EXAMINING FACTORS INFLUENCING USE OF A DECISION AID IN PERSONNEL SELECTION

by

ALEXANDER THOMAS JACKSON

B.S., Oklahoma State University, 2010
M.A., The University of Tulsa, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

In this research, two studies were conducted to examine the factors influencing reliance on a decision aid in personnel selection decisions. Specifically, this study examined the effect of feedback, the validity of selection predictors, and the presence of a decision aid on the use of the decision aid in personnel selection decisions. The results of both studies demonstrate that when people are provided with the decision aid, their predictions were significantly more similar to (but not the same as) the predictions made by the aid than people who were not provided with the decision aid. This suggests that when people are provided with an aid, they will use it at least to some degree. This research also shows that when provided with a decision aid that has high validity, people will increase their reliance on the decision aid over multiple decisions. Finally, this research shows that, in general, there are individual differences that influence how participants weight the different selection predictors.

EXAMINING FACTORS INFLUENCING USE OF A DECISION AID IN PERSONNEL SELECTION

by

ALEXANDER THOMAS JACKSON

B.S., Oklahoma State University, 2010
M.A., The University of Tulsa, 2012

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Patrick A. Knight

Abstract

In this research, two studies were conducted to examine the factors influencing reliance on a decision aid in personnel selection decisions. Specifically, this study examined the effect of feedback, the validity of selection predictors, and the presence of a decision aid on the use of the decision aid in personnel selection decisions. The results of both studies demonstrate that when people are provided with the decision aid, their predictions were significantly more similar to (but not the same as) the predictions made by the aid than people who were not provided with the decision aid. This suggests that when people are provided with an aid, they will use it at least to some degree. This research also shows that when provided with a decision aid that has high validity, people will increase their reliance on the decision aid over multiple decisions. Finally, this research shows that, in general, there are individual differences that influence how participants weight the different selection predictors.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I will try to keep this short and sweet, like the acknowledgements in award shows. First and foremost, I would like to thank Dr. Satoris "Tori" Culbertson, for constantly pushing me to do more and better myself as a student, as a researcher, as an instructor, and even a runner. I would also like to thank her for lending me the book *Judgment and Decision Making at Work.* It is this book that piqued my interest in the present topic. Without that book, I would have likely escalated my own commitment to my escalation of commitment research. Tori has always gone out of her way to help me become a better academic and a better person. Tori for everything you have done to help me, thank you.

I would also like to thank Dr. Patrick Knight. Although he was not my original advisor when I arrive at K-State, Dr. Knight took me under his wing and provided guidance throughout my doctoral studies. His quick feedback has been immensely helpful in terms of shaping the direction and improving the rigorousness of my research as well as in terms of improving my writing. I would also like to thank Dr. Knight for providing the funding I needed to pay participants in this dissertation. This funding enabled me to obtain better samples than I would have obtained without it.

I must also thank Dr. Michael "Mike" Young. He has imbued me with the beginnings of statistical expertise and has helped develop my statistical capabilities throughout the entirety of my tenure at K-State. He has also been immensely patient with me. I have spent many hours in his office discussing research methodology, statistical analyses, job applications, job offers and negotiations, and job decisions. While at times Mike may have had a lot on his plate with the duties of being the department head, he has always took the time to provide guidance in any way he could.

I would also like to thank the remaining members of my dissertation committee, Drs. Gary Brase and Junehee Kwon, for their helpful comments and suggestions on the various stages of my dissertation.

Of course, I would also like to thank my fellow students and friends, (listed alphabetically and in no order of importance – you are all important to me) Lynn & Amy Beer, Aaron Entringer, Lesly Krome, Tony McCoy, Lisa Vangsness, and Steve Van Krevelen, for their feedback, support, and friendship. I would also like to thank you all for the comments, suggestions, and feedback you have provided throughout the various stages of my dissertation. You have helped me become a better researcher, a better writer, and a better friend.

I would not be at writing this without the support of my family, so I would like to thank my dad and mom, Randy and Patricia Jackson, for their support throughout all of my schooling. I want to thank my mom for urging me to be more patient throughout this process. I want to thank my dad for always wanting to be engaged in my research. Anytime I had a major research presentation or a new journal article, you wanted me to send it to you, so you could read it. Wanting to be engaged, including wanting to actually read my dissertation, shows how proud you are of me. Thank you for that. I would also like to thank my father and mother-in-law, Chuck and Cindy Browne, for their support and encouragement throughout my graduate school career. The two of you not only welcomed me with open arms into your family while I was in the middle of graduate school, but you also provided continuous support while I finished working toward my doctorate. Without the support and encouragement that the two of you have provided, this process may have been much more stressful and taken much longer. For this, I thank you.

And finally, my acknowledgements would be incomplete without thanking my beautiful and wonderful wife Laura. I would like to thank her for all of her patience and continuous support. She has endured numerous nights alone while I was neck deep in reading, writing, and analyses. On long road trips, she burdened herself with the driving, so I could spend my time writing. It is her patience and continuous support that has helped keep me moving forward. I would also like to thank her for pushing me. She pushed me to be a better student. She pushed me to be a better researcher. She pushed me to be a better instructor. She pushed me to finish this dissertation. She also pushed me to be a better husband and a better person. To my wife Laura, thank you, and I love you.

I would also like to thank anyone I may have missed that has helped me in some way to finish my dissertation. If I missed you, please do not take it personally. I am simply trying to stick to my original intentions and keep this short and sweet. Queue the award show time limit music…

Chapter 1 - Literature Review

Introduction

Researchers have consistently shown that people are hesitant to rely on decisions aids, especially in the form of a statistical model, when making predictions or decisions (Arkes, Dawes, & Christensen, 1986; Ashton, 1990; Boatsman, Mockel, & Pei, 1997; Diab, Pui, Yankelevich, & Highhouse, 2011; Eastwood, Snook, & Luther, 2012; Fildes & Goodwin, 2007; Goodwin, Fildes, Lawrence, & Nikolopoulos, 2007; Sanders & Manrodt, 1994, 2003). In discussing the use, misuse, disuse, and abuse of decision aids, Parasuraman (1997) identified a variety of factors that influence one's decision to use a decision aid. Specifically, factors such as trust in the aid, task complexity, confidence, workload, decision aid accuracy, skill and risk all interact to influence one's use of a decision aid. Parasuraman also argued that when new decision aids are introduced, people may distrust the decision aid and be resistant to accept and use it. For instance, in a series of studies, Dietvorst, Simmons, and Masey (2015a) demonstrated that when people see a decision aid make a mistake, they distrust it. This is known as algorithm aversion, in which people are less likely to rely on a decision aid and rely more on their own decision processes.

The purpose of this research is to examine the factors influencing an individual's reliance on a decision aid in a personnel selection context. Specifically, I examined whether the presence of a decision aid (e.g., a statistical model), the validity of selection predictors, and feedback regarding one's performance predictions impacts the reliance on a decision aid. Study 1 focused on examining the effect the presence of the decision aid and the validity of the selection predictors on decision aid reliance. In study 1, participants made ten hiring decisions based on data from a real hiring situation. Study 2 aimed to replicate study 1 and extend it by examining the effect of feedback as well as increasing the number of hiring decisions to twenty.

Literature Review

Personnel Selection

Personnel selection is one of the primary functions of human resource management within organizations and refers to "the process employers use to make decisions concerning which individuals from a group to choose for particular jobs or roles within the organization" (Farr & Tippins, 2010, p. 1). At the broadest level, the primary goal of any personnel selection system is to assess job candidates on the physical and psychological attributes (knowledge, skills, abilities, and other characteristics) that have been identified as being required to perform the job in order to identify individuals who will have better performance and improve organizational effectiveness and efficiency (Farr & Tippins, 2010; Robertson & Smith, 2001). According to the *Principles for the Validation and Use of Personnel Selection Procedures*, validity is the most important aspect to consider when developing, evaluating, and utilizing the selection test (Society for Industrial and Organizational Psychology, 2003) because validity refers to the degree to which specific interpretations and uses of test scores are supported by evidence and theory (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 1999). In other words, validity refers to the ability of a selection test to predict an applicant's future job performance, as well as other relevant criteria (Schmidt & Hunter, 1998).

Effectiveness of Selection Methods

Because the ability of a test to predict future performance is paramount to personnel selection, researchers have spent decades investigating the predictive validity of various constructs as well as assessment methods. This has allowed researchers to conduct meta-analytic studies examining the overall effectiveness of various selection tests, such as cognitive ability tests (e.g., Schmidt & Hunter, 1998, 2004), personality tests (e.g., Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991), and aptitude tests (e.g., Schmidt & Hunter, 1998; Vinchur, Schippmann, Switzer III, & Roth, 1998). Furthermore, these meta-analytic studies have enabled researchers to investigate whether some testing methods are more or less effective than others, such as unstructured versus structured interviews (e.g., Huffcut & Arthur, 1994). From these meta-analytic studies, several conclusions can be made regarding the overall effectiveness of various predictors in selection systems. First, some of the best predictors of job performance are specific aptitude tests, work sample tests, such as a sales ability test, and general cognitive ability tests (Schmidt & Hunter, 1998, 2004; Vinchur et al., 1998). For example, Hunter and Hunter (1984) found that work sample tests, which traditionally measure either motor or verbal skills (Asher & Sciarrino, 1974), have an average validity of .54, and cognitive ability tests have an average validity of .53 when predicting job performance. More recent meta-analytic work shows that the operational validity of work sample tests tends to be slightly lower than originally thought with coefficients ranging from .28 to .50 (Roth, Bobko, & McFarland, 2005). Similarly, specific aptitude tests have operational validities ranging from .35 to .50, depending on the specific aptitude being measure (Bertua, Anderson, & Salgado, 2005). On the other hand, general cognitive ability has been repeatedly shown to be one of the best predictors of job performance with operational validity coefficients ranging from .31 to .73 (Schmidt & Hunter, 2004). Therefore, when all else is equal, it appears that cognitive ability is the best overall predictor of job performance.

Second, the operational validities (i.e., the corrected coefficients) of specific personality characteristics can predict job performance much better than previously thought (Barrick & Mount, 1991; Tett et al., 1991) and in some cases nearly as well as cognitive ability. This is especially true when one considers the relevance between specific traits and specific situations (Tett & Guterman, 2000) as well as the effect of bidirectionality on meta-analytic findings (Tett, Jackson, Rothstein, & Reddon, 1999). For example, Barrick, Mount, and Judge (2001) demonstrated that conscientiousness has an operational validity of .31 when predicting supervisor ratings of job performance. However, the 90% credibility interval ranges from .11 to .40, meaning that in some situations, the relationship between conscientiousness and supervisor ratings of job performance is near zero, while in other situations, the relationship is almost as strong as the relationship between cognitive ability and job performance. In contrast, the operational validities for the remaining four factors of the five-factor model were found to be smaller (.13 for extraversion, .13 for emotional stability, .13 for agreeableness, and .07 for openness to experience). Accordingly, in the context of the five-factor model of personality, conscientiousness is the best overall predictor of supervisor ratings of job performance when all else is equal.

Third, and perhaps one of the most important meta-analytic findings about personnel selection predictors, structured interviews are far superior to unstructured interviews when predicting job performance (Huffcutt & Arthur Jr., 1994; Huffcutt, Culbertson, & Weyhrauch, 2014; Huffcutt & Culbertson, 2011). Specifically, the most structured interview format has an operational validity of .70, while the least structured interview format has an operational validity of .20 (Huffcutt et al., 2014).

The fourth and final conclusion that can be made from the meta-analytic evidence is that using multiple valid predictors can improve predictions of job performance. For example, Schmidt and Hunter (1998) demonstrated that simply adding conscientiousness to a measure of cognitive ability increases the predictive validity by 18% (increase from r = .51 to R = .60). Similarly, adding a structured interview to a test of cognitive ability increases the predictive validity by 24% (increase from r = .51 to R = .63). Surprisingly, even adding an unstructured interview increases the predictive validity by 8% (increase from r = .51 to R = .55)[1].

In summary, through meta-analytic studies researchers have demonstrated that when all other factors are equal cognitive ability and conscientiousness are two of the better predictors of job performance. Similarly, structured interviews are far superior to unstructured interviews, and using multiple selection predictors improves the overall validity of selection methods. Unfortunately, human resource managers have misperceptions regarding the effectiveness of various selection methods.

Perceived Effectiveness of Selection Methods

While research has clearly demonstrated the effectiveness of various selection predictors, human resource managers tend to have misperceptions regarding the effectiveness of the various selection methods (Highhouse, 2008; Terpstra, 1996). For example, Terpstra asked 201 human resource professionals to rate various selection methods in terms of how well they predict future job performance. He found that the unstructured interview was perceived to be more effective

---

[1] It is worth noting that while the addition of the interview can increase the predictive validity of a selection system, the interview is a method, not a construct. This distinction is important because the interview is a tool that can be used to assess a variety of different constructs, including cognitive ability. Furthermore, confusion about the distinction between methods and constructs can lead inappropriate conclusions regarding selection tests and constructs. For a detailed argument regarding the importance of the distinction between constructs and methods, see Arthur and Villado (2008).

than structured interviews, assessment centers, specific aptitude tests, personality tests, general cognitive ability tests, and biographical information. Perhaps most surprising about Terpstra's findings is that general cognitive ability tests were rated as the second *worst* selection method (behind biographical data) for predicting job performance. The perceptions of these human resource professionals do not match the reality of the actual effectiveness of these selection predictors. For example, when predicting performance ratings among salespeople, biographical data and general cognitive ability have been meta-analytically shown to be among the *best* predictors of performance (Vinchur et al., 1998), whereas the unstructured interview has been shown to be one of the poorer predictors of job performance (Huffcut & Arthur, 1994; Huffcutt & Culbertson, 2011). Figure 1 illustrates these misperceptions.

Why Managers Have Misperceptions

It is clear that human resource professionals have large misperceptions regarding the effectiveness of a variety of selection methods. These misperceptions arise for several reasons. First, human resource managers may lack knowledge regarding the effectiveness of these predictors, or they may simply be unfamiliar with the predictors altogether (Terpstra & Rozell, 1997). Second, human resource professionals may reject the research regarding the effectiveness of selection methods because they may not believe that the research is relevant to their own hiring situations (Terpstra & Rozell, 1997). For example, meta-analyses and individual studies regarding the effectiveness of selection methods have been conducted for a variety of job contexts, such as professionals, police, managers, salespeople, and skilled workers (Barrick & Mount, 1991), firefighters (Barrett, Polomsky, & Mcdaniel, 1999), administrative employees (Morgeson, Delaney-Klinger, & Hemingway, 2005), and physicians (Lievens & Sackett, 2012). However, if a human resource professional is making hiring decisions for janitorial workers, they may reject the research because it does not directly assess the effectiveness of the predictors in the context of janitorial work. As such, hiring managers may be aware of the research findings regarding selection predictors, yet they may not be fully convinced by the findings (Colbert, Rynes, & Brown, 2005). A third reason why human resource professionals may have misperceptions about the effectiveness of selection predictors is that they believe that good hiring is simply a matter of using and relying on one's intuition (Highhouse, 2008). As such, it is not surprising that hiring managers tend to over-rely on their own decision-making processes, such as intuition, when making hiring decisions, instead of relying on the validated selection methods.

While managers tend to over-rely on their own decision processes to make hiring decisions, the obvious next question is: Why? One possible reason is that managers believe it is possible to achieve perfect prediction. Einhorn (1986) argued that clinical approaches to judgment and decision making rest on the lofty goal of perfect predictability. However, because statistical models actually accept error as a fundamental assumption and characteristic, they tend to produce less error in prediction. Similarly, Highhouse (2008) argued that managers have specific implicit beliefs about personnel selection, one of which is that it is possible to achieve perfect prediction regarding job performance. In other words, managers believe it is possible to predict with 100% accuracy an applicant's job performance based on a handful of selected predictors (e.g., cognitive ability, personality, and a work sample). Indeed, Dawes (1979) claimed that one implicit assumption of predicting graduate school performance is that the criteria of interest are highly predictable. As such, people believe that if one method of prediction, such as a statistical model, fails to achieve perfect prediction, another (human intuition) may more closely achieve perfect prediction. However, the evidence fails to support these notions.

Similar to the assumption that it is possible to achieve perfect prediction, people believe that intuitive expertise for predicting human behavior actually exists and that they can become skilled at making intuitive predictions about performance. Highhouse (2008) called this the "Myth of Expertise." This myth is demonstrated in the wide prevalence of clinical (intuitive) approaches to managerial assessment (Highhouse, 2002). However, the preponderance of evidence highlights that this belief is actually a myth and the experience does not yield better intuitive predictions (Dawes, Faust, & Meehl, 1989; Grove, Zald, Lebow, Snitz, & Nelson, 2000). Interestingly, in some situations, such as financial forecasting, the performance of novices can match, and sometimes exceed, the performance of experts (e.g., Armstrong, 1980; Yates, McDaniel, & Brown, 1991) providing further support that the myth is truly a myth.

Intuition

A recent online book title search on Amazon.com using the term *intuition* produced 5,881 search results. Adding the term *business* to the search produces still 950 search results. Some of the results include *Practical Intuition*, *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work*, *The Invisible Gorilla: How Our Intuitions Deceive Us*, and *Trust Your Gut!: Practical Ways to Develop and Use Your Intuition for Business Success*. Clearly, the buying public shows interest in utilizing intuition, improving intuition and counteracting the potential negative consequences of intuition.

This interest is mirrored among organizational scholars. For example, Highhouse (2008) began an academic discussion regarding the managerial overreliance on intuition in personnel selection decisions. This article produced 10 commentaries discussing the issue (Chorągwicka &

Janta, 2008; Colarelli & Thompson, 2008; Fisher, 2008; Klimoski & Jones, 2008; Kuncel, 2008; Martin, 2008; Mullins & Rogers, 2008; O'Brien, 2008; Phillips & Gully, 2008; Thayer, 2008). Furthermore, cognitive science and judgment and decision-making researchers have examined how people *should* make decisions and how people *actually* make decisions (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1981) showing that people often do rely on intuition to make decisions (e.g., Kahneman & Klein, 2009). For example, in personnel selection, research clearly demonstrates that hiring managers should utilize mechanical prediction (statistical prediction based on validated selection tools) in making hiring decisions (Highhouse, 2008; Schmidt & Hunter, 1998). However, managers tend to ignore information regarding the statistically best prediction regarding who should be hired and over-rely on their intuitive predictions when making selection decisions (Highhouse, 2008; Slaughter & Kausel, 2014). Thus, it comes as no surprise that researchers have often called for research examining ways to train hiring managers to utilize valid predictors of job performance instead of relying on intuition when making selection decisions. For example, Slaughter and Kausel (2014) recommended that researchers provide direct feedback regarding the accuracy of one's decision as a possible method of demonstrating the fallibility of intuitive decision-making processes.

In the following sections, I discuss the ways in which intuition has been defined, dual information processing theory, the role of intuition in prediction, prediction under uncertainty, decision aid use, and the role of feedback in intuitive decision making.

Intuition Defined

As Dane and Pratt (2007) discuss, intuition, especially in the organizational sciences, has been defined and operationalized in many different ways causing some confusion about what intuition is and how it operates. For example, Jung (1923) defined intuition as "that psychological function transmitting perceptions *in an unconscious way*" (emphasis in original, p. 567-568). Further, Jung argued that intuition serves an *irrational* psychological function whereby people can have perceptions of situations, objects, problems, and people without any knowledge or recollection of how the perception originated.

Similarly, Simon (1987) described intuition as the process of recognition. For instance, Simon described how expert chess players can play up to 50 opponents simultaneously because they are able recognize patterns in the layout of the pieces. This intuitive recognition enables grandmaster chess players to make rapid decisions regarding their next move. Interestingly, these expert players have no awareness of how they make their judgments regarding the moves. It seems clear that both Jung (1923) and Simon are describing some form of automatic information processing and judgment. However, Jung described intuition as a psychological function and Simon described intuition as the act of recognition.

More recently, Shapiro and Spence (1997) defined intuition as "a nonconscious, holistic processing mode in which judgments are made with no awareness of the rules of knowledge used for inference and can feel right despite one's inability to articulate the reason" (p. 64). As Dane and Pratt (2007) noted, many researchers have clearly defined intuition in their own intuitive way. Upon reviewing the intuition literature, Dane and Pratt (2007) recognized that many of the definitions converged in one way or another. As such, they argued that intuition has four major characteristics, and that intuition is a "(1) nonconscious process (2) involving holistic associations (3) that are produced rapidly, which (4) result in affectively charged judgments" (p. 36). Because the definition Dane and Pratt (2007) used was based on an extensive review of the literature, this is the definition adopted in this study.

Dual Information Processing Theory

According to the dual information processing theory of reasoning, cognitive processing occurs in one of two ways (Evans, 2003, 2010). The first method of information processing, referred to as system 1, occurs very rapidly and *intuitively* and has a large capacity (Evans, 2010). Furthermore, system 1 processes are holistic, automatic, unconscious and are less cognitively demanding (Stanovich & Toplak, 2012). As can be seen in their definition, intuition overlaps conceptually with system 1. Some have even described system 1 as the intuitive system (Evans, 2010). In contrast, system 2 is much more systematic. System 2 processes are slow, reflective, and have a low capacity. Further, these processes are analytic, controlled, cognitively demanding, and often conscious (Stanovich & Toplak, 2012). System 1 and system 2 processes are thought to be independent (Sinclair & Ashkanasy, 2005). In other words, an individual can use both systems simultaneously or one system in isolation when making a decision. Thus, in the context of personnel selection, individuals who rely on intuition to make predictions about future job performance and to make hiring decisions are relying on the automatic processing associated with system 1. In contrast, when individuals rely on a statistical model to make performance predictions and hiring decisions, these individuals are relying on the systematic processing associated with system 2. This study aims to investigate ways to shift information processing from system 1 to system 2 when making predictions about future performance of job candidates and making hiring decisions.

Prediction and Intuition

Because intuition involves judgments that are automatic, many researchers have been concerned with the accuracy of intuitive judgments. Some argue that, in appropriate situations, intuition leads to effective decision making. For example, when comparing managers' real-time predictions regarding coupon redemption and product sales to predictions made using statistical models, managers' predictions were as accurate as the predictions of the statistical models (Blattberg & Hoch, 1990). Similarly, expert engineers' intuitive judgments about the weight capacity of highways were more accurate than analytical judgments (Hammond, Hamm, Grassia, & Pearson, 1987).

However, in other situations, the reliance on intuition may be, as Jung (1923) stated, irrational and result in suboptimal decision outcomes. Dawes, Faust, and Meehl (1989) reviewed research examining the efficacy of clinical prediction (intuitive human prediction) compared to the efficacy of actuarial prediction (statistical prediction) and demonstrated that actuarial methods of prediction often predict human behavior better than intuitive methods of prediction. Interestingly, this occurred even when experts with greater amounts of information made the intuitive predictions. In a recent meta-analysis, Kuncel, Klieger, Connelly and Ones (2013) reviewed the literature comparing clinical and statistical prediction when making judgments about future job performance or academic performance. They demonstrated that using clinical, intuitive prediction resulted in a substantial loss in predictive validity when predicting an applicant's job performance or academic performance. Thus, while some research shows that intuition can be effective in some situations, the preponderance of evidence suggests that mechanical or statistical prediction often outperforms intuitive prediction (Grove et al., 2000).

In addition to showing that mechanical prediction is often more accurate than clinical prediction, researchers have shown that utilizing a statistical model based on an individual's decision strategies (i.e., a bootstrapped model) often outperforms the individual (Armstrong, 2001). While this finding may seem counterintuitive, the model of an individual's decision making strategy uses the decision cue weights in a consistent fashion, whereas the individual may unknowingly adjust the weights for each decision. The result is that the individual makes predictions using inconsistent weighting of the decision cues (Dawes, 1971). Because the bootstrapped model consistently applies the weighting to the decision cues, error in the prediction is minimized in the predicted values. In contrast, when an individual inconsistently applies his or her weighting of the decision cues, error in the prediction is not minimized. For example, Dawes (1979) demonstrated that even improper linear models that used equal weighting of predictors tend to have higher validity than the average validity of human judges.

Is Intuition Intuition?

To this point, the discussion surrounding intuition has treated intuition as though it is something that truly exists, is observable, and can be manipulated. However, with the present technological and methodological abilities, there is no true way of observing or manipulating intuition. Given the inherent non-cognitive nature of intuition and that intuition falls in the system 1 type of information processing, it is not possible to observe or measure intuition, yet many researchers try to do this ignoring the reality of intuition. For instance, existing scales designed to measure intuition tend to measure intuition as an ability or disposition (e.g., Intuitive Management Survey, Rational-Experiential Inventory, & the International Survey on Intuition) instead of actual use of intuition (Sinclair & Ashkanasy, 2005). When used, retrospective accounts of intuition enable participants to describe their own perceptions of how they made their decision (Dane & Pratt, 2009). However, retrospective measures are flawed in that if intuition exists as a non-cognitive process, an individual has no way of accessing information regarding their own intuition. In attempts to overcome this major measurement issue, other researchers have attempted to infer the use of intuition by using alternative measures of intuition. For instance, Glöckner (2009) recommended the incorporating the time to make a decision as an additional variable to consider for inferring the use of intuition because intuitive processing, by definition, is extremely fast, but so is guessing. Furthermore, when making inferences regarding intuitive processes, one must recognize that intuition is likely often confused with other decision-making processes. In the context of personnel selection, instead of relying on intuition or a decision aid, an individual may have their own "mental model" that they employ when making the decision. An individual may focus all of their attention on a single predictor, such as cognitive ability tests, instead of the combination of predictors. Similarly, an individual may employ a unit weighting strategy, whereby all of the predictors receive the same weight. An individual could also just rely on guessing. All of these different processes can be, and likely are, confused with intuition even among researchers. In fact, it appears as though researchers have lumped all decision processes that deviate from a decision aid or reliance on a statistical model as intuition (e.g., Highhouse, 2008). This is unfortunate for research focusing on intuition because it does not truly further our understanding of intuition as a construct or as a decision-making process.

Decision Aid Use

Researchers have consistently shown that people are hesitant to rely on decisions aids, especially in the form of a statistical model, when making predictions or decisions (Arkes et al., 1986; Ashton, 1990; Boatsman et al., 1997; Diab et al., 2011; Eastwood et al., 2012; Fildes & Goodwin, 2007; Goodwin et al., 2007; Sanders & Manrodt, 1994, 2003). For instance, Parasuraman argued that when new decision aids are introduced, people may be resistant to accept and use the aids, stating that people may dislike and mistrust the decision aid. In a series of studies, Dietvorst, Simmons, and Masey (2015a) demonstrated that when people see a decision aid err, such as a when a statistical model inevitably makes an imperfect prediction, people distrust the decision aid. In fact, Dietvorst and his colleagues found that people are more tolerant of their own *larger* errors than of the model's *smaller* errors. This results in algorithm aversion, in which people are less likely to rely on the model and rely more on their own decision processes.

Impact of Validity on Decision Aid Use

One reason why people are hesitant to rely on decision aids is that people believe that they are capable of perfect prediction (Highhouse, 2008). However, it is quite clear from the research that people, and for that matter statistical models, do not make perfect predictions, especially when predicting human behavior. The reason is simple: the outcomes being predicted (e.g., job performance) are uncertain. Furthermore, people cannot use all of the possible information to make perfect predictions. To do so would result in information overload. Because statistical models are developed by humans who input the information into the models, the models are limited by the information they contain. However, even in the event that a manager (or model) is able to predict employees' job performance with a high degree of accuracy, tragedy, such as equipment failure and injured employees, can still strike rendering future predictions inaccurate. Indeed, the "variance in [employee] success is simply not predictable prior to employment" (Highhouse, 2008, pp. 335–336). For instance, if a manager makes a prediction about an auto technician's performance, then the auto technician later sustains an injury to his hand, it is highly likely that the manager's prediction will be quite inaccurate. Therefore, when predicting human behavior, the outcomes are inevitably uncertain, and there is the guarantee of error in the prediction.

While predicting human behavior is an endeavor containing uncertainty, Agor (1986, 1991) demonstrated that uncertainty is a key factor influencing managerial reliance on intuition.

Specifically, he asked 200 executives how managers actually make decisions. He found that almost all of the executives stated that they use intuition to guide decision-making. Furthermore, the executives stated that they relied on intuition most heavily when a high level of uncertainty exists (Agor, 1986). Managers also relied on intuition when outcomes are less scientifically predictable, when information is limited, when the information available does not provide clear direction on how to proceed, when statistical data have limited utility, and when time pressures are greatest.

Additionally, researchers have demonstrated that the accuracy, or validity, of a decision aid influences the use of the decision aid. For instance, Gomaa, Hunton, Vaassen, and Carree (2011) directly manipulated the validity of a decision aid, such that the decision aid participants were presented with had an accuracy of 50%, 60%, 70%, 80%, or 90%. They found that the validity of the decision aid significantly impacted its use with more valid decision aids being used to a significantly greater extent. However, this study did not manipulate the presence of the decision aid. Across several studies Dietvorst et al. (2015a) manipulated participants' experience with a decision aid by providing the decision aid's previous forecasting performance, their own previous forecasting performance, previous forecasting performance for both the decision aid and their self, or no previous performance. As previously mentioned, Dietvorst et al. (2015a) showed that after viewing the forecasting performance of the decision aid people were less likely to use it because they were less tolerant of the decision aid's smaller errors than their own larger errors. Interestingly, the control condition (no previous performance) consistently utilized the decision aid most frequently suggesting that the mere option to use the decision aid led people to use the decision aid. It is only after receiving information about the performance of the decision aid that people distrust it. All of this information suggests that managers are most likely to rely on a decision aid when it has a higher level of validity. This leads to the following hypothesis:

Hypothesis 1: Participants' hiring choices and performance predictions will more closely match those made by the decision aid when the cues are more valid than when they are less valid.

Presence of a Decision Aid

However, while much of the research shows that people are averse to using decision aids and algorithms, this does not mean that people do *not* use them at all. In fact, in four of Dietvorst et al.'s (2015a) five studies, the control group that had no prior experience with the model used the model 54% - 69% of the time. Further, after seeing the results of human predictions, participants used the model 56% - 76% of the time. Indeed, in the absence of information about the model, participants are using the decision aid (i.e., the statistical model) a majority of the time. It is only when participants are provided with information about the inaccuracy of the model that they elect to use the model less. More convincingly, Dietvorst, Simmons, and Masey (2015b) demonstrated across a different series of studies that when people are allowed to adjust the predictions made by a statistical model, even if the adjustment is as small as two percentiles, people are more likely to use the model. In fact, after having the opportunity to make adjustments to the model's predictions in an initial set of forecasts, participants were more likely to elect to rely entirely on the model in a second set of forecasts than participants who could not adjust the predictions in the first set of forecasts. This suggests that people do utilize decision aids; they are just *underutilized.* As such, people would be expected to rely more on the model than their own decision strategy (i.e. intuition) when simply providing people with information about the statistical model and its predictions. Furthermore, Dietvorst et al. (2015b) convincingly demonstrated that when allowed to adjust the model, they are more likely to rely on the model. Thus, when no explicit restrictions are placed on how one uses the model (e.g., whether one can adjust the predictions or whether one must rely strictly on the model's predictions), people should be more likely to rely on the model; they just might adjust the predictions. This leads to the following hypothesis:

Hypothesis 2: Participants' hiring choice and performance predictions will more closely match the choice and performance predictions made by the decision aid when it is provided.

None of the studies discussed have examined the interactive effects of decision aid presence and the validity of predictors on the reliance on a decision aid. This study will close this gap in the literature by examining this interactive effect, and based on the literature reviewed, I propose the following hypothesis:

Hypothesis 3: The presence of the decision aid will interact with the validity of the cues, such that when the decision aid is present and the cues are more valid, participants' hiring choices and performance predictions will more closely match those made by the decision aid than in all other conditions.

Feedback

Despite the evidence showing the superiority of statistical prediction, managerial decision makers tend to over-rely on intuition when making personnel selection decisions (Highhouse, 2008; Slaughter & Kausel, 2014). Slaughter and Kausel (2014) offered several ways to improve personnel selection decisions. For example, they suggest that because managers may be hesitant to ignore their own intuitions and rely entirely on statistical predictions (e.g., Posthuma, Morgeson, & Campion, 2002), decision makers should be asked to make specific, numerical predictions about job performance. By making such predictions, decision makers can be provided with feedback regarding the decisions they made. Feedback regarding an individual's decision allows for the decision maker to become better calibrated to making personnel selection decisions and learn from the feedback.

Feedback regarding an individual's decision may actually serve as a vital source of information in calibrating one's decision strategies. Feedback regarding an individual's decision is often operationalized in terms of providing information to a decision maker regarding the accuracy (or inaccuracy) of one's decision (e.g., Louie, 1999; Tuttle & Stocks, 1997). Such feedback has been shown to have dramatic and meaningful influences on individuals' decision-making processes. For example, Louie (1999) demonstrated that individuals who receive positive feedback (i.e., the outcome was consistent with an individual's prediction) regarding their decision exhibit a strong hindsight bias, or the tendency to believe the outcome of an event was predictable after learning the outcome of the event (Roese & Vohs, 2012). Additionally, Brown (2006) demonstrated that decision feedback has an interactive effect on the relationship between outcome uncertainty and decision effectiveness. More specifically, when decision outcomes are less uncertain, decision feedback actually leads to decreases in the effectiveness of decision-making strategies. However, when decision outcomes are more uncertain, decision feedback leads to more effective decision-making.

Decision feedback can also influence the future decision-making strategies a person elects to employ. Using a repeated measures design, Wofford and Goodwin (1990) examined the effect of repeated positive feedback and repeated negative feedback on a person's decision-making strategies. When individuals were presented with repeated positive feedback regarding their decision-making strategies, decision strategies were maintained. However, when individuals were presented with repeated negative feedback regarding one's decision-making strategies, individuals tended to change their decision-making strategies and explore alternative strategies. In essence, the feedback was a form of operant conditioning whereby positive feedback reinforced a person's decision strategy and negative feedback punished a decision strategy. Because the negative feedback led to a change in the decision-making strategies individuals used, it would be expected that providing negative feedback in the form of information about the magnitude of one's errors would lead them to utilize different decision-making strategies.

Hypothesis 4: Participants' hiring choice and performance predictions will more closely match those made by the decision aid when negatively-framed feedback is provided regarding their predictions than when no feedback is provided.

Additionally, it is likely that feedback will interact with the cue validity. Specifically, when the cues have lower validity (the statistical model makes larger errors), people who receive feedback may be more likely to rely on their own decision-making processes. For instance, Arkes et al. (1986) examined the effect of different types of feedback on decision aid reliance when the decision aid was 70% accurate, which may be considered highly valid. They found that feedback type had a significant effect on decision aid reliance. However, the validity of the decision aid was not manipulated. Along these lines, Gomaa, Hunton, Vaassen, and Carree (2011) demonstrated that when a decision aid is more valid, people tend to utilize the decision aid to a greater extent. Conversely, when people receive feedback in the form of information about their own forecasting performance, information about a model's forecasting performance, or both, people were more likely to utilize a human forecaster than the model to make additional forecasts. Essentially, after observing a model make mistakes and receiving information about the magnitude of those mistakes, people lose trust in the model and instead rely on their own decision-making processes. Furthermore, researchers have directly examined the interactive effects of future uncertainty and feedback on optimal decision making strategies. Specifically, when provided with feedback when future outcomes were uncertain, people made less prudent decisions than when provided with feedback when the outcomes are not uncertain (Brown, 2006). This leads to the following hypothesis:

Hypothesis 5: The effect of feedback on decision aid reliance will depend on the validity of the cues, such that when the cues are more valid and feedback is provided, participants' hiring choices and performance predictions will more closely match those made by the decision aid than all other conditions.

Chapter 2 - Study 1

Purpose

The purpose study 1 was to examine hypotheses 1 – 3. In this study, participants were presented with applicant information for 10 pairs of job candidates. For each pair, participants were asked to estimate the performance for each candidate and select the candidate who should be hired.

Operationalization of Decision Aid Reliance

I utilized three operationalizations of decision aid reliance. Specifically, I operationalized decision aid reliance as the degree of match between a participant's hire choice and the model's hire choice, the degree of match between a participant's predicted performance and the model's predicted performance, and the participant's degree of cue calibration in regards to the cue weighting that should be employed to make an optimal decision. Because statistical models are empirically derived and represent the optimal weighting of predictors in a given situation, any deviation from the model's predictions would indicate the lack of reliance on the decision aid. Accordingly, the greater the degree of match in the hire choice, the greater the reliance on the decision aid. Similarly, the greater the degree of match in performance predictions, the greater the reliance on the decision aid. It is worth noting that a match in hire choice or performance predictions could result for reasons unrelated to reliance on the decision aid, such as guessing. Therefore, I will also examine participant's cue utilization of the various predictors using Brunswik's (1952) Lens Model to determine whether participants' weighting of the cues represents the optimal weighting strategy.

According to the Lens Model paradigm, peoples' understanding of the natural environment can be revealed by examining the structure of cues people utilize to make inferences about the state of the environment (Brunswik, 1952, 1955; Cooksey, 1996; Todd & Gigerenzer, 2007). As such, proximal cues are probabilistically related to some distal criterion. The relationships between the proximal cues and the distal criterion are known as the ecological validities. By examining the relationship between the values for the proximal cues and an individual's judgments, one can determine the cue utilization validity. Essentially, one creates a bootstrapped model of the individual's judgment and decision-making policies (Armstrong, 2001). One can then compare the individual's policies or cue utilization to the ecological validities to determine the degree to which an individual's cue usage resembles the optimal weighting strategy. For a detailed review of judgmental bootstrapping, see Armstrong (2001).

Method

Participants

One hundred fifty-four participants were recruited from Amazon's Mechanical Turk (MTurk) program. MTurk originated as a market research platform, but it has since evolved and grown into a multidisciplinary research tool. MTurk is an online crowdsourcing platform that enables researchers and organizations to pay people to perform specific tasks, such as transcribing audio recordings and selecting among a variety of potential photographs to use in advertising. Another common use of MTurk is as research system that allows researchers from a variety of disciplines to recruit participants for their research. In exchange for their participation, participants receive monetary compensation that is specific to each study. Previous researchers have shown that MTurk can be an appropriate tool for recruiting participants. Specifically, those participants recruited using MTurk are more representative of the United States population than college students (Paolacci, Chandler, & Ipeirotis, 2010). Researchers have also demonstrated that data obtained through MTurk is similar in quality and reliability to data collected through other means, including undergraduate psychology research pools (Buhrmester, Kwang, & Gosling, 2011). This is particularly the case for basic decision-making processes. For example, Kausel, Culbertson, Leiva, Slaughter, and Jackson (2015) conducted multiple studies examining the effects of narcissism on advice taking. In studies 1 and study 4, Kausel et al. obtained samples from an undergraduate population, and in studies 2 and 3, they data were collected from MTurk. Across the four studies, the results from MTurk samples converged with the results from undergraduate samples.

In the present study, participants were paid one US dollar for their participation. The median completion time for participants was 14 minutes. Approximately 57% of participants were male. The average age of participants was 37.71 (*SD* = 11.82). Seventy-three percent were Caucasian, and 89% were employed. For the employed individuals, the average number of hours worked per week was *M* = 40.27 (*SD* = 9.54).

Several attention check and screening items were used to help identify those participants who were not paying attention and were simply clicking through the survey. The attention check items were, "Are you paying attention right now? If you are paying attention, select no," "What is the cognitive ability test percentile rank for candidate B," "Please solve the following math equation. 2 + 2 = ?," "Please enter today's date," and "This is an attention check question. Please select strongly agree." Participants were also asked what their native language is. Only participants who successfully passed at least 3 of the attention check questions and who listed English as their native language were included in the final data set. Additionally, at the end of the study, participants were asked, "Is there any reason why we should NOT use your data?" Any participant who indicated that his or her data should not be used was also excluded from the analyses. Two participants' native language was not English; two participants indicated that their data should not be used; and three participants did not successfully pass at least three of the attention check questions. Therefore, these participants were excluded from the analyses.

Materials and Procedure

Decision task. The decision task that was used in this study is an adapted version of the decision task used by Kausel, Culbertson and Madrid (under review). In this task, participants were presented with applicant information for the position of a ticket checker with an airline company in which all of the applicants used in this study were eventually hired. The airline company used a test of cognitive ability, a personality test that measured conscientiousness, and an unstructured interview to make their hiring decisions. Therefore, participants were presented with applicants' percentile scores on a cognitive ability measure, a measure of conscientiousness, and an unstructured interview. Participants were asked to predict the performance of both candidates in each decision. More specifically, participants were asked to provide an estimate of the candidates' performance percentile rank from 0 (*will perform worse than all other employees*) to 100 (*will perform better than all other employees*). Participants were then asked to select the candidate that the airline company should hire.

Feedback information. All participants were provided with feedback regarding their predictions and hiring choices after each decision. Participants were shown what their original predictions were. They were also shown the job performance of both candidates once they were hired as determined by the cue validity manipulation. Last, participants were shown what their prediction error was for each candidate's performance. As such, participants were provided with negative performance feedback to the extent that their predictions differed from the candidates' performance. In study 1, I did not manipulate feedback; the feedback manipulation was introduced in study 2.

Cue validity manipulation. In order to examine the effect of cue validity on participants' reliance on the decision aid, participants were randomly assigned to either a high validity condition or a moderate validity condition. In other words, participants were either assigned to a condition in which the job candidates' eventual performance once hired is highly predictable from an appropriate weighting of the three predictors (cognitive ability, conscientiousness, and unstructured interview), or participants were assigned to a condition in which the job candidates' eventual performance is less predictable from an appropriate weighting of the three predictors.

The weighting of the predictors in both conditions was .50 for cognitive ability, .40 for conscientiousness, and .10 for the unstructured interview. These weights were selected based on the results of meta-analyses (e.g., Huffcutt & Arthur, 1994, Schmidt & Hunter, 1998). Furthermore, these weights were selected such that when summed, the result would produce an estimate of performance in percentile units. The weights were then summed, and random error was introduced using a logistic function in which random error was either small or large, depending on the condition. The predicted performance was then rounded to the nearest integer to simplify the solution. Thus, the model used to create the high valididty condition was:

Equation 1

$$y_p = round(logistic\left(logistic\ percent(.50 * x_1 + .40 * x_2 + .10 * x_3) + \frac{x_r \sim N(0,1)}{6}\right) * 100)$$

Where y_p represents the candidate's eventual performance in the high validity condition. Similarly, the model used to create the moderate validity condition was:

Equation 2

$$y_{lp} = round((logistic(logistic\ percent(.50 * x_1 + .40 * x_2 + .10 * x_3) + x_r \sim N(0,1)) * 100)$$

Where y_{lp} represents the candidate's eventual performance in the moderate validity condition. In both equations, x_1 represents the candidate's cognitive ability score, x_2 represents the candidate's conscientiousness score, and x_3 represents the candidate's interview score. Additionally, $x_r \sim N(0,1)$ represents the value randomly sampled from a standard normal distribution.

In order to determine the actual validity of the cues once the random error has been introduced in the eventual performance of the candidates, the candidates' test scores were used to predict their eventual performance. The model used to predict the candidates' eventual performance used the same weighting used in Equations 1 and 2. Therefore, the formula used to predict the candidates' eventual performance was:

Equation 3

$$\hat{y} = .50 * x_1 + .40 * x_2 + .10 * x_3$$

Where $\hat{y}$ = the predicted eventual performance for the candidate. In the high cue validity condition, equation 3 resulted in an $R^2 = .962$. In the moderate cue validity condition, equation 3 resulted in an $R^2 = .504$. This confirms that the conditions represent situations in which the selection predictors are highly valid and moderately valid, respectively.

Decision aid manipulation. Participants were randomly assigned to one of two conditions in which a decision aid was either present or absent. In the decision aid present condition, participants were provided information about the validity of the three predictors and information regarding a statistical model that should be used to predict job performance of the candidates. In the decision aid absent condition, participants did not receive any information regarding the validity of the three selection predictors or the model.

Participants were asked to utilize the candidates' scores to estimate the candidates' performance as well as select one of the candidates to hire. For participants in the decision aid present condition, participants were presented with equation 3, but they were not provided with the results of the calculations for each candidate. Instead, participants were only presented with the result of the validity weights multiplied by the predictor scores. Thus, participants would still be required to add the three weighted predictor scores. The rationale for this was that participants who engaged in more systematic information processing (i.e., relied more on the statistical model's prediction) would actually add these scores. Thus, their predictions should match the predictions made by the model. In contrast, individuals who engaged in more automatic information processing would not rely on the information provided by the model. Instead, they would rely on their own decision-making processes to make their predictions, which would likely result in predictions that do not match the predictions made by the model. The stimuli presented to participants are displayed in Appendix A.

Results

Because reliance on the decision aid is operationalized in three different ways, three separate analyses were conducted.

Match in Hire Choice

In order to examine reliance on the decision aid based on a match between the participant's hire choice and the model's hire choice, a repeated measures logistic regression was conducted using the generalized liner mixed-effects modeling package in *R* (Bates, Maechler, Bolker, & Walker, 2014). The cue validity, decision aid presence, trial, and their interactions were entered as fixed effects. Because it is possible that participant's trial slopes may differ at each decision, two models were compared. In the first model (model 1), trial was entered only as a fixed effect. In the second model (model 1a) trial was entered as a fixed effect and as a random effect. In both models, the intercept was included as a random effect. The match in hire choice was entered as the dependent variable and was coded as 1 = *match*, 0 = *no match*. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. Cue validity and decision aid presence were both centered using effects coding, such that for cue validity 1 = *high* cue validity and -1 = *moderate* cue validity, and for decision aid presence 1 = *decision aid is present* and -1 = *decision aid is not present.* Trial was mean centered.

In order to determine the best fitting model, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the log likelihood model fit indices were examined. For the AIC and BIC, the model with smaller values is a better fitting model. Similarly, for the log likelihood the model with the larger negative value is the better fitting model. Occasionally, the fit indices will provide slightly conflicting information. In the event of conflicting information, the AIC will be used as the criterion of fit because the AIC uses savvy prior probabilities in its calculation, whereas the BIC does not (Burnham & Anderson, 2004). Model 1 (*AIC* = 1340, *BIC* = 1388, *log likelihood* = -661) appeared to be a better fitting than model 1a, *AIC* = 1341, *BIC* = 1400, *log likelihood* = -659. Furthermore, the addition of trial as a random effect did not produce a significantly better fitting model, $\chi^2(2) = 3.276$, $p = .194$. Therefore, the best fitting model does not include trial as a random effect. Accordingly, the results of model 1 are reported.

The results of model 1 are displayed in Figure 2. The results of model 1 showed a significant main effect of model presence, $B = 0.453$, $z = 5.126$, $p < .001$. When model information was provided, participants' hire choices were significantly more likely to match the model's hire choices than when model information was not provided. There was not a significant main effect of cue validity ($B = 0.053$, $z = 0.604$, $p = .546$) or trial ($B = 0.025$, $z =$

0.963, p = .336). Further, none of the interactions were significant, $ps \geq .12$. Table 1 displays the parameter estimates, standard errors, z-values, and p-vales for the model predicting match in hiring choice.

Match in Predicted Performance

In order to examine reliance on the decision aid based on a match between the participant's predictions about the candidates' performance and the model's performance predictions, a repeated measures linear regression was conducted using the linear mixed-effects modeling package in *R* (Bates et al., 2014). Cue validity, decision aid presence, trial, and their interactions were entered as fixed effects. Because it is possible that participant's slopes may differ at each decision, two models were compared. In the first model (model 2), trial was entered only as a fixed effect. In the second model (model 2a) trial was entered as a fixed effect and as a random effect. The absolute value of the difference between participants' performance predictions for each candidate and the model's performance predictions for each candidate was used as the dependent variable. However, because the absolute value of difference scores results in highly positively skewed data, the variable was log transformed using the natural log. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. The same centering procedures that were used for the first analysis were in this second analysis.

Model 2 was a significantly better fitting model than model 2a, $\chi^2(2) = 12.998$, p = .002. Furthermore, by examining the fit indices, it appeared that model 2 (*AIC* = 12091, *BIC* = 12152, *log likelihood* = -6036) was a better fitting than model 2a, *AIC* = 12081, *BIC* = 12154, *log likelihood* = -6029. Therefore, the best fitting model does not include trial as a random effect. Accordingly, the results of model 2 are reported.

The results of model 2 are displayed in Figure 3. The results of model 2 revealed a significant main effect of cue validity, $B = -0.368$, $t(3071) = -2.932$, $p < .05$. Participants in the high validity condition made performance predictions that were significantly more similar to the model's predictions than participants in the moderate validity condition. The results also revealed a significant main effect of decision aid presence ($B = -0.750$, $t(3071) = -5.970$, $p < .05$), such that when provided with the decision aid participants' performance predictions were significantly more similar to the model's predictions than participants who were not provided with the decision aid. Further, these main effects were qualified by a significant interaction, such that the effect of the cue validity is stronger when the decision aid ($B = -0.675$) was provided than when it was not provided, $B = -0.061$. There was not a significant main effect of trial ($B = -0.003$, $t(3071) = -0.340$, $p > .05$), and none of the interactions with trial were significant, *ps* > .05. Table 2 displays the parameter estimates, standard errors, and *t*-values for the model predicting the match in candidate performance predictions.

Cue Utilization

Finally, in order to examine reliance on intuition based on how participants relied on the various cues, a repeated measures logistic regression was conducted using the generalized liner mixed-effects modeling package in *R* (Bates et al., 2014). However, in this final set of analyses, because the goal is to examine *how* participants are utilizing the various predictor cues to make their hiring choice, the differences between the job candidates' selection predictor scores (difference between cognitive ability for candidate A and the cognitive ability for candidate B, difference between conscientiousness for candidate A and the conscientiousness for candidate B, difference between the unstructured interview for candidate A and the unstructured interview for candidate B) were used to predict the participants' hire choice in each decision. Additionally, cue validity, decision aid presence, and their interactions were included as predictors. Because it is possible that participant's slopes may differ at each decision and as a function of the difference between the two candidates' the selection predictors, multiple models were examined. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. The same centering procedures that were used for the first and second analyses were in this final analysis. The differences between the job candidates' selection predictor scores were also mean centered.

A summary of the structure of the two best fitting models is displayed in Table 3. Model 3 (*AIC* = 1024, *BIC* = 1120, *log likelihood* = -494) was a better fitting model than model 3a, *AIC* = 1031, *BIC* = 1169, *log likelihood* = -489. Furthermore, the inclusion of cue validity and its interactions as fixed effects did not produce a significantly better fitting model, $\chi^2(12) = 18.800$, $p = .093$. Therefore, the results for model 3 are reported.

The results of model 3 are displayed in Table 4. As can be seen in the table, the difference between candidates' cognitive ability scores ($B = 0.158$, $z = 5.718$, $p < .001$), conscientiousness scores ($B = 0.114$, $z = 5.639$, $p < .001$), and interview scores ($B = .091$, $z = 5.075$, $p < .001$) all significantly predicted the participant's hiring choice. There was not a significant main effect of decision aid presence in predicting the participant's hiring choice, $B = -0.148$, $z = -0.978$, $p = .328$. However, there was a significant interaction between decision aid presence and the difference between candidates' cognitive ability scores and conscientiousness scores. Specifically, the slope for the difference between candidates' cognitive ability scores was $B = 0.201$ for the decision aid present condition and $B = 0.114$ for the decision aid absent condition. Similarly, the slope for the difference between candidates' conscientiousness scores was $B = .133$ in the decision aid present condition and $B = 0.096$ in the decision aid absent condition. There was not a significant interaction between decision aid presence and the difference between candidates' interview scores.

Based on the results of model 3, it appears that there is differential weighting of the cues based on the presence of the decision aid, at least for cognitive ability and conscientiousness. Because the results of model 3 represent the cue weighting for a binomial choice (hire candidate A vs. hire candidate B), the optimal weights cannot be compared to the optimal weights presented in equation 3. In order to determine if the weights participants are using represent optimal weighting, a Monte Carlo simulation was conducted in which the differences between the three selection predictors for 10,000 job candidate pairs were generated. The differences between the three selection predictors for each pair were then entered into equation 3 to determine the candidate that the statistical model would select. Because the values entered into the formula were difference scores, a positive value would favor candidate A, and a negative value would favor candidate B. The differences between the three selection predictors for the 10,000 pairs were then used to predict the hire choice (candidate A vs. candidate B). The resulting model yields the optimal weighting for the three selection predictors. Specifically, the optimal weights were $B = 0.189$ for cognitive ability, $B = 0.150$ for conscientiousness, and $B = 0.038$ for the unstructured interview.

Figure 4 displays the optimal cue weighting, participants' weighting of the three cues in the decision aid present condition, and the participants' weighting of the three cues in the decision aid absent condition. As can be seen in the figure, the participants who received the decision aid appeared to have a cue utilization that approximates the optimal cue weighting. As such, these participants were well calibrated, suggesting that participants relied on the decision aid that was provided to make their predictions regarding the candidates' performance. In contrast, the figure shows that participants who did not receive the decision aid appear to have a cue utilization that differs largely from the optimal cue weighting. The participants were poorly calibrated utilizing essentially a unit weighting decision strategy, which suggests that participants were relying on their own decision-making processes. This also serves as a comparison group because it shows that people do change their decision-making strategy when provided with a decision aid.

Exploratory Cluster Analysis

Given the width of the error bars around the cue weighting participants' used, an exploratory cluster analysis was conducted to examine whether there may be the presence of individual differences in how participants are weighting the cues. Specifically, a Ward hierarchical cluster analysis was conducted using the JMP Pro 12 software for the decision aid present and decision aid not present conditions. The results of the cluster analysis are displayed in Figures 5 and 6. As can be seen in Figure 5, two clusters emerged when the decision aid was provided. The first cluster contained 58 individuals, and the second cluster contained 17 individuals. The first cluster approximates the optimal weighting of the cues. However, the second cluster uses a cue weighting that is quite different from the optimal weighting, such that cognitive ability had a weighting of $B = 0.044$, conscientiousness had a weighting of $B = 0.025$, and the unstructured interview had a weighting of $B = 0.069$. Thus, participants in cluster 2 used an approximate unit weighting system with the unstructured interview having the most weight.

As can be seen in Figure 6, three clusters emerged when the decision aid was not provided. The first cluster contained 12 individuals, the second cluster contained 48 individuals, and the third cluster contained 19 individuals. The second cluster had weighting that was similar to the optimal weighting of the cues. However, the first and third clusters used cue weights that are quite different from the optimal weighting. For the first cluster, cognitive ability had a weighting of *B* = 0.048, conscientiousness had a weighting of *B* = 0.046, and the unstructured interview had a weighting of *B* = 0.007. The participants in this cluster appeared to weight cognitive ability and conscientiousness equally and weighted the interview lower. Finally, participants in cluster 3 gave the most weight to cognitive ability (*B* = 0.120) and the unstructured interview (*B* = 0.113) and gave the least weight to conscientiousness (*B* = 0.067). Based on the results of the cluster analysis, it is apparent that there are individual differences in how people are weighting the different predictors, regardless of whether the decision aid was provided or not.

Summary of Results and Discussion

The purpose of this study was to examine the interactive effects of decision aid presence and cue validity on reliance on a decision aid over a series of hiring decisions. As such, this study sought to test hypotheses 1 – 3. Hypothesis 1 stated that participants' hiring choices and performance predictions would more closely match those made by the decision aid when the cues were more valid than when they were less valid. In the analysis that examined the match in hiring choices, cue validity was not a significant predictor. However, in the analysis that examined the match in performance predictions, cue validity was a significant predictor, such that when the cues had higher validity, there was a greater degree of match between participants' performance predictions and the model's performance predictions. Therefore, hypothesis 1 was partially supported. Hypothesis 2 stated that participants' hiring choices and performance predictions would more closely match the choice and performance predictions made by the decision aid when it is provided. When examining the degree of match in hiring choices and the degree of match in performance predictions, the presence of the decision aid was a significant predictor, thus supporting hypothesis 2. The third hypothesis was concerned with the interaction between the presence of the decision aid and the validity of the cues. Only when examining the degree of match in performance predictions was the interaction significant, such that the greatest degree of match in performance predictions occurred when the decision aid was provided and the cues were highly valid. Therefore, hypothesis 3 was partially supported.

The results of this study show that, when making a hiring choice, providing a decision aid leads to greater reliance on the decision aid regardless of the validity of the cues. However, the results of this study also demonstrate that when making specific performance predictions, reliance on the decision aid is greater if the validity of the cues is high. These results suggest that simply providing a decision aid for practitioners to use will lead to reliance on the decision aid. The results of the analysis that examined participants' cue weighting suggest that when the decision aid is provided, participants' are using a weighting system that approximates the optimal weighting used by the decision aid. However, participants' weighting of the predictors was not identical to the optimal weighting, which suggests that while participants are using the decision aid, they may be adjusting the weights in some way. Therefore, when provided with a decision aid in a personnel selection context, participants used the aid but they did not rely on it entirely. When examining the results of the cluster analysis, it is quite clear that there are individual differences in how participants are weighting the selection predictors, regardless of whether the decision aid was provided or not.

Interestingly, the results also show that there was no significant effect of decision trial, suggesting that there was no learning effect over time. However, this may be the case because participants were asked to only make ten decisions. Thus, their experience within the context of this study was rather limited. Study 2 used a greater number of trials to examine whether there is any learning effect over time.

Study 1 sought to test hypotheses 1 – 3, excluding the effects of feedback. Study 2 was conducted to replicate study 1 and test hypotheses 4 and 5. In order to examine hypotheses 4 and 5, the study 2 directly manipulated the presence of feedback.

Chapter 3 - Study 2

Study 2 was similar to study 1 with a few notable exceptions. First, study 2 utilized twenty decision trials instead of ten trials. In study 1, small, albeit non-significant, effects of decision trial were present. Therefore, I examined how reliance on the decision aid would shift over a longer period of time. Second, because most selection environments have cues that are collectively much less predictable than R^2 = .962 or .504, I introduced a third cue validity condition, one in which the validity of the cues is much lower and represents the validity of realistic hiring situations, R^2 = .204. Third, I manipulated feedback. In study 1, all participants received feedback regarding their decision in the form of error feedback. In study 2, feedback was manipulated, such that half of the participants received feedback while the other half did not receive any feedback. Thus, the resulting design of the study 2 is a 2 (model information provided versus no model information provided) x 3 (high cue validity, moderate cue validity, realistic cue validity) x 2 (error feedback provided or no error feedback provided) mixed design with trial as a within subjects variable. The fourth and final change from study 1 is that a fourth cue was added: handwriting analysis. Handwriting analysis, also referred to as graphology, has been shown to be an invalid predictor of job performance (Neter & Ben-Shakhar, 1989; Reilly & Chao, 1982; Schmidt & Hunter, 1998). Despite being a poor predictor of job performance, in some European countries, graphology is widely used. For instance, in France approximately 93% of companies use handwriting analysis in personnel selection (Bruchon-Schweitzer & Ferrieux, 1991). The prevalence of the use of graphology in France has led to the emergence of a myth surrounding graphology. Specifically, the myth is that "graphology is a frequently used and valued selection method in European countries" (Bangerter, Ko, Blatti, & Salvisberg, 2009, p. 219). As Bangerter et al. argue, this myth may be based on the notion that some companies require handwritten application letters. However, the evidence suggests that graphology is not as widely used, as one would expect. For instance, since 2000 less than 1% of job ads in Switzerland require a handwritten application letter (Bangerter et al., 2009). Interestingly, 22.2% of the time, the purpose of handwritten application letters is to analyze handwriting. Regardless of whether graphology is used 10% of the time or 90% of the time, the lack of validity evidence supporting its use would suggest than any use of graphology is overuse. Given that handwriting analysis *is* used in some places around the world despite that lack of validity evidence supporting its use, handwriting analysis was added as a fourth predictor to determine whether participants' cue weighting strategies can accommodate a cue that has a near-zero relationship with job performance.

Method

Participants

Five hundred nineteen hiring professionals were recruited using Qualtrics participant panels. In order to participate in this study, participants were required to have a minimum of one year of hiring experience. In this sample, participants had an average of 7.70 (*SD* = 6.69) years of hiring experience. Furthermore, 93% of the sample was currently employed, and those who were employed worked an average of 42.99 (*SD* = 10.01) hours per week. The average age of participants was 38.96 (*SD* = 11.33). Approximately 52% of the participants were female. Eighty percent were Caucasian. The median completion time for participants was 18 minutes.

Like in study 1, several attention check questions and screening questions were used. The same attention check items that were used in study 1 were used in study 2. However, because the data was collected using the Qualtrics participant panels, any participant who failed a single attention check question was excluded. Additionally, any participant whose native language was not English was excluded. Because participants were removed from the data prior to Qualtrics sending the final sample, it is not known how many participants failed attention checks or how many were excluded because English is not their native language. However, like in study 1, participants were asked if there was any reason their data should not be used. Based on their responses, twenty participants were excluded from the analyses.

Materials and Procedure

Decision task. This study used the same decision task that was used in study 1. However, instead of making 10 selection decisions, participants were asked to make 20 selection decisions. The increase in the number of decisions allowed for a better examination of how participants' decision strategies change over time. The ordering of the 20 decisions were randomized to reduce the effects of ordering and to examine the changes over time.

Cue validity manipulation. Like in study 1, participants were randomly assigned to the cue validity conditions. However, because reality is much less predictable, a third condition was added. Schmidt and Hunter (1998) demonstrated that the addition of a single predictor to cognitive ability yielded coefficients ranging from $R^2 = .260$ to $R^2 = .423$. While Schmidt and Hunter demonstrated that the predictive power of multiple predictors approaches an $R^2 = .500$, it is worth noting that their calculations included corrections for attenuation, such as range restriction. When examining the observed validities in a multiple predictor selection system, the validities will be lower. For example, Jacobs, Conte, Day, Silva, and Harris (1996) conducted a job analysis and identified six relevant job predictors. Using a composite of the identified predictors, they demonstrated that the six predictors relatively modest predictive power with R^2 ranging from .01 to .09, depending on the performance criterion used.

Because the predictability of actual employee behavior can be substantially smaller than the moderate validity condition from study 1, participants were randomly assigned to either the high cue validity condition (R^2 = .962), the moderate cue validity condition (R^2 = .504), or the realistic cue validity condition (R^2 = .204). In order to create the realistic cue validity condition, the same procedures used to create the high validity and the moderate validity conditions were used to create the realistic validity condition. However, a greater degree of random error was introduced to produce an environment that more closely matched the validity of predictors of job performance in real business settings. Thus, the formula used to create the realistic cue validity was:

Equation 4

$$y_r = round(logistic(logistic\ percent(.50 * x_1 + .40 * x_2 + .10 * x_3 + .0 * x_4) + 1.5 * (x_r \sim N(0,1))) * 100)$$

Where y_r represents the candidate's eventual performance in the realistic condition, x_1 represents the candidate's cognitive ability score, x_2 represents the candidate's conscientiousness score, x_3 represents the candidate's interview score, and x_4 represents the candidate's handwriting analysis score. Additionally, $x_r \sim N(0,1)$ represents the value randomly sampled from a standard normal distribution.

In order to determine the actual validity of the cues once the random error has been introduced in the eventual performance of the candidates, the candidates' test scores were used to predict their eventual performance. The formula used to predict the candidates' eventual performance was:

Equation 5

$$\hat{y} = .50 * x_1 + .40 * x_2 + .10 * x_3 + .00 * x_4$$

Using equation 5 to predict the eventual performance of candidates in the realistic validity condition resulted in $R^2 = .204$.

The stimuli that were presented to the participants are displayed in Appendix B.

Decision aid presence manipulation. The same decision aid presence manipulation that was utilized in study 1 was used in this study. Thus, participants were randomly assigned to receive decision aid or not receive the decision aid.

Feedback information. One purpose of the present study was to examine the effect of feedback on reliance on intuition. In contrast to study 1, participants in study 2 were randomly assigned to receive feedback regarding their performance predictions and hiring choices after each decision or not receive any feedback regarding their performance predictions and hiring choices. Those assigned to the feedback condition were shown what their original performance predictions were, the actual job performance of both candidates once they were hired, and their prediction error for each candidate's performance. Participants assigned to not receive feedback did not receive any feedback regarding what their original performance predictions were, the actual job performance of both candidates once they were hired, or their prediction error for each candidate's performance.

Results

To test the hypotheses, the analytical approach was similar to the approach used in study 1, with three exceptions. The number of trials increased from ten to twenty. Study 2 included the additional cue validity condition, and feedback was manipulated. Accordingly, these changes were incorporated into the analytical procedures.

Match in Hire Choice

In order to examine utilization of the decision aid based on a match between the participant's hire choice and the hire choice made by the model, a repeated measures logistic regression was conducted using the generalized liner mixed-effects modeling package in *R* (Bates et al., 2014). Cue validity, model presence, the presence of feedback, trial, and their interactions were entered as fixed effects. While study 1 demonstrated that participants' slopes did not differ at each decision, this may have occurred because there were not enough decisions to demonstrate any learning effects. Therefore, because it is possible that participant's slopes may differ at each decision with the increased number of decisions, three models were compared. In the first model (model 4), trial was not included in the model because trial was not a significant predictor in study 1. Additionally, the exclusion of trial from the model creates a model with fewer parameters, which should reduce the risk of overfitting the model. In the second model (model 4a), trial was entered as a fixed effect, and in the third model (model 4b) trial was entered as a fixed effect and as a random effect. The match in hire choice was entered as the dependent variable and was coded as 1 = *match*, 0 = *no match*. In order to reduce the effects of multicollinearity between the main effects and the interactions, the categorical predictors were centered using effects coding, and trial was mean centered.

Like in study 1, the AIC, BIC, and log likelihood model fit indices were examined to determine the best fitting model. Because these indices may occasionally provide conflicting information, the model with the smallest AIC values was selected as the better fitting model (see Burnham & Anderson, 2004). Model 4 (*AIC* = 10502, *BIC* = 10596, *log likelihood* = -5238) appeared to be a better fitting than model 4a (*AIC* = 10511, *BIC* = 10692, *log likelihood* = -5231) or model 4b, *AIC* = 10512, *BIC* = 10708, *log likelihood* = -5229. The addition of trial as a fixed effect did not produce a significantly better fitting model ($\chi^2(12) = 14.546$, $p = .267$), nor did the addition of trial as a fixed effect and random effect produce a significantly better fitting model, $\chi^2(2) = 2.775$, $p = .250$. Therefore, the best fitting model does not include trial. Accordingly, the results of model 4 are reported.

The results of model 4 are displayed in Figure 7. The results of model 4 showed that there was not a significant main effect of cue validity on whether participants' hiring choices matched the model's hiring choices, $F(2, 507) = 0.286$, $p = .752$. However, there was a significant main effect of decision aid presence, $B = 0.152$, $z = 3.95$, $p < .001$. When the decision aid was provided, participants' hire choices were significantly more likely to match the model's hire choices than when model information was not provided. Additionally, there was a significant main effect of feedback on whether participants' hiring choices matched the model's choices, $B = -0.085$, $z = -2.22$, $p = .027$. When feedback was provided, participants' hiring choices were significantly less likely to match the model's choices. Finally, none of the interactions were significant, $ps \geq .067$. Table 5 displays the parameter estimates, standard errors, *z*-values, and *p*-vales for the model predicting match in hiring choice.

Match in Performance Predictions

In order to examine reliance on the decision aid based on a match between the participant's predictions about the candidates' performance and the model's performance predictions, a repeated measures linear regression was conducted using the linear mixed-effects modeling package in *R* (Bates et al., 2014). Cue validity, decision aid presence, feedback, trial, and their interactions were entered as fixed effects. As with the previous analyses, three models were examined. In model 5, trial (and its interactions) was not included as either a fixed effect or random effect. In model 5a, trial was included as a fixed effect only, and in model 5b, trial was included as both a fixed effect and random effect. The absolute value of the difference between participants' performance predictions for each candidate and the model's performance predictions for each candidate was used as the dependent variable. However, because the absolute value of difference scores results in highly positively skewed data, the variable was log transformed using the natural log. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model using the same procedures in the previous analysis.

Model 5a was a significantly better fitting model than model 5 ($\chi^2(12) = 170.51$, $p < .001$) and model 5b was a significantly better fitting model than model 5a, $\chi^2(2) = 534.09$, $p < .001$. Furthermore, by examining the fit indices, it appeared that model 5b (*AIC* = 72611, *BIC* = 72833, *log likelihood* = -36278) was a better fitting than model 5a (*AIC* = 73141, *BIC* = 73348, *log likelihood* = -36545) and model 5, *AIC* = 73288, *BIC* = 73399, *log likelihood* = -36630. Therefore, the best fitting model includes trial as both a fixed effect and a random effect. Accordingly, the results of model 5b are reported.

As in the analyses examining match in hire choice, the results showed that there was not a significant effect of cue validity on the degree of similarity in participants' performance predictions and the model's performance predictions, $F(2, 507) = 1.221$, $p = .296$. In contrast to the analyses examining the match in hiring choice, there was not a significant main effect of feedback, $B = .045$, $t(20719) = 0.86$, $p > .05$. However, the results did reveal a significant main effect of decision aid presence ($B = -0.534$, $t(20719) = -10.13$, $p < .05$), such that when provided with the decision aid participants' performance predictions were significantly more similar to the model's performance predictions than participants who were not provided with the decision aid. There was also a significant main effect of trial ($B = -0.015$, $t(20719) = -4.78$, $p < .05$), such that as participants progressed through the study, their predictions regarding the candidates' performance became more similar to the model's predictions. However, it is worth noting that this effect is quite small. Interestingly, there were several significant interactions; therefore, *post hoc* comparisons were conducted.

All *post hoc* comparisons were conducted using Bonferroni corrected p-values. There was a significant interaction between the cue validity and decision aid presence, $F(2, 507) = 3.566$, $p = .029$. The results of this interaction are displayed in Figure 8. When the decision aid was not provided, there was no significant difference between the high validity ($B = 0.552$), the moderate validity ($B = 0.525$), and the realistic validity ($B = 0.526$) conditions. Furthermore, when the decision aid was provided, there was no significant difference between the moderate validity ($B = -0.652$) and the high validity ($B = -0.809$, $z = -0.729$, $p = .968$) or the realistic validity ($B = -0.142$, $z = -2.413$, $p = .110$) conditions. However, when the decision aid was provided, there was a significant difference between the high validity ($B = -0.809$) and the realistic validity ($B = -0.142$) conditions, $z = -3.403$, $p = .006$.

There was a significant interaction between cue validity and trial ($F(2, 507) = 4.135$, $p = .017$), such that the slope of the high validity condition ($B = -0.015$) was significantly different from the slope of the moderate validity condition ($B = 0.004$, $z = -2.527$, $p = .031$) and significantly different from the slope of the realistic validity condition, $B = 0.011$, $z = -3.550$, $p = .001$. However, the slope of the moderate validity condition was not significantly different from the slope of the realistic validity condition, $z = -0.798$, $p = .704$. Finally, there was a significant three-way interaction between cue validity, decision aid presence, and trial ($F(2, 507) = 7.211$, $p < .001$). Figure 9 displays the results of this interaction. Post hoc analyses revealed that for the high validity predictors condition, the slope when the decision aid was provided ($B =$

-0.030) was significantly different than the slope when the decision aid was not provided, $B <$ 0.001, $z = -2.931$, $p = .027$. Additionally, when the model was provided, the slope of the high validity condition ($B = -.030$) was significantly different from the slope of the moderate validity condition ($B = 0.006$, $z = -2.888$, $p = .030$) and significantly different from the slope of the realistic validity condition, $B = .022$, z -4.547, $p < .001$). All other differences examined were not significant.

Table 6 displays the parameter estimates, standard errors, and *t*-values for the model predicting the match in candidate performance predictions.

Cue Utilization

Finally, in order to examine reliance on the decision aid based on how participants weighed the various cues, a repeated measures logistic regression was conducted using the generalized liner mixed-effects modeling package in *R* (Bates et al., 2014). Like in study 1, the differences between the job candidates' selection predictor scores (difference between cognitive ability for candidate A and the cognitive ability for candidate B, difference between conscientiousness for candidate A and the conscientiousness for candidate B, difference between the unstructured interview for candidate A and the unstructured interview for candidate B, and the difference between the handwriting analysis score for candidate A and the handwriting analysis score for candidate B) were used to predict the participants' hire choice in each decision. Additionally, the cue validity, decision aid presence, feedback were included as predictors. Because it is possible that participant's slopes may differ at each decision and as a function of the difference between the two candidates' the selection predictors, multiple models were examined. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. The same centering procedures that were used for the first and second analyses were in this final analysis. The differences between the job candidates' selection predictor scores were also mean centered.

While cue validity and feedback were originally included in the model, their effects were not significant. Given the complexity of the model structure, these variables were removed to create a more parsimonious model and to avoid overfitting. A summary of the structure of the two models examined is displayed in Table 7. Model 6a was a significantly better fitting model than model 6, $\chi^2(4) = 19.638$, $p < .001$. Furthermore, an examination of the fit indices suggests, especially the AIC, that model 6a (*AIC* = 8911, *BIC* = 9093, *log likelihood* = -4431) was a better fitting model than model 6, *AIC* = 8923, *BIC* = 9075, *log likelihood* = -4441. Therefore, the results for model 6a are reported.

Because the results of model 6a represent the cue weighting for a binomial choice (hire candidate A vs. hire candidate B), the optimal weights cannot be compared to the optimal weights presented in equation 5. In order to determine if the weights participants are using represent optimal weighting, the same Monte Carlo simulation procedures that were used in study 1 were used with the addition of the addition of the handwriting analysis cue. The resulting Monte Carlo simulation model yields the following optimal weighting: $B = 0.189$ for cognitive ability, $B = 0.150$ for conscientiousness, $B = 0.038$ for the unstructured interview, and $B = 4.533*10^{-5}$ for the handwriting analysis. Notice that these optimal weights are the same as the optimal weights used in study 1, with the addition of the handwriting analysis.

The results of model 6a are displayed in Table 8. As can be seen in the table, the difference between candidates' cognitive ability scores ($B = 0.043$, $z = 21.693$, $p < .001$), conscientiousness scores ($B = 0.031$, $z = 25.206$, $p < .001$), interview scores ($B = .018$, $z = 9.918$, $p < .001$), handwriting analysis scores ($B = 0.017$, $z = 11.406$, $p < .001$), and the presence of the decision aid ($B = -0.249$, $z = -6.372$, $p < .001$) all significantly predicted the participant's hiring choice. Furthermore, there was a significant interaction between decision aid presence and the difference between candidates' handwriting analysis scores. Specifically, when the decision aid was provided the slope was $B = 0.014$, and when the decision aid was not provided the slope was $B = 0.020$. While this interaction is quite small, it does indicate that when the decision aid is provided participants provide less weight on a cue that has a miniscule relationship with performance. The remaining interactions were not significant. However, for the sake of comparison with study 1, Figure 10 displays the optimal cue weighting, participants' cue weighting when the decision aid is provided, and participants' cue weighting when the decision aid is not provided.

As can be seen in Figure 10, neither participants in the decision aid provided nor the decision aid not provided conditions used a cue weighting that approximated the optimal cue weighting. It does appear that participants are weighting cognitive ability more than conscientious and weighting conscientiousness more than either the unstructured interview or the handwriting analysis. Therefore, it appears that participants are using a weighting system that somewhat appropriately ranks the importance of the cues. From the figure, it does appear that this rank ordering of the weights is slightly better when participants are provided with the decision aid. However, recall that the interaction was not significant.

Exploratory Cluster Analysis

Like in study 1, an exploratory cluster analysis was conducted to examine whether there may be the presence of individual differences in how participants are weighting the cues. Specifically, the Ward hierarchical cluster analysis was conducted using the JMP Pro 12 software for the decision aid present and decision aid not present conditions. The results of the cluster analysis are displayed in Figures 11 and 12. As can be seen in Figure 11, four clusters emerged when the decision aid was provided. The first cluster contained 36 individuals, the second cluster contained 68 individuals, the third cluster contained 53 individuals, and the fourth cluster contained 27 individuals. From these four clusters, a clear pattern emerges. Specifically, clusters 2 and 3 have different weighting for the predictors, but they have an interestingly similar pattern in the weighting. Likewise, cluster 1 and 4 have different weighting for the predictors, but they too have an interestingly similar pattern. In clusters 2 and 3, cognitive ability is weighted more than conscientiousness, which is weighted more than both the unstructured interview and handwriting analysis. In contrast, clusters 1 and 4 use what approximates a unit weighting system with each of the predictors being weighed approximately the same.

As can be seen in Figure 12, a four-cluster solution emerged. The first cluster contained 163 individuals, the second cluster contained 58 individuals, the third cluster contained 48 individuals, and the fourth cluster contained 66 individuals. From these four clusters, one can see that an almost identical pattern of weighting emerged for the four clusters when the decision aid was *not* provided as when the decision aid was provided. Specifically, clusters 2 and 3 have different weighting for the predictors, but they have an interestingly similar pattern in the weighting. Likewise, cluster 1 and 4 have different weighting for the predictors, but they too have an interestingly similar pattern. In clusters 2 and 3, cognitive ability is weighted more than conscientiousness, which is weighted more than both the unstructured interview and handwriting analysis. In contrast, clusters 1 and 4 use what approximates a unit weighting system with each of the predictors being weighed approximately the same. Based on the results of the cluster analysis, it is apparent that there are individual differences in how people are weighting the different predictors, regardless of whether the decision aid was provided or not.

Summary of Results and Discussion

Study 2 was conducted to replicate the findings of study 1 as well as test hypotheses 4 and 5. All three analyses in study 2 showed that there was not a significant main effect of cue validity on the degree to which participants' hiring choices and performance predictions match the hiring choices and performance predictions made by the model. Therefore, hypothesis 1 was not supported. In contrast, all three analyses *did* show a significant main effect of the presence of the decision aid on the degree to which participants' hiring choices and performance predictions matched those made by the model, supporting hypothesis 2. Hypothesis 3 stated that the presence of the decision aid would interact with the validity of the cues, such that when the decision aid is present and the cues are more valid, participants' hiring choices and performance predictions will more closely match those made by the decision aid than in all other conditions. There was not a significant interaction when predicting the match in hiring choice. However, when predicting similarity in performance predictions, there was a significant interaction between the presence of the decision aid and the validity of the cues. Specifically, when the decision aid was provide and the cues had high validity, participants performance predictions more closely matched the predictions made by the model than when the validity of the cues was realistic, but not when they had a moderate level of validity. Therefore, hypothesis 3 was only partially supported. The presence of feedback was only a significant predictor when examining the match between participants' hiring choices and the model's hiring choices. Furthermore, the effect was in the opposite direction than what was predicted. Therefore, hypothesis 4 was *not* supported. Hypothesis 5 stated that the effect of feedback on decision aid reliance would depend on the validity of the cues, such that when the cues are more valid and feedback is provided, participants' hiring choices and performance predictions will more closely match those made by the decision aid than all other conditions. However, across all three analyses in study 2, there was not a significant interaction between the presence of feedback and the validity of the cues. Therefore, hypothesis 5 was also *not* supported.

The results of the analysis examining the participants' cue utilization were quite surprising and did not replicate the findings from study 1. The results showed that there was no interaction between the presence of the decision aid and the cue utilization. In fact, it appeared that, regardless of whether the decision aid was present or not, participants seemed to be using almost a unit weighting system. An examination of the cluster analysis revealed that this is indeed what some groups of people were doing. Furthermore, by using a low unit weighting system participants consistently underweighted the importance of cognitive ability and conscientiousness. While study 2 did not replicate the cue utilization results of study 1, it is important to note that in some situations a unit weighting system can be useful (Dawes, 1979).

A secondary purpose of study 2 was to increase the number of decisions participants made in order to better determine whether any learning effects were present. In contrast to study 1, there was a significant three-way interaction between the validity of the cues, the presence of the decision aid and trial (see Figure 9). While there was not an explicit hypothesis regarding learning, this interaction provides meaningful information about how participants learned over the course of twenty hiring decisions. Specifically, participants experienced the greatest degree of learning when the decision aid was provided and the cues were highly valid. As the validity of the cues decreased, learning decreased. When the validity of the cues was realistic, meaning that it resembled the validity one would expect in a real hiring context, no learning occurred over the twenty trials. There are two possible conclusions from this finding. One is that this may have occurred because there were too few of decisions for participants to discern any type of pattern and learn from their decisions. After all, increasing the number of decisions from ten to twenty allowed for participants to learn when the validity of the cues was almost perfect, $R^2 = .962$. The other obvious possibility is that when the cues have realistic levels of validity (or even moderate validity), there may be so much error involved that people are not able to discern any sort of pattern.

Chapter 4 - General Discussion

The purpose of these two studies was to examine the conditions under which people will utilize decision aids in a personnel selection context. Specifically, this study sought to examine whether a) the mere presence of a decision aid will lead people to rely on the decision aid, b) the validity of the predictors used in the selection context influence reliance on a decision aid, c) the presence of feedback regarding one's predictions of a candidate's performance, and d) the interactions between these factors influence reliance on a decision aid. In this study, the decision aid took the form of a statistical model that should be used to select the candidate to be hired. In both study 1 and study 2, the evidence clearly demonstrated that the mere presence of a decision aid leads people to rely on the decision aid. While this is not an overly profound finding, it does have its own merit. By having a comparison group (those who did not receive the decision aid), I was able to examine whether participants were actually relying on the decision aid. Furthermore, by examining the cue weighting that participants used, I was able to examine whether a match between participants' predictions and choices and the model's predictions and choices was the result of reliance on the decision aid or the use of some alternative decision-making strategy.

The finding that participants rely, to some extent, on a decision aid when it is provided, also has practical importance. Previous research has found that people are often resistant to relying on and underutilize decision aids (Arkes et al., 1986; Ashton, 1990; Dawes et al., 1989; Diab et al., 2011; Dietvorst et al., 2015a; Highhouse, 2008; Parasuraman & Riley, 1997). The results of the present research suggest that in a personnel selection context, when provided with a decision aid people will actually utilize it, at least to some degree. Therefore, organizations should provide individuals responsible for making hiring decisions with a decision aid. While people may not rely entirely on the decision aid when making their hiring decisions, they may utilize it to some degree. This should ultimately make their performance predictions and hiring choices more accurate.

A second major finding in the present research is that the validity of the cues interacts with the presence of a decision aid to influence reliance on the decision aid when making performance predictions. In both study 1 and study 2, the validity of the cues interacted with the presence of the decision aid, such that there was the greatest degree of match between participants' predictions of candidates' performance and the model's predictions of the candidates' performance when the decision aid was provided and the validity of the cues was high. The importance of this finding is inherent in nearly all personnel selection research. Specifically, personnel selection research aims to identify and develop methods of assessment that maximize the relationship between selection tests and future job performance. This research demonstrated that reliance on the decision aid was greatest when the validity of the predictors was greatest. Unfortunately, the observed validity of selection predictors more closely resembles the realistic validity condition (Jacobs et al., 1996; Schmidt & Hunter, 1998). Therefore, a practical reason why people are hesitant to rely on decision aids is that decision aids do err. This leads people to distrust decision aids (e.g., Dietvorst et al., 2015a). This is especially apparent in Figure 9. In the high validity condition, people saw the accuracy of the decision aid, which lead to an increase in the reliance over time. However, in the realistic validity condition, people saw the decision aid err, which led to a slight (non-significant) decrease in the reliance over time.

The third way in which decision aid reliance was operationalized was an appropriate weighting of the relevant predictors in making the hiring choice and performance predictions. In study 1, participants who received the decision aid used cue weights that had an appropriate rank ordering of the cue weights. Further, the weights used were in the approximate numerical vicinity of the optimal cue weights. Participants who did not received the decision aid used cue weighting that much more closely resembled a unit weighting system, and the cues were each weighed less than when participants did receive the decision aid. However, in study 2, regardless of whether the decision aid was provided or not, participants used a weighting system that quite closely resembled a unit weighting system. Further, on average, the participants' weighting was lower for each predictor than in study 1. One possibility for the differences in cue utilization between study 1 and study 2 is the participants. Specifically, participants in study 2 were required to have a minimum of one year of hiring experience in order to participate, whereas there was not such restriction in study 1. Accordingly, participants' average number of years of hiring experience in study 2 was 7.70 years. Therefore, the participants in study 2 may be collectively considered as hiring experts. Because participants in study 1 were not required to have hiring experience, the number of years of hiring experience is unknown. However, it may be assumed that the participants in study 1 are not collectively considered experts. Previous research has demonstrated that in some situations, such as financial forecasting, the performance of novice decision makers can match, and sometimes exceed, the performance of expert decision makers (e.g., Armstrong, 1980; Yates et al., 1991). Interestingly, Dew, Read, Sarasvathy, and Wiltbank (2009) demonstrated that in entrepreneurial decision making, experts used qualitatively different decision-making strategies than novices, who tended to use more "by-the-book" strategies. Therefore, in the present findings, the difference between the cue utilization between study 1 and study 2 may be the result of differences in novice versus expert decision-making strategies. Indeed, it was the participants in study 1 whose cue utilization was more "by-the-book" than the participants in study 2.

This research also sought to answer the call by researchers to examine the effect of immediate feedback on reliance on a decision aid in a personnel selection context (Slaughter & Kausel, 2014). The results of study 2 showed that feedback did not have a significant effect on reliance on the decision aid. Nor did feedback interact with trial, decision aid presence, or the validity of the cues to influence reliance on the decision aid. This is surprising, especially given the three-way interaction between trial, decision aid presence, and cue validity. The significant interaction would suggest that for the high validity condition, people are able to learn to use the decision aid when it is provided. However, people cannot learn about the validity of the decision aid without feedback. It may be the case that the form and content of feedback may influence the reliance on a decision aid.

Limitations and Future Directions

One limitation of the present research involves the nature of the task. When asked to make the hiring decisions participants simply saw the information regarding the candidates' scores. This may not resemble real hiring decisions. In real hiring decisions, managers may and likely do have more information regarding the candidates than simply their scores. Furthermore, managers may have actually been the ones who conducted the interview (regardless of the structure) with the candidate. Therefore, they would have more information than a numerical score. Additionally, managers may have access to biographical data, integrity tests, background tests, drug tests, previous employment performance records, etc. In the context of the present research, participants' information was quite limited, which may have lowered the psychological fidelity of the hiring situation. Despite this limitation, one may be able to place some faith in the results of this study given that the purpose of the study was to examine the factors that influence reliance on a decision aid. Future research should modify this study in order to increase the psychological fidelity of the research. For example, researchers could have video recordings of the interviews and allow participants to rate the candidates' performance in the interview prior to making their hiring choice and performance predictions. Alternatively, future researchers should partner with organizations to develop organizationally specific decision aids and conduct the research in actual hiring contexts.

The non-significant findings associate with feedback warrant further investigation. As stated previously, in order for participants to learn the validity of a decision aid, they must have feedback regarding its performance. Therefore, it may be the form and content of feedback that influences reliance on the decision aid. In this research, feedback was provided regarding participants' predictions only. Feedback was not provided regarding the predictions of the decision aid. Previous research has shown that providing feedback regarding the decision aid's performance can lead people to distrust the decision aid, especially when people see the decision aid err (Dietvorst et al., 2015a). However, previous research has not examined whether providing feedback interacts with the validity of the decision aid. Therefore, future research should investigate whether providing feedback regarding the decision aid's performance interacts with the validity of the decision aid to influence reliance on the decision aid. Similarly, previous researchers have also argued that resistance to using decision aids stems from a lack of trust in the decision aid (e.g., Dietvorst et al., 2015a; Parasuraman & Riley, 1997). Therefore, future research should actually assess participants' trust in a decision aid and how it changes over a series of decisions.

Across both studies, the results of an exploratory cluster analysis revealed that there are clear differences in how people are weighting the different cues. In study 1, the clusters that emerged differed based on whether the decision aid was present. In study 2, the clusters that emerged did not differ based on whether the decision aid was present. What is consistent across study 1 and 2 is that there are clear clusters that emerge. This suggests the presence of individual difference variables that may be influencing the decision strategies that people choose to employ. Therefore, future research should explore which individual difference variables and under what conditions those individual difference variables influence participants' weighting of cues and reliance on a decision aid.

The stark difference between participants' cue utilization in study 1 compared to study 2 warrants investigation. As previously mentioned, this may be the result of differences between the samples in terms of expertise. Accordingly, future research should directly examine whether hiring novices use qualitatively different strategies compared to hiring experts.

A final direction for future research is to examine the effects of cue validity, decision aid presence, and feedback over a larger number decisions and over time. In study 1, there was no effect of trial. It was only when the number of decision trials was increased to twenty was learning observed, and it was only observed in the high validity condition when the decision aid was provided. Therefore, it is quite possible that the twenty decisions were simply not enough to determine whether participants in the moderate validity or realistic validity conditions are actually learning. While it is possible that the amount of error in the moderate validity and realistic validity situations is simply too large for people to discern any type of relationship between the cues and the eventual performance of candidates, it is also possible that people need to make more decisions in order to discern the relationship. Accordingly, future research should extend the number of decisions in an attempt to discover whether learning can occur in moderate to low validity contexts. In a related vein, future research should examine whether people are able to retain the information they learned about the validity of the decision aid over time. For instance, study 2 showed that people learned that the when the decision aid has a high level of validity it is a useful tool, which lead people to increase their reliance on the decision aid over multiple decisions. However, future research should examine the lagged effects of this learning. Will people continue to rely on the decision aid after some time delay, such as a week, a month, six months, or even a year?

Conclusions

This research sought to examine the effects of cue validity, presence of a decision aid, and feedback on reliance on a decision aid in a personnel selection context. Simply providing a decision aid to people leads to reliance on the decision aid, at least to some degree. Further, when the cues have high validity and the decision aid is provided, people increase their reliance on the decision aid over time. Finally, it is clear that there are individual difference variables that influence how people weight decision cues.

References

Agor, W. H. (1986). The logic of intuition: How top executives make important decisions. *Organizational Dynamics*, *14*(3), 5–18. doi:10.1016/0090-2616(86)90028-8

Agor, W. H. (1991). How intuition can be used to enhance creativity in organizations. *Journal of Creative Behavior*, *25*(1), 11–19.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC.

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, *37*(1), 93–110. doi:10.1016/0749-5978(86)90046-4

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, *83*(7), 16–24.

Armstrong, J. S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting. In J. S. Scott (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 1–18). Norwell, MA: Kluwer Academic Publishers.

Arthur Jr., W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*(2), 435–442. doi:10.1037/0021-9010.93.2.435

Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, *27*(4), 519–533. doi:10.1111/j.1744-6570.1974.tb01173.x

Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, *28*(S), 148–180. doi:10.2307/2491253

Bangerter, A., Ko, C. J., Blatti, S., & Salvisberg, A. (2009). How widespread is graphology in personnel selection practice? A case study of a job market myth. *International Journal of Selection and Assessment*, *17*(2), 219–230.

Barrett, G. V., Polomsky, M. D., & Mcdaniel, M. A. (1999). Selection tests for firefighters: A comprehensive review and meta analysis. *Journal of Business and Psychology*, *13*(4), 507–513.

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? *Personality and Performance*, *9*(1/2), 9–30.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 [Computer software]. Retrieved from http://cran.r-project.org/package=lme4

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*(3), 387–409. doi:10.1348/096317905X26994

Blattberg, R. C., & Hoch, S. J. (1990). Database models and managerial intuition: 50% model + 50% manager. *Management Science*, *36*(8), 887–899.

Boatsman, J. R., Mockel, C., & Pei, B. K. (1997). The effects of decision consequences on auditors' reliance on decision aids in audit planning. *Organizational Behavior and Human Decision Processes*, *71*(2), 211–247.

Brown, J. C. (2006). The effects of behavioral and outcome feedback on prudent decision-making under conditions of present and future uncertainty. *Judgment and Decision Making*, *1*(1), 76–85.

Bruchon-Schweitzer, M., & Ferrieux, D. (1991). Une enquête sur le recrutement en France [A survey of selection in France]. *European Review of Applied Psychology*, *41*(9-17).

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago, IL: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. doi:10.1037/h0047470

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. doi:10.1177/1745691610393980

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. doi:10.1177/0049124104268644

Chorągwicka, B., & Janta, B. (2008). Why Is It So Hard to Apply Professional Selection Methods in Business Practice? *Industrial and Organizational Psychology*, *1*(3), 355–358.

Colarelli, S. M., & Thompson, M. (2008). Stubborn reliance on human nature in employee selection: Statistical decision aids are evolutionarily novel. *Industrial and Organizational Psychology*, *1*(3), 347–351. doi:10.1111/j.1754-9434.2008.00060.x

Colbert, A. E., Rynes, S. L., & Brown, K. G. (2005). Who believes us? Understanding managers' agreement with human resource research findings. *The Journal of Applied Behavioral Science*, *41*(3), 304–325. doi:10.1177/0021886305275799

Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking and Reasoning*, *2*(2-3), 141–174. doi:10.1080/135467896394483

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *The Academy of Management Review*, *32*(1), 33–54.

Dane, E., & Pratt, M. G. (2009). Conceptualizing and measuring intuition: A review of recent trends. In G. P. Hodgkinson & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 24). Chichester, West Sussex, UK: John Wiley & Sons Ltd.

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, *26*(2), 180–188. doi:10.1037/h0030868

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. doi:10.1037/0003-066X.34.7.571

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical Versus Actuarial Judgement. *Science*, *243*, 1668–1674.

Dew, N., Read, S., Sarasvathy, S. D., & Wiltbank, R. (2009). Effectual versus predictive logics in entrepreneurial decision-making: Differences between experts and novices. *Journal of Business Venturing*, *24*(4), 287–309. doi:10.1016/j.jbusvent.2008.02.002

Diab, D. L., Pui, S., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids and US and non-US samples. *International Journal of Selection and Assessment*, *19*(2), 209–216. doi:10.1111/j.1468-2389.2011.00548.x

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015a). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. doi:10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015b). *Overcoming algorithm aversion: People will use algorithms if they can (even slightly) modify them*. Manuscript submitted for publication.

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*(20), 458–468. doi:10.1002/bdm.741

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, *50*(3), 387–395. doi:10.1207/s15327752jpa5003_8

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), 313–326. doi:10.1080/1047840X.2010.521057

Farr, J. L., & Tippins, N. T. (2010). Handbook of Employee Selection: An Introduction and Overview. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of Employee Selection*. New York, NY: Taylor & Francis Group.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576. doi:10.1287/inte.1070.0309

Fisher, C. D. (2008). Why don't they learn? *Industrial and Organizational Psychology*, *1*(3), 364–366. doi:10.1111/j.1754-9434.2008.00065.x

Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making*, *4*(3), 186–199.

Gomaa, M. I., Hunton, J. E., Vaassen, E. H. J., & Carree, M. A. (2011). Decision aid reliance: Modeling the effects of decision aid reliability and pressures to perform on reliance behavior. *International Journal of Accounting Information Systems*, *12*(3), 206–224. doi:10.1016/j.accinf.2011.02.001

Goodwin, P., Fildes, R., Lawrence, M., & Nikolopoulos, K. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, *23*(3), 391–404. doi:10.1016/j.ijforecast.2007.05.016

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. doi:10.1037//1040-3590.12.1.19

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*(5), 753–770. doi:10.1109/TSMC.1987.6499282

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology*, *55*, 363–397. doi:10.1111/j.1744-6570.2002.tb00114.x

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. doi:10.1111/j.1754-9434.2008.00058.x

Huffcutt, A. I., & Arthur Jr., W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*(2), 184–190.

Huffcutt, A. I., & Culbertson, S. S. (2011). Interviews. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology, Vol 2: Selecting and Developing Members for* *the Organization* (pp. 185–203). Washington, DC: American Psychological Association. doi:10.1037/12170-006

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment*, *22*(3), 297–309. doi:10.1111/ijsa.12078

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. doi:10.1037/0033-2909.96.1.72

Jacobs, R. R., Conte, J. M., Day, D. V., Silva, J. M., & Harris, R. (1996). Selecting bus drivers: Multiple predictors, multiple perspectives on validity, and multiple estimates of utility. *Human Performance*, *9*(3), 199–217.

Jung, C. G. (1923). *Psychological types*. London, UK: Routledge & Kegan Paul.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. doi:10.1037/a0016755

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kausel, E. E., Culbertson, S. S., Leiva, P. I., Slaughter, J. E., & Jackson, A. T. (2015). Too arrogant for their own good? Why and when narcissists dismiss advice. *Organizational Behavior and Human Decision Processes*, *131*, 33–50. doi:http://dx.doi.org/10.1016/j.obhdp.2015.07.006

Kausel, E. E., Culbertson, S. S., & Madrid, H. (n.d.). Overconfidence and accuracy in recruiter predictions of applicant performance. *Personnel Psychology*.

Klimoski, R., & Jones, R. G. (2008). Intuiting the selection context. *Industrial and*

Organizational Psychology, *1*(3), 352–354. doi:10.1111/j.1754-9434.2008.00061.x

Kuncel, N. R. (2008). Some new (and old) suggestions for improving personnel selection. *Industrial and Organizational Psychology*, *1*(3), 343–346. doi:10.1111/j.1754-9434.2008.00059.x

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060–1072. doi:10.1037/a0034156

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*(2), 460–468. doi:10.1037/a0025741

Louie, T. A. (1999). Decision makers' hindsight bias after receiving favorable and unfavorable feedback. *Journal of Applied Psychology*, *84*(1), 29–41.

Martin, S. L. (2008). Managers also overrely on tests. *Industrial and Organizational Psychology*, *1*(3), 359–360. doi:10.1111/j.1754-9434.2008.00063.x

Morgeson, F. P., Delaney-Klinger, K., & Hemingway, M. A. (2005). The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *Journal of Applied Psychology*, *90*(2), 399–406. doi:10.1037/0021-9010.90.2.399

Mullins, M. E., & Rogers, C. (2008). Reliance on intuition and faculty hiring. *Industrial and Organizational Psychology*, *1*(3), 370–371. doi:10.1111/j.1754-9434.2008.00067.x

Neter, E., & Ben-Shakhar, G. (1989). The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences*, *10*, 737–745.

O'Brien, J. (2008). Interviewer resistance to structure. *Industrial and Organizational*

Psychology, *1*(3), 367–369. doi:10.1111/j.1754-9434.2008.00066.x

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. doi:10.1518/001872097778543886

Phillips, J. M., & Gully, S. M. (2008). The role of perceptions versus reality in managers' choice of selection decision aids. *Industrial and Organizational Psychology*, *1*(3), 361–363. doi:10.1111/j.1754-9434.2008.00064.x

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*, 1–81.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, *35*(2), 1–63. doi:10.1111/j.1744-6570.1982.tb02184.x

Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441–472.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. doi:10.1177/1745691612454303

Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*(4), 1009–1037. doi:10.1111/j.1744-6570.2005.00714.x

Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: Survey results. *Interfaces*, *24*(2), 92–100. doi:10.1287/inte.24.2.92

Sanders, N. R., & Manrodt, K. B. (2003). Forecasting software in practice: Use, satisfaction and performance. *Interfaces*, *33*(5), 90–93. doi:10.1287/inte.33.5.90.19251

Schmidt, F. L., & Hunter, J. (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), 262–274. doi:10.1037//0033-2909.124.2.262

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. doi:10.1037/0022-3514.86.1.162

Shapiro, S., & Spence, M. T. (1997). Managerial intuition: A conceptual and operational framework. *Business Horizons*, *40*(1), 63–68.

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *The Academy of Management Executive*, *1*(1), 57–64. doi:10.5465/AME.1987.4275905

Sinclair, M., & Ashkanasy, N. M. (2005). Intuition: Myth or a decision-making tool? *Management Learning*, *36*(3), 353–370. doi:10.1177/1350507605055351

Slaughter, J. E., & Kausel, E. E. (2014). Employee Selection Decisions. In S. Highhouse, R. S. Dalal, & E. Salas (Eds.), *Judgment and Decision Making at Work* (pp. 57–79). New York, NY: Routledge.

Society for Industrial and Organizational Psychology. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind and Society*, *11*(1), 3–13. doi:10.1007/s11299-011-0093-6

Terpstra, D. E. (1996). The search for effective methods. *HR Focus*, *73*(5), 16–18.

Terpstra, D. E., & Rozell, E. J. (1997). Sources of human resource information and the link to organizational profitability. *Journal of Applied Behavioral Science*, *33*(1), 66–83.

doi:10.1177/0021886397331006

Tett, R. P., & Guterman, H. A. (2000). Situation Trait Relevance, Trait Expression, and Cross-Situational Consistency: Testing a Principle of Trait Activation. *Journal of Research in Personality*, *34*(4), 397–423. doi:10.1006/jrpe.2000.2292

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703–742.

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. *Human Performance*, *12*(1), 1–29.

Thayer, P. W. (2008). That's not the only problem. *Industrial and Organizational Psychology*, *1*(3), 372. doi:10.1111/j.1754-9434.2008.00068.x

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*(3), 167–171. doi:10.1111/j.1467-8721.2007.00497.x

Tuttle, B., & Stocks, M. H. (1997). The effects of task information and outcome feedback on individuals' insight into their decision models. *Decision Sciences*, *28*(2), 421–442.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Vinchur, A. J., Schippmann, J. S., Switzer III, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, *83*(4), 586–597. doi:10.1037/0021-9010.83.4.586

Wofford, J. C., & Goodwin, V. L. (1990). Effects of feedback on cognitive processing and choice of decision style. *Journal of Applied Psychology*, *75*(6), 603–612. doi:10.1037//0021-9010.75.6.603

Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *49*(1), 60–79. doi:10.1016/0749-5978(91)90042-R

Appendix A - Study 1 Materials

Instructions (all conditions)
Thanks for participating in this study. One of the major objectives of personnel selection is to predict candidates' performance based on available information.

In this study, we are interested in how people make hiring decisions using *limited information*. As such, your opinions are very important to us.

The following is from a large airline company. The firm was validating their selection procedures for the Ticket Agent job. As such, more than 200 applicants took a standardized personality test (conscientiousness factor), standardized cognitive ability test, and completed an unstructured interview before being hired. Three months after being hired, these same individuals were assessed by their supervisors in terms of their general performance.

On the following pages, you'll be presented with pre-hiring information of 10 pairs of applicants. Based on this information, for each pair, we ask you to

- Make a prediction of each candidate's potential job performance as rated by his or her supervisor, and
- Choose which candidate should be hired.

Information about the decision aid (decision aid present condition)
According to research examining various selection procedures, scores on standardized cognitive ability tests are good predictors of future job performance. Scores on the conscientiousness factor of standardized personality tests are moderate predictors of future job performance. Lastly, scores on unstructured interviews are weak predictors of future job performance. Based on this information, one can use the following equation to estimate a candidate's job performance

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance

For example, if an individual's scores were cognitive ability = 50, conscientiousness = 100 and unstructured interview = 75, then

$$0.50 \times (50) + 0.40 \times (100) + 0.10 \times (75) = \text{Predicted Job Performance}$$
$$25+40+7.5 = \text{Predicted Job Performance}$$
$$72.5 = \text{Predicted Job Performance}$$

Decision 1
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating

Candidate A	85	95	50
Candidate B	82	09	70

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 5 = Predicted Job Performance

Candidate B:
41 + 3.6 + 7 = Predicted Job Performance

Decision 2
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	66	11	50
Candidate B	85	07	30

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 4.4 + 5 = Predicted Job Performance

Candidate B:

42.5 + 2.8 + 3 = Predicted Job Performance

Decision 3
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	51	23	90
Candidate B	41	28	30

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
25.5 + 9.2 + 9 = Predicted Job Performance

Candidate B:
20.5 + 11.2 + 3 = Predicted Job Performance

Decision 4
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	26	73	70
Candidate B	26	35	90

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
13 + 29.2 + 7 = Predicted Job Performance

Candidate B:
13 + 14 + 9 = Predicted Job Performance

Decision 5
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	36	28	50
Candidate B	46	86	70

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
18 + 11.2 + 5 = Predicted Job Performance

Candidate B:
23 + 34.4 + 7 = Predicted Job Performance

Decision 6
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	33	35	90

Candidate B	12	68	90

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
16.5 + 14 + 9 = Predicted Job Performance

Candidate B:
6 + 27.2 + 9 = Predicted Job Performance

Decision 7
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	85	95	70
Candidate B	61	28	50

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 7 = Predicted Job Performance

Candidate B:
30.5 + 11.2 + 5 = Predicted Job Performance

Decision 8

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	09	51	30
Candidate B	93	86	50

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition

Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:

4.5 + 20.4 + 3 = Predicted Job Performance

Candidate B:

46.5 + 34.4 + 5 = Predicted Job Performance

Decision 9

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	66	51	90
Candidate B	04	23	30

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition

Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 20.4 + 9 = Predicted Job Performance

Candidate B:
2 + 9.2 + 3 = Predicted Job Performance

Decision 10
Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating
Candidate A	41	35	70
Candidate B	57	35	50

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
20.5 + 14 + 7 = Predicted Job Performance

Candidate B:
28.5 + 14 + 5 = Predicted Job Performance

Appendix B - Study 2 Materials

Instructions (all conditions)
Thanks for participating in this study. One of the major objectives of personnel selection is to predict candidates' performance based on available information.

In this study, we are interested in how people make hiring decisions using *limited information*. As such, your opinions are very important to us.

The following is from a large airline company. The firm was validating their selection procedures for the Ticket Agent job. As such, more than 200 applicants took a standardized personality test (conscientiousness factor), standardized cognitive ability test, and completed an unstructured interview before being hired. Three months after being hired, these same individuals were assessed by their supervisors in terms of their general performance.

On the following pages, you'll be presented with pre-hiring information of 20 pairs of applicants. Based on this information, for each pair, we ask you to

- Make a prediction of each candidate's potential job performance as rated by his or her supervisor, and
- Choose which candidate should be hired.

Information about the decision aid (decision aid present condition)
According to research examining various selection procedures, scores on standardized cognitive ability tests are good predictors of future job performance. Scores on the conscientiousness factor of standardized personality tests are moderate predictors of future job performance. Lastly, scores on unstructured interviews are weak predictors of future job performance. Lastly, scores on the handwriting analysis do not predict future job performance. Based on this information, one can use the following equation to estimate a candidate's job performance

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis) = Predicted Job Performance

For example, if an individual's scores were cognitive ability = 50, conscientiousness = 100, unstructured interview = 75, and handwriting analysis = 65, then

$$0.50 \times (50) + 0.40 \times (100) + 0.10 \times (75) + 0.00 \times (65) = \text{Predicted Job Performance}$$
$$25 + 40 + 7.5 + 0 = \text{Predicted Job Performance}$$
$$72.5 = \text{Predicted Job Performance}$$

Decision 1

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	85	95	50	54
Candidate B	82	09	70	62

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance
Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 5 + 0 = Predicted Job Performance

Candidate B:
41 + 3.6 + 7 + 0 = Predicted Job Performance

Decision 2

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	66	11	50	41
Candidate B	85	7	30	29

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 4.4 + 5 + 0 = Predicted Job Performance

Candidate B:
42.5 + 2.8 + 3 + 0 = Predicted Job Performance

Decision 3

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	51	23	90	86
Candidate B	41	28	30	18

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
25.5 + 9.2 + 9 + 0 = Predicted Job Performance

Candidate B:
20.5 + 11.2 + 3 + 0 = Predicted Job Performance

Decision 4

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	26	73	70	87
Candidate B	26	35	90	36

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
13 + 29.2 + 7 + 0 = Predicted Job Performance

Candidate B:
13 + 14 + 9 + 0 = Predicted Job Performance

Decision 5

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	36	28	50	50
Candidate B	46	86	70	51

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
18 + 11.2 + 5 + 0 = Predicted Job Performance

Candidate B:
23 + 34.4 + 7 + 0 = Predicted Job Performance

Decision 6

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	33	35	90	99
Candidate B	12	68	90	93

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
16.5 + 14 + 9 + 0 = Predicted Job Performance

Candidate B:
6 + 27.2 + 9 + 0 = Predicted Job Performance

Decision 7

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	85	95	70	60
Candidate B	61	28	50	31

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 7 + 0 = Predicted Job Performance

Candidate B:
30.5 + 11.2 + 5 + 0 = Predicted Job Performance

Decision 8

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	9	51	30	63
Candidate B	93	86	50	91

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
4.5 + 20.4 + 3 + 0 = Predicted Job Performance

Candidate B:
46.5 + 34.4 + 5 + 0 = Predicted Job Performance

Decision 9

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	66	51	90	77
Candidate B	4	23	30	70

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 20.4 + 9 + 0 = Predicted Job Performance

Candidate B:
2 + 9.2 + 3 + 0 = Predicted Job Performance

Decision 10

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	41	35	70	61
Candidate B	57	35	50	68

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
20.5 + 14 + 7 + 0 = Predicted Job Performance

Candidate B:
28.5 + 14 + 5 + 0 = Predicted Job Performance

Decision 11

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	66	11	50	41
Candidate B	4	23	30	70

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 4.4 + 5 + 0 = Predicted Job Performance

Candidate B:
2 + 9.2 + 3 + 0 = Predicted Job Performance

Decision 12

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	26	73	70	87
Candidate B	61	28	50	31

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
13 + 29.2 + 7 + 0 = Predicted Job Performance

Candidate B:
30.5 + 11.2 + 5 + 0 = Predicted Job Performance

Decision 13

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	33	35	90	99
Candidate B	46	86	70	51

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
16.5 + 14 + 9 + 0 = Predicted Job Performance

Candidate B:
23 + 34.4 + 7 + 0 = Predicted Job Performance

Decision 14

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	9	51	30	63
Candidate B	41	28	30	18

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
4.5 + 20.4 + 3 + 0 = Predicted Job Performance

Candidate B:
20.5 + 11.2 + 3 + 0 = Predicted Job Performance

Decision 15

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	41	35	70	61
Candidate B	82	9	70	62

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
20.5 + 14 + 7 + 0 = Predicted Job Performance

Candidate B:
41 + 3.6 + 7 + 0 = Predicted Job Performance

Decision 16

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	85	95	50	54
Candidate B	57	35	50	68

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 5 + 0 = Predicted Job Performance

Candidate B:
28.5 + 14 + 5 + 0 = Predicted Job Performance

Decision 17

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	51	23	90	86
Candidate B	93	86	50	91

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
25.5 + 9.2 + 9 + 0 = Predicted Job Performance

Candidate B:
46.5 + 34.4 + 5 + 0 = Predicted Job Performance

Decision 18

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	36	28	50	50
Candidate B	12	68	90	93

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
18 + 11.2 + 5 + 0 = Predicted Job Performance

Candidate B:
6 + 27.2 + 9 + 0 = Predicted Job Performance

Decision 19

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	85	95	70	60
Candidate B	26	35	90	36

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
42.5 + 38 + 7 + 0 = Predicted Job Performance

Candidate B:
13 + 14 + 9 + 0 = Predicted Job Performance

Decision 20

Below is the information for two candidates. Use this information to predict each applicant's job performance and identify which candidate the organization should hire.

	Cognitive Ability Test Percentile Rank	Conscientiousness Test Percentile Rank	Unstructured Interview Rating	Handwriting Analysis Percentile Rank
Candidate A	66	51	90	77
Candidate B	85	7	30	29

(Note: *Percentile* is the percentage of individuals who score less than the candidate. For example, a percentile score of 50 on the cognitive ability test means that the candidate performed better than 50% of the other individuals).

Participants assigned to decision aid condition
Recall that the prediction formula was:

0.50 x (cognitive ability score) + 0.40 x (conscientiousness score) + 0.10 x (unstructured interview score) + 0.00 x (handwriting analysis score) = Predicted Job Performance

Based on the scores for each candidate, the formula for each candidate is:

Candidate A:
33 + 20.4 + 9 + 0 = Predicted Job Performance

Candidate B:
42.5 + 2.8 + 3 + 0 = Predicted Job Performance

Appendix C - Figures

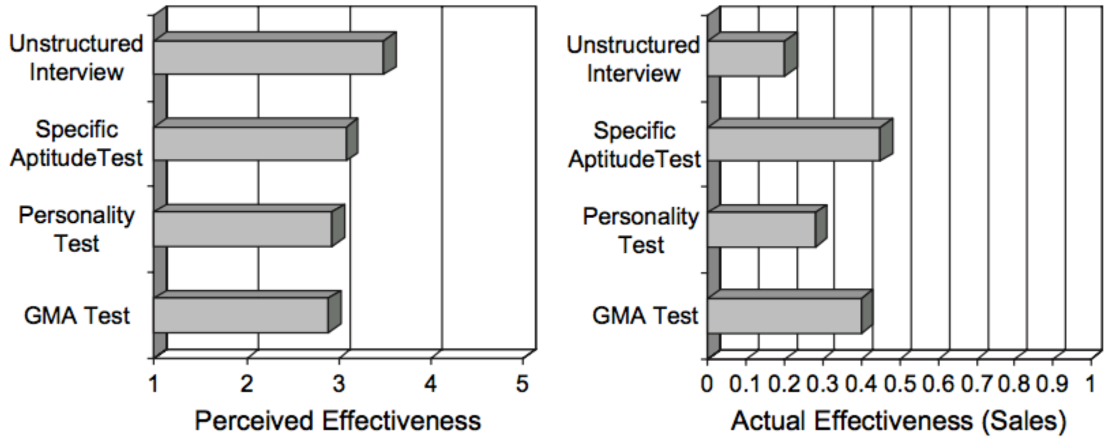

Figure 1. Perceived versus actual effectiveness of various selection methods. From "Stubborn Reliance on Intuition and Subjectivity in Employee Selection" by S. Highhouse, 2008, *Industrial and Organizational Psychology, 1,* p. 334. Reprinted with permission.

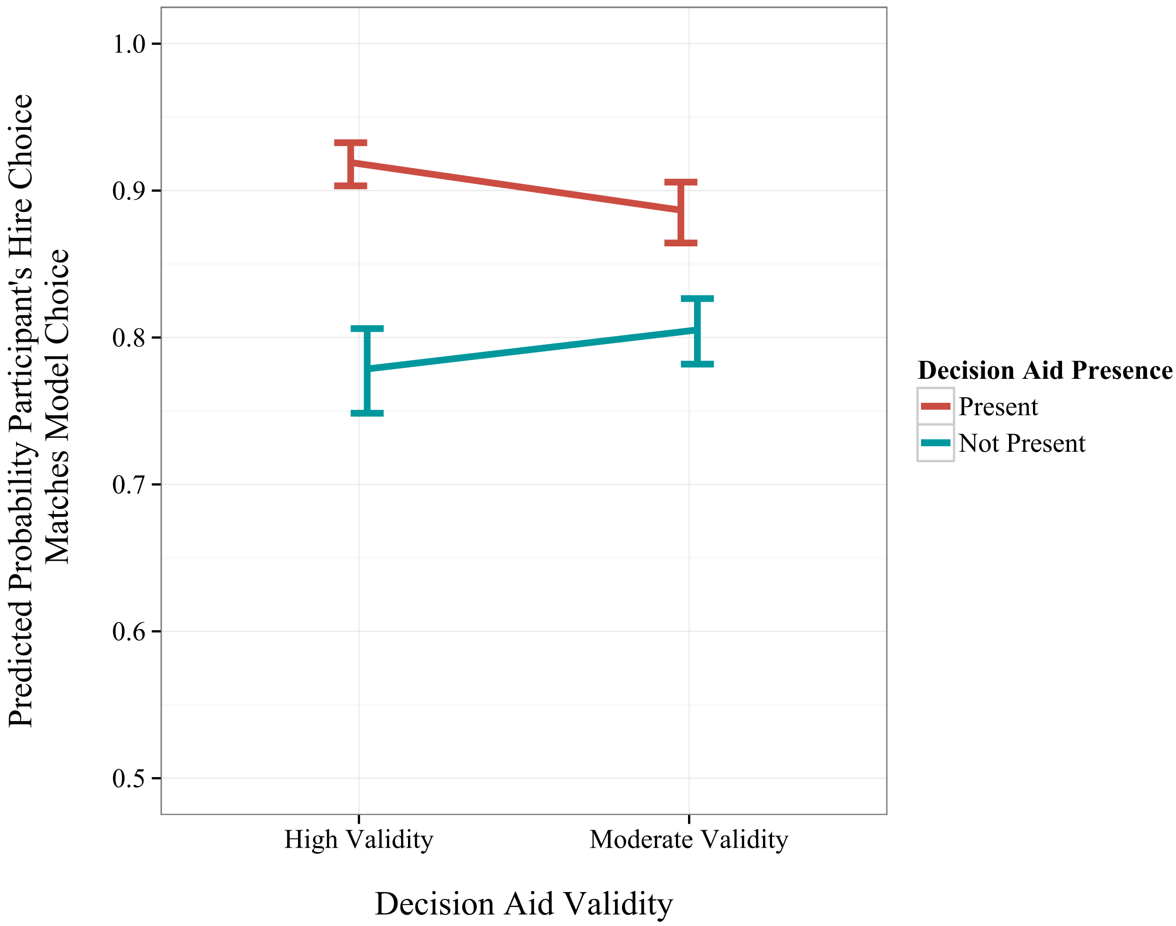

Figure 2. Predicted match in hiring choice in study 1.

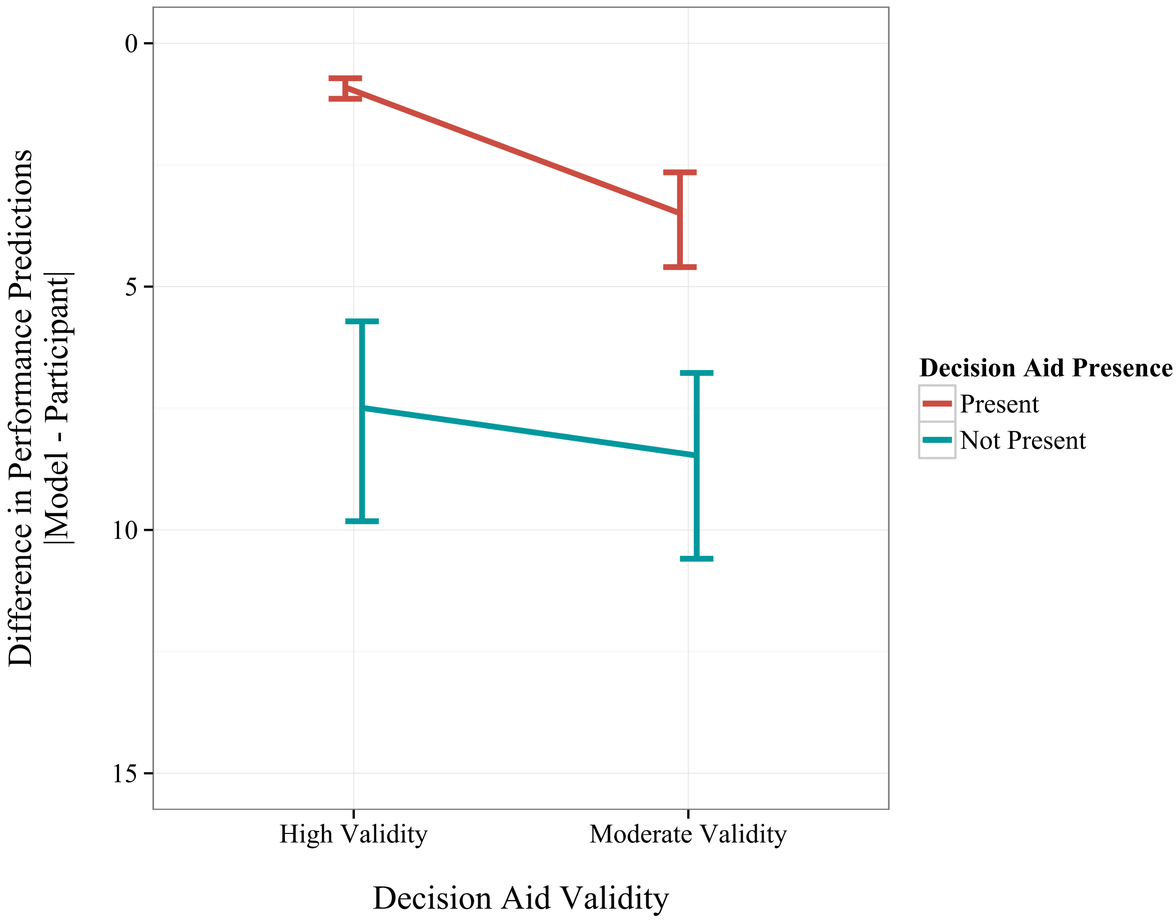

Figure 3. Predicted match in performance predictions in study 1. Note that the y-axis has been inverted to ease comparison across operationalizations of decision aid reliance.

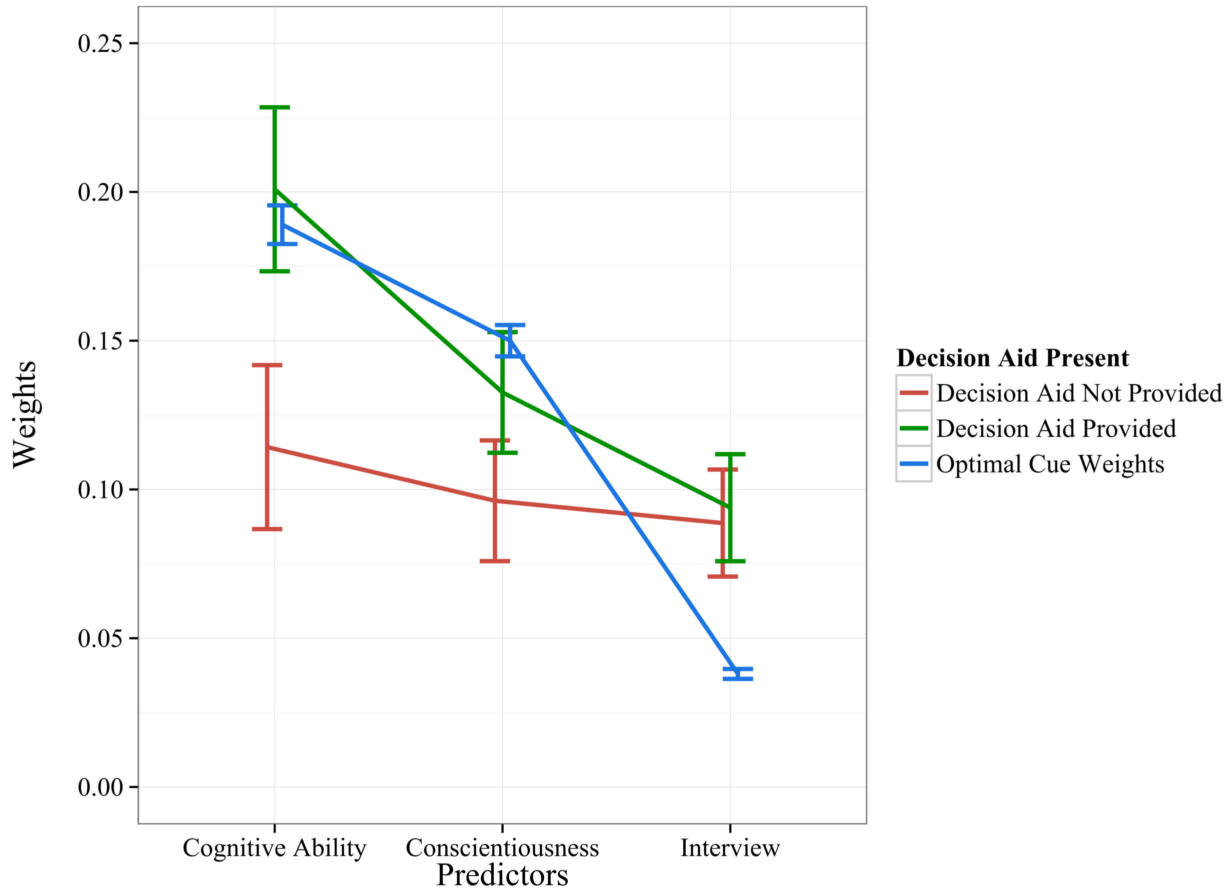

Figure 4. Cue utilization in study 1

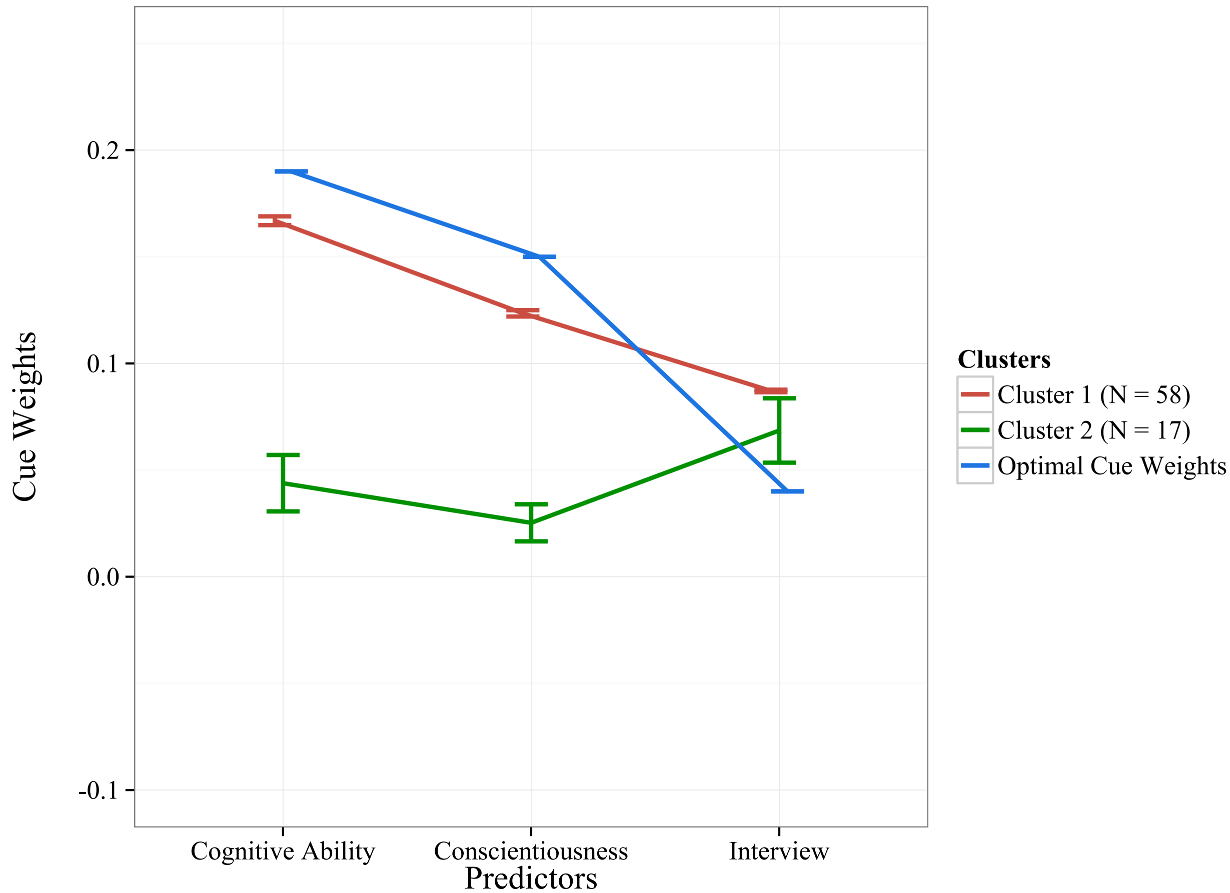

Figure 5. Cluster analysis of random effects of participants' cue weighting when the decision aid is provided. Error bars represent +/- 1 standard deviation.

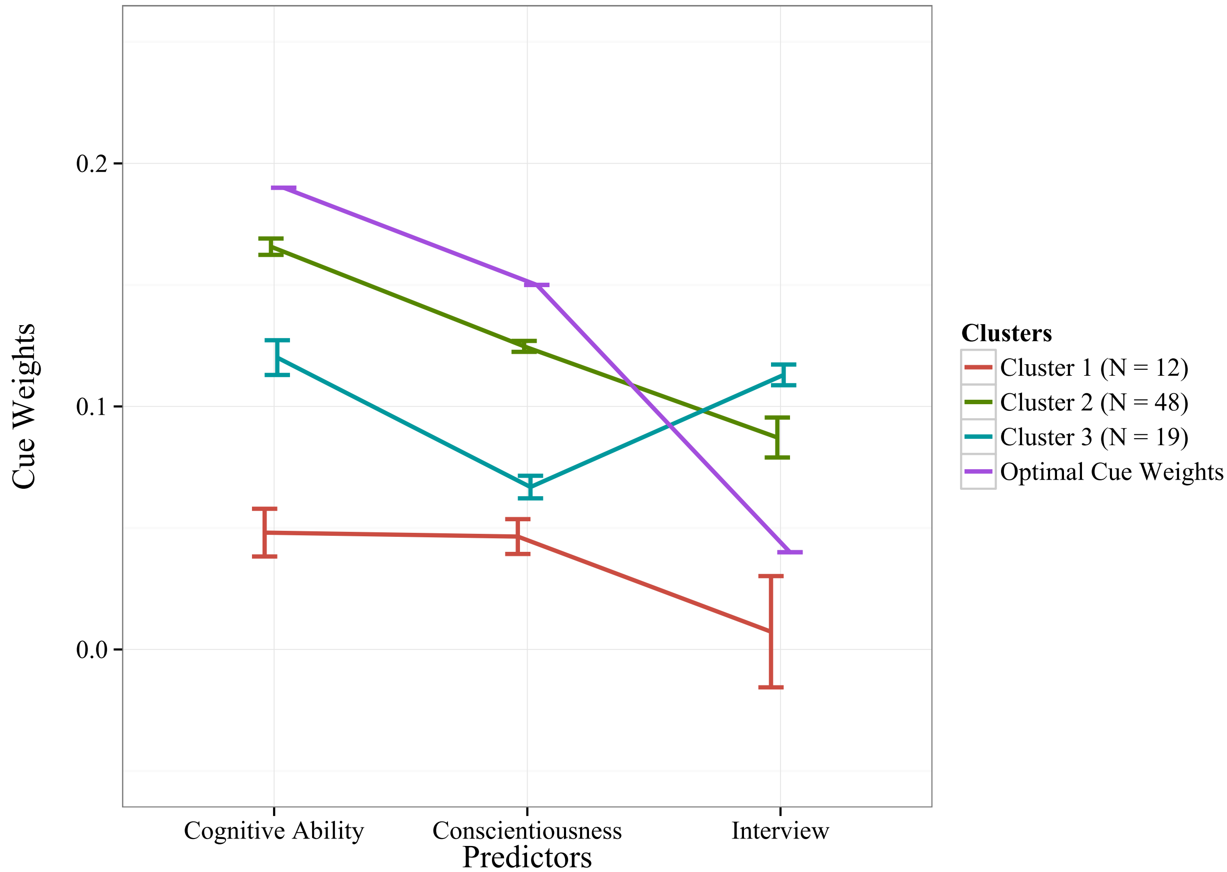

Figure 6. Cluster analysis of random effects of participants' cue weighting when the decision aid is not provided. Error bars represent +/- 1 standard error.

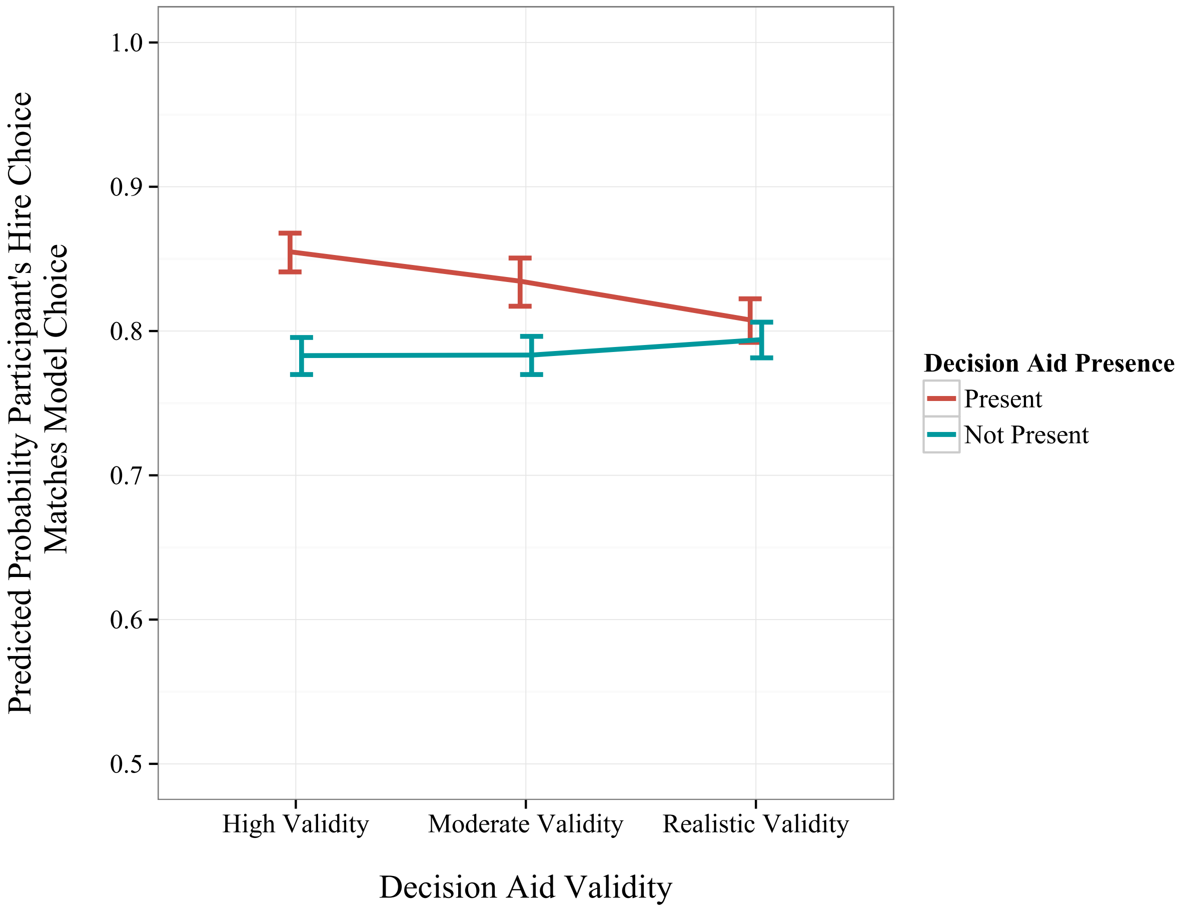

Figure 7. Predicted match in hiring choice in study 2.

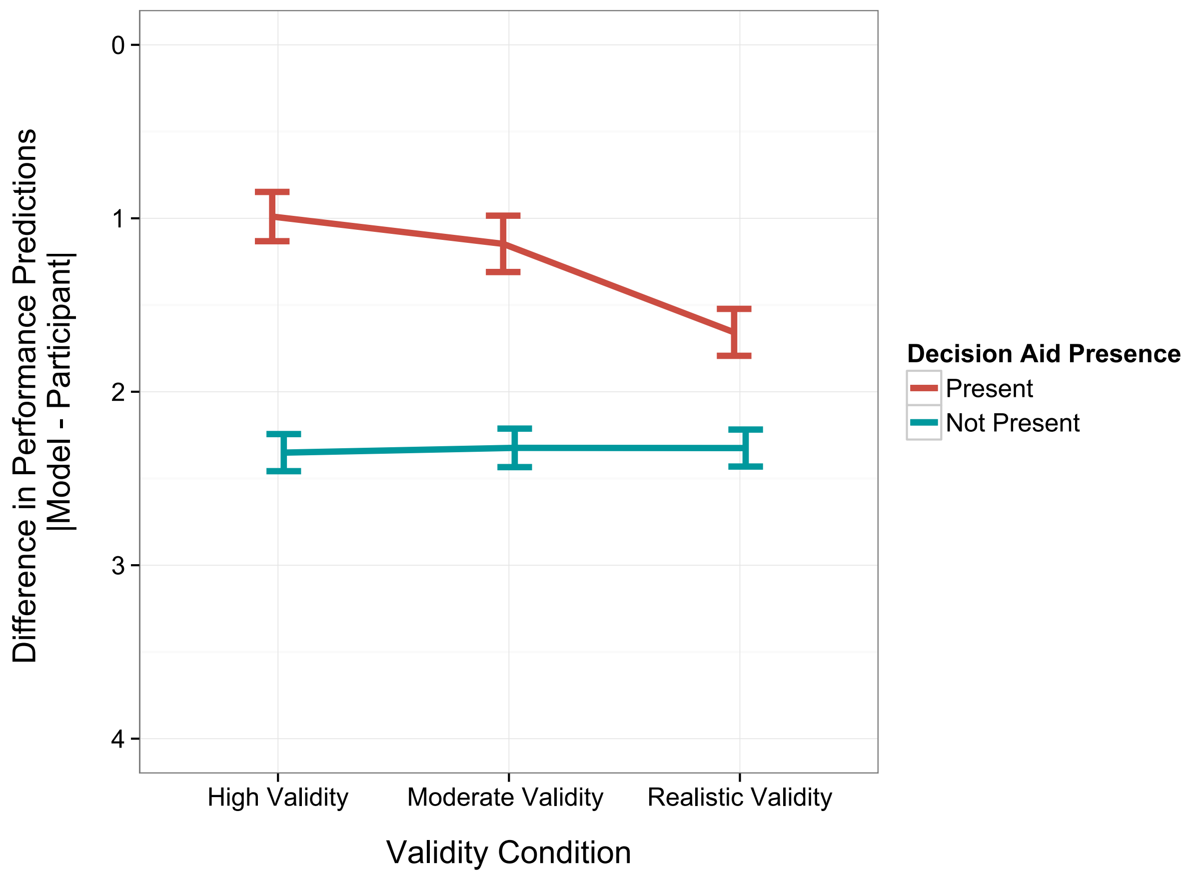

Figure 8. Predicted match in performance predictions in study 2. Note that the y-axis has been inverted to ease comparison across operationalizations of decision aid reliance.

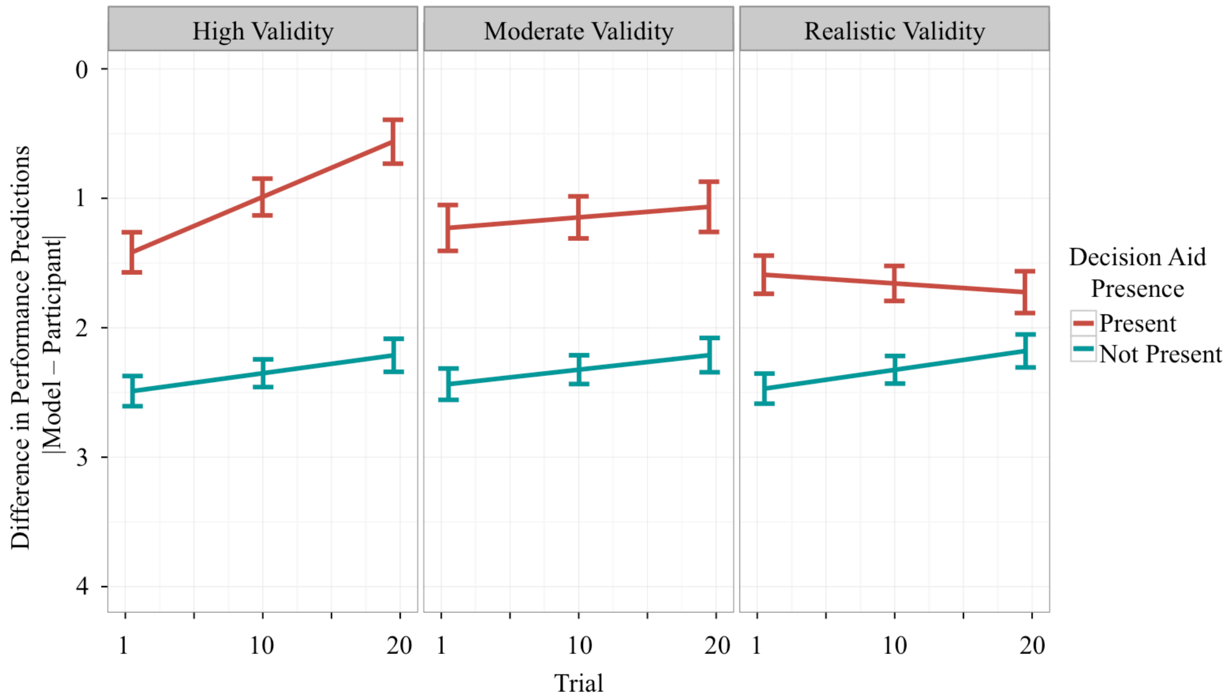

Figure 9. Three-way interaction between cue validity, decision aid presence, and trial. Note that the y-axis has been inverted for ease of comparison across operationalizations of decision aid reliance.

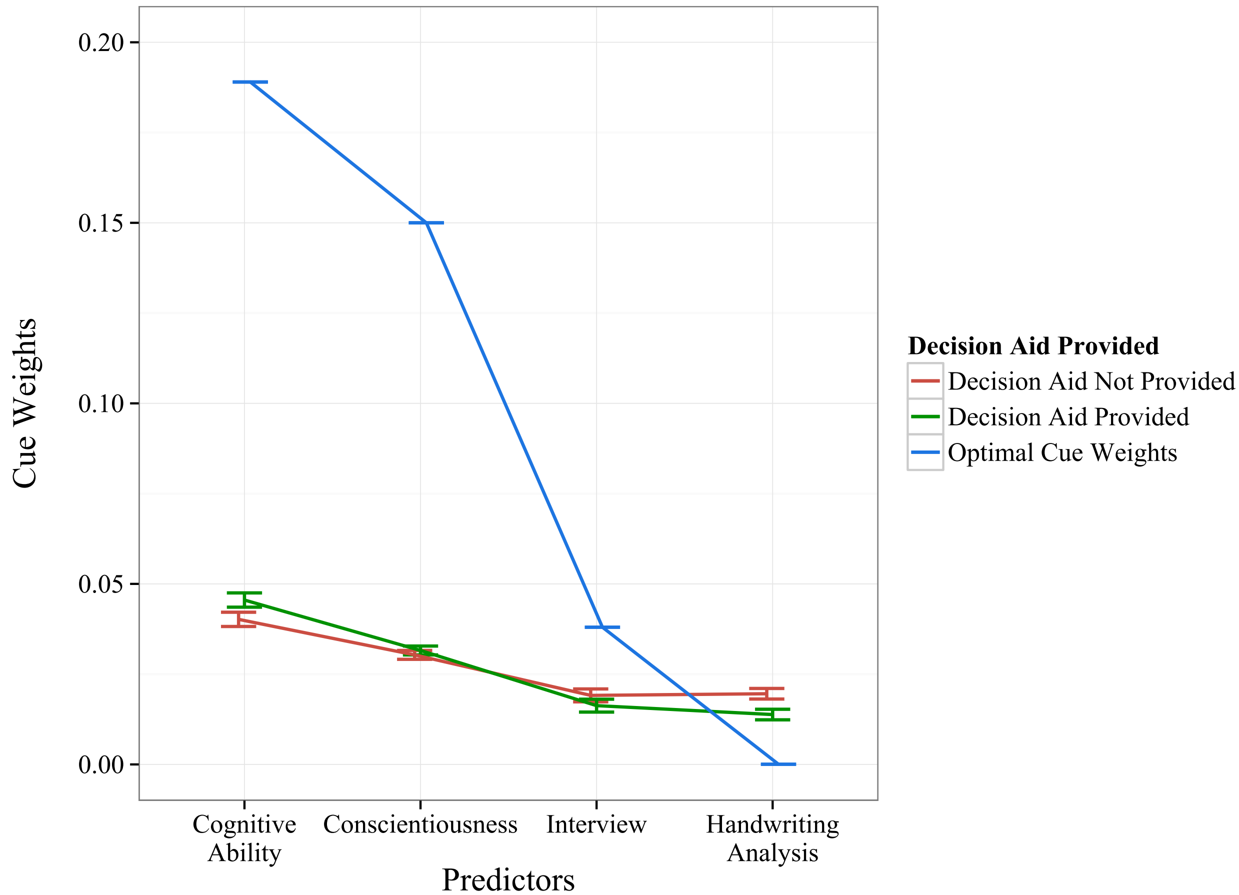

Figure 10. Cue utilization in study 2

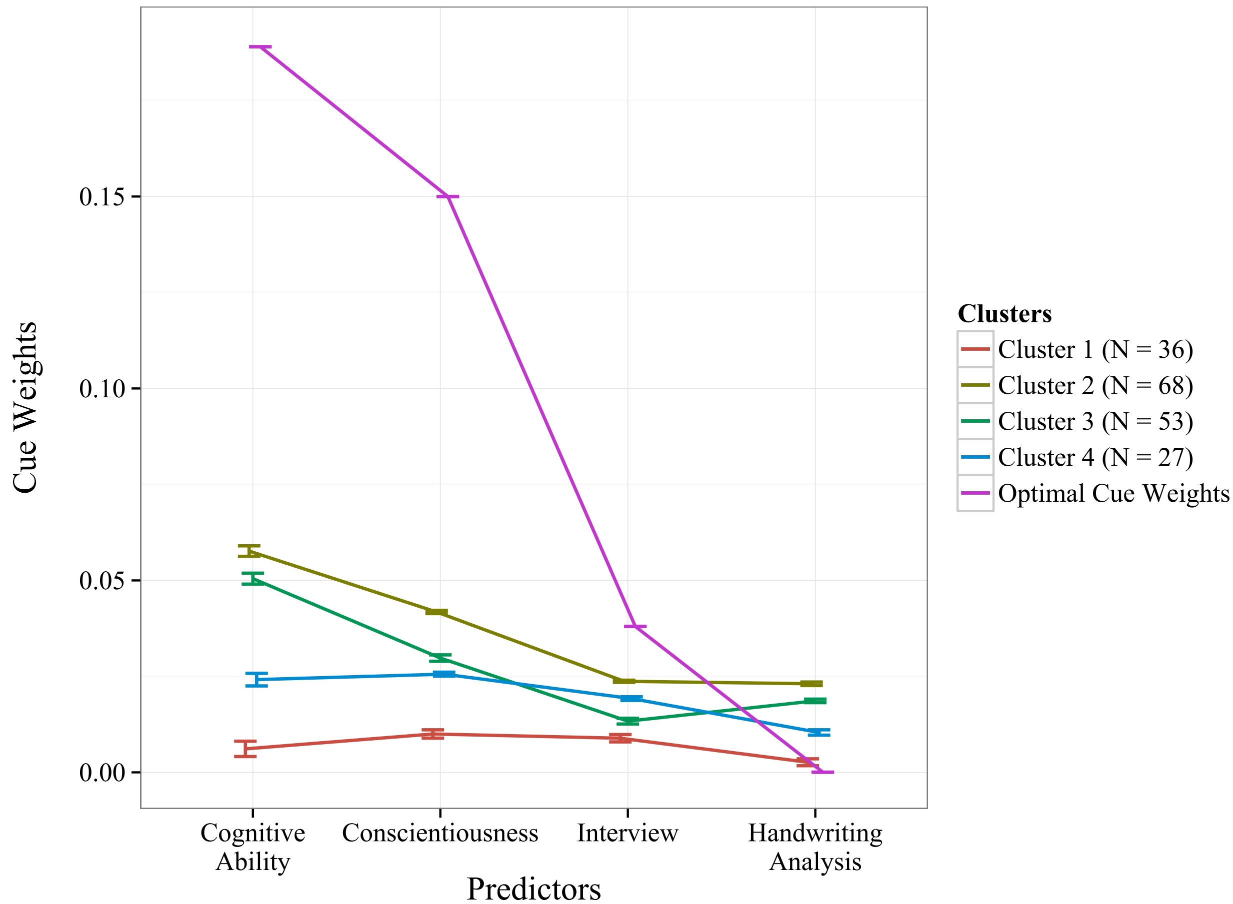

Figure 11. Cluster analysis of random effects of participants' cue weighting when the decision aid is provided. Error bars represent +/- 1 standard error.

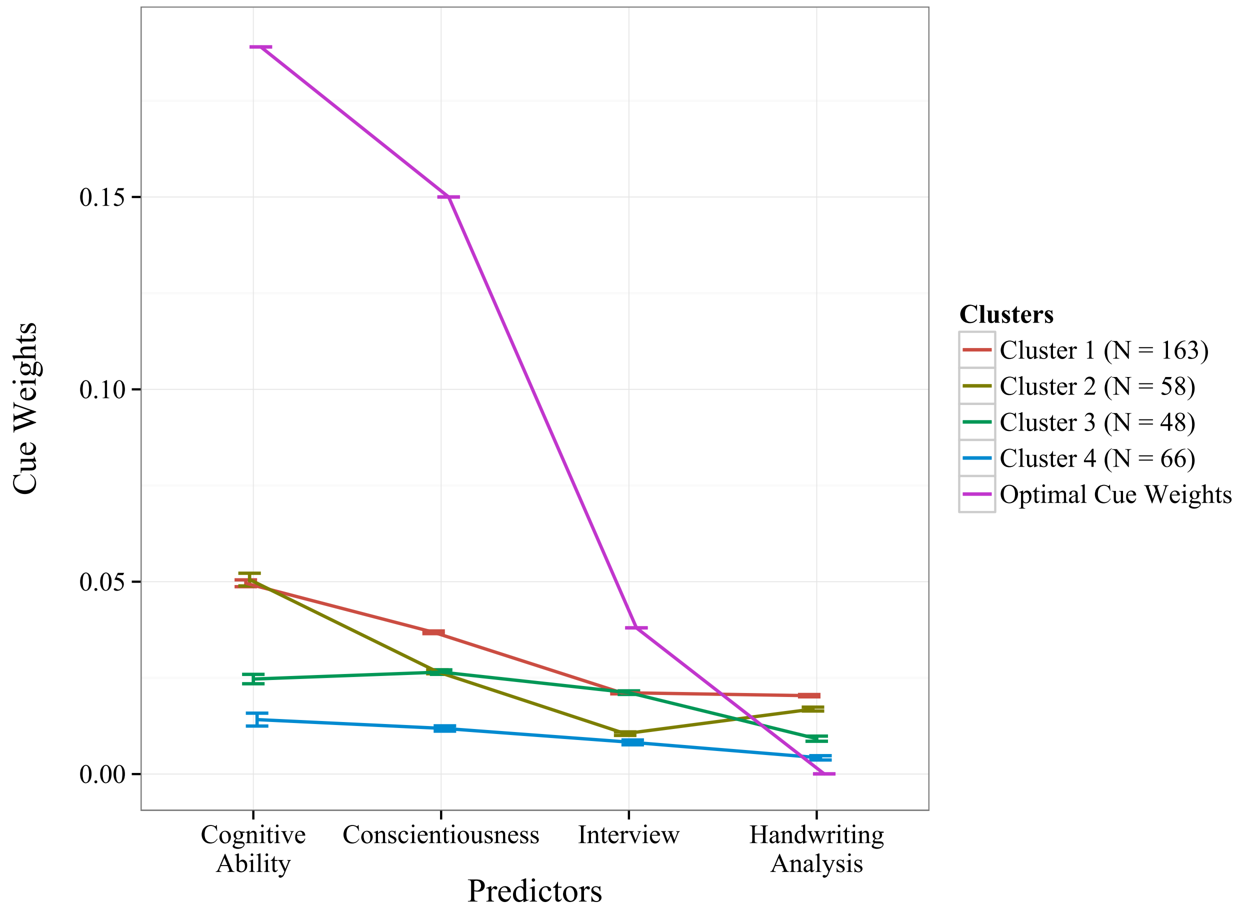

Figure 12. Cluster analysis of random effects of participants' cue weighting when the decision aid is not provided. Error bars represent +/- 1 standard error.

Appendix D - Tables

Table 1. Model effects predicting match in hiring choice

	B	*SEB*	*z*	*p*
Intercept	1.791	0.096	12.728	< .001
Cue Validity	0.053	0.088	0.604	.546
Decision Aid Presence	0.453	0.088	5.126	< .001
Trial	0.025	0.026	0.963	.336
Cue Validity * Decision Aid Presence	0.134	0.088	1.526	.127
Cue Validity * Trial	0.042	0.026	1.587	.113
Decision Aid Presence * Trial	0.002	0.026	0.062	.951
Cue Validity * Decision Aid Presence * Trial	0.035	0.026	1.343	.179

Table 2. Model effects predicting match in candidate performance predictions

	B	*SEB*	*t*
Intercept	1.325	0.126	10.555
Cue Validity	-0.368	0.126	-2.932
Decision Aid Presence	-0.750	0.126	-5.970
Trial	-0.003	0.010	-0.340
Cue Validity * Decision Aid Presence	-0.307	0.126	-2.443
Cue Validity * Trial	-0.009	0.010	-0.914
Decision Aid Presence * Trial	0.010	0.010	1.029
Cue Validity * Decision Aid Presence * Trial	0.019	0.010	1.898

Table 3. Summary of model structures examining cue utilization

Model	Fixed Effects	Random Effects
Model 3	D.Cog.Ability	D.Cog.Ability
	D.Conscientiousness	D.Conscientiousness
	D.Interview	D.Interview
	Decision Aid Presence	
	D.Cog.Ability * Decision Aid Presence	
	D.Conscientiousness * Decision Aid Presence	
	D.Interview * Decision Aid Presence	
Model 3a	D.Cog.Ability	D.Cog.Ability
	D.Conscientiousness	D.Conscientiousness
	D.Interview	D.Interview
	Decision Aid Presence	
	Cue Validity	
	Decision Aid Presence * Cue Validity	
	D.Cog.Ability * Decision Aid Presence	
	D.Cog.Ability * Cue Validity	
	D.Cog.Ability * Decision Aid Presence * Cue Validity	
	D.Conscientiousness * Decision Aid Presence	
	D.Conscientiousness * Cue Validity	
	D.Conscientiousness * Decision Aid Presence * Cue Validity	
	D.Interview * Decision Aid Presence	
	D.Interview * Cue Validity	
	D.Interview * Decision Aid Presence * Cue Validity	

Note. D.Cog.Ability = difference between candidate A and candidate B on cognitive ability, D.Conscientiousness = difference between candidate A and candidate B on conscientiousness, D.Interview = difference between candidate A and candidate B on unstructured interview

Table 4. Model summary for determining cue utilization

	B	*SEB*	*Z*	*p*
(Intercept)	1.869	0.428	4.364	< .001
D.Cog.Ability	0.158	0.028	5.718	< .001
D.Conscientiousness	0.114	0.020	5.639	< .001
D.Interview	0.091	0.018	5.075	< .001
Model Presence	-0.148	0.151	-0.978	.328
D.Cog.Ability*Model Presence	0.043	0.012	3.650	< .001
D.Conscientiousness*Model Presence	0.018	0.008	2.303	.021
D.Interview*Model Presence	0.003	0.009	0.281	.778

Note. D.Cog.Ability = difference between candidate A and candidate B on cognitive ability, D.Conscientiousness = difference between candidate A and candidate B on conscientiousness, D.Interview = difference between candidate A and candidate B on unstructured interview

Table 5. Model effects predicting match in hiring choice

	B	*SEB*	*Z*	*p*
(Intercept)	1.458	.039	37.030	< .001
Cue Validity 1	0.071	.054	1.320	.188
Cue Validity 2	-0.006	.056	-0.100	.920
Decision Aid Presence	0.152	.038	3.950	< .001
Feedback	-0.085	.038	-2.220	.027
Cue Validity 1 * Decision Aid Presence	0.094	.054	1.740	.081
Cue Validity 2 * Decision Aid Presence	0.015	.056	0.270	.791
Cue Validity 1 * Feedback	0.059	.054	1.100	.270
Cue Validity 2 * Feedback	-0.073	.056	-1.290	.196
Decision Aid Presence * Feedback	0.011	.038	0.300	.764
Cue Validity 1 * Decision Aid Presence * Feedback	0.043	.054	0.800	.424
Cue Validity 2 * Decision Aid Presence * Feedback	-0.067	.056	-1.180	.237

Note. Variables were coded using effects coding. Cue validity 1 was coded as 1 = highly valid cues, 0 = moderately valid cues, -1 = low validity cues. Cue validity 2 was coded as 0 = highly valid cues, 1 = moderately valid cues, -1 = low validity cues. Decision aid presence was coded as 1 = decision aid present, -1 = decision aid not present. Feedback was coded as 1 = feedback provided, -1 = feedback not provided.

Table 6. Model effects predicting match in candidate performance predictions

	B	*SEB*	*t*
(Intercept)	1.799	.053	34.130
Cue Validity1	-0.128	.074	-1.750
Cue Validity2	-0.064	.077	-0.820
Decision Aid Presence	-0.534	.053	-10.130
Feedback	0.045	.053	0.860
Trial	-0.015	.003	-4.780
Cue Validity1 * Decision Aid Presence	-0.146	.074	-1.990
Cue Validity2 * Decision Aid Presence	-0.054	.077	-0.700
Cue Validity1 * Feedback	-0.105	.074	-1.430
Cue Validity2 * Feedback	0.076	.077	0.980
Decision Aid Presence * Feedback	0.084	.053	1.590
Cue Validity1 * Trial	-0.015	.004	-3.510
Cue Validity2 * Trial	0.004	.005	0.990
Decision Aid Presence * Trial	-0.001	.003	-0.260
Feedback * Trial	0.002	.003	0.590
Cue Validity1 * Decision Aid Presence * Feedback	-0.135	.074	-1.830
Cue Validity2 * Decision Aid Presence * Feedback	0.104	.077	1.340
Cue Validity1 * Decision Aid Presence * Trial	-0.014	.004	-3.360
Cue Validity2 * Decision Aid Presence * Trial	0.002	.005	0.530
Cue Validity1 * Feedback * Trial	-0.009	.004	-2.000
Cue Validity2 * Feedback * Trial	0.005	.005	1.000
Decision Aid Presence * Feedback * Trial	0.002	.003	0.620
Cue Validity1 * Decision Aid Presence * Feedback * Trial	-0.007	.004	-1.630
Cue Validity2 * Decision Aid Presence * Feedback * Trial	0.003	.005	0.610

Table 7. Summary of model structures examining cue utilization

Model	Fixed Effects	Random Effects
Model 6	D.Cog.Ability	D.Cog.Ability
	D.Conscientiousness	D.Conscientiousness
	D.Interview	D.Interview
	D.Handwriting	D.Handwriting
	Model Presence	
Model 6a	D.Cog.Ability	D.Cog.Ability
	D.Conscientiousness	D.Conscientiousness
	D.Interview	D.Interview
	D.Handwriting	D.Handwriting
	Model Presence	
	D.Cog.Ability*Model Presence	
	D.Conscientiousness*Model Presence	
	D.Interview*Model Presence	
	D.Handwriting*Model Presence	

Note. D.Cog.Ability = difference between candidate A and candidate B on cognitive ability, D.Conscientiousness = difference between candidate A and candidate B on conscientiousness, D.Interview = difference between candidate A and candidate B on unstructured interview, D.Handwriting = difference between candidate A and candidate B on handwriting analysis score.

Table 8. Model summary for determining cue utilization

	B	*SEB*	*Z*	*p*
(Intercept)	0.587	.042	13.890	< .001
D.Cog.Ability	0.043	.002	21.693	< .001
Decision Aid Presence	-0.249	.039	-6.372	< .001
D.Conscientiousness	0.031	.001	25.206	< .001
D.Interview	0.018	.002	9.918	< .001
D.Handwriting	0.017	.001	11.406	< .001
D.Cog.Ability * Decision Aid Presence	0.003	.002	1.537	0.124
D.Conscientiousness * Decision Aid Presence	0.001	.001	0.558	0.577
D.Interview * Decision Aid Presence	-0.001	.001	-0.963	0.336
D.Handwriting * Decision Aid Presence	-0.003	.001	-2.167	0.030

Note. D.Cog.Ability = difference between candidate A and candidate B on cognitive ability, D.Conscientiousness = difference between candidate A and candidate B on conscientiousness, D.Interview = difference between candidate A and candidate B on unstructured interview, and D.Handwriting = difference between candidate A and candidate B on handwriting analysis