberts, M. R., & Dombkowski, D. A. (1999). Resume characteristics as predictors of an invitation to interview. *Journal of Business and Psychology, 13*(3), 339-356.

Unsal, P., & Caliskur, A. (2004). The effect of selectors' neuroticism on job application outcomes. *Personality and Individual Differences, 37*(6), 1313-1318.

Wilkin, C. L. & Connelly, C. E., (2012). Do I look like someone who cares? Recruiters' ratings of applicants' paid and volunteer experience. *International Journal of Selection and Assessment, 20*(3), 308-318.