THE GARGANTUAN GAP: A MODEL OF USER REACTIONS TOWARD AND BELIEFS ABOUT EMPLOYEE SELECTION PROCEDURES

By

Connor Eichenauer

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2024

ABSTRACT

Candidate results from multiple employee selection procedures are most often combined clinically using hiring managers' judgment, but evidence suggests this approach attenuates predictive validity compared to mechanical data combination approaches due to hiring managers' misconceptions about selection procedures. The present research proposes and tests a model that explains how and why hiring managers determine the extent to which they will utilize candidate results from various selection procedures or sources of candidate information. Specifically, the model posits that utilization is driven by user beliefs about the predictiveness and fairness of procedures, which are in turn informed by nine dimensions of "user reactions" or perceptions of properties of predictor methods (procedural autonomy, evaluation autonomy, fidelity, fakability, evaluation consistency, and transparency) and predictor constructs (job relatedness, malleability, and development equity). Study 1 tested the model by asking a sample of hiring managers to view and rate selection procedures manipulated into nine predictor method/construct combinations. Results indicated that users' predictiveness beliefs are a stronger driver of utilization than fairness beliefs; additionally, most dimensions of user reactions predicted utilization intentions in the hypothesized direction. Study 2 evaluated the efficacy of autonomy-based interventions by having a sample of hiring managers view and rate a structured interview or computerized assessment manipulated into high and low levels of autonomy. Results suggested that slightly increasing hiring manager autonomy within standardized selection procedures led to increases in predictiveness beliefs and utilization intentions. Findings have implications for designing selection procedures and tailoring hiring manager communication and training efforts, both with the goal of encouraging utilization of more valid predictors in employee selection decision-making contexts.

TABLE OF CONTENTS

Introduction

Employee selection is one of the oldest and most commonly studied topics in the field of industrial and organizational (I-O) psychology (Ployhart et al., 2017). Broadly, the primary goal of selection systems is to enable decision-makers[1] (e.g., hiring managers in employment selection settings or admissions officers in academic admissions settings[2]) to predict the future success of candidates with greater accuracy than chance; doing so can result in competitive advantage and positive financial outcomes for organizations (Ployhart & Weekley, 2017). To aid in predicting which candidates will be successful as employees, organizations commonly collect information[3] about candidates via selection procedures. These selection procedures, or predictors, measure candidate attributes that are designed to predict a criterion of interest, such as future job performance or turnover, and are used to help guide decisions about which candidates are most likely to succeed. Accordingly, a primary focus of selection research has been to develop selection procedures that predict outcomes to the greatest degree possible.

The way in which selection decision-makers choose to combine or utilize candidate data from selection procedures when making overall candidate judgments can vary. In some selection systems, candidate data from multiple selection procedures is combined mechanically[4] such that candidate information from different predictors is weighted in a standardized manner, such as optimal weighting (i.e., based on estimates from local validity studies or validity generalization methods, predictors are weighted based on their relative relationships with outcomes of interest).

[1] For simplicity, the terms hiring managers, decision-makers, and users will be used relatively interchangeably, all referring to those who evaluate candidates and selection information in selection contexts.
[2] The primary focus of this dissertation is on personnel selection contexts although academic admissions decisions contexts are also relevant to this topic.
[3] In this dissertation, the terms "candidate information" or simply "information" will be used to refer to individual pieces of data about job candidates derived from selection procedures or predictors (e.g., a score from a cognitive ability assessment).
[4] For the purposes of this dissertation, the terms mechanical judgment and mechanical data combination will refer to standardized data integration processes that do not involve manual integration by humans.

In this way, mechanical data combination systems utilize algorithms to produce overall candidate judgments. In other systems, candidate data is combined via decision makers' clinical judgment[5] (i.e., users determine how to weight candidate information into overall candidate evaluations as they see fit). Importantly, meta-analytic evidence suggests that mechanical data combination produces superior candidate predictions compared to clinical judgments in personnel selection judgment contexts (Kuncel et al., 2013) and other contexts (e.g., Grove et al., 2000). This means that when clinical judgment processes are relied upon to combine candidate data into overall evaluations, the predictive validity of selection procedures that produce these pieces of candidate data is likely attenuated compared to if candidate data from multiple predictors were combined using a mechanical method. However, evidence suggests that clinical data combination is most common in organizational selection settings (Neumann et al., 2023a).

While mechanical data combination systems optimize the accuracy of selection decisions, it may not be feasible to implement them in organizational settings due to administrative complexities (e.g., decentralized hiring, use of different selection procedures and predictors across roles) and users' disdain for mechanical processes (Guion, 2011; Neumann et al., 2023a). Acknowledging these realities, one useful avenue of research is to investigate how the predictive validity of selection procedures could be maintained when candidate data is combined clinically by selection decision makers. That is, how do hiring managers consume and integrate different pieces of candidate information into overall evaluations when given the latitude to do so themselves? Which selection procedures yield information that decision-makers prefer to give more weight to in overall candidate judgments, and why? As noted by Highhouse (2008), despite

[5] For the purposes of this paper, the terms "clinical" or "intuitive" judgment or data combination will refer to data integration processes in which humans manually integrate data themselves into overall evaluations. Other authors in the decision-making literature commonly use terms such as "holistic" judgment or "expert" judgment synonymously.

an abundance of selection research considering how *applicant reactions* can have consequences for selection outcomes, far less attention has been dedicated to understanding implications of *user reactions* and beliefs on selection decisions and selection system validity outcomes. Anderson (2005) also bemoaned the lack of attention dedicated to user influences on selection system validity: "In fact, the whole area of practitioner beliefs about selection methods and processes is a gargantuan one which research has made little or no inroads into" (p. 19).

As will be reviewed, a small body of literature has emerged to answer questions such as which selection procedures users believe are most valid and which they tend to utilize. For example, preliminary evidence suggests users often hold negative perceptions of selection procedures that have higher predictive validity than alternative procedures (e.g., preference for unstructured over structured interviews; e.g., Rynes et al., 2002). However, despite some understanding of what users believe, to date there is no comprehensive framework describing why users believe certain selection procedures are more valid than others. Beyond how favorably users view various selection procedures, if a rich understanding could be gained of why selection decision-makers choose to utilize candidate data from certain selection procedures over others, it is possible that users' utilization tendencies could be systematically influenced to increase alignment with empirical evidence (i.e., so that greater weight given to more valid predictors). The lack of investigation into causal mechanisms, combined with the treatment of predictors as monoliths (i.e., not considering combinations of predictor constructs and predictor methods), has limited this literature's ability to yield useful insights that could be used to design interventions influencing how users combine candidate information from selection procedures into overall judgments.

To address these shortcomings, the purpose of this dissertation is to develop and test a model of user reactions toward and beliefs about selection procedures (see Figure 1). In doing so, several contributions will be made. The model extends research examining user favorability perceptions to uncover why users believe certain selection procedures are better than others. Specifically, the model proposes two components of how users evaluate selection procedures– predictiveness beliefs and fairness beliefs, the latter of which has not been previously considered. The model also proposes nine dimensions of user reactions or perceptions of selection procedures that drive predictiveness and fairness beliefs. Additionally, consistent with the modular approach to evaluating selection procedures (Lievens & Sackett, 2017), the model seeks to disentangle influences of user reactions toward predictor methods and predictor constructs, which are utilized in various combinations to form selection procedures. Understanding how different properties of predictor methods and constructs might independently influence user beliefs will lead to richer understanding and the ability to design targeted interventions to influence how users view selection procedures. After testing this model of user reactions and beliefs, a second study will investigate the efficacy of autonomy-enhancing interventions within predictor methods to influence user perceptions of autonomy and, ultimately, beliefs about the predictiveness of selection procedures. Such interventions could be implemented to reduce gaps between how practitioners utilize candidate information from selection procedures and empirical validity evidence.

In what follows, I first review literature related to why individuals resist mechanical data combination methods, a primary reason why clinical judgment is common in organizational selection contexts. Next, I examine the role of clinical judgment in selection decision-making and discuss how Brunswik's Lens Model (1955) is a useful framework through which to view

clinical data combination in selection. After this, I summarize fragmented research investigating how hiring managers and HR practitioners view selection procedures, especially in terms of predictive validity beliefs. Despite a lack of research directly considering why selection decision-makers believe certain predictors are better than others, I then briefly review two potential reasons (i.e., naivety about evidence-based selection practices and preference for intuition-based methods) as well as findings from applicant reactions literature that could help answer this question. Based on extant research and adjacent literature, I introduce a model of user reactions toward and beliefs about selection procedures, upon which a set of hypotheses and model propositions are based. Finally, I propose two empirical studies to test the conceptual model and interventions designed to influence user reactions and beliefs.

Resistance Against Mechanical Data Combination

As noted, some selection systems are set up in ways that utilize algorithms, or a mechanical combination of data points, to aid in decision making by combining individual pieces of candidate information into overall judgments using a standardized method. For example, such systems might input candidates' predictor scores from an interview, cognitive ability test, and personality assessment and produce an overall score reflecting an optimally weighted combination of the inputs to predict candidates' likelihood of success. The purpose of such algorithmic systems is to aid users in decision-making and maximize the predictive validity of candidate judgments during the selection process by reducing influences of fallibilities inherent in human judgment processes (Kahneman et al., 2021). Indeed, a recent meta-analysis of employment selection predictions found that on average, combining candidate data mechanically improved predictions of candidate job performance by over 50% compared to judgments made using clinical data combination procedures (Kuncel et al., 2013). Similarly, a meta-analysis of

judgments in the selection of executives showed that mechanical judgments using only information about candidates' cognitive ability exhibited better prediction than clinical judgments based on cognitive ability scores and additional candidate information (Morris et al., 2015). These findings replicate studies from other contexts that have similarly concluded that data combination processes relying on clinical judgment result in inferior predictions (Grove & Meehl, 1994; Grove et al., 2000; Dawes et al., 1989). The superiority of mechanical judgment systems seems to hold even when they utilize suboptimal weighting systems (e.g., unit weighting, consistent random weighting; Yu & Kuncel, 2020; 2022).

Despite the demonstrated superiority of mechanical judgment, an emerging body of evidence on "algorithm[6] aversion" suggests that individuals hold negative views about mechanical data combination in many contexts, including towards algorithmic-based decision aids in selection contexts (Burton et al., 2020; Lacroux & Martin-Lacroux, 2022). This has clear implications for how users make overall candidate judgments for selection decisions. For example, when presented with a system that provides algorithmically based recommendations or rankings of candidates, users' perceptions of that system will affect the extent to which they choose to follow the algorithm's recommendations over their own judgments. Another example of this could be observed within a single assessment. Some assessments, for example, provide dimension or competency sub-scores (e.g., scores for each construct assessed within an interview) in addition to an overall score that reflects the optimally weighted combined prediction based on the sub-scores. The extent to which users utilize the overall score instead of computing their own overall judgements based on their interpretation and integration of the sub-scores will influence the predictive accuracy of their judgment.

[6] The term "algorithm" can refer to different things. For the purposes of this dissertation, algorithms are discussed simply as a method of mechanical data combination.

Burton et al.'s (2020) review of algorithm aversion identifies several reasons why people harbor negative reactions toward algorithms. Two factors that go hand in hand are 1) users' false expectations and misconceptions about the accuracy of algorithms, and 2) users' false beliefs in the superiority of intuition-based decision-making compared to mechanical decision making. For example, one study found that after observing them make prediction errors in specific cases, individuals' perceptions of algorithmic decision aids become substantially more negative. Most notably, compared to when individuals were informed of errors they made using their own clinical judgments, the study found that the drop in perceptions of the algorithm's accuracy was much larger than the drop in perceived accuracy of their own intuitive judgment abilities (Dietvorst et al., 2015). This indicates that people hold algorithms to a higher standard than their own intuitive judgment abilities, the latter of which is less accurate in making predictions in the first place.

According to the review by Burton and colleagues (2020), another reason why users avoid utilizing algorithms is due to perceived losses in decision-making autonomy. Users feel less autonomy when asked to rely upon an algorithmic decision aid over their own decision-making abilities (Neumann et al., 2021), and it is not surprising that autonomy losses lead to negative outcomes given the critical role of autonomy as a fundamental need and driver of motivation and well-being in workplace contexts (Ryan & Deci, 2000; Deci et al., 2017). Further, use of mechanical decision-making methods can undermine perceptions of an individual's decision-making competence (Arkes et al., 2007; Diab et al., 2011; Nolan et al., 2016). Importantly, individuals' latitude to make clinical judgments when forming overall candidate evaluations is severely threatened when mechanical or algorithmic based data combination techniques are utilized (Neumann et al., 2021). To help alleviate this issue, some

have attempted to mitigate autonomy losses by giving individuals some degree of input or involvement in processes involving algorithms. For example, one study found that use of algorithms increases if users are able to slightly modify weights attached to each component (Dietvorst et al., 2018).

In sum, a growing literature exists on user attitudes towards and use of algorithms and algorithmic decision aids, including in employment selection contexts. Because meta-analytic evidence suggests that mechanical data combination is superior to clinical data combination in selection contexts (Kuncel et al., 2013), additional investigation of prediction accuracy from clinical versus mechanical data combination does not seem warranted and is thus outside the scope of this dissertation. Instead, because clinical judgment processes are more commonly utilized by individual hiring managers in organizational contexts (Guion, 2011; Highhouse, 2008; Neumann et al., 2023a), this dissertation focuses on ways in which the validity of selection systems can be maintained when candidate data is combined clinically by decision makers. In other words, the primary goal is to better understand processes underlying how selection decision-makers form overall candidate judgments when given the latitude to utilize and combine different types and sources of information about candidates as they see fit. In the following section, I summarize human judgment processes in selection, including applying the Lens Model as a framework that can be used to understand how individuals make clinical candidate judgments based on candidate data gathered from selection procedures.

Clinical Judgment in Selection

Selection research has given considerable attention to the importance of hiring managers' judgment abilities within selection *data collection* in achieving optimal levels of predictive validity. For example, prior research has considered how the validity of specific selection

procedures that utilize clinical judgment (e.g., interviews, assessment centers) are highly dependent on hiring managers' ability to make accurate, consistent evaluations of candidates within those procedures (e.g., Huffcutt et al., 2013; Woehr & Arthur, 2003). To address this concern, researchers have developed methods such as interview and assessment center rater training programs to reduce biasing influences of human judgment idiosyncrasies on candidate evaluations within those procedures (e.g., Powell & Goffin, 2009). However, far less research attention has been dedicated to hiring manager judgments within clinical selection *data combination* situations, or the processes by which hiring managers integrate or combine previously gathered candidate information from multiple selection procedures into overall candidate judgments. Understanding how selection decision-makers combine candidate data using their own clinical judgment is important as one recent study of U.S.-based hiring managers found that 82% claimed to use a clinical process as their most frequent method of combining candidate data compared to one of several possible mechanical processes (Neumann et al., 2023a).

One popular framework that has been applied to the study of human prediction and judgment processes is Brunswik's Lens Model (1955). The Lens Model is most useful in situations where individuals are making predictions based on probabilistic environmental cues (i.e., when using available information to predict uncertain outcomes). For example, the Lens Model could be applied to situations in which investors predict future stock prices based on cues such as past and current financial indicators, or when physicians provide initial diagnoses based upon cues derived from patient symptoms and medical history. In each of these situations, the statistical relationship between the environmental cues and the outcome of interest across many iterations can be calculated, but a certain amount of unexplained and unpredictable variance is

also present. Similarly, personnel selection involves a probabilistic relationship between candidate information gathered during the selection process and that candidate's future job performance if they are hired. Thus, the Lens Model provides a relevant framework to study judgment processes in personnel selection.

The Lens Model outlines several processes describing how information from the environment is used to form judgments and predictions. First, the criterion of interest is objectively able to be predicted by available environmental cues varies to a certain degree (i.e., some situations have outcomes that are more predictable than others). The extent to which an outcome can be predicted (i.e., environmental predictability) is based on an optimally weighted model of these available cues. Next, the model recognizes that individuals perceive and mentally attach implicit weights to available cues to make judgmental predictions about an outcome of interest (i.e., subject judgments). The Lens Model then outlines how subjective judgments can be compared to the optimally weighted predictions (i.e., environmental predictability) to evaluate the accuracy of the subjective judgments (i.e., achievement index).

Consider the following example to illustrate the applicability of the Lens Model framework to employment selection contexts. To make a hiring decision reflecting a prediction about candidates' future job performance, a hiring manager may collect information (i.e., cues) about candidates by conducting an interview to evaluate candidates' job-related knowledge and reviewing candidate resumes to evaluate job experience. Using these two cues (i.e., job knowledge ratings and years of experience), the hiring manager combines this information in some way to make overall judgments of each candidate so that they can be compared. If data were to be collected on these cues for each candidate, the managers' overall judgments of each candidate, and a criterion of interest (e.g., each hired candidates' future job performance), it

would be possible to then calculate, over time, 1) the objective relationship between these cues (when optimally weighted) and future job performance, and 2) the accuracy (i.e., achievement) in which the hiring manager combines these cues to form overall predictions of future job performance in reality.

The strength of the Lens Model is that it offers a framework to evaluate the accuracy of clinical judgment processes compared to the objective predictability of the environment. By considering pieces of information about job candidates derived from selection procedures as "cues" that decision-makers use to form judgments about candidates, the Lens Model is a useful framework to study clinical judgment in selection contexts. Specifically, the Lens Model illustrates how candidate information can be weighted into overall judgments as more or less important compared to the objective relationship between that predictor and an outcome (i.e., empirical evidence of predictor-criterion relationship strength).

The meta-analysis by Kuncel and colleagues (2013) suggested that one likely reason clinical data combination in selection contexts results in less accurate candidate predictions is due to low reliability and high levels of unsystematic error in clinical judgments. However, outside of gathering additional individual judgments from multiple raters and aggregating them to reduce error variance, efforts to train decision-makers to improve the accuracy of individual judgments using clinical data combination have been largely unsuccessful (Kahneman et al., 2021). Instead, it may be more useful to dedicate attention to a second factor contributing to inferior prediction accuracy in clinical data combination: systematic error attributable to sub-optimal mental weights users attach to various cues in the judgment process. That is, as will be reviewed in subsequent sections, selection decision-makers often systematically attach sub-optimal weights (compared to empirical validity evidence) to candidate information from certain

selection procedures due to inaccurate beliefs about these procedures. In support of this idea, Kuncel and colleagues (2013) call for future research that investigates how and why users choose to utilize or ignore different pieces of candidate information when making overall candidate judgments.

To improve the accuracy of users' clinical selection judgments, it is therefore necessary to understand which pieces of candidate information tend to be systematically weighted as more or less important than others. That is, which selection procedures do users believe are the best predictors of candidate outcomes? User beliefs about selection procedures are an important factor influencing the extent to which they intend to utilize candidate information from procedures in their overall candidate judgments. The subsequent sections summarize the limited research that has investigated how users perceive various selection procedures in terms of predictive validity beliefs and how user beliefs are often inconsistent with empirical evidence.

User Perceptions of Selection Procedures

Users who have inaccurate beliefs about the relative predictive validity of selection predictors (i.e., deviations from empirical, meta-analytic evidence) will likely integrate candidate information from these sources into overall judgments in a sub-optimal manner. While the selection literature offers observed, meta-analytic criterion-related validity estimates of different predictors (e.g., Sackett et al., 2022), decision-maker beliefs about the extent to which different procedures predict future performance can be inconsistent with meta-analytic estimates (e.g., Di Milia, 2004; Furnham, 2008; König et al. 2010; Lievens & De Paepe 2004; Rynes et al., 2002; Taylor et al. 2002; Zibarras & Woods 2010). In this vein, one line of research has measured hiring manager favorability perceptions toward, importance beliefs about, and actual utilization of information from different selection procedures when making hiring judgments. That is,

among the various different types and sources of information that might exist about candidates, what predictors do hiring managers actually believe are useful?

Before reviewing this literature, it is important to note the majority of research examining user reactions and perceptions of selection procedures does not distinguish between *predictor methods* and *predictor constructs*. This shortcoming is not unique to the literature on user perceptions of selection procedures. Prominent meta-analyses of selection predictor validity (e.g., Sackett et al., 2022) often produce estimates for broad methods, such as structured interviews, that are administered to measure many different constructs, and for general constructs, such as personality traits, which can be measured using several different methods (e.g., paper/pencil or online assessments, interviews, etc.). The problem with viewing broad methods or constructs as homogeneous is that different combinations of methods and constructs are often used in practice, and factors related to each can influence outcomes such as validity, applicant reactions, and user reactions. As argued by Lievens and Sackett (2017), it is more appropriate to view predictors through a modular lens that reflects the reality in which different combinations of methods and constructs are utilized. Unfortunately, with a few exceptions, most studies examining user perceptions of selection procedures do not take a modular approach, resulting in less precision in our understanding of what drives their beliefs.

To evaluate the accuracy of users' predictive validity beliefs about various selection procedures, Terpstra (1996) asked HR professionals to rate the effectiveness (i.e., predictive validity) of several selection procedures. One major finding from this study is that practitioners tended to prefer unstructured interviewing methods and viewed them as more effective than structured interviews. Similar conclusions related to practitioners' favorability perceptions, utility perceptions, and use of unstructured interviews is well-documented (e.g., Chapman &

Zweig, 2005; Dipboye, 1997; Highhouse et al., 2017; Lievens & De Paepe, 2004; Rynes et al., 2002; Terpstra & Rozell, 1997; Van der Zee et al., 2002) and mirrors findings that candidates themselves believe that information gathered from unstructured interviews is more accurate than that gathered from structured interviews (Steiner & Gilliland, 1996; 2001). This perception represents a critical science-practice gap as it is inconsistent with empirical evidence demonstrating the superiority of structured interviewing methods over less structured ones in terms of both predictive validity and other outcomes such as group differences (Huffcutt et al., 2014; Sackett et al., 2022).

Results from Terpstra (1996) suggested that users viewed personality traits as more predictive constructs than cognitive ability, highlighting a gap between what HR professionals believe and meta-analytic validity estimates of that era (Schmidt & Hunter, 1998; Huffcutt & Arthur, 1994) and more recently (Sackett et al., 2022). In different samples, several lab studies that gave information about hypothetical candidates' levels of cognitive ability and Big Five personality traits found that hiring managers gave the most weight to cognitive ability and conscientiousness (Dunn et al., 1995; Ones & Viswesvaran, 1999), which is more consistent with what meta-analytic validity estimates would recommend as optimal. Other experiments found that users attached greater weights than supported by empirical evidence to personality constructs such as agreeableness, extraversion, and openness (Williams et al., 1995).

In another prominent study, Rynes and colleagues (2002) conducted a large survey asking HR practitioners about their attitudes and beliefs toward different topics related to evidence-based management. Related to employee selection practices, their results found that users viewed predictors such as personality traits and cognitive ability far less favorably than validity evidence would recommend as optimal. Conversely, users viewed unstructured

interviews highly favorably. A 2016 study asked the same questions to a sample of Spanish, South Korean, and Finnish HR professionals; some cross-cultural differences were found, but overall results suggested these beliefs were also held but HR professionals in other countries (Tenhiälä et al., 2016). Results from the Rynes et al. study were also largely replicated in a recent study of HR professional beliefs (Fisher et al., 2021). Results from the study by Fisher and colleagues suggested that in the 20 years since the original study, science-practice gaps on some selection topics had shrunk (e.g., relating to the validity of structured vs. unstructured interviews) but gaps on other topics had increased (e.g., relating to the value of cognitive ability as a valid predictor of job performance).

In another study, sub-samples of lay individuals and HR professionals were asked to rank twelve selection procedures in terms of their relative predictive validity (Jackson et al., 2018). Researchers then compared these rankings to rankings they developed based on meta-analytic evidence. Lay individuals' top five most predictive procedures rankings were 1) years of work experience (ranked 7th most predictive by the study authors), 2) interviews (2nd), 3) work samples (3rd), 4) tests of cognitive ability (1st), and 5) years of education (12th). HR professionals fared a bit better than lay individuals in their rankings: 1) years of experience (7th), 2) interviews (2nd), 3) tests of cognitive ability (1st), 4) work samples (3rd), and 5) assessment centers (4th). Overall, respondents correctly placed high value in interviews (although this study did not distinguish interview methods by level of structure), significantly overvalued years of experience and education as predictors, and undervalued tests of cognitive ability, assessment centers, biodata, and occupational interests (Jackson et al., 2018).

Two studies were identified that differentiated between predictor constructs and predictor methods when examining how hiring managers perceive and utilize selection information. Both

considered how user perceptions of the importance of candidate cognitive ability and personality trait information differed by whether these constructs were assessed via interviews or paper-and-pencil assessments. Importantly, both studies concluded that users preferred using information about candidates' cognitive ability and personality traits when it was gathered using interviews over when it was gathered using paper-and-pencil assessments (Lievens et al., 2005; Topor et al., 2007). These results highlight the importance of distinguishing between predictor constructs and methods and suggest that additional in-depth investigation of the influence of predictor methods and method factors on user perceptions of selection procedures is warranted.

Others have extended research into selection validity misconceptions by considering who is more likely to hold certain beliefs about selection procedures. For example, Lodato et al. (2011) constructed a profile of individuals who are more likely to prefer clinical judgment-based hiring procedures (e.g., unstructured interviews) over procedures that rely less on user judgments (e.g., standardized assessments). The profile of traits the study identified included experiential thinking style (i.e., those who tend to make everyday decisions based on feelings) and factors such as having less hiring experience, working for smaller organizations, and a lack of advanced professional certification. Another study looked at individual differences in structured interviewing method adoption rates and found that older individuals are less likely to use structured interviews than younger individuals (Langhammer et al., 2012). Hiring experience was positively related to importance attached to tests of cognitive ability in another study (Lievens et al., 2005).

A qualitative approach was taken by König and colleagues (2011) in an attempt to discern reasons why HR practitioners prefer certain selection procedures over others. In this approach, interviewees produced preferred characteristics of selection procedures. Across 40

respondents, 39 different constructs were produced; only five were mentioned by at least half of participants: oral (versus written), examining several aspects (versus examining only a small number of aspects), measuring personality (versus measuring mental abilities), internal administration (versus external administration), and standardized (versus unstandardized). However, there was limited agreement across participants in producing constructs nor in terms of which pole of the construct was preferred (e.g., nearly half who mentioned the standardized construct preferred the unstandardized pole). The study is also limited by vagueness of selection procedure characteristics and the inclusion of concepts related to practical considerations within selection processes but not procedures themselves (e.g., "for all positions versus only for managers"). Finally, the study asked participants about preferred characteristics but did not consider their actions in a realistic hiring scenario, which may have affected results. For example, more participants said they preferred standardized (versus unstandardized) and objective (versus subjective) procedures, which is inconsistent with evidence that users resist standardization and objective methods when actually facing these situations (Highhouse, 2008). Practitioners might say they prefer standardization and objectivity, but which procedures do they view as having these characteristics, and more importantly, do they actually choose more objective and standardized procedures when given the option in actual hiring situations?

Overall, this stream of research has primarily focused on descriptively investigating user beliefs about certain selection procedures, specifically focusing on the occurrence of selection procedure validity misconceptions. A secondary focus in this literature has been identifying individual differences associated with misconceptions. A broad conclusion from this research is that many selection decision-makers hold beliefs about the predictive validity of selection predictors and methods that are inconsistent with empirical evidence, leading to large science-

practice gaps. In fact, this gap has been described by several as the largest among any topic in the field of I-O psychology (Rousseau & Brands, 2011; Rynes, 2012). Further investigation of this gap is needed to gain more precise understanding, including by taking a modular approach to examine user perceptions and beliefs about selection procedures with different combinations of predictor constructs and predictor methods.

Fairness Perceptions

In addition to user perceptions about the predictive validity of selection procedures, perceptions of the fairness of selection procedures might also be a factor that influences the extent to which users actually utilize candidate information derived from those procedures. According to the *Standards for Educational and Psychological Testing* authored by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, fairness in selection contexts could refer to equal group outcomes, equitable treatment of groups during selection processes, equitable access to constructs measured by a selection procedure, and lack of statistical bias (American, 2014). Perceptions of fairness might also include concepts from organizational justice literature, which has been applied to the study of applicant reactions to selection procedures (Gilliland, 1993; Ployhart & Ryan, 1997). The concept of selection procedure fairness is becoming increasingly relevant as organizations and higher education institutions grapple with social justice movements and advance diversity, equity, and inclusion (DEI) initiatives (e.g., Hu, 2020). In light of these concerns, some organizations and educational institutions have even dropped standardized testing requirements due to equity concerns (Einhorn, 2022). In addition to legitimate concerns reflecting actual group differences in performance on certain selection procedures (Hough et al., 2001), misinformation from various sources (e.g., the 2021 HBO documentary titled “Persona:

The dark truth behind personality tests") has begun to sow a culture of doubt about the utility and fairness of assessments used for employment selection (SIOP, 2021).

Despite the likely importance of fairness perceptions in users' utilization of information from different selection procedures, surprisingly little research has examined this issue. Only one relevant study was identified, which asked HR professionals to rank a list of selection procedures in terms of fairness to candidates with fairness defined narrowly in terms of procedural fairness (Jackson et al., 2018). With this definition in mind, results indicated that HR professionals ranked the following as the fairest procedures: 1) years of work experience, 2) interviews, 3) cognitive ability tests, 4) work samples, and 5) assessment centers. Conversely, respondents rated assessments of occupational interests and personality traits among the least fair selection procedures. However, a deeper understanding of factors driving fairness perceptions is lacking, and predictor method and construct disentanglement is needed here as well.

Together, findings from research investigating user perceptions of selection procedures indicate that users hold misconceptions to a certain degree about the predictive validity of many predictors. Users who hold these inaccurate beliefs will understandably manually integrate candidate information from these sources into overall judgments in a sub-optimal manner. While less research has examined user perceptions of selection procedure fairness, it seems likely that user beliefs about the fairness of various procedures may also influence their utilization of candidate information from those procedures. However, the degree to which users' fairness perceptions influence their utilization of candidate information derived from a procedure, and the extent to which users hold misconceptions about selection procedure fairness, are not well understood. The next logical question is why users hold misconceptions about selection procedures. In other words, user beliefs about selection procedures are likely to influence the

extent to which they utilize information from that procedure in candidate judgments, but what features or properties of selection procedures drive user beliefs about selection procedure usefulness and fairness?

Why Users Hold Misconceptions About Selection Procedures

To date, little research has directly focused on uncovering factors that drive user beliefs about selection procedures or utilization of information from those procedures. Several streams of adjacent research provide indirect insights to this question. One perspective is that selection decision-makers are simply misinformed or naïve about the predictive validity of different selection procedures, which has inspired research that seeks to better communicate validity evidence to users to influence their judgment processes. Another perspective is that users prefer and believe in the predictive superiority of selection procedures that provide greater amounts of autonomy and rely on users' intuition-based judgments, thus influencing their utilization of these procedures. Finally, considering the candidate perspective and applicant reactions research may provide some insights that can be applied to better understand selection decision-makers' beliefs, especially as it relates to selection procedure fairness beliefs. Each of these perspectives will be discussed in the subsequent sections.

Lack of User Understanding

One explanation of users' misconceptions about selection procedures is that hiring managers lack awareness and understanding of evidence relating to the relative predictive validity of selection procedures. This perspective suggests that users are not necessarily stubborn in their reliance on less valid predictors, but rather are naïve to evidence-based practices and/or lack the numeracy skills (i.e., understanding of correlations or other numerical effect sizes) needed to interpret validity information such as correlations. Indeed, even practitioners in the HR

domain tend to not read academic journal articles which contain information about evidence-based selection practices, and the practitioner outlets they do read contain very little information about selection procedure validity (Rynes et al., 2007). If large selection science-practice gaps exist in the beliefs of HR professionals, who in theory should have more knowledge of evidence-based HR practices than hiring managers in other domains, it seems likely that hiring managers outside of HR have even less knowledge of selection procedure validity. From this perspective, a small body of literature has considered ways to decrease the science-practice gap in user perceptions towards and use of selection procedures through strategies to better communicate validity information to users.

For example, several studies have considered how altering selection procedure utility analysis information provided to hiring managers or re-ordering how this information is presented affects managers' favorability reactions (e.g., Hazer & Highhouse, 1997; Highhouse et al., 2017). Another example was illustrated by Zhang et al. (2018) who showed that presenting users with graphical visual aids to communicate the validity of structured interviews was effective in increasing their use over less valid unstructured interviews. Others have tested the use of different types of effect sizes, such as comparative effect sizes, to influence perceptions of predictive validity among users who do not have backgrounds in statistics (e.g., Childers et al., 2021). Additionally, Lee et al. (2021) investigated how user comprehension of incremental validity information is affected by displaying expectancy charts instead of correlations. A broad conclusion from this literature is that the way in which predictor information is presented to selection users and decision-makers influences their attitudes towards predictor constructs and methods. Thus, it may be useful to consider different approaches to making the validity of various predictor methods and constructs clearer to users, who are often naïve to evidence-based

selection decision-making practices and have varying degrees of numeracy. Doing so may reveal potential interventions that could effectively increase user understanding and intention to use more valid predictors over less valid ones.

However, as noted by Guion (2011), selection decision-makers often do not have access to statistical validity information in any form, and even if they do, may disregard this information in favor of their own selection procedure validity inferences. This, along with research summarized in the next section that users' stubborn beliefs rather than naivety are to blame for misconceptions, indicates that simply communicating the validity of selection procedures may be insufficient in overcoming users' existing validity beliefs. Instead, it could be more efficacious to design communication strategies that target root causes of why users believe certain procedures are better than others. Taking this approach would first require an understanding of what drives user beliefs about selection procedures, a primary focus of this dissertation.

Preference For Intuition-Based Predictor Methods

Another perspective on why users often utilize candidate information from predictors with sub-optimal predictive validity blames users' stubbornness in preference for intuition-based or clinical predictor methods over methods that involve higher degrees of standardization (Highhouse, 2008). The primary reason posited to drive this preference is individuals' need for autonomy, or the ability to have control over their own behavior (Ryan & Deci, 2000; Dipboye, 1997), similar to arguments made about user preferences for clinical data combination (Neumann et al., 2021). For example, compared to the use of assessment-based methods to gather candidate information, interviews tend to provide more autonomy for users to rely on their own judgment to evaluate candidates. Within interviewing methods, unstructured interviews typically give

interviewers more control over which questions are asked and how candidates are evaluated, better fulfilling decision-makers' need for autonomy compared to more standardized processes inherent in structured interviews.

Similar to the role of autonomy in influencing user perceptions of mechanical data combination methods, autonomy is proposed to play a critical role in users' perceptions of candidate data collection methods in selection contexts. In a test of this perspective, Nolan and Highhouse (2014) conducted a study to investigate users' perceptions of autonomy and utilization intentions of interview methods with varying levels of structure. Their first experiment confirmed that users perceived lower degrees of autonomy in interview methods that contained more elements of structure (e.g., standardized questions, limited use of small talk and follow-up questions, use of rating scales). In their second study, results indicated that allowing flexibility within individual elements of the structured interview method increased autonomy perceptions and use intentions. Thus, evidence suggests that perceptions of autonomy within a selection method are an important factor in driving use intentions.

In sum, altering certain aspects of selection procedures to increase users' autonomy perceptions is likely a useful approach to enhance user beliefs about the predictive validity of selection procedures and use of candidate information from these procedures. Unfortunately, as noted previously, methods that allow users more flexibility to utilize intuition-based clinical judgment (and higher degrees of autonomy) tend to offer lesser degrees of predictive validity. Managing this tradeoff seems to pose a challenge to increasing users' utilization of valid predictors. Discovering additional factors outside of autonomy perceptions that influence how users form beliefs about selection procedures may offer alternative strategies that could be applied to alter user beliefs and utilization.

Insights from Applicant Reactions Literature

While not from the perspective of hiring managers or selection decision-makers, the literature on applicant reactions to selection procedures provides some insight into what drives individuals' views of different procedures, especially in terms of fairness beliefs. For example, Gilliland's (1993) model of applicant reactions applies an organizational justice lens to propose factors about selection processes and procedures that lead to candidate beliefs in the fairness of the selection process. Gilliland's model was derived from a larger literature on justice rules that underlie perceptions of fairness in many different decision-making contexts (see Colquitt et al., 2001; 2003 for a review). Some of these procedural justice rules have been applied to help explain candidate perceptions of hiring procedures, such as the ability to appeal a decision and perceptions of job relatedness, the opportunity to perform, and consistency of administration (e.g., Gilliland, 2003; Hausknecht et al., 2004). In doing so, applicant reactions research helpfully suggests factors that drive *candidates' perceptions* of selection procedures, especially in terms of explaining candidate evaluations of the fairness of selection processes. While not previously considered, I propose that several of these factors are also relevant to *selection decision-makers' perceptions* of selection procedures as they represent general beliefs people have about these procedures. However, to date a gap exists in our understanding of what influences how selection decision-makers perceive selection procedures, including the role of fairness perceptions in their decisions about the extent to which they should utilize candidate information from a procedure in their overall candidate judgments. Considering the relevance of certain dimensions of applicant reactions to how users view selection procedures could provide useful insights.

In summary, users often make choices about which information about candidates to utilize in their overall candidate judgments in ways that are sub-optimal due to perceptions about the utility of certain selection procedures that are inconsistent with scientific evidence. As noted, two potential reasons why users hold misconceptions are due to a lack of awareness of selection research indicating predictors' relative validities, and users' tendency to prefer procedures that involve intuition-based judgments and offer greater degrees of autonomy. Additional insights on why users hold certain beliefs about selection procedures may be provided from adjacent literature on applicant reactions. Based on this research, I next introduce a conceptual model that describes the process by which users perceive and develop beliefs about selection procedures.

A Model of User Reactions Toward and Beliefs About Selection Procedures

In this dissertation, I introduce a conceptual model that describes the process by which users perceive and evaluate selection procedures (see Figure 1). The model outlines multiple dimensions of *user reactions toward selection procedures*, defined as users' perceptions about properties of predictor methods and constructs. User reactions are proposed as causal mechanisms driving user beliefs about selection procedures. In turn, users' beliefs about selection procedures, comprised of *predictiveness beliefs* and *fairness beliefs*, are proposed to influence the extent to which users intend to utilize candidate information from these different sources when forming overall candidate judgments. In what follows, I present a series of hypotheses based on the model that will be empirically tested in the present research.

User Beliefs About Selection Procedures and Use of Information

The primary objective of selection procedures is to enhance the prediction of candidates' future performance based on measured attributes of candidates in the application stage. As noted by Guion, hiring managers often do not have access to (or, if available, ignore) statistical

estimates of criterion-related validity for selection procedures. Instead, they rely on quasirational judgments of expected validity, which is a matter of user inference (Guion, 2011). Assuming hiring managers are attempting to hire the best candidates as a primary motivation, it naturally follows that the extent to which they believe that selection procedures used to collect candidate information are accurate predictors of candidate outcomes will inform the extent to which they intend to use candidate information from these procedures in overall hiring judgments. Thus, the following is expected:

> H1a: User beliefs about the predictive validity of a selection procedure will be positively related to their intentions to utilize information from that procedure in overall candidate judgments.

Predictiveness beliefs are expected to be the primary driver of user utilization, but they are likely not the only factor. Selection procedure fairness is likely important to hiring managers for several reasons. For one, employment law prohibits the use of selection procedures that result in adverse impact (i.e., group differences on a predictor) for which validity evidence has not been demonstrated or disparate treatment based on protected classes. In addition to legal considerations, some evidence suggests positive organizational outcomes may be associated with membership diversity (see Roberson et al., 2017 for a review); the circulation of these studies in the popular press and recent social justice movements advocating for organizational DEI efforts both could appeal to hiring managers' desire to use selection procedures that result in more equitable outcomes to enhance diversity. Supporting this view, Jones and Cunningham (2023) report that organizations are increasingly seeking to design selection systems that produce minimal group differences and are willing to sacrifice some validity to enhance fairness outcomes and/or perceptions of fairness. Hiring managers might also care about selection

procedure fairness for reasons related to candidate experience. To date, a considerable body of research has examined effects of applicants' fairness perceptions on selection outcomes (Anderson et al., 2010; Steiner & Gilliland, 1996). Candidates may not have much insight into selection criteria, but to the extent that they do, hiring managers are likely motivated to utilize candidate information in a way they perceive as fair to appease candidates' reactions. The fairness of selection procedures has come under scrutiny in light of recent social movements, causing some to drop standardized testing procedures that some perceive as unfair (Einhorn, 2022). This suggests that selection procedure fairness beliefs may be more important than ever. Together, it seems likely that selection procedure fairness is an outcome that users take into consideration.

As stated previously, little research has examined selection procedure fairness beliefs from the perspective of selection decision-makers. One study with this focus was identified (Jackson et al., 2018), which found that HR practitioners' beliefs about selection procedure fairness varied, although fairness was defined narrowly in terms of procedural fairness. Further, users' fairness beliefs were related to intentions to utilize information from those procedures in candidate judgments. Thus, in addition to predictive validity beliefs, user beliefs about the fairness of selection procedures are hypothesized to influence the extent to which users intend to utilize candidate information derived from those procedures.

H1b: User beliefs about the fairness of a selection procedure will be positively related to their intentions to utilize information from that procedure in global candidate judgments.

Antecedents of Users' Predictive Validity Beliefs

Few studies investigating user perceptions of selection procedures or utilization intentions have examined *why* users have these beliefs. As noted by Guion (2011), users make

judgments of the validity of selection procedures, which are often not based on provided estimates of criterion-related validity but on personal inferences. However, influences of users' selection procedure validity inferences are not well understood. That is, what elements or properties of predictor methods and constructs explain users' beliefs about the predictiveness of these procedures? Additionally, what factors influence user beliefs about the fairness of these procedures? Just as the selection literature has identified dimensions of *applicant reactions* that influence job candidates' overall views of selection procedures, identification of *user reactions* that influence users' beliefs about the predictiveness and fairness of selection procedures is needed. In doing so, the model draws attention to the fact that user beliefs about selection procedures are dependent on the construct assessed and the method used to assess it (Sackett & Lievens, 2017).

One previously reviewed perspective (e.g., Dipboye, 1997; Highhouse, 2008) argues that users prefer to utilize intuition-based predictor methods that rely on clinical judgment over methods that involve higher degrees of standardization because of individuals' need for autonomy. The model proposes that users react to two aspects of autonomy within a predictor method. First, the degree to which users have control over procedural elements of predictor methods can vary. For example, compared to assessment-based methods, interview methods provide users with greater procedural autonomy to do things like determine how to measure the predictor construct (e.g., selecting interview questions to ask). Second, the degree to which users have autonomy or involvement in the evaluation or scoring of candidates can vary. For example, compared to assessment-based methods that automatically provide candidate ratings, interview-based methods allow interviewers to evaluate candidates based on their own judgment. Research on source credibility suggests that individuals tend to believe their own judgment over that of

others (Petty & Wegener, 1988; Yaniv, 2004), so it can be inferred that the level of user autonomy in candidate evaluation is important to them. Additionally, evidence suggests that users resist selection decision-making aids that integrate data across procedures (e.g., algorithmic decision aids) or within procedures (e.g., decomposing interview judgments into structured rating dimensions); instead, they prefer to make autonomous, intuition-based clinical judgments (Arkes et al., 2006; 2010).

Supporting this perspective on users' need for autonomy within selection procedures, one study examining properties of interview methods found that users perceived lesser degrees of autonomy in interviews that contained higher levels of structure (e.g., standardized questions, use of rating scales). Further, intervening by slightly relaxing rigid requirements of the structured interview method (e.g., allowing users to select interview questions from an approved list) increased users' autonomy perceptions and use intentions (Nolan & Highhouse, 2014). This study highlighted the importance of autonomy in both the process of conducting the selection procedure and how candidates are evaluated. Thus, user perceptions of *user procedural autonomy* and *user evaluation autonomy* in a predictor method are expected to be two dimensions of user reactions that influences users' predictive validity beliefs.

H2a: User perceptions of a predictor method's level of user procedural autonomy will be positively related to utilization intentions; predictive validity beliefs will mediate this relationship.

H2b: User perceptions of a predictor method's level of user evaluation autonomy will be positively related to utilization intentions; predictive validity beliefs will mediate this relationship.

In addition to perceptions of user autonomy, other dimensions of user reactions are proposed. The first is perceptions of a predictor method's level of *fidelity,* or the extent to which a selection procedure reflects a role's context and behavioral requirements (Goldstein et al., 1993). It is well established that the level of fidelity within selection procedures will vary by design (Viswesvaran & Ones, 2018). For example, asking candidates to demonstrate their skills through a role play or work sample more closely mimics behaviors that candidates will need to display on-the-job compared to taking an online assessment, even if that assessment measures similar job-related knowledge or skills (e.g., a situational judgment test). Additionally, perceptions may vary based on the role in question. Asking candidates to give an oral presentation showcasing their role-related knowledge will likely be viewed as more relevant for roles that actually require oral presentations than roles that do not. In line with applicant reactions research suggesting that fidelity is positively related to candidate perceptions of selection procedures (e.g., Chan & Schmitt, 1997), it is expected that selection decision-makers will also view predictor methods with higher levels of fidelity as more predictive because the link between behaviors displayed during the procedure and on-the-job behaviors is clearer to them.

H2c: User perceptions of a predictor method's fidelity will be positively related to utilization intentions; predictive validity beliefs will mediate this relationship.

Much has been written about the importance of transparency to increase user acceptance and use of algorithms (e.g., Burton et al., 2020) and artificial intelligence (Glikson & Woolley, 2020), both of which are often viewed as "black boxes" due to a lack of explanation of scoring in some of these systems. Similarly, the *transparency* of predictor methods used to produce candidate data (i.e., candidate evaluations on a construct from a single selection procedure) is

likely to matter to users. Research on source credibility and use of advice supports this perspective, highlighting how individuals are more inclined to trust their own judgment due in part to lack of awareness of how others developed their judgment (Yaniv, 2004). Predictor methods can vary in the level of evaluation transparency; for example, users may have less awareness of how certain assessment-based methods evaluate candidates and produce scores. The level of transparency in how candidate data is produced by predictor methods is hypothesized to positively influence predictiveness beliefs due to reduced uncertainty in source credibility.

H2d: User perceptions of a predictor method's transparency to users will be positively related to utilization intentions; predictive validity beliefs will mediate this relationship.

Candidates' ability to distort responses to be deceptive or dishonest within selection procedures has received considerable attention, including the effect of a selection procedure's perceived fakability on applicant reactions (e.g., Rolland & Steiner, 2007; Vasilopoulos & Cucina, 2006). Surprisingly, the effect of a selection procedure's fakability on user perceptions has not received much research attention to date. In a qualitative study asking HR practitioners about their perceptions of selection procedures, several mentioned the fakability of the procedure being important, presumably because they view less fakable procedures as yielding more accurate information about candidates (König et al., 2011). Although there is debate about the effect of candidate faking behaviors on the actual validity of selection procedures (Melchers et al., 2020; Hogan et al., 2007), user perceptions of the *fakability* of a predictor method are proposed to negatively influence beliefs about selection procedure predictiveness due to lack of confidence in the accuracy of candidate information from that procedure.

H2e: User perceptions of a predictor method's fakability will be negatively related to utilization intentions; predictive validity beliefs will mediate this relationship.

The aforementioned five dimensions of user reactions toward *predictor methods* are proposed to influence users' selection procedure predictiveness beliefs regardless of the *predictor construct* being measured. Dimensions of user reactions toward predictor constructs are also proposed to separately influence beliefs about selection procedure predictiveness. Predictiveness beliefs are first proposed to be influenced by perceptions of a predictor construct's *job relatedness*, or the extent to which predictor constructs are viewed as related to job requirements. Job relatedness has been proposed as a dimension of applicant reactions (Gilliland, 1993), and evidence suggests it is related to outcomes such as organizational attractiveness (Bauer et al., 1998). Research has not considered perceptions of selection procedure job relatedness from the perspective of selection decision makers, but it is expected that the extent to which users view a predictor construct as job relevant will influence their views of the predictiveness of candidate information from that construct.

H3a: User perceptions of a predictor construct's job relatedness will be positively related to utilization intentions; predictive validity beliefs will mediate this relationship.

Predictiveness beliefs are also proposed to be influenced by a dimension of user reactions called *malleability*. Malleability refers to perceptions of the stability or change over time in individuals' standing on a latent construct. In other words, in the view of users, to what extent are individuals able to change their standing on a construct via developmental efforts? Malleability is related to implicit theories of ability (Dweck et al., 1995) that suggest people have differing views of the malleability of personal attributes. Specifically, entity theorists believe that traits are relatively fixed while incremental theorists believe that personal attributes are highly

malleable and can change through developmental efforts. Some have investigated this topic in the context of applicant reactions (e.g., Reeder et al., 2012), but it has not been considered in terms of how users perceive selection predictor constructs. User perceptions of the malleability of individuals' standing on a construct measured by a selection procedure are proposed to negatively influence predictiveness beliefs as they are likely to believe that more malleable attributes (e.g., job knowledge) can be learned on-the-job and are therefore less predictive of success than more stable attributes (e.g., trait achievement motivation).

H3b: User perceptions of a predictor construct's malleability will be negatively related to utilization intentions; predictive validity beliefs will mediate this relationship.

Antecedents of Users' Fairness Beliefs

As noted, in addition to predictiveness beliefs, the extent to which users believe selection procedures are fair is proposed to influence their intentions to utilize candidate information from that procedure. First, fairness beliefs are proposed to be influenced by two dimensions of user reactions toward predictor methods. Similar to the concept of candidate perceptions of consistency of administration (e.g., Arvey & Sackett, 1993; Gilliland, 1993), user perceptions of a predictor method's *evaluation consistency*, or the degree of stability in its measurement of a candidate attribute, are expected to be positively related to fairness beliefs. In other words, predictor methods that users believe produce consistent results are expected to be viewed as fairer to candidates. Evaluation consistency is similar to the psychometric concept of reliability, which can differ across selection procedures (e.g., Huffcutt et al., 2013). Users may not be able to articulate the concept of reliability, as evidenced by a study finding that giving applicants information about the reliability of a selection procedure did not affect their perceptions of it in most cases (Lievens et al., 2003). However, research from other contexts suggests that when the

concept of reliability is explained in lay terms, individuals' perceptions of reliability influence how they view assessments (Barg-Walkow & Rogers, 2016; Chamberlain, 2013). In line with research on drivers of candidates' procedural justice perceptions (i.e., the related concept of perceptions of consistency in administration; Bauer et al., 1998; Gilliland, 1993), it is expected that users' fairness beliefs will be positively influenced by perceptions of a predictor method's evaluation consistency.

H4a: User perceptions of a predictor method's evaluation consistency will be positively related to utilization intentions; fairness beliefs will mediate this relationship.

The second dimension of user reactions toward predictor methods proposed to influence fairness beliefs are perceptions of a predictor method's *fakability*. Arvey and Sackett (1993) and Gilliland (1995) suggested that perceived fakability could influence applicant perceptions of selection procedure fairness, and subsequent evidence has confirmed this relationship (Rolland & Steiner, 2007). Thus, the following is hypothesized:

H4b: User perceptions of a predictor method's fakability will be negatively related to utilization intentions; fairness beliefs will mediate this relationship.

It is proposed that user beliefs about selection procedure fairness are also driven by two dimensions of user reactions toward predictor constructs. The first is perceptions of a construct's *job relatedness.* In addition to selection decision-makers considering the job relatedness of the predictor construct in determining their beliefs of the predictive validity of a procedure, they also are likely to consider job relatedness in terms of the fairness of a selection procedure. Job relatedness perceptions are an important driver of candidate views of the fairness of selection processes (Bauer et al., 1998; Gilliland, 1993), and managers are also expected to view

procedures measuring constructs that are more closely job related as fairer than constructs that are less clearly relevant to the role in question. Thus, the following is hypothesized:

H5a: User perceptions of a predictor construct's job relatedness will be positively related to utilization intentions; fairness beliefs will mediate this relationship.

A final dimension of user reactions proposed to influence fairness beliefs is termed *construct development equity*, defined as the extent to which users perceive individuals to have equal opportunities to acquire or develop standing on a construct (i.e., knowledge, skill, ability, or other attributes) regardless of personal background or barriers. For example, individuals in higher socioeconomic classes may have received a higher quality of education, resulting in higher scores on ability constructs or constructs reflecting job-related knowledge or skills. This dimension also arises from the fact that mean differences in measurement of predictor constructs by race, age, and gender have been observed (cf. Hough et al., 2001). While mean differences reflect a combination of factors (e.g., measurement inequivalence), one factor is that there are real differences in the distributions of standing between groups on some constructs (Guion, 2011). For example, racial differences exist in many tests of verbal ability, a construct that is influenced by factors such as quality of educational opportunities and socioeconomic status (Outtz & Newman, 2010); because the development of verbal ability can be inhibited by factors such as these, a selection procedure measuring verbal ability might be viewed as less fair. In contrast to these examples, constructs such as personality traits, occupational interests, and integrity might be viewed as having more equitable opportunities to possess or develop. Because organizations seem to be placing increased importance on selection procedures that produce smaller group differences (Jones & Cunningham, 2023), it is expected that perceptions of a construct's development equity will influence user beliefs about construct fairness.

H5b: User perceptions of a predictor construct's development equity will be positively related to utilization intentions; fairness beliefs will mediate this relationship.

Notably, two dimensions of user reactions (fakability and job relatedness) are hypothesized to influence user beliefs of both selection procedure predictiveness and fairness while the rest are hypothesized to act as distinct drivers. It is possible that additional dimensions of user reactions will influence both types of user beliefs. Some uncertainty arises from the possibility that user beliefs of predictiveness and fairness may not be independent of each other. The relationship between predictiveness and fairness beliefs will be evaluated, and additional connections between user reactions dimensions and user beliefs will be reported.

As depicted by the front end of the model, there is expected to be variance in user perceptions of and beliefs about selection procedures. For the purposes of testing this model, it is not practically feasible to hypothesize favorability differences (or the lack thereof) in each proposed dimension of user perceptions, as well as predictiveness beliefs and fairness beliefs, between a complement of all possible combinations of selection procedures by construct and method. Instead, to provide a test of the model with some generalizability across methods and constructs, a set of three predictor methods and three predictor constructs commonly used in organizations will be tested. Additionally, Study 1 will thoroughly investigate and summarize results comparing each of these predictor methods and constructs in terms of users' relative favorability in user reactions dimensions, beliefs about predictiveness and fairness, and intentions to utilize information from these procedures in overall candidate judgments.

The proposed model makes several contributions. First, the model goes beyond extant research examining favorability perceptions to propose two components of user beliefs about selection procedures (i.e., predictiveness beliefs and fairness beliefs) that drive users' utilization

of candidate information derived from selection procedures. The model also proposes nine dimensions of user reactions that together form how users perceive selection procedures, including specifying links between user reaction dimensions and beliefs about the predictiveness and fairness of procedures. Among the dimensions of user reactions, most have not been explicitly considered as factors related to how selection decision makers view selection procedures. Further, while user autonomy has been previously considered, the model distinguishes between procedural autonomy and evaluation autonomy as factors influencing user beliefs about predictor methods.

Consistent with the modular approach to studying selection procedures (Lievens & Sackett, 2017), the model also makes a contribution by disentangling influences of predictor method and predictor construct properties in influencing user reactions toward selection procedures. Further, the model delineates paths through which user reactions separately drive beliefs about selection procedure predictiveness or fairness, or in some cases, both. Understanding how different properties of predictor methods and constructs might independently influence user beliefs about predictiveness and fairness will lead to richer understanding and the ability to design targeted interventions to influence how users view selection procedures.

Study 1

Study 1 Method

Study 1 tested the conceptual model among a sample of hiring managers by measuring user reactions, predictiveness and fairness beliefs, and intentions to utilize candidate information from a set of selection procedures presented in experimental vignettes. Hiring manager perceptions and beliefs were assessed after presenting participants with different sets of selection

procedures manipulated into combinations of three common predictor methods and three common predictor constructs.

Study 1 Sample

A sample of hiring managers from Prolific was recruited for this study. Sampling criteria stipulated that participants must live and work full-time in the U.S. and be in a hiring manager role in which they have experience making employee selection decisions. Participants who completed the study and passed all data quality checks were compensated $6. The average survey completion time was nearly 35 minutes, resulting in an hourly compensation rate of approximately $10.50.

The initial sample consisted of 395 participants. Of these, 17 were excluded because they indicated they did not have prior experience as a hiring manager, contrary to Prolific's internal screening procedures. Additionally, 21 participants failed at least one of the three attention checks, and another 67 participants failed at least one of the two vignette comprehension checks. The final sample consisted of 290 participants. The final sample was gender balanced (female = 50.2%), and most (73.8%) were White. The average age was just over 41 years (SD = 11.3). Participants' total years of previous hiring experience ranged from one to 42 years (M = 8.69, SD = 7.64), and estimates of the total number of employees they had hired in their career ranged from one to 1000 (M = 42.72, SD = 105.64). Most (63.1%) had a bachelor's degree or higher with nearly 22% having a master's degree or higher.

Study 1 Procedure and Manipulations

After agreeing to the informed consent, participants were shown a vignette describing a fictional hiring scenario (see Appendix C). The first section of the vignette described the scenario and told participants to imagine that they were on a hiring committee for a human

resources (HR) generalist role in a fictional organization. A specific role was articulated in the scenario to add realism compared to asking participants to make hiring decisions for a vague role. This role was chosen because 1) it is highly general and performs a variety of tasks; 2) it is likely that most people would have some basic familiarity with the role as HR roles exist in nearly all organizations across industries; 3) it is likely that selection procedures are commonly utilized in hiring processes for HR roles because they are professional roles and selection processes are often administered organization-wide by internal HR functions (i.e., hiring processes for HR roles likely utilize tools they implement and oversee); and 4) data from LinkedIn indicates that turnover among HR roles is high compared to other roles (Lewis & Sorongon, 2022), necessitating greater hiring volumes. A brief description of job duties was presented in the vignette (see Appendix C).

The vignette explained that recruiting staff had already screened the initial batch of applicants for minimum required qualifications to produce a shortlist of candidates, and that it is now time for the hiring committee to determine the top candidate from this shortlist. Participants were told that to do this, each hiring committee member would independently provide overall evaluations of each candidate. Situating participants as a hiring committee member rather than the hiring manager puts them in a similar position to independently evaluate candidates and selection procedures while adding realism in that hiring committees often include individuals who do not work within the functional group of the role being hired (i.e., participants may feel unqualified if asked to serve as a hiring manager for a role with which they do not work closely in real life). Participants were told that to help the hiring committee evaluate candidates, the organization requires that all finalist candidates be subjected to three selection procedures (i.e., two interviews and an assessment) to collect data about the candidates and help predict their

likelihood of success. The vignette explained that participants will have access to candidate resumes and candidate results from these three selection procedures and that they can use any of this information as they see fit when making overall evaluations of each candidate.

Participants were each shown descriptions of three selection procedures used in the fictional hiring process (see Appendix D). Predictor method and construct combinations shown to participants were manipulated across conditions (i.e., cognitive ability, conscientiousness, or job knowledge measured via unstructured interview, structured interview, or standardized assessment). The three predictor methods and constructs were chosen due to their prevalence in organizational selection contexts (Risavy et al., 2019), because they are the subject of large bodies of empirical validity research (Sackett et al., 2022), and because they make logical sense when applied modularly with other constructs or methods to form a selection procedure. Descriptions of predictor methods and constructs shown to participants were loosely based on selection handbook definitions of common predictors (e.g., Pulakos, 2005). Descriptions of structured and unstructured interview methods incorporated properties of interview structure as defined by Huffcutt et al. (2014).

Participants viewed three procedures but did not see the same method or construct more than once. Thus, there were six total conditions from the possible non-repeating predictor method and construct combinations (see Table 1). The presentation order of the three procedures within each condition were randomized to eliminate potential ordering effects.

Participants were each shown descriptions of three selection procedures based on the predictor method and construct combinations but were not shown candidate scores on any of the procedures as would be done in a policy capturing study. This is because including candidate scores would introduce noise and require dozens of additional experimental conditions;

additionally, the present research is interested in examining hiring decision-maker perceptions of selection procedures themselves rather than candidate profiles or scores. Instead, participants read detailed descriptions about each selection procedure (see Appendix D), after which they responded to measures assessing the extent to which they intended to utilize candidate information from the procedure in their overall candidate evaluations, their beliefs about the predictiveness and fairness of the procedure, and their perceptions of the procedure along dimensions of user reactions. Participants were also asked the extent to which they would utilize two pieces of resume-based candidate information in their overall evaluations: educational background and employment experience. Finally, participants responded to measures assessing individual differences and sample characteristics.

Study 1 Measures

All measures were rated on a five-point Likert agreement scale (1 = strongly disagree, 5 = strongly agree) unless otherwise noted. Participants responded to measures assessing utilization intentions, predictiveness and fairness beliefs, and dimensions of user reactions for each of the three selection procedures they viewed. The referent for distal outcome variables (i.e., predictiveness beliefs, fairness beliefs, and utilization intentions) was "this selection procedure" while the referents for measures of user reactions dimensions were specific methods or constructs (e.g., "this type of interview/assessment…" or "conscientiousness/cognitive ability/job knowledge…"). In addition to the below measures, three simple attention check items were included in the study. Two comprehension check items were also included to ensure that participants adequately read the vignette and understood what they were being instructed to do in the scenario.

Intentions to Utilize Information from a Procedure. The extent to which hiring managers intended to utilize information from a selection procedure (and education and experience from candidate resumes) was assessed using a three-item measure adapted from Nolan and Highhouse (2014). Items included, "To what extent would you rely on results from this selection procedure to evaluate candidates?", "To what extent would you compare candidates using results from this selection procedure?", and "When making a hiring decision, to what extent would you take into consideration results from this selection procedure?" Responses were rated on a six-point scale (1 = not at all, 2 = to a very small extent, 3 = to a small extent, 4 = to a moderate extent, 5 = to a great extent, 6 = to a very great extent). This measure demonstrated good internal consistency reliability ($\alpha = .94$).

Intentions to utilize information from each of the three selection procedures as well as candidate education and experience from resumes were also measured by asking participants to attach weights to each piece of information in terms of relative importance in informing their hiring decision (i.e., numbers between 0-100 in which the five weights sum to 100). Participants were given the following instructions:

> "When making hiring decisions for this role, how much weight would you give to candidate information from each of the following? Weights attached to each piece of information can range from 0 to 100 (lower numbers = less importance, higher numbers = greater importance in your overall decision). Weights across the five pieces of information below must sum to 100. For example, 0 means I would ignore this information completely, and 100 means I would base hiring decisions on this information alone. Weights of 20 would indicate equal weighting of each of the five pieces of

information. You can select different numbers for each piece of information as you prefer, including 0."

Predictiveness Beliefs. Selection procedure predictiveness beliefs were measured using a five-item scale adapted from Smither et al. (1993): "Candidates who perform well on this selection procedure are more likely to perform well in this job than candidates who perform poorly on this selection procedure", "Failing to do well on this selection procedure indicates that a candidate can't do this job", "This selection procedure can predict how well a candidate will perform in this job", "A candidate's performance on this selection procedure is a good indicator of their ability to do this job", and "Employers can tell a lot about an applicant's ability to do this job from their results on this selection procedure." This measure exhibited adequate reliability (α = .90).

Fairness Beliefs. Selection procedure fairness beliefs were measured using a scale that included items related to both procedural and outcome fairness. Four items were related to procedural fairness: "All candidates would be treated the same way in this selection procedure", "The way in which this selection procedure is conducted is fair to candidates", "Candidates from certain groups would be treated differently in this selection procedure" (reverse-scored), and "The process candidates go through in this selection procedure is fair regardless of candidates' backgrounds." Four items were related to outcome fairness and group differences: "Candidates from certain groups would be at a disadvantage to do well on this selection procedure" (reverse-scored), "Candidates from certain groups are likely to do better than others on this selection procedure (reverse-scored)", "Candidates from different groups have equal chances to do well on this selection procedure", and "Using this selection procedure to make hiring decisions would result in unequal outcomes for candidates from certain groups" (reverse-scored).

CFA was conducted to examine the dimensionality of this measure. Results indicated that fit was substantially better for a two-factor solution in which procedural and outcome fairness items were separated into distinct factors (χ^2 = 387.3, SRMR = .07, RMSEA = .18, CFI = .93, TLI = .88) compared to a single factor measure including all eight items (χ^2 = 934.56, SRMR = .07, RMSEA = .24, CFI = .82, TLI = .75). Thus, procedural fairness beliefs (α = .86) and outcome fairness beliefs (α = .91) were considered as two separate measures of selection procedure fairness beliefs.

Procedural Autonomy Perceptions. Hiring manager perceptions of procedural autonomy within a predictor method were assessed using six items adapted from Nolan & Highhouse (2014). This section included an introductory sentence explaining that the next few items ask about their perceptions of the way a selection procedure is administered; each item began with the stem, "*The way in which the [interview/assessment] is conducted* gives me a sense of…" followed by six words: control, choice, free will, influence, self-sufficiency, and freedom. This scale had high reliability (α = .94).

Evaluation Autonomy Perceptions. Hiring manager perceptions of evaluation autonomy within a predictor method were assessed using six items adapted from the same Nolan and Highhouse (2014) scale. This section included a similar introductory sentence explaining that the next few items ask about their perceptions of the way candidate information from a selection procedure is scored; each item began with the stem, "*The way in which candidates are evaluated or scored in this [interview/assessment]* gives me a sense of…" followed by six words: control, choice, free will, influence, self-sufficiency, and freedom. This scale also had high reliability (α = .95).

Fidelity Perceptions. User perceptions of a predictor method's level of fidelity were measured using a self-developed measure based on existing definitions of fidelity as relevant to selection procedures (e.g., Motowidlo et al., 1990). Items in this self-developed measure included the following: "This [interview/assessment] provides a realistic simulation of this job", "This [interview/assessment] is representative of tasks required in this job", "This [interview/assessment] is representative of behaviors required in this job", "This [interview/assessment] requires people to do what they would do in this job", "This [interview/assessment] requires people to show what they would do in this job", "This [interview/assessment] closely mimics the real job context of this job", and "This [interview/assessment] requires candidates to demonstrate actual job behaviors." This measure exhibited high reliability ($\alpha = .95$).

Transparency Perceptions. Hiring manager perceptions of the transparency of a predictor method were measured using a three-item scale adapted from Smither et al. (1993): "It would be clear to me how candidates were scored in this [interview/assessment]", "I would know exactly on what aspects of the [interview/assessment] the candidate performed well and poorly", and "I would clearly know how a candidate is evaluated in this [interview/assessment]." The transparency scale had high reliability ($\alpha = .95$).

Fakability Perceptions. Perceptions of a predictor method's fakability were measured using three items adapted from Gilliland and Honig (1994). Items included "Candidates would be able to distort their responses in this [interview/assessment] to make themselves look better", "It would be easy for candidates to be dishonest when answering questions in this [interview/assessment] to make themselves look better", and "I think smart candidates could beat

this [interview/assessment] by giving the answers it is looking for." This measure has adequate reliability (α = .93).

Evaluation Consistency Perceptions. User perceptions of the consistency with which a predictor method evaluates candidates were measured using three items adapted from Bauer et al.'s (2001) measure of consistency of administration: "This [interview/assessment] scores all candidates in the same way", "There are no inconsistencies in how this [interview/assessment] evaluates candidates", and "This [interview/assessment] is consistent in how it scores candidates." This scale demonstrated high reliability (α = .95).

Job Relatedness Perceptions. Perceptions of a predictor construct's job relatedness were measured using three items; two were adapted from Bauer et al. (2001): "It would be clear to anyone that [conscientiousness/job knowledge/cognitive ability] is related to the HR generalist job" and "[Conscientiousness/Job knowledge/Cognitive ability] is clearly related to the HR generalist job" (α = .87). The third item was adapted from Gilliland's (1993) definition of job relatedness and read "[Conscientiousness/Job knowledge/Cognitive ability] appears to be relevant to the HR generalist's job situation." The job relatedness measure had high reliability (α = .93).

Malleability Perceptions. User perceptions of the malleability of constructs assessed in selection procedures were measured using four relevant items adapted from a larger scale by Chiu et al. (1997): "A person's level of [conscientiousness/job knowledge/cognitive ability] is very basic and people can't change it very much", "People can change their level of [conscientiousness/job knowledge/cognitive ability]" (reverse-scored), "Everyone has a certain amount of [conscientiousness/job knowledge/cognitive ability] and there is not much they can do to really change it", and "People can significantly change their level of [conscientiousness/job

knowledge/cognitive ability]" (reverse-scored). The malleability scale exhibited good reliability (α = .91).

Construct Development Equity Perceptions. The extent to which users perceive constructs as having developmental equity was measured using an eight-item self-developed measure. Items included, "Candidates have equal opportunities to possess high levels of [conscientiousness/job knowledge/cognitive ability]", "Having high levels of [conscientiousness/job knowledge/cognitive ability] is dependent on candidates' backgrounds" (reverse-scored), "All candidates could possess high amounts of [conscientiousness/job knowledge/cognitive ability] regardless of barriers", "All candidates have a reasonable opportunity to acquire high levels of [conscientiousness/job knowledge/cognitive ability]", "Candidates having high levels of [conscientiousness/job knowledge/cognitive ability] is a matter of equal opportunities" (reverse-scored), "Some candidates might not have a reasonable opportunity to develop high amounts of [conscientiousness/job knowledge/cognitive ability]" (reverse-scored), "There are excusable reasons why some candidates do not have high levels of [conscientiousness/job knowledge/cognitive ability]" (reverse-scored), and "Candidates have equal chances to have developed high amounts of [conscientiousness/job knowledge/cognitive ability]." Psychometric analysis indicated that the fifth item (i.e., "Candidates having high levels of [conscientiousness/job knowledge/cognitive ability] is a matter of equal opportunities") exhibited very low inter-item correlations and was thus dropped from the scale. The final, seven-item measure demonstrated adequate internal consistency reliability (α = .86).

Sample Characteristics. Participants were also asked to indicate their gender, age, racial/ethnic group, education level, years of hiring experience, and job type.

Exploratory Measures.

Prior Experience with a Predictor. The amount of prior experience hiring managers have with specific predictor methods may influence their beliefs about selection procedures and utilization intentions and was thus considered as a variable that might predict these outcomes. After participants completed the experimental vignette, they were asked to rate the extent to which they had prior experience evaluating candidates using structured interviews, unstructured interviews, and computerized assessments (all assessing any construct). Additionally, hiring managers were asked the extent to which had prior experience evaluating candidates using information about their cognitive ability, conscientiousness, and job knowledge (all assessed by any method) using the same scale. Items read, "To what extent do you have prior experience using [structured interviews/unstructured interviews/automated assessments/cognitive ability/conscientiousness/job knowledge] when evaluating job candidates?" and were rated on a six-point scale (1 = not at all, 2 = to a very small extent, 3 = to a small extent, 4 = to a moderate extent, 5 = to a great extent, 6 = to a very great extent).

Occupational Field Norms. Another variable that might influence hiring manager beliefs about selection procedures and utilization intentions is the extent to which it is common to use a specific predictor method or construct in hiring processes within their occupational field. After participants completed the experimental vignette, they were asked to estimate how common it is in the occupational field for roles they typically hire to evaluate candidates using each of the three predictor methods and each of the three predictor constructs they viewed. Candidate education and experience were included as two additional predictor constructs, and resume screening was included as an additional predictor method. Specifically, participants were asked, "Think about the role you most commonly hire for. How commonly in this occupational field

(across organizations) are job candidates evaluated using [the following candidate attributes]/[information from the following methods or tools]?" Responses were rated on a 0-100 slider indicating percent of the time.

Preference for Intuition-Based Hiring. Another individual difference variable that was included was a measure of hiring managers' preferences to base hiring decisions on their intuition. As summarized previously, some prior research has considered whether individual differences among selection decision-makers influences how they make selection decisions, which could be relevant to which selection procedures they utilize, and which properties of selection procedures are important to them. One of these variables is preference for intuition-based hiring. A six-item scale from Lodato et al. (2011) was administered in the present study, which contained the following items: "I believe it is important to rely on your 'gut' when hiring employees", "It is important to rely on your instincts when hiring an employee", "I believe it is important to rely on your intuition when hiring employees", "Hiring an employee is more of an art than a science", "You can't always explain why a candidate is the best one – you just know it", and "You can 'read between the lines' to detect whether someone is suitable to hire." This measure had high reliability ($\alpha = .90$).

Social Dominance Orientation. Users' individual attitudes and values relating to fairness and egalitarianism might inform the extent to which they give weight to selection procedure fairness beliefs in the extent to which they intend to use different pieces of selection information in overall candidate judgments. Thus, a measure of social dominance orientation (SDO) was included as an exploratory moderator variable of the relationship between fairness beliefs and utilization intentions. SDO was measured using the four items from con-trait anti-egalitarianism dimension of the SDO-7 scale (Ho et al., 2015). Items included, "We should work to give all

groups an equal chance to succeed", "We should do what we can to equalize conditions for different groups", "No matter how much effort it takes, we ought to strive to ensure that all groups have the same chance in life", and "Group equality should be our ideal." This measure demonstrated high reliability ($\alpha = .92$). All items were reverse scored such that higher values indicated higher SDO.

Study 1 Analytic Approach

Due to the nested experimental design (i.e., each participant viewed and rated three selection procedures on the same measures), it was a concern that non-independence of observations would subsequently bias tests using the general linear model. Indeed, intraclass correlations indicated some degree of clustering effects; most intraclass correlation values were less than .20 but they ranged up to .41 (see Table 2). To manage effects of non-independence, cluster-robust standard errors were computed within MANOVA tests and path models. This approach is recommended over multilevel modeling to deal with clustering effects when they are a nuisance due to study design (McNeish et al., 2017). Further, simulation studies (e.g., McNeish, 2014) indicate that tests relying on cluster-robust standard errors are less biased than multilevel model tests in situations with sparse data clusters such as in this study (i.e., there were only three observations per cluster or individual). All MANOVA results and path model results reported below were conducted utilizing cluster robust standard errors to determine *p* values for significance testing to mitigate effects of non-independent observations due to study design. Cluster robust standard errors were computed using the Huber-White procedure (also known as the sandwich estimator) which takes clustering into account and assumes independence only among cluster units, not individual observations.

Study 1 Results

Study 1 Descriptive Results

Table 3 displays descriptive statistics and bivariate correlations among Study 1 variables. Table 4 displays hiring managers' ratings of intentions to utilize information from each predictor. Table 4 also displays results from when participants were asked to rate the extent to which they would weight candidate information from the five different sources in the hiring decision (i.e., the three selection procedures from the experiment as well as resume-based experience and education). Table 5 displays ratings of predictiveness beliefs and fairness beliefs (outcome and procedural fairness) by predictor method and construct in aggregate and by specific method-construct combinations.

MANOVA tests were conducted to determine if ratings of utilization intentions, predictiveness beliefs, outcome fairness beliefs, and procedural fairness beliefs (see Tables 4-5) differed by *predictor method* condition (i.e., aggregated across *predictor construct* conditions). Results suggested that ratings of intentions to utilize differed by predictor method, $F(2,867) = 12.17$, $MSE = 1.04$, $p < .001$, $\eta^2 = .027$. Post hoc tests showed that structured interviews received higher ratings of intentions to utilize than unstructured interviews ($p < .01$) and computerized assessments ($p < .001$), but there was not a significant difference between unstructured interviews and computerized assessments ($p = .19$). For the next outcome variable, results suggested that ratings of predictiveness beliefs also differed by predictor method, $F(2,867) = 7.14$, $MSE = .76$, $p < .001$, $\eta^2 = .021$. Post hoc tests revealed that structured interviews received higher ratings of predictiveness than unstructured interviews ($p < .05$) and computerized assessments ($p < .001$), but there was no difference between unstructured interviews and computerized assessments ($p = .27$). Next, results showed that outcome fairness beliefs also

differed by predictor method, $F(2,867) = 19.34$, $MSE = 1.25$, $p < .001$, $\eta^2 = .043$. Post hoc tests showed that there was no difference between structured interviews and computerized assessments ($p = .96$), which received the highest ratings of outcome fairness; unstructured interviews received lower outcome fairness ratings than structured interviews ($p < .001$) and computerized assessments ($p < .001$). Finally, procedural fairness beliefs also differed by method, $F(2,867) = 65.39$, $MSE = .90$, $p < .001$, $\eta^2 = .144$. Post hoc tests revelated unstructured interviews received lower ratings than structured interviews ($p < .001$) and computerized assessments ($p < .001$) but there was no difference between structured interviews and computerized assessments ($p = .69$).

Another set of MANOVA tests were conducted to determine if ratings of utilization intentions, predictiveness beliefs, outcome fairness beliefs, and procedural fairness beliefs differed by *predictor construct* condition (i.e., aggregated across *methods* conditions; see Tables 4-5). Results suggested that ratings of intentions to utilize differed by predictor construct, $F(2,867) = 22.22$, $MSE = 1.02$, $p < .001$, $\eta^2 = .049$. Post hoc tests revelated that job knowledge received higher utilization intentions ratings than cognitive ability ($p < .01$) and conscientiousness ($p < .001$). Further, cognitive ability received higher ratings of intentions to utilize than conscientiousness ($p < .01$). Next, results showed that predictiveness beliefs also differed by predictor construct, $F(2,867) = 16.32$, $MSE = .75$, $p < .001$, $\eta^2 = .036$. Post hoc tests revealed that job knowledge received higher predictiveness ratings than cognitive ability ($p < .05$) and conscientiousness ($p < .001$) and that cognitive ability received higher predictiveness ratings than conscientiousness ($p < .01$). Results suggested, however, that ratings of outcome fairness did not differ by predictor construct, $F(2,869) = 2.53$, $MSE = 1.30$, $p = .08$, $\eta^2 = .006$. Finally, ratings of procedural fairness differed by predictor construct, $F(2,867) = 5.06$, $MSE =$

1.03, $p < .01$, $\eta^2 = .012$. Post hoc tests showed that conscientiousness received lower outcome fairness ratings than job knowledge ($p < .01$) and cognitive ability ($p < .05$), the latter two of which did not differ ($p = .37$).

Tables 4-5 also display means and standard deviations for ratings of utilization intentions, predictiveness beliefs, and fairness beliefs for predictor method and construct condition combinations. Among these combinations, structured interviews measuring job knowledge received the highest ratings of intentions to utilize and predictiveness beliefs, and among the highest ratings of outcome and procedural fairness beliefs. Computerized assessments measuring conscientiousness received the lowest ratings of intentions to utilize and predictiveness, while unstructured interviews measuring conscientiousness received the lowest ratings of both outcome and procedural fairness beliefs. Compared to the selection procedures manipulated in the experimental design, prior work experience from candidate resumes received among the highest ratings of intentions to utilize, but educational background from resumes received the lowest ratings of intentions to utilize.

Table 6 displays mean ratings of user reactions dimensions by predictor constructs and methods conditions in aggregate. First, MANOVA tests were conducted to test for mean differences by predictor method conditions across predictor construct conditions. Results suggested that procedural autonomy perceptions differed by predictor method, $F(2,867) = 48.09$, $MSE = 1.13$, $p < .001$, $\eta^2 = .100$. Post hoc follow up testing indicated that unstructured interviews received higher ratings of procedural autonomy than structured interviews ($p < .001$) and assessments ($p < .001$). Structured interviews received higher ratings of procedural autonomy than computerized assessments ($p < .001$). Results also suggested that evaluation autonomy perceptions differed by predictor method, $F(2,867) = 51.35$, $MSE = 1.24$, $p < .001$, η^2

= .106. Post hoc tests revealed similar results; unstructured interviews received higher ratings of evaluation autonomy than structured interviews ($p < .001$) and assessments ($p < .001$), and structured interviews higher than assessments ($p < .001$). Transparency perceptions also differed by method, $F(2,867) = 68.74$, $MSE = 1.09$, $p < .001$, $\eta^2 = .137$. Unstructured interviews received lower ratings of transparency than structured interviews ($p < .001$) and computerized assessments ($p < .001$). Structured interviews were not different from computerized assessments ($p = .99$). Next, fakability perceptions differed by predictor method, $F(2,867) = 26.17$, $MSE = 1.63$, $p < .001$, $\eta^2 = .057$. Unstructured interviews received higher fakability ratings than structured interviews ($p < .001$) and assessments ($p < .001$), and structured interviews higher than assessments ($p < .05$). Perceptions of fidelity also differed by predictor method, $F(2,867) = 7.67$, $MSE = 1.14$, $p < .001$, $\eta^2 = .017$; structured interviews received higher fidelity ratings than assessments ($p < .001$) but were not different from unstructured interviews ($p = .22$). Unstructured interviews did not receive different fidelity ratings than assessments ($p = .07$). Finally, evaluation consistency perceptions differed by method, $F(2,867) = 143.66$, $MSE = 1.12$, $p < .001$, $\eta^2 = .249$. Assessments received higher ratings than structured interviews ($p < .01$) and unstructured interviews ($p < .001$), and structured interviews received higher ratings of evaluation consistency than unstructured interviews ($p < .001$).

MANOVA tests were also conducted to test for mean differences in user reactions dimensions by predictor construct conditions across predictor method conditions. First, job relatedness perceptions differed by predictor construct, $F(2,867) = 23.26$, $MSE = .71$, $p < .001$, $\eta^2 = .051$. Post hoc tests revelated that job knowledge received higher ratings of job relatedness than cognitive ability ($p < .001$) and conscientiousness ($p < .001$); however, cognitive ability was not different from conscientiousness in job relatedness perceptions ($p = .63$). Next, malleability

perceptions differed by predictor construct, $F(2,867) = 93.47$, $MSE = .91$, $p < .001$, $\eta^2 = .177$. Job knowledge received higher ratings of malleability than conscientiousness ($p < .001$) and cognitive ability ($p < .001$), and conscientiousness was rated higher in malleability than cognitive ability ($p < .05$). Finally, construct development equity perceptions also differed by predictor construct, $F(2,867) = 9.38$, $MSE = .79$, $p < .001$, $\eta^2 = .021$. Conscientiousness received higher development equity ratings than cognitive ability ($p < .001$) but not different from job knowledge ($p = .27$), and job knowledge received higher ratings than cognitive ability ($p < .05$).

Overall, ratings of outcome variables and user reactions dimensions varied across predictor method and construct conditions. Among predictor methods, structured interviews tended to receive the most favorable ratings for outcomes and many user reactions dimensions. Among predictor constructs, job knowledge tended to be viewed most favorably. Unsurprisingly, structured interviews measuring job knowledge emerged as the method/construct combination hiring managers said they would utilize to the greatest extent of the nine possible combinations. Notably, this most favorable selection procedure was rivaled by prior experience from candidate resumes in ratings of utilization intentions and weighting in overall judgments.

Table 7 displays means and standard deviations of the extent to which hiring managers reported having prior experience with predictor methods and constructs included in this study. As shown in Table 3, within the corresponding predictor method conditions (across constructs), prior experience with a predictor method was positively related to utilization intentions ($r = .28$, $p < .001$) and predictiveness beliefs ($r = .26$, $p < .001$) but not related to outcome fairness beliefs ($r = .05$, $p = .17$) or procedural fairness beliefs ($r = .02$, $p = .58$). Within the corresponding predictor construct conditions (across predictor methods), prior experience with a predictor construct was positively related to utilization intentions ($r = .30$, $p < .001$), predictiveness beliefs

(r = .28, p < .001), outcome fairness beliefs (r = .11, p < .01), and procedural fairness beliefs (r = .13, p < .001).

Table 7 also displays means and standard deviations for the extent to which hiring managers felt it was a norm to use a predictor method or construct in hiring processes within hiring managers' occupational fields. Within the corresponding predictor method conditions (across constructs), occupational norms using a predictor method were positively related to utilization intentions (r = .30, p < .001), predictiveness beliefs (r = .27, p < .001), and outcome fairness beliefs (r = .07, p < .05), but not related to procedural fairness beliefs (r = .05, p = .13). Within the corresponding predictor construct conditions (across predictor methods), occupational norms using a predictor construct were positively related to utilization intentions (r = .37, p < .001), predictiveness beliefs (r = .32, p < .001), outcome fairness beliefs (r = .10, p < .01), and procedural fairness beliefs (r = .15, p < .001).

Significant relationships between both norms using predictor methods and constructs and prior experience using predictor methods and constructs with utilization intentions, predictiveness beliefs, and fairness beliefs suggests they should be considered as potential control variables within the conceptual model. Because norms and previous experience using predictor methods and constructs were highly correlated, and because previous experience with predictors is likely more proximal (i.e., norms likely lead to previous experience using predictors), only prior experience with predictor methods and constructs were considered as control variables when testing the conceptual model.

Study 1 Hypothesis Tests

Next, the conceptual model was tested. A path modeling approach using Mplus was utilized to allow for the testing of model fit to compare alternative models. As previously

mentioned, cluster robust standard errors were computed using the Huber-White or sandwich procedure (i.e., using the "MLR" estimator in Mplus specifying the person-level variable as the cluster variable). Model results reported below represent standardized estimates. Because CFA results indicated that outcome fairness and procedural fairness beliefs were distinct factors, separate models were tested with each of these dimensions of fairness beliefs. Post hoc estimates of power for the overall hypothesized structural model using the Muthen and Muthen (2002) approach was 1.00.

Models with Outcome Fairness Beliefs. The first model estimated the paths in the hypothesized conceptual model using the outcome fairness beliefs variable (i.e., predictiveness beliefs and outcome fairness beliefs predicting intentions to utilize; procedural autonomy, evaluation autonomy, fidelity, fakability, transparency, job relatedness, and malleability perceptions predicting predictiveness beliefs; job relatedness, fakability, evaluation consistency, and construct development equity predicting outcome fairness beliefs). Additionally, direct paths from each perceptions variable to utilization intentions were estimated to test for mediation effects. This initial model demonstrated good fit (χ^2 = 55.66, RMSEA = .08, SRMR = .02, CFI = .97, TLI = .89; R^2 estimate = .53). Residuals and modification indices recommended adding a path from predictiveness beliefs to outcome fairness beliefs. Because the hypothesized model was exploratory in nature and this change made sense conceptually, it was added in the second version of the model. This second version of the model displayed excellent fit (χ^2 = 3.75, RMSEA < .01, SRMR = .01, CFI > .99, TLI > .99; R^2 estimate = .54). Results from this model are displayed in Table 8.

Table 8 also displays results for mediation analyses. Standard errors were computed using the bootstrapping procedure approach with 1,000 iterations. First, results showed that

utilization intentions were positively predicted by predictiveness beliefs (β = .531, *SE* = .038, p < .001) supporting H1a, and by outcome fairness beliefs (β = .088, *SE* = .034, p < .05), supporting H1b. Predictiveness beliefs were positively related to outcome fairness beliefs (β = .147, *SE* = .036, p < .001), and there was a significant indirect effect between predictiveness beliefs and utilization intentions via outcome fairness beliefs (β = .013, *SE* = .006, p < .05). A strong direct effect from predictiveness beliefs to utilization intentions remained.

Below is a summary of user reactions dimensions hypothesized to influence utilization intentions via predictiveness beliefs (see full model results in Table 8). Procedural autonomy perceptions were positively related to utilization intentions, and this effect was partially mediated by predictiveness beliefs, supporting H2a; a direct effect on utilization intentions remained. Evaluation autonomy perceptions were negatively related to utilization intentions, and there was no indirect effect via predictiveness beliefs. As such, H2b was not supported. Fidelity perceptions were positively related to utilization intentions, and this effect was fully mediated through predictiveness beliefs, supporting H2c. Transparency perceptions were positively related to utilization intentions, and this effect was partially mediated by predictiveness beliefs, which supported H2d; a direct effect on utilization intentions remained. Fakability perceptions were negatively related to utilization intentions, and this effect was fully mediated via predictiveness beliefs; thus, H2e was supported. Job relatedness perceptions were positively related to utilization intentions; however, this effect was not significantly mediated through predictiveness beliefs, and a direct effect remained. This indicates that H3a was only partially supported. Finally, malleability perceptions did not have a significant total effect on utilization intentions; however, malleability was negatively related to predictiveness beliefs, and the indirect effect on

predictiveness was significant, providing partial support for H3b. A positive direct effect between malleability perceptions and utilization intentions also existed.

Next is a summary of user reactions dimensions hypothesized to influence utilization intentions via outcome fairness beliefs. Evaluation consistency perceptions did not have a total effect on utilization intentions, but there was a significant indirect effect between evaluation consistency perceptions and utilization intentions via outcome fairness beliefs, providing partial support for H4a. Fakability perceptions were negatively related to utilization intentions, and outcome fairness beliefs had a significant indirect effect in this relationship; thus, H4b was supported. Job relatedness perceptions were positively related to utilization intentions, but this effect was not mediated by outcome fairness beliefs and a direct effect remained, suggesting only partial support for H5a. Construct development equity perceptions did not have a total effect on utilization intentions, but there was a significant indirect effect between construct development equity perceptions and utilization intentions via outcome fairness beliefs, partially supporting H5b. A positive direct effect between construct development equity perceptions and utilization intentions existed.

An additional exploratory model was run in which all paths not hypothesized between user reactions dimensions variables and predictiveness and fairness beliefs were added. This model showed that predictiveness beliefs were not predicted by evaluation consistency perceptions (β = .004, *SE* = .095, p = .63) nor construct development equity perceptions (β = .043, *SE* = .026, p = .11). Additionally, outcome fairness beliefs were not predicted by procedural autonomy perceptions (β = .037, *SE* = .056, p = .51), evaluation autonomy perceptions (β = -.018, *SE* = .054, p = .74), fidelity perceptions (β = -.025, *SE* = .045, p = .59),

malleability perceptions (β = .006, *SE* = .031, p = .84), nor transparency perceptions (β = -.036, *SE* = .045, p = .42).

Finally, an additional model was run in which hiring managers' experience with predictor constructs and methods were added to the proposed conceptual model as control variables. This was done because both of these variables were significantly related to predictiveness beliefs and utilization intentions (see Table 3). Residuals and modification indices again recommended adding a path from predictiveness beliefs to outcome fairness beliefs, so this path was added to the next model iteration. Results from this model showed that hiring managers' experience with a predictor method was positively related to utilization intentions of that selection procedure (β = .091, *SE* = .027, p < .01). This relationship was partially mediated by predictiveness beliefs ("a" path β = .089, *SE* = .034, p < .01; indirect effect β = .045, *SE* = .018, p < .05), and a significant direct effect on utilization intentions remained (β = .091, *SE* = .027, p < .01). There was not a significant relationship between experience with predictor methods and outcome fairness beliefs (β = .044, *SE* = .029, p = .13), and there was not a significant indirect effect via outcome fairness beliefs. Next, results showed that hiring managers' experience with a predictor construct was not significantly related to utilization intentions of that selection procedure (β = .045, *SE* = .031, p = .15). Experience with predictor constructs was not related to selection procedure predictiveness beliefs (β = .041, *SE* = .032, p = .20) nor to outcome fairness beliefs (β = -.024, *SE* = .031, p = .43).

Importantly, no model results changed substantially after including experience with predictor constructs and methods as control variables. The effect of predictiveness beliefs on utilization intentions decreased only slightly (β = .514 vs. β = .531). Effects of user beliefs and user reactions dimensions on utilization that were previously significant remained so (and vice

versa), all effects on predictiveness beliefs that were previously significant remained so (and vice versa), and all mediation effects remained the same. This suggests that prior experience with predictor methods does explain some variance in utilization intentions, but the effects of this variable are relatively small compared to effects of user reactions dimensions and beliefs on utilization and do not change model results.

Overall, most of the proposed elements of the conceptual model were supported (see Figure 2). Predictiveness and outcome fairness beliefs were both related to utilization intentions although predictiveness beliefs had a much larger effect. Hypotheses for all proposed dimensions of user reactions received at least partial support except for evaluation autonomy, which had a small negative effect on utilization intentions. Malleability, evaluation consistency, and development equity did not have significant total effects on utilization intentions (but had significant indirect effects as hypothesized), and job relatedness had a significant effect on utilization intentions but was not mediated by predictiveness beliefs or outcome fairness beliefs. Procedural autonomy, fidelity, fakability, and transparency perceptions all had significant effects on utilization intentions in the hypothesized directions and were mediated by predictiveness or outcome fairness beliefs as hypothesized. Further, exploratory analyses showed that none of the user reactions dimensions had relationships with predictiveness or fairness beliefs that were not hypothesized, providing additional support for the conceptual model.

Models with Procedural Fairness Beliefs. The next set of models also began with the paths in the hypothesized conceptual model but used the procedural fairness beliefs variable instead of the outcome fairness beliefs variable. Direct paths from each perceptions variable to utilization intentions were again estimated to test for mediation effects. This initial model demonstrated good fit (χ^2 = 55.66, RMSEA = .08, SRMR = .02, CFI = .97, TLI = .89; R^2 estimate = .53).

Similar to the model with outcome fairness beliefs, residuals and modification indices recommended adding a path from predictiveness beliefs to procedural fairness beliefs. This second version of the model demonstrated excellent fit (χ^2 = 16.68, RMSEA = .04, SRMR = .01, CFI = .99, TLI = .97; R^2 estimate = .54), and model results are also displayed in Table 9.

Table 9 also displays results for mediation analyses. Standard errors were again computed using the bootstrapping procedure approach with 1,000 iterations. First, results showed that utilization intentions were positively predicted by predictiveness beliefs (β = .534, *SE* = .039, $p < .001$). However, procedural fairness beliefs were not a significant predictor of utilization intentions (β = .070, *SE* = .040, p = .08). Predictiveness beliefs were positively related to procedural fairness beliefs (β = .170, *SE* = .029, $p < .001$), but there was not a significant indirect effect between predictiveness beliefs and utilization intentions via procedural fairness beliefs (β = .012, *SE* = .007, p = .11).

Relationships between user reactions dimensions and predictiveness beliefs are shown in Table 9. Next is a summary of user reactions dimensions hypothesized to influence utilization intentions via *procedural fairness beliefs* (i.e., a test of H4 and H5 with a different fairness measure). Evaluation consistency perceptions did not have a total effect on utilization intentions, and there was not a significant indirect effect between evaluation consistency and utilization intentions via procedural fairness beliefs, failing to support H4a. Fakability perceptions were negatively related to utilization intentions, but procedural fairness beliefs did not mediate this relationship, suggesting only partial support for H4b. Job relatedness perceptions were positively related to utilization intentions, but this effect was not mediated by procedural fairness beliefs and a direct effect remained, suggesting only partial support for H5a. Construct development equity perceptions did not have a total effect on utilization intentions, and there was not a

significant indirect effect between construct development equity and utilization intentions via procedural fairness beliefs, providing only partial support for H5b.

An additional exploratory model was run in which all paths not hypothesized between user reactions dimensions variables and predictiveness and fairness beliefs were added (non-hypothesized relationships between user reactions dimensions and predictiveness beliefs were reported in the previous section). This model showed that procedural fairness beliefs were not predicted by procedural autonomy perceptions (β = .018, *SE* = .045, p = .70), evaluation autonomy perceptions (β = -.042, *SE* = .046, p = .36), fidelity perceptions (β = -.022, *SE* = .035, p = .53), nor malleability perceptions (β = -.025, *SE* = .025, p = .32). However, transparency perceptions were positively related to procedural fairness beliefs (β = .113, *SE* = .040, $p < .01$).

An additional model was again run in which hiring managers' experience with predictor constructs and methods were added to the proposed conceptual model as control variables. Residuals and modification indices again recommended adding a path from predictiveness beliefs to procedural fairness beliefs, so this path was added to the next model iteration. Effects from these two variables on utilization intentions and predictiveness beliefs were previously reported. There was not a significant relationship between experience with predictor methods and procedural fairness beliefs (β = .012, *SE* = .017, p = .48), and there was not a significant indirect effect via procedural fairness beliefs. Experience with predictor constructs was also not related to procedural fairness beliefs (β = -.013, *SE* = .021, p = .54). Once again, results suggested that no model results changed substantially after including experience with predictor constructs and methods as control variables.

Overall, the model with procedural fairness received less support compared to the model with outcome fairness. First, procedural fairness beliefs were not significantly related to

utilization intentions. The four dimensions of user reactions proposed to predict procedural fairness beliefs were significant in doing so; perceptions of two dimensions (evaluation consistency and job relatedness) were more strongly related to procedural fairness beliefs than outcome fairness beliefs but the other two (fakability and development equity) were more strongly related to outcome fairness beliefs. Importantly, while three of these four dimensions were partially or fully mediated by outcome fairness beliefs, none of the four had significant indirect effects on utilization intentions via procedural fairness beliefs, suggesting procedural fairness beliefs do not play a mediating role.

Study 1 Exploratory Analyses

One set of exploratory analyses relates to the influence of hiring managers' endorsement of intuition-based hiring decision-making practices on relationships between autonomy perceptions and predictiveness beliefs. In a regression model with procedural autonomy perceptions and intuition hiring predicting predictiveness beliefs, procedural autonomy perceptions (ß = .34, t = 10.74, p < .001) were positively related to predictiveness beliefs but intuition hiring (ß = .05, t = 1.43, p = .15) was not. In a second model, the interaction between procedural autonomy perceptions and intuition hiring explained significant variance over the original model, ΔR^2 = .011, ΔF = 11.17, p < .001. Figure 3 depicts the interaction via simple slopes and shows that the procedural autonomy perceptions-predictiveness beliefs relationship is more strongly positive among hiring managers who highly endorse intuition hiring practices.

In a similar regression model but with evaluation autonomy perceptions, evaluation autonomy perceptions (ß = .32, t = 9.77, p < .001) were positively related to predictiveness beliefs but intuition hiring (ß = .04, t = 1.19, p = .23) was not. In a second model, the interaction between evaluation autonomy perceptions and intuition hiring explained significant variance

over the original model, $\Delta R^2 = .014$, $\Delta F = 13.49$, $p < .001$. Figure 4 depicts a similar interaction effect; the evaluation autonomy perceptions-predictiveness beliefs relationship is more strongly positive among hiring managers who highly endorse intuition hiring practices.

Another set of exploratory analyses explored the influence of hiring managers' prior hiring experience. As shown in Table 3, hiring managers' years of hiring experience was positively related to intentions to utilize ($r = .08$, $p < .05$), predictiveness beliefs ($r = .09$, $p < .05$), and intuition hiring tendencies ($r = .07$, $p < .05$). Next, prior hiring experience was tested as a potential moderator variable of relationships between user reactions dimensions and user beliefs, and of relationships between user beliefs and utilization intentions. In moderated regression tests, prior hiring experience did not moderate relationships between predictiveness beliefs, procedural fairness beliefs, nor outcome fairness beliefs, respectively, with utilization intentions. In nine additional moderated regression tests, prior hiring experience did not moderate relationships between any user reactions dimensions and utilization intentions with one exception. In the model with malleability perceptions predicting utilization intentions, malleability perceptions ($ß = .09$, $t = 2.79$, $p < .01$) and hiring experience ($ß = .08$, $t = 2.30$, $p < .05$) both positively predicted utilization intentions in the first model. In a second model, the malleability perceptions-hiring experience interaction ($ß = .29$, $t = 2.10$, $p < .05$) explained significant variance over the first model, $\Delta R^2 = .005$, $\Delta F = 4.42$, $p < .05$. Simple slopes depicting the interaction are shown in Figure 5; among hiring managers with low levels of hiring experience, malleability perceptions are unrelated to utilization intentions whereas malleability perceptions are positively related to utilization intentions among hiring managers with higher levels of hiring experience.

SDO was considered as a moderator of the relationship between fairness beliefs and utilization intentions. The first set of models was run with the procedural fairness beliefs variable. In the first model, procedural fairness beliefs (β = .42, t = 13.31, p < .001) and SDO (β = .10, t = 3.03, p < .01) were both significant predictors of utilization intentions. In a second step, the procedural fairness beliefs-SDO interaction did not explain significant variance over and above the first model, ΔR^2 < .001, ΔF = .15, p = .69. A second set of models was run with the outcome fairness beliefs variable. In the first model, outcome fairness beliefs (β = .33, t = 11.55, p < .001) and SDO (β = .11, t = 3.12, p < .01) were both significant predictors of utilization intentions. In a second step, the outcome fairness beliefs-SDO interaction did not explain significant variance over and above the first model, ΔR^2 = .003, ΔF = 2.91, p = .09. Thus, there was no evidence of SDO moderating the fairness beliefs-utilization intentions relationship.

Study 1 Discussion

Overall, results largely supported the proposed conceptual model. Interestingly, predictiveness beliefs were much more strongly related to utilization intentions than outcome or procedural fairness beliefs. While causality cannot be concluded from this study design, results also indicated that predictiveness beliefs are related to fairness beliefs. The precondition of validity for there to be fairness is an argument that makes sense logically and theoretically but was not hypothesized given alternative perspectives that validity and fairness can also be viewed as conceptually unrelated (e.g., Guion (2011) notes that flipping a coin to make selection decisions is not valid but could be viewed as fair). This relationship between predictiveness beliefs and fairness beliefs in how hiring managers view selection procedures is noteworthy and in line with applicant reactions research that suggests validity is a precondition of fairness from candidates' perspective.

Most of the proposed dimensions of user reactions predicted utilization intentions in the hypothesized direction and were mediated by predictiveness beliefs. Predictor method-related dimensions of user reactions tended to have larger effects on predictiveness beliefs than construct-related dimensions. In particular, the strongest predictors of predictiveness beliefs and utilization intentions were perceptions of predictor method fidelity (β = .48 and .32), procedural autonomy (β = .18 and .24), and transparency (β = .14 and .20). Fakability perceptions (β = -.13 and -.09) had negative effects on predictiveness beliefs and utilization intentions as hypothesized but to a smaller degree. Surprisingly, evaluation autonomy perceptions had a small negative total effect on utilization intentions and no relationship with predictiveness beliefs. Predictor construct job relatedness perceptions had a small effect on utilization intentions but were not mediated by predictiveness beliefs, and construct malleability perceptions had small effects on utilization intentions and predictiveness beliefs.

As noted, fairness beliefs had a smaller relationship with utilization intentions than predictiveness beliefs (β = .09 (outcome fairness) and β = .07 (procedural fairness) vs. β = .53 for predictiveness beliefs), and all four dimensions of user reactions purported to influence utilization intentions via fairness beliefs had smaller effects (βs ranging from -.03 to .14). Evaluation consistency perceptions were far and away the strongest predictor of procedural fairness beliefs (β = .56), while both evaluation consistency and development equity perceptions were strong predictors of outcome fairness beliefs (β = .33 and .31) Procedural fairness beliefs did not mediate any hypothesized relationships between user reactions dimensions and utilization intentions, while outcome fairness beliefs mediated these relationships for fakability, evaluation consistency, and development equity perceptions (but not job relatedness perceptions). Overall, outcome fairness beliefs seemed to fit better than procedural fairness beliefs within the

conceptual model as evidenced by mediation results and marginally better path model fit statistics. This indicates that hiring managers may be more concerned with the fairness of the end result of selection procedures than the fairness of processes within selection procedures.

One notable result from Study 1 was that among predictor methods, structured interviews (aggregated across predictor construct conditions) received the highest ratings of intentions to utilize and predictiveness beliefs. This finding is surprising and runs counter to previous findings suggesting that hiring managers largely dislike and resist using structured interviews (e.g., Chapman & Zweig, 2005; Dipboye, 1997; Highhouse et al., 2017; Lievens & De Paepe, 2004; Rynes et al., 2002; Terpstra, 1996; Terpstra & Rozell, 1997; Van der Zee et al., 2002). Structured interviews did receive relatively higher ratings of fidelity, which was a strong predictor of outcomes. Although unstructured interviews received the highest ratings of procedural and evaluation autonomy, as expected, hiring managers also rated unstructured interviews as having the highest levels of fakability and lowest levels of evaluation consistency. Beyond properties of the interview methods, it is possible that structured interviews are now being used more commonly than in the past. This explanation is plausible as many studies comparing user reactions to structured and unstructured interviews were conducted more than 20 years ago. Indeed, hiring managers reported greater levels of prior experience with structured interviews than unstructured interviews; further, they reported greater norms for use of structured interviews. It is also possible that hiring managers do not think about structure within interviews and thus were not able to accurately categorize their past experiences with interview procedures.

Also related to predictor methods conditions, computerized assessments received the lowest ratings of utilization intentions and predictiveness beliefs. Part of this reason could be that hiring managers reported having the least experience with computerized assessments as a

predictor method by a wide margin, a factor that was related to predictiveness and utilization intentions as discussed below. However, hiring managers also acknowledged that computerized assessments had relatively high levels of procedural and outcome fairness, high levels of transparency and evaluation consistency, and low levels of fakability. It was also interesting that hiring managers rated unstructured interviews as the least fair (in terms of process and outcomes) by wide margins.

Among predictor constructs (aggregated across methods), job knowledge had the highest utilization intentions and predictiveness beliefs, and among the highest fairness beliefs. It is possible that it was rated the best or near the best of the predictor constructs because job knowledge represents a more proximal predictor construct, as evidenced by it receiving by far the highest ratings of job relatedness. It is also possible that it was viewed so positively because it may be viewed as a broader construct compared to conscientiousness and cognitive ability. An alternative takeaway could be that hiring managers reported much less experience using selection procedures measuring cognitive ability and conscientiousness and also rated them much less positively in job relatedness. Conscientiousness, in particular, received surprisingly low ratings of utilization intentions, predictiveness, and outcome and procedural fairness.

However, another takeaway from this study is that procedures (methods aggregated across constructs, constructs aggregated across methods, and specific method/construct combinations) that were even rated the lowest relative to others still received mean ratings greater than four (i.e., "to a moderate extent") on the six-point scale of utilization intentions. This suggests that on average, hiring managers still appreciated having candidate information from these procedures and indicated they would use results from these procedures over nothing, or at least would not ignore these results completely in favor of results from alternative procedures.

Hiring manager perceptions of selection procedures along user reactions dimensions were logical or objectively "accurate" in many cases. For example, unstructured interviews were rated as the highest in procedural autonomy and evaluation autonomy and lowest in transparency and evaluation consistency. Computerized assessments were rated highly in transparency and evaluation consistency but were lowest in procedural and evaluation autonomy. Malleability ratings were highest for job knowledge followed by conscientiousness and cognitive ability. Each of these results is supported by logic and/or empirical evidence, and it deserves mention that hiring manager perceptions of selection procedures were uncontroversial and unsurprising in this regard. It is also interesting to consider how hiring managers perceived these selection procedures along dimensions in which results were uncertain or subject to individual judgment. For example, it is interesting that hiring managers perceived unstructured interviews as the most fakable and computerized assessments as the least fakable. It is also interesting that job knowledge received the highest ratings of job relatedness by a wide margin. Another noteworthy result is that cognitive ability received the lowest ratings of development equity.

Notably, except for interacting with malleability perceptions to predict utilization intentions, hiring experience did not predict user perceptions or beliefs or moderate any relationships between perceptions, beliefs, and utilization intentions. This was a bit surprising given prior evidence that hiring experience may play an important role in how hiring managers view and use selection procedures (e.g., Lodato et al., 2011). However, hiring managers' prior experience with predictor methods and constructs was related to beliefs about the predictiveness of procedures and utilization intentions. When entered as a control variable into path models, prior experience with predictor methods remained a significant predictor, but prior experience with predictor constructs did not, suggesting that prior experience with predictor methods may be

a more important factor. Importantly, prior experience with predictor methods and constructs had small overall effects on predictiveness and utilization intentions compared to user reactions dimensions, and the addition of these variables into path models did not alter estimates of relationships between user reactions dimensions and predictiveness beliefs or utilization intentions. These findings undercut a potential alternative explanation that perhaps prior experience with predictors is a more important driver of outcomes than hiring manager perceptions of predictors' properties. Rather, results suggest that prior experience with predictor methods influences hiring managers' utilization of information from selection procedures to a certain extent, but their perceptions of the properties of the selection procedures matter to a greater degree.

Procedural and evaluation autonomy perceptions were both more strongly related to predictiveness beliefs among hiring managers who endorsed intuition-based hiring practices to a greater degree. This finding is perhaps unsurprising and indicates that preferences for intuition-based hiring practices is an important individual difference variable that influences the extent to which hiring managers place importance upon autonomy as a property of selection procedures. Despite a plausible argument that hiring managers with greater levels of SDO would place less importance on fairness beliefs in their selection procedure utilization intentions, SDO did not moderate relationships between procedural and outcome fairness beliefs, respectively, with utilization intentions. Perhaps a moderation effect in line with this argument could be captured if an individual difference construct that is more closely related to personal values or beliefs about the importance of equity, diversity, or fairness specifically within organizational hiring practices was measured rather than a more general measure like SDO. Any measure attempting to capture this concept would likely be plagued by social desirability response bias.

Study 2

A follow up question to Study 1 relates to how organizations might be able to intervene to influence how users choose to utilize candidate information from different selection procedures. Successful interventions could reduce gaps between empirical evidence guiding optimal use of selection information and how practitioners actually utilize selection information. As summarized previously, one perspective about what drives user preferences of selection procedures suggests that autonomy plays a major role. In support of this idea, some researchers have examined how tweaking predictor methods to include greater levels of user autonomy might influence how these procedures are viewed. These strategies have been described as "autonomy-enhancing procedures" (Neumann et al., 2023b) and have most often been examined in the context of increasing autonomy within mechanical data combination systems to increase favorability of these systems (e.g., Dietvorst et al., 2015; 2018; Neumann et al., 2022; 2023b).

Autonomy-enhancing procedures could also be applied to predictor methods to influence how they are viewed within clinical data combination systems. For example, Nolan and Highhouse (2014) designed an intervention to increase use of structured interviews by adding small amounts of autonomy to a structured interview procedure; however, this intervention by itself failed to increase use intentions. If successful, these types of interventions have the potential to be incredibly useful to organizations as interventions that could be institutionalized within selection procedures. However, far less attention has been dedicated to enhancing perceptions of user autonomy within selection data collection procedures, especially for methods other than structured interviews. Therefore, Study 2 tests the efficacy of implementing autonomy-enhancing procedures within structured interview and computerized assessment procedures to influence user beliefs about the predictive validity of these procedures. Study 1

provided some preliminary support for these interventions as procedural autonomy perceptions were positively related to utilization intentions via predictiveness beliefs. Evaluation autonomy perceptions were not, which was surprising and warrants further investigation given the preponderance of previous assertions that autonomy influences hiring manager utilization of selection procedures. Based on extant research on autonomy-enhancing interventions in selection data combination situations, the following is hypothesized about selection data collection procedures in clinical data combination contexts:

H6a: Enacting process autonomy-enhancing interventions within predictor methods will increase user perceptions of procedural autonomy, thereby leading to more positive beliefs about selection procedure predictiveness and increased utilization intentions.

H6b: Enacting scoring autonomy-enhancing interventions within predictor methods will increase user perceptions of evaluation autonomy, thereby leading to more positive beliefs about selection procedure predictiveness and increased utilization intentions.

Study 2 Method

Study 2 tested influences of enacting autonomy-enhancing interventions within predictor methods on user reactions of evaluation and procedural autonomy, selection procedure predictiveness beliefs, and utilization intentions. To do this, a sample of hiring managers participated in an experimental procedure similar to that of Study 1 but only viewed a single structured interview or computerized assessment manipulated into high or low levels of evaluation autonomy and procedural autonomy.

As indicated by the conceptual model, it is not expected that enhancing user autonomy within selection procedures will influence users' fairness beliefs. Because this has not been previously examined, however, it is not entirely impossible that user autonomy perceptions and

fairness beliefs are related (e.g., perhaps users view themselves as fair evaluators and that giving them more involvement in predictor method evaluation process would result in more fair outcomes for candidates). User fairness beliefs were also be measured in Study 2 as an exploratory variable to test this possibility and provide a more complete picture.

Similar to in Study 1, there is expected to be variance in user reactions dimensions and beliefs variables, especially for perceptions of autonomy which relate to the proposed interventions in Study 2. To provide a test of the efficacy of autonomy-enhancing interventions with some generalizability across methods, interventions were enacted for two common predictor methods (i.e., structured interviews and computerized assessments) as they are generally highly standardized methods and were thus expected to be perceived as having lower levels of user autonomy.

Study 2 Sample

A sample of hiring managers from Prolific was recruited for this study. Sampling criteria stipulated that participants must live and work full-time in the U.S. and be employed in a hiring manager role in which they have experience making employee selection decisions. Participants who completed the study and passed all data quality checks were compensated $2.60. The average survey completion time was about 10 and one-half minutes, resulting in an hourly compensation rate of nearly $15. A priori power analysis indicated that the required sample size to detect medium sized effects and achieve 80% power within the experimental design described below was 240 participants. 360 participants would be required to achieve 95% power.

The initial sample consisted of 487 participants. Of these, 36 participants indicated they did not have prior hiring experience, counter to Prolific's internal screening procedures. Additionally, 14 participants failed at least one of the two attention checks and 69 participants

failed at least one of the two vignette comprehension checks. The final sample consisted of 368 participants. The final sample was gender balanced (male = 50.3%), and most (72.6%) were White. Participants' average age was just over 41 years (SD = 10.62). Years of prior hiring experience ranged from one to 37 (M = 7.89, SD = 6.88), and the estimated number of employees hired over their career ranged from one to 1250 (M = 33.80, SD = 84.49). Over 70% of the sample had a bachelor's degree or higher, and just over 25% had a master's degree or higher.

Study 2 Procedure and Manipulations

After agreeing to the informed consent, participants were shown a vignette describing a fictional hiring scenario (see Appendix E). The first part of the vignette told participants to imagine that they were the hiring manager for a role they typically hire for in their actual career as a hiring manager. The vignette explained that recruiting staff had already screened the initial batch of applicants to produce a shortlist of candidates, and that it was now up to the hiring manager to provide overall evaluations of each candidate to determine the top candidate from this list. To help them make a decision, the organization required that all finalist candidates be subjected to a selection procedure to collect data about the candidates and help predict their likelihood of success. Participants were told that when making overall evaluations of each candidate, they would be able to use data from this selection procedure and data from the candidate's resume (i.e., education background and prior work experience).

Half of participants were shown a description of a structured interview measuring a battery of personality traits and the other half were shown a description of a computerized assessment measuring the same battery of traits. These two predictor methods were chosen due to the expectation that each will be viewed as having lower levels of user autonomy than other

common predictor methods. Each of these two predictor methods were manipulated into low or high levels of user evaluation autonomy and user procedural autonomy across conditions. Thus, the study design was a 2 (structured interview v. assessment method) x 2 (high v. low evaluation autonomy) x 2 (high v. low procedural autonomy) between-subjects experimental manipulation, resulting in eight total conditions, and participants were randomly assigned into one of these conditions. Procedure descriptions were written in a way that provided either no or a moderate degree of hiring manager control over the way in which the interview or assessment is administered (e.g., no choice in interview questions or the ability to select questions from a bank of options) and the way in which a candidate's overall performance score from that procedure is computed (e.g., being automatically provided an overall score or having the ability to give an overall score based on dimension scores). Importantly, the interview procedure described in the high procedural autonomy and high evaluation autonomy condition would still be considered a structured interview (i.e., structure level 3 out of 4) based on classifications of interview structure provided by Huffcutt et al. (2014).

After reading a description about the selection procedure available to help them hire a candidate in the hiring scenario (see selection procedure descriptions in Appendix F), participants responded to measures assessing the extent to which they would utilize information from this procedure in overall candidate judgments, their beliefs about the predictiveness and fairness of the procedure, as well as their perceptions of the procedure along four dimensions of user reactions toward predictor methods relevant to the experimental manipulation of predictor method autonomy levels: evaluation autonomy, procedural autonomy, transparency, and evaluation consistency. Participants were also asked to rate the extent to which they would utilize information related to candidates' educational background and employment experience in

their overall evaluations and the extent to which they believe this type of candidate information is predictive and fair. Finally, participants responded to measures assessing individual differences and sample characteristics.

Study 2 Measures

The same scales from Study 1 were used to measure the following primary variables of interest: utilization intentions (α = .90), predictiveness beliefs (α = .91), evaluation autonomy perceptions (α = .97), and procedural autonomy perceptions (α = .96). Also included as exploratory measures using the same scales from Study 1 were procedural fairness beliefs (α = .81) and outcome fairness beliefs (α = .81), preference for intuition hiring (α = .89), previous hiring experience (number of years), and two other user reactions dimensions relevant to the manipulation of predictor methods in this study (i.e., transparency perceptions (α = .90) and evaluation consistency perceptions (α = .90)). Two attention check items were included within the survey.

Individual Need for Autonomy. Users' need for autonomy at work is an individual difference that was explored as a potential moderator of relationships between procedural and evaluation autonomy perceptions and predictiveness beliefs or utilization intentions (i.e., relationships might be more strongly positive among those with greater autonomy needs). Individual need for autonomy was measured using six items adapted from the same Nolan and Highhouse (2014) scale used to measure autonomy perceptions. The item stems were adapted to state, "At work and in your job generally, it is important for me to feel a sense of…" followed by six items: control, choice, free will, influence, self-sufficiency, and freedom. This measure had adequate reliability (α = .83).

Study 2 Pilot Study

A pilot study was conducted to test the adequacy of the vignette experimental manipulations. Participants were masters of human resources students at a large midwestern university and were provided with course extra credit for participation. The initial sample included 95 participants, but 32 were excluded for failing data quality checks; thus, the final sample consisted of 62 participants. Participants were introduced to the fictional scenario and told to imagine they were hiring managers. Next, participants were shown the vignettes for each of the eight experimental conditions. After reading each vignette, participants rated their perceptions of the level of procedural autonomy and evaluation autonomy within the interview or assessment procedure provided to them as the hiring manager.

Results suggested the manipulations in the assessment conditions were successful. Procedural autonomy ratings were greater in the "high" procedural autonomy condition (M = 3.95, SD = .66) compared to the "low" condition (M = 3.16, SD = .96); this difference was significant, $t(61) = 5.73$, $p < .001$, and large in magnitude (d = 1.09). Further, evaluation autonomy ratings were greater in the "high" evaluation autonomy condition (M = 3.58, SD = .71) compared to the "low" condition (M = 2.98, SD = .85); this difference was also significant, $t(61) = 4.61$, $p < .001$, and large in magnitude (d = 1.03).

Results also suggested that manipulations in the interview conditions were also successful. Ratings of procedural autonomy were greater in the "high" procedural autonomy condition (M = 4.15, SD = .70) compared to the "low" condition (M = 3.06, SD = .98), a difference that was significant, $t(61) = 6.97$, $p < .001$, and large in magnitude, (d = 1.23). Additionally, ratings of evaluation autonomy were greater in the "high" evaluation autonomy

condition (M = 3.86, SD = .80) compared to the "low" condition (M = 3.26, SD = .95). This difference was significant, $t(61) = 4.09$, $p < .001$, and large (d = 1.17).

Study 2 Main Study Results

Table 10 displays descriptive statistics and bivariate correlations for variables measured in Study 2.

Study 2 Hypothesis Tests

MANOVA tests were conducted to determine whether interventions to enhance user autonomy via procedural and scoring control influence user perceptions of predictor method procedural and evaluation autonomy, respectively, as well as predictiveness beliefs, fairness beliefs, and intentions to utilize information from the procedure. Mediation tests utilized the bootstrapping procedure approach with 1,000 iterations.

H6a suggested that interventions to enhance process autonomy would result in increased perceptions of procedural autonomy as well as increased predictiveness beliefs and utilization intentions. In a MANOVA test as a function of process autonomy condition, procedural autonomy perceptions were higher in the high process autonomy condition (M = 3.84, SD = .84) than the low process autonomy condition (M = 2.62, SD = 1.14), $F(1,366) = 136.67$, MSE = .99, $p < .001$, d = 1.22. The interaction between procedure condition (i.e., assessment vs. interview) and process autonomy condition was significant, $F(1,364) = 10.92$, MSE = .91, $p < .01$, suggesting this relationship differed between the assessment and interview conditions. As shown in Figure 6, the difference in procedural autonomy perceptions between process autonomy conditions was greater for the interview condition (high process autonomy condition M = 3.75, SD = .87; low process autonomy condition M = 2.20, SD = .97; d = 1.68) than the assessment condition (high process autonomy condition M = 3.92, SD = .80; low process autonomy

condition $M = 3.04$, $SD = 1.14$; $d = .90$). Mediation testing suggested that procedural autonomy perceptions were positively related to utilization intentions (total $B = .31$, $SE = .04$, $z = 6.92$, $p < .001$) and that this effect was significantly mediated by predictiveness beliefs (indirect $B = .19$, $SE = .03$, $z = 6.51$, $p < .001$). Results suggested a partial mediation effect as a direct effect between procedural autonomy perceptions and utilization intentions remained ($B = .12$, $SE = .04$, $z = 3.07$, $p < .01$).

Additionally, predictiveness beliefs were higher in the high process autonomy condition (M = 3.21, SD = .88) than the low process autonomy condition (M = 2.99, SD = .96), $F(1,366) = 5.32$, $MSE = .85$, $p < .05$, $d = .24$; the procedure condition-process autonomy condition interaction was not significant in predicting predictiveness beliefs, indicating this relationship was similar in both the assessment and interview conditions. However, intentions to utilize did not differ between conditions, $F(1,366) = 3.39$, $MSE = .93$, $p = .066$, $d = .10$). The procedure condition-process autonomy condition interaction was also not significant in predicting intention to utilize. Overall, H6a was partially supported.

H6b suggested that interventions to enhance scoring autonomy would result in higher perceptions of evaluation autonomy as well as increased predictiveness beliefs and utilization intentions. In a MANOVA test as a function of scoring autonomy condition, evaluation autonomy perceptions were higher in the high scoring autonomy condition ($M = 3.83$, $SD = .92$) than the low scoring autonomy condition (M = 2.58, SD = 1.17), $F(1,366) = 125.01$, $MSE = 1.13$, $p < .001$, $d = 1.19$. The procedure condition-scoring autonomy condition interaction was not significant, indicating this relationship operated similarly in both the interview and assessment conditions. Mediation testing suggested that evaluation autonomy perceptions were positively related to utilization intentions (total $B = .32$, $SE = .04$, $z = 7.50$, $p < .001$) and that this effect

was significantly mediated by predictiveness beliefs (indirect $B = .17$, $SE = .02$, $z = 6.75$, $p < .001$). Results suggested a partial mediation effect as a direct effect between evaluation autonomy perceptions and utilization intentions remained ($B = .15$, $SE = .04$, $z = 3.82$, $p < .001$).

Predictiveness beliefs did not differ between conditions, $F(1,366) = 3.31$, $MSE = .85$, $p = .07$, $d = 19$. The procedure condition-scoring autonomy condition interaction was also not significant. However, intentions to utilize were higher in the high scoring autonomy condition ($M = 4.27$, $SD = .96$) than the low scoring autonomy condition ($M = 4.00$, SD = .96), $F(1,366) = 7.15$, $MSE = .92$, $p < .01$, $d = .28$. The procedure condition-scoring autonomy condition interaction was not significant in predicting intentions to utilize. Overall, H6b was partially supported.

In summary, the interventions had expected effects on procedural and evaluation autonomy perceptions, respectively. Further, both procedural and evaluation autonomy perceptions were significantly related to utilization intentions, and these relationships were mediated by predictiveness beliefs. As a function of process autonomy condition, predictiveness beliefs were higher but utilization intentions did not differ significantly. As a function of scoring autonomy condition, predictiveness beliefs did not differ significantly but utilization intentions were higher.

Study 2 Exploratory Analyses

Procedural and outcome fairness beliefs were also considered as exploratory outcomes of the process and scoring autonomy interventions. First, procedural fairness beliefs did not differ by process autonomy condition, $F(1,366) = 3.66$, $MSE = .90$, $p = .057$). Additionally, the procedure condition-process autonomy condition interaction was also not significant in predicting procedural fairness beliefs. However, outcome fairness beliefs were higher in the high

process autonomy conditions (M = 3.44, SD = 1.07) than in the low process autonomy conditions (M = 3.12, SD = 1.06), $F(1, 366) = 8.50$, MSE = 1.13, $p < .01$. The process autonomy condition-procedure condition interaction was not significant in predicting outcome fairness beliefs. Next, procedural fairness beliefs also did not differ by scoring autonomy condition, $F(1, 366) = .17$, MSE = .91, p = .68. Additionally, the interaction between scoring autonomy condition and procedure condition was not significant in predicting procedural fairness beliefs. Outcome fairness beliefs were also not different as a function of scoring autonomy condition, $F(1, 366) = 1.72$, MSE = 1.15, p = .19, and the scoring autonomy-procedure condition interaction was not significant in predicting outcome fairness beliefs. In sum, procedural fairness beliefs did not differ as a function of process or scoring autonomy conditions, and outcome fairness beliefs differed by process autonomy condition but not by scoring autonomy condition.

MANOVA tests also considered how ratings of user reactions dimensions may have differed as a function of procedural and scoring autonomy conditions. Transparency perceptions did not significantly differ by process autonomy condition, $F(1,368) = 3.32$, MSE = 1.23, p = .07, and the process autonomy condition-procedure condition interaction was not significant in predicting transparency perceptions. Additionally, evaluation consistency perceptions did not significantly differ by process autonomy condition, $F(1,368) = .29$, MSE = 1.08, p = .59, and the process autonomy condition-procedure condition interaction was not significant.

Transparency perceptions did differ as a function of scoring autonomy condition; they were higher in the high scoring autonomy conditions (M = 3.79, SD = 1.04) than in the low scoring autonomy conditions (M = 3.34, SD = 1.14), $F(1,368) = 15.36$, MSE = 1.18, $p < .001$, d = .41. The scoring autonomy condition-procedure condition interaction was not significant in predicting transparency perceptions. Additionally, evaluation consistency perceptions were lower

in the high scoring autonomy conditions (M = 3.66, SD = 1.17) than in the low scoring autonomy conditions (M = 3.94, SD = .90), $F(1,368) = 7.10$, MSE = 1.06, $p < .01$, d = .27. The scoring autonomy condition-procedure condition interaction was not significant in predicting evaluation consistency perceptions.

Relationships between user reactions dimensions measured in this study (i.e., procedural autonomy, evaluation autonomy, transparency, and evaluation consistency) and predictiveness beliefs, fairness beliefs, and utilization intentions were tested via regression. Testing these relationships is similar to tests of the conceptual model in Study 1, although not all user reactions dimensions were measured in Study 2 as not all were relevant to manipulations that were conducted on levels of user autonomy. Some relationships in Study 2 may differ because fewer dimensions are included in these models than in Study 1. Relationships may also differ because of differences in variance observed in certain variables due to the fact that experimental manipulations in Study 1 and Study 2 are fundamentally different. Nevertheless, exploratory tests of relationships between user reactions dimensions measured in Study 2 and predictiveness beliefs, fairness beliefs, and utilization intentions were tested to provide a complete picture.

First, predictiveness beliefs and fairness beliefs were considered as predictors of utilization intentions. A regression model considering procedural fairness beliefs and predictiveness beliefs as predictors of utilization intentions was significant, $F(2,365) = 148.22$, MSE = .52, $p < .001$, R^2 = .45. Predictiveness beliefs (β = .46, t = 10.31, $p < .001$) and procedural fairness beliefs (β = .32, t = 7.16, $p < .001$) were both positively related to utilization intentions. Relative weights analysis (Tonidandel & LeBreton, 2011) indicated that the predictors were not significantly different from each other (predictiveness beliefs raw relative weight 95% CI = .19-.33; procedural fairness beliefs raw relative weight 95% CI = .12-.24). A similar model but with

outcome fairness beliefs and predictiveness beliefs was also significant, $F(2,365) = 127.27$, $MSE = .56$, $p < .001$, $R^2 = .41$. Predictiveness beliefs ($\beta = .52$, $t = 11.71$, $p < .001$) and outcome fairness beliefs ($\beta = .22$, $t = 4.99$, $p < .001$) were both positively related to utilization intentions. In this case, relative weights analysis indicated that these predictors were significantly different from each other (predictiveness beliefs raw relative weight 95% CI = .22-.37; outcome fairness beliefs raw relative weight 95% CI = .07-.17).

Next, a model with a model with procedural autonomy, evaluation autonomy, transparency, and evaluation consistency as predictors of utilization intentions was significant, $F(4,363) = 37.18$, $MSE = .67$, $p < .001$, $R^2 = .29$. All predictors were positively related to utilization intentions in the following order of magnitude: transparency perceptions ($\beta = .27$, $t = 5.33$, $p < .001$), evaluation autonomy perceptions ($\beta = .18$, $t = 2.99$, $p < .01$), procedural autonomy perceptions ($\beta = .17$, $t = 2.82$, $p < .01$), and evaluation consistency perceptions ($\beta = .14$, $t = 3.06$, $p < .01$). A model with these four predictors and predictiveness beliefs as the outcome variable was also significant, $F(4,363) = 41.43$, $MSE = .60$, $p < .001$, $R^2 = .31$. All predictors were positively related to predictiveness beliefs in the following order of magnitude: transparency perceptions ($\beta = .26$, $t = 5.23$, $p < .001$), procedural autonomy perceptions ($\beta = .26$, $t = 4.45$, $p < .001$), evaluation consistency perceptions ($\beta = .16$, $t = 3.43$, $p < .001$), and evaluation autonomy perceptions ($\beta = .12$, $t = 2.04$, $p < .05$).

A model with procedural fairness beliefs as the outcome variable was also tested with evaluation consistency perceptions as a model-driven predictor and procedural and evaluation autonomy perceptions as exploratory predictors; this model was significant, $F(3,364) = 76.12$, $MSE = .56$, $p < .001$, $R^2 = .38$. All three predictors were positively related to procedural fairness beliefs in the following order of magnitude: evaluation consistency perceptions ($\beta = .57$, $t =$

13.76, $p < .001$), evaluation autonomy perceptions ($\beta = .14$, $t = 2.64$, $p < .05$), and procedural autonomy perceptions ($\beta = .13$, $t = 2.32$, $p < .05$). Finally, these same three predictors were entered in a model predicting outcome fairness beliefs, which was also significant, $F(3,364) = 39.52$, $MSE = .88$, $p < .001$, $R^2 = .24$. Evaluation consistency perceptions ($\beta = .41$, $t = 8.93$, $p < .001$) and procedural autonomy perceptions ($\beta = .19$, $t = 3.18$, $p < .01$) were positively related to outcome fairness beliefs but evaluation autonomy perceptions ($\beta = .11$, $t = 1.81$, $p = .07$) were not.

Preference for intuition hiring practices was again tested as a potential moderator variable. First, preference for intuition hiring was entered as a predictor along with procedural autonomy perceptions in a model with predictiveness beliefs as the outcome variable. In the first step of this model, procedural autonomy perceptions were positively related to predictiveness beliefs ($\beta = .42$, $t = 8.85$, $p < .001$) but intuition hiring preferences were not ($\beta = -.08$, $t = -1.72$, $p = .09$). In a second step, the procedural autonomy perceptions-intuition hiring preferences interaction did not explain additional variance beyond the first step, $\Delta R^2 = .001$, $\Delta F = .28$, $p = .60$. Next, preference for intuition hiring was entered as a predictor along with evaluation autonomy perceptions in a model with predictiveness beliefs as the outcome variable. In the first step of this model, evaluation autonomy perceptions were positively related to predictiveness beliefs ($\beta = .30$, $t = 8.30$, $p < .001$) but intuition hiring preferences were negatively related to predictiveness beliefs ($\beta = -.10$, $t = -2.02$, $p < .05$). In a second step, the evaluation autonomy perceptions-intuition hiring preferences interaction did not explain additional variance beyond the first step, $\Delta R^2 = .001$, $\Delta F = .50$, $p = .48$. Thus, there was no evidence of intuition hiring practices endorsement moderating relationships between autonomy perceptions and predictiveness beliefs in Study 2.

Hiring managers' need for autonomy was explored as a potential moderator variable between autonomy perceptions and predictiveness beliefs. First, a model was run with procedural autonomy perceptions and need for autonomy as predictors and predictiveness beliefs as the outcome. In the first step, procedural autonomy perceptions (β = .43, t = 8.86, p < .001) were positively related to predictiveness beliefs but need for autonomy was not (β = -.03, t = -.52, p = .60). In the second step of the model, the procedural autonomy perceptions-need for autonomy interaction did not explain variance in predictiveness beliefs over the first step, ΔR^2 = .002, ΔF = .76, p = .38. A similar set of models was run but with utilization intentions as the outcome variable. In the first step of this model, procedural autonomy perceptions were again positively related to utilization intentions (β = .37, t = 7.46, p < .001) but need for autonomy was not (β = .03, t = .62, p = .53). In the second step of the model, the procedural autonomy perceptions-need for autonomy interaction did not explain variance in utilization intentions over the first step, ΔR^2 = .001, ΔF = .53, p = .47.

Next, need for autonomy was explored as a moderator of between evaluation autonomy perceptions and predictiveness beliefs. In the first step of this model, evaluation autonomy perceptions (β = .40, t = 8.21, p < .001) were positively related to predictiveness beliefs but need for autonomy was not (β < .01, t = .05, p = .96). In the second step of the model, the evaluation autonomy perceptions-need for autonomy interaction did not explain variance in predictiveness beliefs over the first step, ΔR^2 = .001, ΔF = .49, p = .49. A similar set of models was run but with utilization intentions as the outcome variable. In the first step of this model, evaluation autonomy perceptions were again positively related to utilization intentions (β = .40, t = 8.22, p < .001) but need for autonomy was not (β = .05, t = .99, p = .32). In the second step of the model, the evaluation autonomy perceptions-need for autonomy interaction did not explain variance in

utilization intentions over the first step, $\Delta R^2 = .005$, $\Delta F = 2.30$, $p = .13$. In conclusion, there was no evidence of hiring managers' need for autonomy moderating relationships between procedural or evaluation autonomy perceptions with either predictiveness beliefs or utilization intentions.

Another set of exploratory analyses relates to participants' number of years of prior hiring experience. Years of hiring experience was not significantly related to utilization intentions, predictiveness beliefs, procedural or outcome fairness beliefs, or any of the four user reactions dimensions. Next, prior hiring experience was tested as a potential moderator variable of relationships between user reactions dimensions and user beliefs, and of relationships between user beliefs and utilization intentions. In moderated regression tests, prior hiring experience did not moderate relationships between utilization intentions and predictiveness beliefs, procedural fairness beliefs, nor outcome fairness beliefs, respectively. Further, hiring experience did not moderate the following relationships: procedural autonomy perceptions with predictiveness beliefs or utilization intentions; evaluation autonomy perceptions with predictiveness beliefs or utilization intentions; transparency perceptions with predictiveness beliefs or utilization intentions; nor evaluation consistency perceptions with procedural fairness beliefs, evaluation fairness beliefs, predictiveness beliefs, or utilization intentions.

Study 2 Discussion

As hypothesized, interventions to enhance levels of process and scoring autonomy within predictor methods resulted in increased perceptions of procedural and evaluation autonomy, respectively. Additionally, procedural and evaluation autonomy perceptions were positively related to utilization intentions and were mediated by predictiveness beliefs in both cases. However, while MANOVA results indicated that predictiveness beliefs were higher in the high

process autonomy compared to the low process autonomy condition, utilization intentions did not differ significantly by process autonomy condition despite a small mean difference in the expected direction. Conversely, MANOVA results by scoring autonomy condition indicated that utilization intentions were higher in the high scoring autonomy compared to the low scoring autonomy condition, but predictiveness beliefs did not differ significantly by scoring autonomy condition despite a small mean difference in the expected direction. Overall, mean differences in procedural and evaluation autonomy perceptions by the corresponding autonomy condition were large in magnitude (*d* values of 1.19 and 1.22), but mean differences by autonomy conditions were smaller for predictiveness beliefs (*d* values of .19 and .24) and utilization intentions (*d* values of .10 and .28). This indicates that while interventions to increase autonomy within predictor methods led to much higher perceptions of both procedural and evaluation autonomy, these interventions resulted in smaller effects on distal outcomes. Nevertheless, an argument could be made that interventions yielding effect sizes of this magnitude would be practically meaningful in terms of utility for organizations. That is, even small influences on hiring manager utilization of candidate results from certain selection procedures (e.g., greater utilization of procedures with larger validity coefficients) over many hiring decisions could have meaningful impacts on organizational outcomes such as performance and turnover.

In exploratory analyses, neither procedural nor outcome fairness beliefs differed as a function of scoring autonomy condition. Procedural fairness beliefs also did not differ by process autonomy condition, but outcome fairness beliefs were higher in the high process autonomy condition than the low process autonomy condition. This pattern of findings is interesting in that process-related autonomy and scoring-related autonomy interventions did not produce differences in theoretically corresponding fairness-related beliefs (i.e., process autonomy and

procedural fairness; scoring autonomy and outcome fairness), but process-related autonomy interventions produced differences in outcome fairness beliefs. In the other two predictor methods-related user reactions dimensions relevant to these interventions, transparency perceptions and evaluation consistency perceptions did not differ by process autonomy condition, but transparency perceptions were higher and evaluation consistency perceptions were lower in the "high" scoring autonomy condition than the "low" scoring autonomy condition. Hiring managers' perceptions of greater levels of transparency in how candidates are evaluated when they have more control in determining overall evaluations is unsurprising. However, hiring managers also seem to be aware that in these scenarios, the consistency with which candidates will be evaluated may decrease. This awareness and acknowledgement is noteworthy.

Predictiveness beliefs and procedural fairness beliefs were both positively related to utilization intentions in a regression model, and relative weights analysis indicated they were not different in magnitude. While outcome fairness beliefs were also a significant predictor of utilization intentions, they were less strong of a predictor compared to predictiveness beliefs. In both cases, the relative difference in magnitude between predictiveness beliefs and fairness beliefs was smaller in Study 2 than in Study 1, where predictiveness beliefs dominated in predicting utilization intentions. These results from Study 2 should be interpreted with caution as experimental manipulations focused on fewer selection procedures (i.e., only two predictor methods and a single predictor construct) in which autonomy was manipulated, a variable that was not proposed to strongly influence fairness beliefs in the first place.

In additional regression models, all four dimensions of user reactions measured in this study (i.e., procedural and evaluation autonomy perceptions, transparency perceptions, and evaluation autonomy perceptions) were positively related to predictiveness beliefs and utilization

intentions, with transparency perceptions being the strongest predictor of both. These results are noteworthy as evaluation autonomy perceptions were not positively related to these outcomes in Study 1. Evaluation consistency perceptions were the strongest predictor of both procedural and outcome fairness beliefs. Interestingly, procedural and evaluation autonomy perceptions were both positively related to procedural fairness beliefs, and procedural (but not evaluation) autonomy perceptions were significantly related to outcome fairness beliefs. Neither procedural nor evaluation autonomy perceptions were related to procedural or outcome fairness beliefs in Study 1. It is possible that autonomy was a significant predictor of outcomes in Study 2 simply because the experimental manipulations called attention to user autonomy and thus made it salient for participants.

Results did not suggest that prior hiring experience was related to utilization intentions, predictiveness or fairness beliefs, or the four user reactions dimensions. It also did not moderate relationships between predictiveness beliefs or fairness beliefs with utilization intentions, user reactions dimensions with predictiveness or fairness beliefs, or user reactions dimensions with utilization intentions. These results are counter to previous findings that hiring managers with more hiring experience were more likely to prefer procedures that involved greater levels of autonomy (e.g., Lodato et al., 2011). These results also do not support a plausible logical possibility that hiring managers with more experience want to leverage this experience by preferring less structured procedures that rely on intuition.

Hiring managers' reports of their need for autonomy at work did not moderate relationships between autonomy perceptions and predictiveness beliefs or utilization intentions, respectively. These results were unexpected; one potential explanation is that general autonomy needs at work could be less related to autonomy needs within hiring procedures as a hiring

decision-maker. It could also be the case that hiring managers generally have a high degree of autonomy to make many decisions at work and thus their frame of reference for this measure is oriented differently than many employees.

Hiring manager preferences for intuition-based hiring procedures also did not moderate relationships between autonomy perceptions and predictiveness beliefs or utilization intentions, respectively. These results were also counter to expectations. One potential reason why preferences for intuition-based hiring practices was not a significant moderator is that even "high" procedural and scoring autonomy conditions in this experimental design represented moderately structured hiring procedures (e.g., the high procedural and high scoring autonomy conditions would still be classified as a structured interview according to Huffcutt et al.'s (2014) guidelines). In other words, perhaps hiring managers' preference for intuition-based hiring practices did not moderate relationships in the present study because none of the procedures presented in vignettes truly offered a high degree of opportunity to rely on intuition.

General Discussion

The present research makes several contributions. First, prior research on how hiring managers view selection procedures in terms of favorability was extended to examine how they view them in terms of nine dimensions of predictor method and construct properties and why they would use candidate information from them. In line with hypotheses, hiring managers' beliefs about the predictiveness of selection procedures were strongly related to their intentions to utilize candidate results from these procedures. Hiring manager beliefs about the outcome fairness of these procedures were also positively related to utilization intentions to a much lesser degree, while procedural fairness beliefs were not related to utilization intentions. The finding that predictiveness beliefs seem to be the primary driver of utilization intentions is noteworthy

and supports related lines of research focusing on communicating selection procedure validity evidence to hiring managers. This finding is also notable because it suggests that while organizations are increasingly considering fairness outcomes when designing selection systems (Jones & Cunningham, 2023), hiring managers in the present research gave relatively little weight to their beliefs about the fairness of selection procedures when determining the extent to which they would utilize candidate results from these procedures. While fairness beliefs were not nearly as important as predictiveness beliefs, one novel finding from this research is that predictiveness beliefs and fairness beliefs are correlated. Conceptually, predictiveness beliefs are likely antecedents to utilization intentions, although future research should further examine this relationship.

Additional insight into the "why" behind hiring manager utilization of selection procedures was provided through proposing and testing a model of nine antecedent dimensions of user reactions toward selection procedures, or perceptions of selection procedure properties. Specifically, six properties of predictor methods and three properties of predictor constructs were proposed and tested as factors that influence how hiring managers view selection procedures in terms of predictiveness and fairness. Overall, predictor method-related dimensions of user reactions had larger effects than construct-related dimensions on utilization intentions and mediating mechanisms (i.e., predictiveness beliefs and fairness beliefs). These findings highlight the importance of adopting a modular view of selection procedures and thinking about the distinct influences predictor methods and constructs may have on why hiring managers hold certain beliefs about or choose to utilize (or not) candidate results from certain selection procedures.

Theoretical Implications

It has previously been asserted that the amount of autonomy within selection procedures is a significant determinant of hiring manager use of these procedures (e.g., Nolan & Highhouse, 2014). As discussed, the present research contributes by suggesting that multiple factors beyond user autonomy influence how hiring managers view selection procedures. Another important contribution, however, was made by taking a deeper look at how user autonomy may influence hiring manager views of selection procedures by separating procedural autonomy and evaluation autonomy as separate factors. In the Study 1 path model, procedural autonomy was positively related to predictiveness beliefs and utilization intentions, while evaluation autonomy was negatively related to utilization intentions and unrelated to predictiveness beliefs. When levels of autonomy were specifically manipulated in Study 2, results suggested that increasing hiring manager levels of both procedural and evaluation autonomy within predictor methods had small, positive effects on hiring manager perceptions of predictiveness and utilization. Similar to prior research demonstrating the effectiveness of autonomy-enhancing interventions to increase use of mechanical data combination approaches in selection decision-making contexts (e.g., algorithms and decision aids; Dietvorst et al., 2018; Neumann et al., 2022; 2023b), the present research suggests that designing predictor methods to have higher levels of autonomy can influence hiring managers' predictiveness beliefs and utilization intentions. As noted, these effects were small but could yield large benefits in terms of selection system utility if applied as an intervention to increase use of candidate information from selection procedures with large validity coefficients. Future research should further consider the role of different aspects of autonomy in hiring managers' use of candidate information from selection procedures.

Results from the present research also inspire additional conceptual questions about key variables. For example, the finding that predictiveness beliefs and fairness beliefs are correlated suggests the need for research that definitively determines the directional nature of this relationship. The idea that hiring managers' beliefs about the predictiveness (or validity) of selection procedures is a precondition for their beliefs about the fairness of these procedures is also a relationship that needs to be probed further.

Practical Implications

The present research has significant implications for potentially influencing how hiring managers use results from selection procedures within clinical data combination contexts. Taking the perspective that users are naïve or misinformed to a certain extent to evidence about selection procedures, one intervention strategy is to reduce lack of awareness by communicating evidence about selection procedures to hiring managers. Research in the previously summarized literature on validity communication has considered how presenting selection procedure validity evidence to decision-makers in different ways can influence their views of and intentions to utilize these procedures (e.g., Childers et al., 2021; Highhouse et al., 2017; Lee et al., 2021; Zhang et al., 2018). However, presenting validity evidence that can be comprehended by lay persons is challenging and subject to cognitive biases, as demonstrated in this body of research. Instead of focusing communication strategies broadly on validity evidence as in prior studies, another strategy would be to focus communication efforts on more specific factors that drive validity beliefs, such as the dimensions of user reactions from the present research. While some dimensions of user reactions represent subjective perceptions of selection procedures that are likely less flexible (e.g., user procedural and evaluation autonomy), perceptions of other dimensions could be altered by communication strategies based on scientific evidence. For

example, empirical evidence about the evaluation consistency or fakability of various selection procedures could be presented to users to correct misconceptions that may exist. Communicating several dimensions of user reactions could therefore alter perceptions of selection procedures, thereby influencing selection procedure predictiveness and fairness beliefs. Results from the present research suggest that such interventions to increase use of predictors should be particularly aimed at dimensions that may drive user beliefs about the predictiveness of selection procedures given that fairness beliefs may be less strongly related to utilization. Findings from Study 2 also indicate that simply making salient certain properties of selection procedures may also increase the importance of these properties to hiring managers.

In addition to influencing communication efforts, findings inform which properties of selection procedures should be given greater consideration in selection procedure design efforts. For example, because fidelity perceptions emerged as a strong antecedent of predictiveness beliefs and utilization intentions, assessment vendors should consider creating and organizations should consider implementing predictor methods with greater levels of fidelity (e.g., simulations, case study interviews, work samples). Another factor that seems important to users is the transparency of predictor methods in producing candidate scores. While some methods are inherently limited in their level of scoring transparency (e.g., unstructured interviews), others could likely enhance transparency. For example, computerized assessments commonly produce overall scores based on multiple factor or dimension sub-scores; explaining how overall scores were calculated could increase transparency to hiring managers. Additionally, structured interview processes could better explain how overall scores are generated based on sub-scores from specific questions or constructs/dimensions.

Another practical implication of this research relevant to selection procedure design is related to the level of procedural autonomy within selection procedures. Because procedural autonomy seems to matter to hiring managers, it may be worth seeking ways in which procedural autonomy can be enhanced to a certain degree while maintaining sufficient levels of standardization. Within interview procedures, for example, hiring managers could be given a list of options from which to select interview questions to assess a construct. Hiring managers could also be given the option to select which constructs or competencies to assess within interviews from options that have been validated and approved. Similarly, perhaps assessment vendors could design computerized assessment procedures in a modular fashion to enhance procedural autonomy for hiring managers. For example, hiring managers could be given the option to select certain constructs to assess candidates against for a specific role (e.g., selecting only several specific cognitive abilities, certain personality traits, or certain competencies), and the assessment would then be composed only of items measuring these concepts. In addition to potentially enhancing hiring manager perceptions of procedural autonomy and use of candidate results from these procedures, it would also enhance the candidate experience by shortening assessments.

Finally, predictor method transparency emerged as a strong predictor of selection procedure predictiveness beliefs and utilization intentions. This has important implications for selection procedures that rely on emerging technologies utilizing machine learning and artificial intelligence. These methods are often described as "black boxes" due to the lack of insight into how candidates are being evaluated, which has inspired research into the explainability of artificial intelligence-based predictor methods (e.g., Langer & König, 2022). Results from the present research suggest that transparency and explainability in how predictor methods evaluate

and score candidates is an important factor to hiring managers when determining the extent to which they will use candidate results from these procedures. While predictor methods relying on emerging technologies such as artificial intelligence were not specifically studied in the present research, this finding clearly has significant implications for hiring managers' potential adoption of these methods in organizational contexts.

Limitations and Future Directions

As with all empirical research, the present studies contained limitations. Both utilized cross-sectional study designs, which raises concerns inferring direction and causality, although the experimental design utilized by Study 2 provides some confidence in inferring causal effects. Relatedly, hiring managers responded to measures about selection procedures in the same order in both studies: first, about utilization intentions; next, about predictiveness and fairness beliefs; and finally, about perceptions of procedures along dimensions of user reactions. The order in which participants responded to these measures may have influenced results; for example, perhaps participants' responses to the user reactions dimensions measures were made in a way that attempted to justify their responses to the utilization intentions and predictiveness beliefs measures. However, ordering measures in the opposite direction would have also been problematic; this could have led to utilization intentions being more significantly influenced by user reactions dimensions after being prompted to consider certain properties of the selection procedures.

Despite sampling actual hiring managers, both studies offered limited degrees of realism; however, the use of experimental designs allowed for the controlling of potential confounding variables. In Study 1, the experimental vignette was contextualized within a hiring process for a human resources role to control for confounds such as job level, role type, and industry, whereas

in Study 2 hiring managers were told to imagine they were hiring for a role from their actual jobs as hiring managers in real life. Future research should consider how relationships and processes investigated in this study may operate differently for different types of jobs. For example, the job level of the role in question may interact with autonomy in driving predictiveness beliefs or utilization intentions; autonomy perceptions could be more important in influencing users' predictiveness beliefs in higher level roles. This is because for higher level roles (e.g., executive selection), hiring managers may feel more pressure to make accurate hiring judgments given the greater consequences of making a poor hiring decision. Under these conditions, selection decision-makers may be motivated by greater levels of self-preservation (Nolan et al., 2016) or simply default to information from methods they perceive as credible (Yaniv, 2004), both resulting in placing more weight on candidate information derived from methods over which they have more procedural control or that utilize their own intuition-based evaluative judgment. Thus, the role of job level should be explored.

Another contextual factor related to job level that may interact with user perceptions of method autonomy to produce predictiveness beliefs is the hiring volume of the role in question. Specifically, hiring managers for roles with greater levels of hiring volume (e.g., hourly roles in larger organizations) may place less importance in predictor method autonomy when evaluating procedures compared to hiring managers for roles with lower hiring volumes. In addition to the aforementioned rationale based on research on hiring manager self-preservation motivations (Nolan et al., 2016) and source credibility (Yaniv, 2004), this assertion is also based upon the practicality of hiring for roles with different levels of hiring volume. That is, it is likely to be more difficult for hiring managers to dedicate significant amounts of time to procedural or evaluation involvement in hiring processes for high volume roles. Under these conditions, when

candidate information is available from methods with lower levels of hiring manager involvement and autonomy (e.g., computerized assessments), the importance of autonomy when evaluating these procedures may decrease.

Another limitation of this research is that interview procedures and assessment procedures in Study 1 only measured a single construct whereas interviews and assessments in practical settings often measure multiple constructs or competencies within a single interview procedure. This difference may have influenced predictiveness and fairness evaluations of these procedures. Relatedly, the unstructured interview procedure in Study 1 was described as measuring specific constructs. According to the interview structure paradigm introduced by Huffcutt et al. (2014), unstructured interviews can measure specific constructs but be classified as unstructured due to lacking consistency in specific interview questions and lacking standardized rating scales. However, many unstructured interviews may not specify constructs or specify only broad domains (e.g., competencies). Future research should thus consider how hiring managers view procedures that measure not only different numbers of constructs but also do not define constructs.

Additionally, interview procedures in practical settings may measure different constructs than assessments. For example, it is perhaps less likely that cognitive ability would be assessed in interviews, whereas personality traits and job knowledge may be more commonly assessed via interviews and assessments. While some previous research has described how commonly hiring managers report using various predictors within organizational hiring processes (e.g., Risavy et al.. 2019; 2021), it is currently unknown how commonly predictors are used in various possible method/construct combinations. Future research should consider this question.

Another future research question is investigating the extent to which hiring decision-makers actually separate constructs and methods in how they think about selection procedures. Within the experimental procedure and measures in Study 1, hiring managers were prompted to think about different properties of selection procedures specifically relating to the predictor method or construct. It is possible that hiring managers do not separate these concepts naturally when thinking about and evaluating selection procedures.

The present research considered intentions to utilize candidate results from selection procedures as the ultimate outcome variable. Future research should employ a policy capturing design study in an attempt to replicate results from the present research while measuring actual utilization of information (i.e., relationships between predictor scores and overall candidate evaluation scores). Doing so could also offer the opportunity to study additional variables that influence how hiring managers utilize information from selection procedures, such as ordering effects in candidate information presentation and the consistency in which hiring managers are able to apply weights attached to different pieces of information.

As alluded to previously, future attention should be dedicated to testing communication or education-related interventions that could influence hiring manager perceptions and beliefs about selection procedures. The present research identifies the strongest correlates of utilization intentions in terms of broader beliefs (e.g., predictiveness beliefs) and specific dimensions of user reactions (e.g., fidelity perceptions). However, the extent to which strategies aimed at influencing information utilization via educational interventions would be successful is unknown. For example, perhaps hiring managers' perceptions or beliefs about some dimensions of user reactions are more amenable to change than others. Future research could answer this

question and lead to the design of successful interventions to influence hiring manager decision-making.

Concluding Thoughts

It is critical to investigate how hiring managers combine candidate data from multiple procedures into selection decisions when left to do so using their own judgment, as is most common in organizations, because misconceptions can systematically attenuate validity. The present research advances our understanding of how hiring managers view selection procedures by identifying multiple dimensions of selection procedure properties that drive hiring managers' intentions to utilize candidate results from these procedures. Findings have implications for designing selection procedures that are viewed positively by hiring managers and for focusing communications or training-based interventions on properties of procedures that are most important to hiring managers. Interventions based on this research could be tailored to influence hiring managers in ways that increase (decrease) use of more (less) valid predictors and sources of selection information, ultimately decreasing science-practice gaps in hiring manager judgement and decision-making in employee selection contexts.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anderson, N. (2005). Relationship between practice and research in personnel selection: Does the left hand know what the right is doing? In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 1–24). Malden, MA: Blackwell.

Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*(3), 291-304.

Arkes, H. R., González-Vallejo, C., Bonham, A. J., Kung, Y. H., & Bailey, N. (2010). Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making, 23*(3), 250-270. https://doi.org/10.1002/bdm.655

Arkes, H. R., Shaffer, V. A., & Dawes, R. M. (2006). Comparing holistic and disaggregated ratings in the evaluation of scientific presentations. *Journal of Behavioral Decision Making, 19*(5), 429-439. https://doi.org/10.1002/bdm.503

Arkes, H. R., Shaffer, V. A., & Medow, M. A. (2007). Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making, 27*(2), 189-202. https://doi.org/10.1177/0272989X06297391

Arvey, R. D., & Sackett, P. R. (1993). Fairness in selection: Current developments and perspectives. In N. Schmitt, & W. Borman (Eds.), *Personnel Selection* Jossey-Bass.

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Human factors, 58*(2), 242-260. https://doi.org/10.1177/0018720815610271

Bauer, T. N., Maertz, C. P., Jr., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology, 83*(6), 892–903. https://doi.org/10.1037/0021-9010.83.6.892

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*(2), 387–419. https://doi.org/10.1111/j.1744-6570.2001.tb00097.x

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

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Chamberlain, S. (2013). Qualification users' perceptions and experiences of assessment reliability. *Research Papers in Education*, *28*(1), 118-133. https://doi.org/10.1080/02671522.2012.754231

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*(1), 143–159. https://doi.org/10.1037/0021-9010.82.1.143

Chapman, D. S., & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology, 58*, 673-702. doi:10.1111/j.1744- 6570.2005.00516.x

Childers, M., Highhouse, S., & Brooks, M. E. (2021). Apples, oranges, and ironing boards: Comparative effect sizes influence lay impressions of test validity. *International Journal of Selection and Assessment*, ijsa.12367. https://doi.org/10.1111/ijsa.12367

Chiu, C., Hong, Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73*, 19–30. https://doi.org/10.1037/0022-3514.73.1.19

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O., & Ng, K. Y. (2001). Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of applied psychology, 86*(3), 425–445. https://doi.org/10.1037/0021-9010.86.3.425

Colquitt, J. A., Greenberg, J., & Greenberg, J. (2003). Organizational justice: A fair assessment of the state of the literature. *Organizational behavior: The state of the science*, 159-200.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674. DOI: 10.1126/science.2648573

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual review of organizational psychology and organizational behavior, 4*, 19-43. https://doi.org/10.1146/annurev-orgpsych-032516-113108

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

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

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Di Milia, L. (2004). Australian management selection practices: Closing the gap between research findings and practice. *Asia Pacific Journal of Human Resources*, *42*(2), 214-228. https://doi.org/10.1177/1038411104045364

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

Dunn, W. S., Mount, M. K., Barrick, M. R., & Ones, D. S. (1995). Relative importance of personality and general mental ability in managers' judgments of applicant qualifications. *Journal of Applied Psychology, 80*, 500–509.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological inquiry*, *6*(4), 267-285. https://doi.org/10.1207/s15327965pli0604_1

Einhorn, E. (2022). *Inside the vast national experiment in test-optional college admissions*. NBC News. Retrieved from https://www.nbcnews.com/news/us-news/college-admissions-test-sat-act-rcna23574

Fisher, P. A., Risavy, S. D., Robie, C., König, C. J., Christiansen, N. D., Tett, R. P., & Simonet, D. V. (2021). Selection Myths: A Conceptual Replication of HR Professionals' Beliefs About Effective Human Resource Practices in the US and Canada. *Journal of Personnel Psychology*, *20*(2), 51–60. https://doi.org/10.1027/1866-5888/a000263

Furnham, A. (2008). HR Professionals' Beliefs About, and Knowledge of, Assessment Techniques and Psychometric Tests. *International Journal of Selection and Assessment*, *16*(3), 300–305. https://doi.org/10.1111/j.1468-2389.2008.00436.x

Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *Academy of Management Review*, *18*(4), 694–734. https://doi.org/10.5465/amr.1993.9402210155

Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment*, *3*(1), 11-18. https://doi.org/10.1111/j.1468-2389.1995.tb00002.x

Gilliland, S. W., & Honig, H. (1994). *Development of the selection fairness survey.* Paper presented at the 9th Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Goldstein, I. L., Zedeck, S., & Schneider, B. (1993). An exploration of the job analysis–content validity process (pp. 2–34). In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 3–34). San Francisco, CA: Jossey-Bass.

Grove, W. M., & Meehl, P. E. (1994). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction. *Psychological Assessment, 12*, 19–30.

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions*. Taylor & Francis.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel psychology*, *57*(3), 639-683.

Hazer, J. T., & Highhouse, S. (1997). Factors influencing managers' reactions to utility analysis: Effects of SDy method, information frame, and focal intervention. *Journal of Applied Psychology, 82*(1), 104–112. https://doi.org/10.1037/0021-9010.82.1.104

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

Highhouse, S., Brooks, M. E., Nesnidol, S., & Sim, S. (2017). Is a .51 validity coefficient good? Value sensitivity for interview validity. *International Journal of Selection and Assessment, 25*(4), 383–389. https://doi.org/10.1111/ijsa.12192

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO_7 scale. *Journal of Personality and Social Psychology, 109*(6), 1003–1028. https://doi.org/10.1037/pspi0000033

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*(5), 1270–1285. https://doi.org/10.1037/0021-9010.92.5.1270

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1-2), 152-194.

Hu, J. C. (2020). Online GRE test heightens equity concerns. *Science, 368*(6498), 1414. https://doi.org/10.1126/science.368.6498.1414

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184–190. https://doi.org/10.1037/0021-9010.79.2.184

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment, 21*(3), 264-276.

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. https://doi.org/10.1111/ijsa.12078

Jackson, D. J. R., Dewberry, C., Gallagher, J., & Close, L. (2018). A comparative study of practitioner perceptions of selection methods in the United Kingdom. *Journal of Occupational and Organizational Psychology*, *91*(1), 33–56. https://doi.org/10.1111/joop.12187

Jones, J. W., & Cunningham, M. R. (2023). Going beyond a validity focus to accommodate megatrends in selection system design. *Industrial and Organizational Psychology, 16*(3), 336-340. doi:10.1017/iop.2023.28

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment.* New York: Little, Brown Spark.

König, C. J., Jöri, E., & Knüsel, P. (2011). The amazing diversity of thought: A qualitative study on how human resource practitioners perceive selection procedures. *Journal of Business and Psychology*, *26*, 437-452. https://doi.org/10.1007/s10869-010-9199-9

König, C. J., Klehe, U. C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment, 18*(1), 17-27. https://doi.org/10.1111/j.1468-2389.2010.00485.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. https://doi.org/10.1037/a0034156

Lacroux, A., & Martin-Lacroux, C. (2022). Should I Trust the Artificial Intelligence to Recruit? Recruiters' Perceptions and Behavior When Faced With Algorithm-Based Recommendation Systems During Resume Screening. *Frontiers in Psychology*, *13*, 1–13. https://doi.org/10.3389/fpsyg.2022.895997

Langer, M., & König, C. (2022). Explainability of artificial intelligence in human resources. In S. Strohmeier (Ed.), *Handbook of Research on Artificial Intelligence in Human Resource Management* (pp. 285-302). Cheltenham, UK: Edward Elgar Publishing.

Langhammer, K., Bernhard-Oettel, C., & Hellgren, J. (2012). Individual determinants of behavioral intentions: What tells us that practitioners really want to change hiring strategies? *International Journal of Selection and Assessment, 20*(4), 453–463.

Lee, S. D., Walmsley, P. T., Sackett, P. R., & Kuncel, N. (2021). A Method for Displaying Incremental Validity with Expectancy Charts. *Applied Measurement in Education*, *34*(4), 251–261. https://doi.org/10.1080/08957347.2021.1987902

Lewis, G., & Sorongon, J. (2022). *The Jobs with the Highest Turnover Rates, According to LinkedIn Data*. LinkedIn Talent Blog. Retrieved from https://www.linkedin.com/business/talent/blog/talent-analytics/types-of-jobs-with-most-turnover?trk=daily-blog-newsletter

Lievens, F., De Corte, W., & Brysse, K. (2003). Applicant perceptions of selection procedures: The role of selection information, belief in tests, and comparative anxiety. *International Journal of Selection and Assessment, 11*(1), 67-77. https://doi.org/10.1111/1468-2389.00227

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *25*(1), 29-46. https://doi.org/10.1002/job.246

Lievens, F., Highhouse, S., & Corte, W. (2005). The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment. *Journal of Occupational and Organizational Psychology*, *78*(3), 453–470. https://doi.org/10.1348/096317905X26093

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*(1), 43–66. https://doi.org/10.1037/apl0000160

Lodato, M. A., Highhouse, S., & Brooks, M. E. (2011). Predicting professional preferences for intuition-based hiring. *Journal of Managerial Psychology*, *26*(5), 352–365. https://doi.org/10.1108/02683941111138985

McNeish, D. M. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods, 19*(4), 552–563. https://doi.org/10.1037/met0000024

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114-140. https://doi.org/10.1037/met0000078

Melchers, K. G., Roulin, N., & Buehl, A. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment*, *28*(2), 123–142. https://doi.org/10.1111/ijsa.12280

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Morris, S. B., Daisley, R. L., Wheeler, M., & Boyer, P. (2015). A meta-analysis of the relationship between individual assessments and job performance. *Journal of Applied Psychology*, *100*(1), 5. https://doi.org/10.1037/a0036938

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647. https://doi.org/10.1037/0021-9010.75.6.640

Neumann, M., Niessen, A. S. M., Hurks, P. P. M., & Meijer, R. R. (2023a). Holistic and mechanical combination in psychological assessment: Why algorithms are underutilized and what is needed to increase their use. *International Journal of Selection and Assessment*, ijsa.12416. https://doi.org/10.1111/ijsa.12416

Neumann, M., Niessen, A. S. M., Linde, M., Tendeiro, J. N., & Meijer, R. R. (2023b). "Adding an egg" in algorithmic decision making: Improving stakeholder and user perceptions, and predictive validity by enhancing autonomy. *European Journal of Work and Organizational Psychology*, 1–18. https://doi.org/10.1080/1359432X.2023.2260540

Neumann, M., Niessen, A. S. M., & Meijer, R. R. (2021). Implementing evidence-based assessment and selection in organizations: A review and an agenda for future research. *Organizational Psychology Review*, *11*(3), 205–239. https://doi.org/10.1177/2041386620983419

Neumann, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2022). The autonomy-validity dilemma in mechanical prediction procedures: The quest for a compromise. *Journal of Behavioral Decision Making*, *35*(4), e2270. https://doi.org/10.1002/bdm.2270

Nolan, K. P., Carter, N. T., & Dalal, D. K. (2016). Threat of technological unemployment: Are hiring managers discounted for using standardized employee selection practices?. *Personnel Assessment and Decisions*, *2*(1), 30–47. https://doi.org/10.25035/pad.2016.004

Nolan, K. P., & Highhouse, S. (2014). Need for Autonomy and Resistance to Standardized Employee Selection Practices. *Human Performance*, *27*(4), 328–346. https://doi.org/10.1080/08959285.2014.929691

Ones, D. S., & Viswesvaran, C. (1999). Relative importance of personality dimensions for expatriate selection: A policy capturing study. *Human Performance*, *12*, 275–294.

Outtz, J. L., & Newman, D. A. (2010). A theory of adverse impact. In Outtz, J. L. (Ed.), *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection* (pp. 80-121). Routledge.

Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin*, *24*(3), 227-240.

Ployhart, R. E., & Ryan, A. M. (1997). Toward an explanation of applicant reactions: An examination of organizational justice and attribution frameworks. *Organizational behavior and human decision processes*, *72*(3), 308-335.

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the *Supreme Problem*: 100 years of selection and recruitment at the *Journal of Applied Psychology*. *Journal of Applied Psychology*, *102*(3), 291–304. https://doi.org/10.1037/apl0000081

Ployhart, R. E., & Weekley, J. A. (2017). Strategy, selection, and sustained competitive advantage. In *Handbook of employee selection* (pp. 115-133). Routledge.

Powell, D. M., & Goffin, R. D. (2009). Assessing personality in the employment interview: The impact of training on rater accuracy. *Human performance*, *22*(5), 450-465. https://doi.org/10.1080/08959280903248450

Pulakos, E. D. (2005). Selection assessment methods: A guide to implementing formal assessments to build a high-quality workforce. In *SHRM Foundation's Effective Practice Guidelines*. SHRM Foundation.

Reeder, M. C., Powers, C. L., Ryan, A. M., & Gibby, R. E. (2012). The Role of Person Characteristics in Perceptions of the Validity of Cognitive Ability Testing: The Role of Person Characteristics. *International Journal of Selection and Assessment*, *20*(1), 53–64. https://doi.org/10.1111/j.1468-2389.2012.00579.x

Risavy, S., Fisher, P., Robie, C., & Konig, C. (2019). Selection Tool Use: A Focus on Personality Testing in Canada, the United States, and Germany. *Personnel Assessment and Decisions, 5*(1). https://doi.org/10.25035/pad.2019.01.004

Risavy, S. D., Robie, C., Fisher, P. A., Komar, J., & Perossa, A. (2021). Selection tool use in Canadian tech companies: Assessing and explaining the research–practice gap. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 53*(4), 445–455. https://doi.org/10.1037/cbs0000263

Roberson, Q., Holmes IV, O., & Perry, J. L. (2017). Transforming research on diversity and firm performance: A dynamic capabilities perspective. *Academy of Management Annals, 11*(1), 189-216. https://doi.org/10.5465/annals.2014.0019

Rolland, F., & Steiner, D. D. (2007). Test-Taker Reactions to the Selection Process: Effects of Outcome Favorability, Explanations, and Voice on Fairness Perceptions. *Journal of Applied Social Psychology, 37*(12), 2800–2826. https://doi.org/10.1111/j.1559-1816.2007.00282.x

Rousseau, D. M., & Barends, E. G. (2011). Becoming an evidence-based HR practitioner. *Human Resource Management Journal, 21*(3), 221-235.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Rynes, S. L. (2012). 13 The Research-Practice Gap in I/O Psychology and Related Fields: Challenges and Potential Solutions. *The Oxford Handbook of Organizational Psychology, Volume 1, 1*, 409.

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR Professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*(2), 149–174. https://doi.org/10.1002/hrm.10029

Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The Very Separate Worlds of Academic and Practitioner Periodicals in Human Resource Management: Implications for Evidence-Based Management. *Academy of Management Journal, 50*(5), 987–1008. https://doi.org/10.5465/amj.2007.27151939

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

SIOP (2021). *HBO Max Documentary on Personality Tests Fails: SIOP Responds*. Society for Industrial and Organizational Psychology. Retrieved from https://www.siop.org/Portals/84/PDFs/SIOP%20Response%20to%20HBO%20Max%20Persona%20Film.pdf?ver=NJZrkm6P5SK4KM142FC6Hw%3d%3d

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*, 49-76. https://doi.org/10.1111/j.1744-6570.1993.tb00867.x

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*(2), 134–141. https://doi.org/10.1037/0021-9010.81.2.134

Steiner, D. D., & Gilliland, S. W. (2001). Procedural justice in personnel selection: International and cross-cultural perspectives. *International Journal of Selection And Assessment, 9*(1-2), 124–137. https://doi. org/10.1111/1468-2389.00169

Tenhiälä, A., Giluk, T. L., Kepes, S., Simón, C., Oh, I.-S., & Kim, S. (2016). The Research-Practice Gap in Human Resource Management: A Cross-Cultural Study. *Human Resource Management, 55*(2), 179–200. https://doi.org/10.1002/hrm.21656

Terpstra, D. (1996). Recruitment and selection: The search for effective methods. *HR Focus, 73*(5), 16-18.

Terpstra, D. E., & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*(4), 483-495. https://doi.org/10.1177/009102609702600405

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology, 26*, 1-9. https://doi.org/10.1007/s10869-010-9204-3

Topor, D. J., Colarelli, S. M., & Han, K. (2007). Influences of traits and assessment methods on human resource practitioners' evaluations of job applicants. *Journal of Business and Psychology, 21*(3), 361–376. https://doi.org/10.1007/s10869-006-9033-6

Van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology, 87*(1), 176–184. doi:10.1037/0021-9010.87.1.176.

Vasilopoulos, N. L., & Cucina, J. M. (2006). Faking on noncognitive measures: The interaction of cognitive ability and test characteristics. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 305−332). Information Age Publishing.

Viswesvaran, C., & Ones, D. S. (2018). Non-test methods and techniques used in employee selection. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology: Personnel psychology and employee performance* (pp. 451–473). Sage Reference.

Williams, J. E., Munick, M. L., Saiz, J. L., & FormyDuval, D. L. (1995). Psychological importance of the Big Five: Impression formation and context effects. *Personality and Social Psychology Bulletin, 21,* 818–826.

Woehr, D. J., & Arthur Jr, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258. https://doi.org/10.1016/S0149-2063(02)00216-7

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational behavior and human decision processes, 93*(1), 1-13. https://doi.org/10.1016/j.obhdp.2003.08.002

Yu, M., & Kuncel, N. (2020). Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments. *Personnel Assessment and Decisions, 6*(2). https://doi.org/10.25035/pad.2020.02.002

Yu, M., & Kuncel, N. R. (2022). Testing the value of expert insight: Comparing local versus general expert judgment models. *International Journal of Selection and Assessment*. https://doi.org/10.1111/ijsa.12356

Zhang, D. C., Highhouse, S., Brooks, M. E., & Zhang, Y. (2018). Communicating the validity of structured job interviews with graphical visual aids. *International Journal of Selection and Assessment,* ijsa.12220. https://doi.org/10.1111/ijsa.12220

Zibarras, L. D., & Woods, S. A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology, 83*(2), 499-511. https://doi.org/10.1348/096317909X42520

APPENDIX A: Tables

Table 1: Study 1 Selection Procedure Experimental Conditions

		Selection Procedure 1	Selection Procedure 2	Selection Procedure 3
Condition 1	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Cognitive ability	Conscientiousness	Job knowledge
Condition 2	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Cognitive ability	Job knowledge	Conscientiousness
Condition 3	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Conscientiousness	Cognitive ability	Job knowledge
Condition 4	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Conscientiousness	Job knowledge	Cognitive ability
Condition 5	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Job knowledge	Cognitive ability	Conscientiousness
Condition 6	**M**	Computerized assessment	Structured interview	Unstructured interview
	C	Job knowledge	Conscientiousness	Cognitive ability

Note. M = predictor method, C = predictor construct. Table shows experimental conditions, which contain all non-repeating predictor method/construct combinations. Within each condition, the order of procedures shown to participants was randomized.

Table 2: Study 1 Intraclass Correlations

Outcome variable	**Intraclass correlation (ρ)**
Utilization Intentions	.15
Predictiveness Beliefs	.21
Procedural Fairness Beliefs	.08
Outcome Fairness Beliefs	.34
Procedural Autonomy perceptions	.15
Evaluation Autonomy perceptions	.15
Transparency perceptions	<.01
Fakability perceptions	.02
Fidelity perceptions	.18
Evaluation Consistency perceptions	<.01
Job Relatedness perceptions	.26
Malleability perceptions	.19
Development Equity perceptions	.41

Note. Intraclass correlation values indicate the amount of variance in the outcome that is explained by the cluster variable (i.e., participant).

Table 3: Study 1 Descriptive Statistics and Bivariate Correlations

Variable	*M(SD)*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Utilization Intentions	4.67(1.03)	.94																	
2. Predictiveness	3.71(0.87)	.70*	.90																
3. Procedural Fairness	3.97(1.05)	.40*	.43*	.86															
4. Outcome Fairness	3.53(1.14)	.36*	.37*	.76*	.91														
5. Procedural Autonomy	3.38(1.12)	.30*	.34*	.02	.05	.94													
6. Evaluation Autonomy	3.33(1.17)	.26*	.31*	.00	.04	.88*	.95												
7. Transparency	3.95(1.12)	.42*	.40*	.64*	.44*	.05	.04	.95											
8. Eval. Consistency	3.79(1.22)	.30*	.30*	.73*	.55*	-.09*	-.09*	.74*	.95										
9. Fakability	2.96(1.31)	-.24*	-.26*	-.43*	-.41*	.07*	.11*	-.43*	-.44*	.93									
10. Fidelity	3.51(1.08)	.53*	.64*	.34*	.29*	.39*	.40*	.36*	.26*	-.17*	.95								
11. Job Relatedness	4.28(0.86)	.43*	.40*	.39*	.30*	.20*	.20*	.41*	.31*	-.15*	.52*	.93							
12. Development Equity	3.32(0.90)	.16*	.22*	.41*	.48*	.08*	.08*	.21*	.30*	-.20*	.22*	.19*	.86						
13. Malleability	3.87(1.05)	.10*	-.02	.03	.06	.00	-.01	.04	.03	-.07*	.08*	.20*	.07*	.91					
14. Intuition Hiring	3.46(0.95)	.02	.05	.04	-.03	.01	.03	.07*	.03	.05	.09*	.09*	-.02	-.11*	.90				
15. SDO	3.31(0.94)	.06	.00	-.09*	-.11*	.04	.05	.04	-.03	.02	.06	.07*	-.17*	.07*	.00	.92			
16. Hiring Exp. (Years)	8.69(7.63)	.08*	.09*	.07*	.03	-.01	-.01	.01	.02	-.05	.03	.06	.05	.00	.07*	-.06	-		
17. Exp. w/ Method	4.05(1.52)	.28*	.26*	.02	.05	.31*	.29*	.00	-.09*	.12*	.27*	.19*	.05	-.04	.10*	-.03	.14*	-	
18. Exp. w/ Construct	4.42(1.31)	.30*	.28*	.13*	.11*	.18*	.19*	.16*	.10*	-.04	.33*	.25*	.09*	.11*	.19*	-.04	.15*	.37*	-

Table 3 (cont'd)

Variable	*M(SD)*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
19. Method Norms	52.18 (30.80)	.30*	.27*	.05	.07*	.32*	.30*	.03	-.06	.09*	.28*	.21*	.06	.04	.02	-.03	.06	.64*	.24*	-
20. Construct Norms	62.50 (28.60)	.37*	.32*	.15*	.10*	.18*	.16*	.18*	.12*	-10*	.33*	.26*	.09*	.09*	.07*	-.01	.10*	.18*	.55*	.37*

Note. *$p < .05$, n = 270. Hiring experience indicated in years. All variables were measured on 5-point scales except for utilization intentions and experience with methods and constructs (6-point scales), method and construct norms (100-point scale indicating percent of the time), and hiring experience (years), with higher values indicating greater magnitude. Alphas reported on the diagonal.

Table 4: Study 1 Mean Ratings for Utilization Intentions and Weighting by Predictor Method and Construct Conditions

		Utilization Intentions		Weighting	
Method	**Construct**	***M***	***SD***	***M***	***SD***
By Predictor Methods (across all constructs)					
Structured interview		4.90_a	.94	21.6	9.5
Unstructured interview		4.63_b	1.01	19.6	10.5
Computerized assessment		4.49_b	1.10	17.7	9.9
By Predictor Constructs (across all methods)					
	Job Knowledge	4.95_a	.96	23.4	10.9
	Cognitive Ability	4.68_b	1.02	19.3	9.4
	Conscientiousness	4.39_c	1.05	16.2	8.6
By Predictor Method/Construct combinations					
Structured interview	Job knowledge	5.15	.86	25.0	9.4
Resume	Work experience	5.08	.81	26.5	12.1
Structured interview	Cognitive ability	4.91	.98	20.9	10.0
Unstructured interview	Job knowledge	4.86	.96	23.0	12.1
Computerized assessment	Job knowledge	4.82	1.03	22.1	10.6
Unstructured interview	Cognitive ability	4.69	.99	19.5	8.3
Structured interview	Conscientiousness	4.65	.92	18.9	8.0
Computerized assessment	Cognitive ability	4.47	1.05	17.5	9.4
Unstructured interview	Conscientiousness	4.30	1.02	15.7	9.2
Computerized assessment	Conscientiousness	4.20	1.16	13.8	7.9
Resume	Education	4.08	1.10	14.7	7.6

Note. Utilization intentions were measured on a 6-point scale with higher values indicating greater utilization. Subscripts indicate if mean differences in utilization intentions were significant in MANOVA tests as a function of method and construct conditions. Weight values represent weights hiring managers would attach to results from each predictor out of five choices available to them (i.e., three selection procedures and resume-based experience and education) such that the five weights must sum to 100.

Table 5: Study 1 Mean Ratings for Predictiveness Beliefs and Fairness Beliefs by Predictor Method and Construct Conditions

		Predictiveness Beliefs		Outcome Fairness Beliefs		Procedural Fairness Beliefs	
Method	**Construct**	***M***	***SD***	***M***	***SD***	***M***	***SD***
By Predictor Methods (across all constructs)							
Structured interview		3.88_a	.80	3.71_a	1.08	4.12_a	.86
Unstructured interview		3.68_b	.91	3.20_b	1.15	3.31_b	1.02
Computerized assessment		3.57_b	.90	3.69_a	1.12	4.15_a	.86
By Predictor Constructs (across all methods)							
	Job Knowledge	3.91_a	.88	3.65_a	1.13	3.97_a	.97
	Cognitive Ability	3.73_b	.83	3.50_a	1.13	3.90_a	1.00
	Conscientiousness	3.50_c	.88	3.45_a	1.15	3.71_b	1.08
By Predictor Method/Construct combinations							
Structured interview	Job knowledge	4.08	.76	3.84	1.08	4.20	.87
Structured interview	Cognitive ability	3.87	.84	3.61	1.15	4.16	.86
Unstructured interview	Job knowledge	3.86	.93	3.29	1.14	3.48	1.02
Computerized assessment	Job knowledge	3.79	.91	3.87	1.08	4.31	.78
Unstructured interview	Cognitive ability	3.74	.82	3.32	1.17	3.37	1.11
Structured interview	Conscientiousness	3.71	.77	3.68	1.00	3.99	.83
Computerized assessment	Cognitive ability	3.59	.82	3.57	1.06	4.13	.81
Unstructured interview	Conscientiousness	3.42	.93	2.98	1.13	3.06	1.15
Computerized assessment	Conscientiousness	3.36	.92	3.64	1.20	4.02	.97

Note. Predictiveness beliefs, outcome fairness beliefs, and procedural fairness beliefs were all measured on five-point scales. Subscripts indicate if mean differences in outcomes were significant in MANOVA tests as a function of method and construct conditions.

Table 6: Study 1 Mean Ratings for User Reactions Dimensions by Predictor Method and Construct Conditions

		By Predictor Methods (across all constructs)		
Predictor Method **User Reaction Dimension**		**Structured Interview**	**Unstructured Interview**	**Computerized Assessment**
Procedural Autonomy	M	3.32_b	3.84_a	2.99_c
	SD	1.06	.94	1.17
Evaluation Autonomy	M	3.26_b	3.82_a	2.90_c
	SD	1.10	.97	1.25
Transparency	M	4.24_a	3.26_b	4.24_a
	SD	.85	1.25	1.00
Fakability	M	2.91_b	3.36_a	2.60_c
	SD	1.25	1.22	1.35
Fidelity	M	3.68_a	3.53_a	3.33_b
	SD	1.00	1.05	1.04
Evaluation Consistency	M	4.08_b	2.94_c	4.35_a
	SD	.94	1.32	.86
		By Predictor Constructs (across all methods)		
Predictor Construct **User Reaction Dimension**		**Job Knowledge**	**Cognitive Ability**	**Conscientiousness**
Job Relatedness	M	4.55_a	3.45_b	3.68_b
	SD	.77	.86	.88
Malleability	M	4.48_a	3.45_c	3.68_b
	SD	.75	1.07	1.01
Development Equity	M	3.35_a	3.15_b	3.47_a
	SD	.95	.89	.82

Note. All variables were measured on five-point scales. Subscripts indicate if mean differences in outcomes were significant in MANOVA tests as a function of method and construct conditions.

Table 7: Study 1 Descriptive Statistics for Hiring Manager Experience with Predictors and Norms

		Prior Experience Using		Norms	
Method	**Construct**	***M***	***SD***	***M***	***SD***
By Predictor Methods (across all constructs)					
Structured interview		4.69	1.15	67.79	26.00
Unstructured interview		4.32	1.34	54.48	28.27
Computerized assessment		3.14	1.57	34.16	28.23
By Predictor Constructs (across all methods)					
	Job Knowledge	5.02	.96	71.88	24.82
	Cognitive Ability	4.13	1.38	60.16	28.87
	Conscientiousness	4.12	1.34	55.43	29.35

Note. Prior experience was rated on a 6-point scale (higher values indicating more experience) and norms were rated on a 100-point scale indicating percent of the time.

Table 8: Study 1 Path Model Predicting Utilization Intentions: Model with Outcome Fairness Beliefs

		Total Effect	**"a" Path**	**Indirect Effect**	**Direct Effect**
Predictiveness Beliefs	*β*	.531*			
	SE	.038			
Outcome Fairness Beliefs	*β*	.088*			
	SE	.034			
		Via Predictiveness Beliefs			
Procedural Autonomy	*β*	.240*	.176*	.094*	.146*
	SE	.047	.052	.028	.044
Evaluation Autonomy	*β*	-.106*	-.039	-.021	-.085
	SE	.048	.053	.028	.044
Transparency	*β*	.196*	.139*	.074*	.122*
	SE	.049	.036	.020	.042
Fakability	*β*	-.088*	-.125*	-.066*	-.022
	SE	.035	.034	.019	.030
Fidelity	*β*	.324*	.482*	.256*	.068
	SE	.040	.040	.029	.036
Job Relatedness	*β*	.136*	.067	.036	.100*
	SE	.041	.036	.019	.033
Malleability	*β*	.029	-.083*	-.044*	.073*
	SE	.028	.028	.015	.024
		Via Outcome Fairness Beliefs			
Fakability	*β*	-.088*	-.155*	-.014*	-.074*
	SE	.035	.037	.006	.030
Evaluation Consistency	*β*	.010	.332*	.029*	-.019
	SE	.038	.037	.012	.041
Job Relatedness	*β*	.136*	.050	.004	.132*
	SE	.041	.030	.003	.033
Development Equity	*β*	-.030	.307*	.027*	-.057*
	SE	.026	.033	.011	.027
Predictiveness Beliefs	*β*	.544*	.147*	.013*	.531*
	SE	.037	.036	.006	.038

Note. *$p < .05$. Standardized estimates are reported. Total effect and direct effect values indicate total effects and direct effects on utilization intentions. "A" path values indicate parameter estimates between user reactions dimensions and predictiveness or fairness beliefs, as indicated.

Table 9: Study 1 Path Model Predicting Utilization Intentions: Model with Procedural Fairness Beliefs

		Total Effect	**"a" Path**	**Indirect Effect**	**Direct Effect**
Predictiveness Beliefs	*β*	.534*			
	SE	.039			
Procedural Fairness Beliefs	*β*	.070			
	SE	.040			
Via Predictiveness Beliefs					
Procedural Autonomy	*β*	.240*	.176*	.094*	.146*
	SE	.047	.052	.028	.044
Evaluation Autonomy	*β*	-.106*	-.039	-.021	-.085
	SE	.048	.053	.028	.044
Transparency	*β*	.196*	.139*	.074*	.122*
	SE	.049	.036	.020	.042
Fakability	*β*	-.088*	-.125*	-.066*	-.022
	SE	.035	.034	.019	.030
Fidelity	*β*	.324*	.482*	.256*	.068
	SE	.040	.040	.029	.036
Job Relatedness	*β*	.136*	.067	.036	.100*
	SE	.041	.036	.019	.033
Malleability	*β*	.029	-.083*	-.044*	.073*
	SE	.028	.028	.015	.024
Via Procedural Fairness Beliefs					
Fakability	*β*	.090*	-.094*	-.007	-.083*
	SE	.035	.025	.004	.030
Evaluation Consistency	*β*	.019	.561*	.039	-.020
	SE	.038	.029	.023	.042
Job Relatedness	*β*	.137*	.098*	.007	.130*
	SE	.041	.025	.004	.032
Development Equity	*β*	-.031	.171*	.012	-.043
	SE	.026	.026	.007	.027
Predictiveness Beliefs	*β*	.545*	.163*	.011	.534*
	SE	.037	.025	.007	.039

Note. *$p < .05$. Standardized estimates are reported. Total effect and direct effect values indicate total effects and direct effects on utilization intentions. "A" path values indicate parameter estimates between user reactions dimensions and predictiveness or fairness beliefs, as indicated.

Table 10: Study 2 Descriptive Statistics and Bivariate Correlations

Variable	*M(SD)*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Procedure Condition	1.50(.50)	-												
2. Process Autonomy Condition	1.50(.50)	-.01	-											
3. Scoring Autonomy Condition	1.44(.50)	-.04	.01	-										
4. Utilization Intentions	4.11(.97)	.13*	.10	.14*	.90									
5. Predictiveness Beliefs	3.10(.93)	.00	.12*	.10	.61*	.91								
6. Procedural Autonomy Perceptions	3.23(1.17)	-.22*	.52*	.31*	.37*	.42*	.96							
7. Evaluation Autonomy Perceptions	3.13(1.23)	-.08	.35*	.51*	.40*	.40*	.65*	.97						
8. Procedural Fairness Beliefs	3.59(.95)	-.01	.10	.02	.54*	.48*	.23*	.24*	.81					
9. Outcome Fairness Beliefs	3.28(1.07)	-.04	.15*	.07	.44*	.42*	.27*	.24*	.79*	.91				
10. Transparency Perceptions	3.54(1.12)	.10	.09	.20*	.45*	.45*	.32*	.41*	.34*	.24*	.90			
11. Evaluation Consistency Perceptions	3.82(1.04)	-.05	-.03	-.14*	.24*	.25*	.02	.02	.57*	.41*	.31*	.90		
12. HM Intuition Hiring Preference	3.41(.90)	.04	.00	-.01	-.04	-.10*	-.05	-.02	.03	-.04	.04	.05	.89	
13. HM Hiring Experience (Years)	7.89(6.88)	-.09	.01	-.08	-.05	-.04	-.03	-.10	-.01	-.02	-.05	.08	.05	.83
14. HM Need for Autonomy	4.21(.56)	.01	.05	.02	.10	.06	.19*	.13*	.13*	.11*	.20*	.10	.05	.07

Note. $*p < .05$, n = 368. Procedure condition 1 = assessment, 2 = interview. Process autonomy and scoring autonomy conditions 1 = low, 2 = high. Hiring experience indicated in years. All other variables were measured on 5-point scales except for utilization intentions and need for autonomy (6-point scales) with higher values indicating greater magnitude. Alphas reported on the diagonal.

APPENDIX B: Figures

Figure 1: Hypothesized Conceptual Model

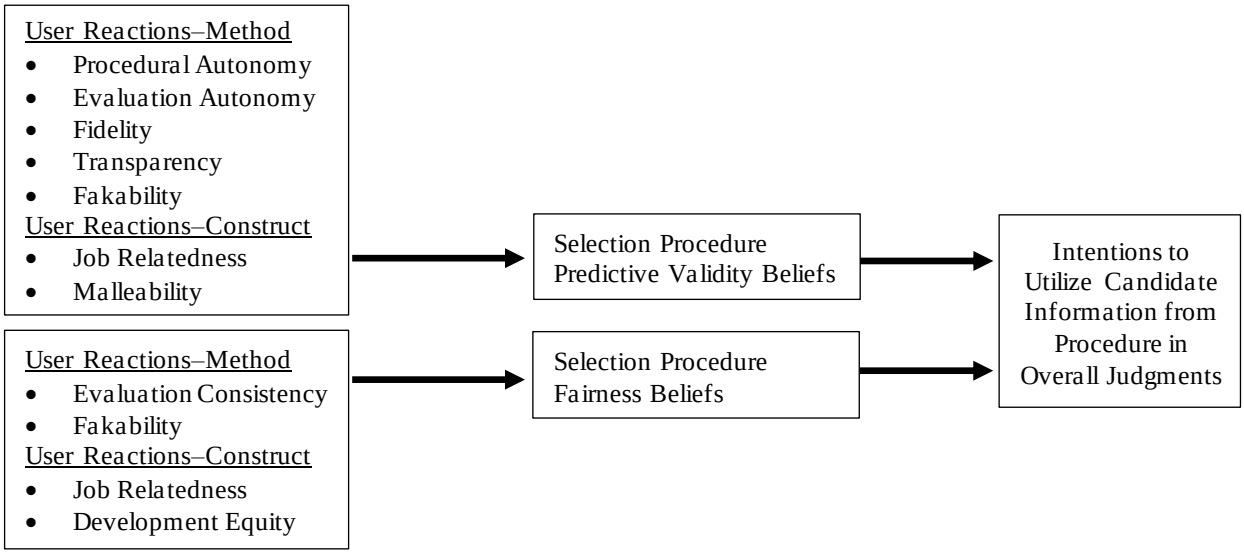
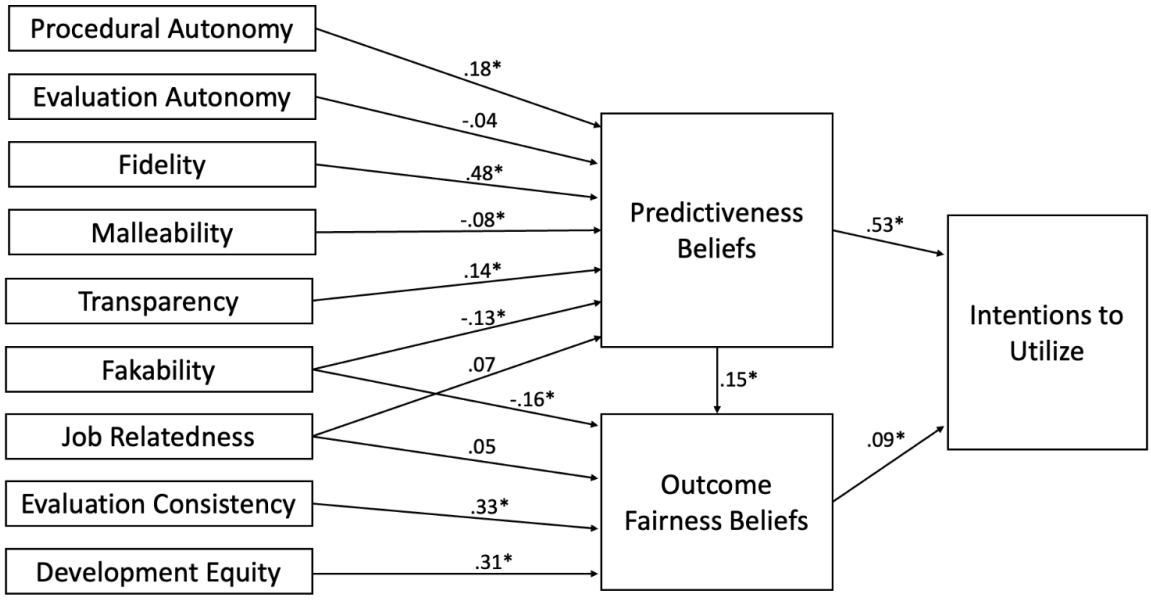
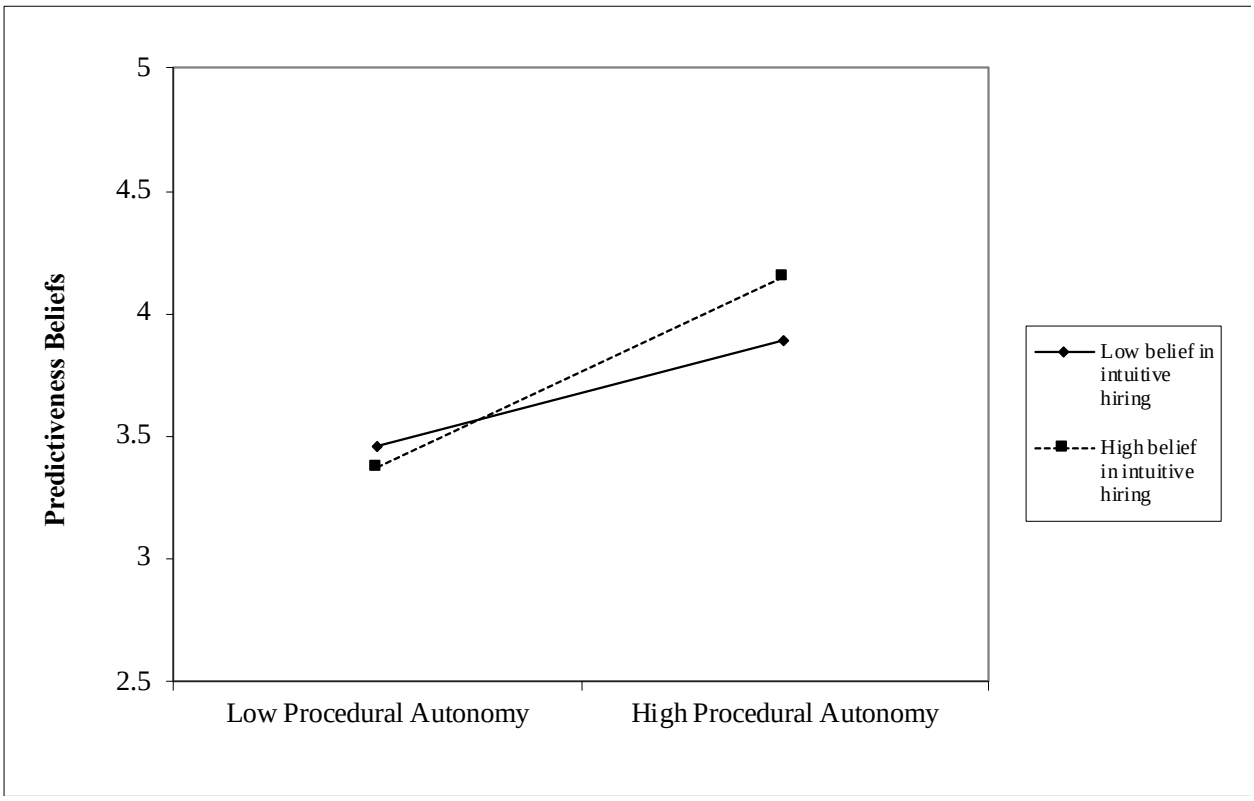


Figure 2: Conceptual Model Results from Path Analysis



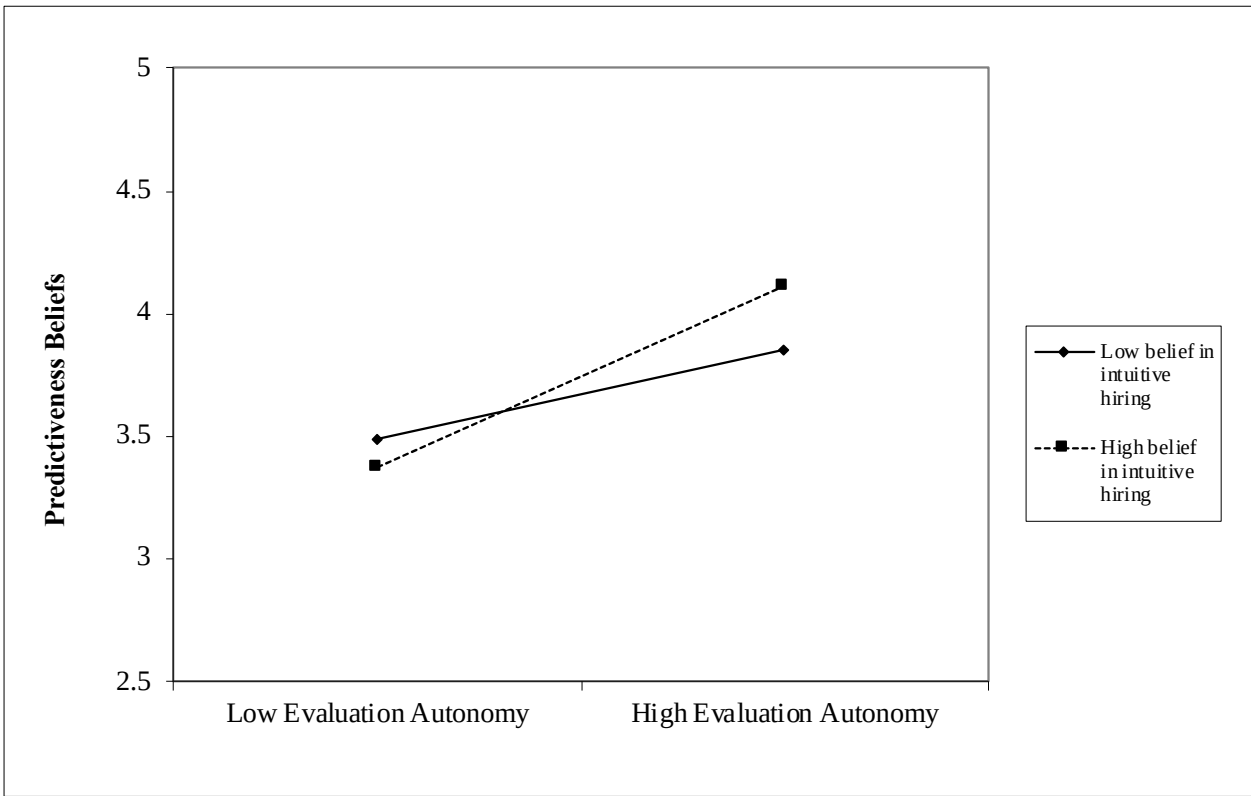
Note. *$p < .05$. Standardized coefficients shown in figure. Table 8 displays full path model results.

Figure 3: Study 1 Interaction Between Procedural Autonomy Perceptions and Intuition Hiring Tendencies Predicting Predictiveness Beliefs



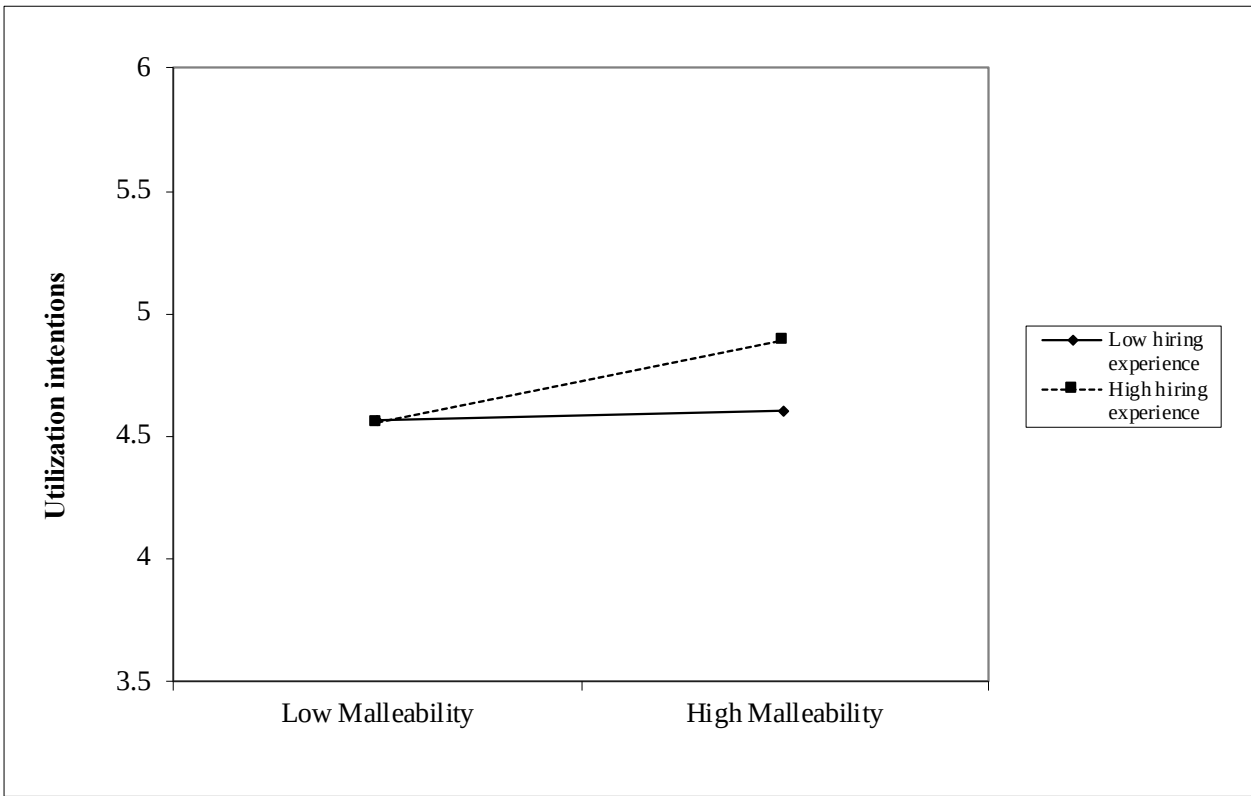
Note. Low and high levels depict one standard deviation below and above means.

Figure 4: Study 1 Interaction Between Evaluation Autonomy Perceptions and Intuition Hiring Tendencies Predicting Predictiveness Beliefs



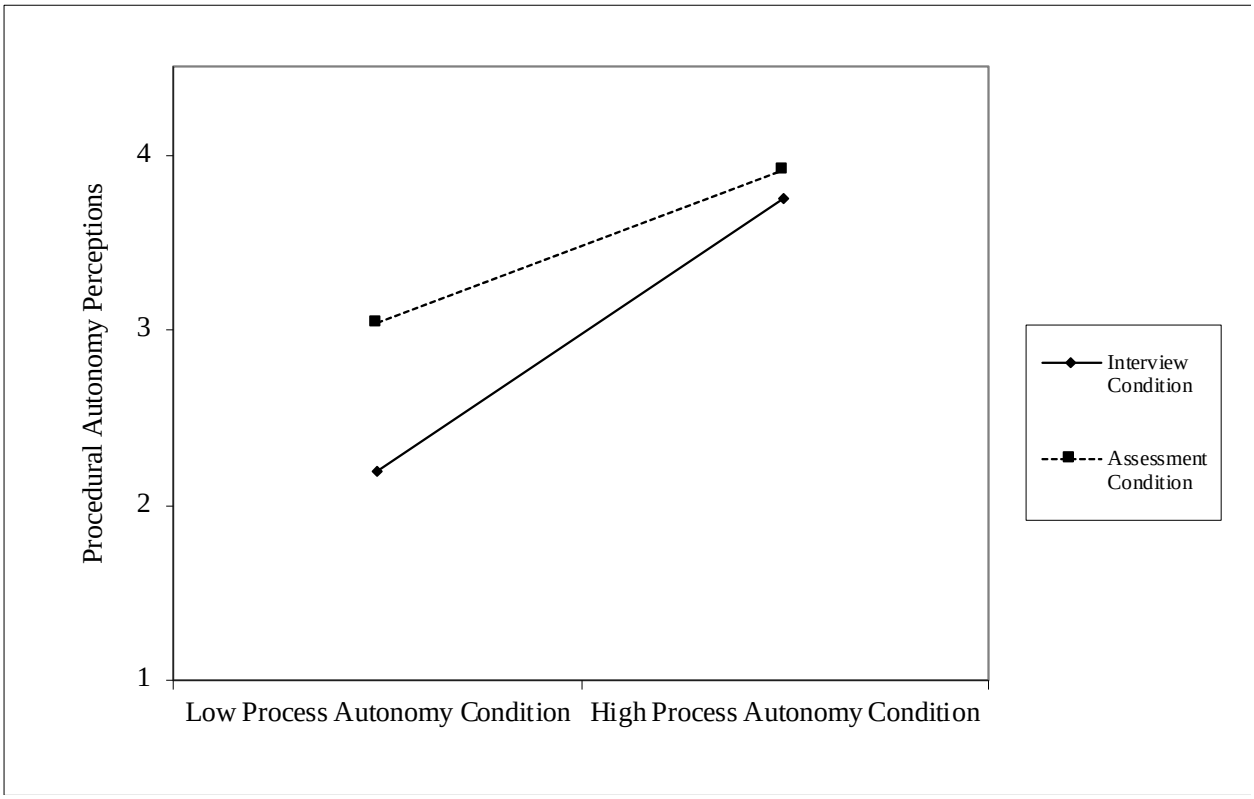
Note. Low and high levels depict one standard deviation below and above means.

Figure 5: Study 1 Interaction Between Malleability Perceptions and Hiring Experience Predicting Utilization Intentions



Note. Low and high levels depict one standard deviation below and above means.

Figure 6: Study 2 Interaction Between Process Autonomy Condition and Procedure Condition on Procedural Autonomy Perceptions



APPENDIX C: Study 1 Vignette

Imagine that you are in the following scenario:

You are an employee at Acme Corporation. Acme is hiring a human resources (HR) generalist, and you have been asked to serve as a member of the hiring committee. As a hiring committee member, you will evaluate candidates to help decide which candidate is most likely to succeed in this role.

Below is a summary of the responsibilities for the HR generalist role:

The HR generalist will be responsible for performing a variety of HR functions, including duties related to:

- recruitment and staffing
- compensation and benefits
- employee onboarding and training
- ensuring employment laws and regulations are upheld

Recruiters have already screened applicants for minimum qualification requirements. Based on this initial screening, the hiring committee has been given a shortlist of finalist candidates. To determine the best candidate, each hiring committee member (including you) will independently provide overall rating evaluations of each finalist.

Here is some additional information about the situation.

- Acme requires that all finalist candidates participate in three selection procedures.
- Selection procedures are tools that collect information about candidates to help hiring decision-makers predict their likelihood of success as employees.
- Acme uses three (3) selection procedures for this role: a structured interview, an unstructured interview, and a computerized assessment. Details about these selection procedures will be provided.
- When evaluating candidates, you will have access to information about candidates' results on each of the three selection procedures (candidate scores from two interviews and one assessment).
- You will also have access to information about each candidate's education and prior experience.
- IMPORTANT: You can use any of this information as you see fit to help make your candidate evaluation ratings and predict their likelihood of success.

The next pages describe the three selection procedures (two interviews and one assessment) used for this role. Carefully read the brief descriptions about each selection procedure. Afterwards, you will be asked to provide your opinions about each procedure as a hiring committee member. It is recommended to read each description at least twice.

APPENDIX D: Study 1 Selection Procedure Descriptions

Online assessment – cognitive ability

- This computerized test automatically scores candidate responses to multiple-choice questions.
- The test is designed to measure candidates' level of cognitive ability. Cognitive ability is a general mental ability to reason, solve problems, plan, think, comprehend complex ideas, and learn from experience.

Online assessment – conscientiousness

- This computerized test automatically scores candidate responses to multiple-choice questions.
- The test is designed to measure candidates' level of conscientiousness. Conscientiousness is a personality trait representing a tendency to think and behave in a way that is organized, planned, persistent, deliberate, and disciplined.

Online assessment – job knowledge

- This computerized test automatically scores candidate responses to multiple-choice questions.
- The test is designed to measure candidates' level of job knowledge, such as knowledge of labor laws and best practices related to recruitment and staffing, compensation and benefits, and onboarding and training.

Structured interview – cognitive ability

- This interview asks all candidates the same questions for consistency. Candidate responses to each question are evaluated using standardized rating scales.
- The interview's objective is to measure candidates' level of cognitive ability. Cognitive ability is a general mental ability to reason, solve problems, plan, think, comprehend complex ideas, and learn from experience.

Structured interview – conscientiousness

- This interview asks all candidates the same questions for consistency. Candidate responses to each question are evaluated using standardized rating scales.
- The interview's objective is to measure candidates' level of conscientiousness. Conscientiousness is a personality trait representing a tendency to think and behave in a way that is organized, planned, persistent, deliberate, and disciplined.

Structured interview – job knowledge

- This interview asks all candidates the same questions for consistency. Candidate responses to each question are evaluated using standardized rating scales.
- The interview's objective is to measure candidates' level of job knowledge, such as knowledge of labor laws and best practices related to recruitment and staffing, compensation and benefits, and onboarding and training.

Unstructured interview – cognitive ability

- This interview asks candidates individualized questions chosen by the interviewer. Candidates are given one overall interview rating at the end.
- The interview's objective is to measure candidates' level of cognitive ability. Cognitive ability is a general mental ability to reason, solve problems, plan, think, comprehend complex ideas, and learn from experience.

Unstructured interview – conscientiousness

- This interview asks candidates individualized questions chosen by the interviewer. Candidates are given one overall interview rating at the end.
- The interview's objective is to measure candidates' level of conscientiousness. Conscientiousness is a personality trait representing a tendency to think and behave in a way that is organized, planned, persistent, deliberate, and disciplined.

Unstructured interview – job knowledge

- This interview asks candidates individualized questions chosen by the interviewer. Candidates are given one overall interview rating at the end.
- The interview's objective is to measure candidates' level of job knowledge, such as knowledge of labor laws and best practices related to recruitment and staffing, compensation and benefits, and onboarding and training.

APPENDIX E: Study 2 Vignette

Imagine that you are in the following fictional scenario:

You are a hiring manager at a fictional company called Acme Organization. In the final stage of hiring processes at Acme, all finalist candidates are required to participate in an [assessment/interview] to help predict their likelihood of success. The [assessment/interview] is detailed on the next page.

When making hiring decisions, hiring managers have access to candidates' results on this [assessment/interview] as well as resume-based information such as their education and prior experience. Hiring managers can use any of this information as they see fit to compare and evaluate candidates.

The [assessment/interview] is designed to measure candidates' personality fit with the Acme 8 Behaviors, a model of organizational culture that encourages high performance. The next section provides more details about how the [assessment/interview] works.

Carefully read the description of the [assessment/interview]. Afterwards, you will be asked your opinions about it, including the extent to which you (as a hiring manager) would rely on results from this [assessment/interview] when evaluating candidates.

APPENDIX F: Study 2 Selection Procedure Descriptions

Online assessment conditions intro

- This is a computerized assessment designed to measure the fit between candidates' personality traits, or behavioral tendencies, with the Acme Organization Behaviors. It assesses candidates against four of the eight Acme Organization Behaviors.

Online assessment – low procedural autonomy – low evaluation autonomy

- The four Acme Behaviors measured by the assessment were decided by HR based on internal research.
- The assessment automatically scores candidate responses to multiple-choice questions. An overall score of fit with Acme Organization is produced by the assessment's scoring system.

Online assessment – low procedural autonomy – high evaluation autonomy

- The four Acme Behaviors measured by the assessment were decided by HR based on internal research.
- The assessment automatically scores candidate responses to multiple-choice questions. The assessment's scoring system produces scores for each of the four Acme Behaviors; based on these, you will determine an overall score of fit with Acme Organization using your judgment.

Online assessment – high procedural autonomy – low evaluation autonomy

- As the hiring manager, you get to choose the four Acme Behaviors that will be measured by the assessment for this hiring process.
- The assessment automatically scores candidate responses to multiple-choice questions. An overall score of fit with Acme Organization is produced by the assessment's scoring system.

Online assessment – high procedural autonomy – high evaluation autonomy

- As the hiring manager, you get to choose the four Acme Behaviors that will be measured by the assessment for this hiring process.
- The assessment automatically scores candidate responses to multiple-choice questions. The assessment's scoring system produces scores for each of the four Acme Behaviors; based on these, you will determine an overall score of fit with Acme Organization using your judgment.

Structured interview conditions intro

- This is an interview designed to measure the fit between candidates' personality traits, or behavioral tendencies, with the Acme Organization Behaviors. To be consistent, this interview will ask all candidates for the same role the same eight questions.

Structured interview – low procedural autonomy – low evaluation autonomy

- You are required to ask one specific question about each of the eight Acme Behaviors; the interview cannot deviate from this list of standard questions.
- You will rate candidate responses to each question and input question rating scores into an interview guide spreadsheet. Based on these scores, the spreadsheet's formula will calculate an overall score of fit with Acme Organization.

Structured interview – low procedural autonomy – high evaluation autonomy

- You are required to ask one specific question about each of the eight Acme Behaviors; the interview cannot deviate from this list of standard questions.
- You will rate candidate responses to each question and input question rating scores into an interview guide spreadsheet. Based on these scores, you will calculate an overall score of fit with Acme Organization using your judgment.

Structured interview – high procedural autonomy – low evaluation autonomy

- As the hiring manager, you get to choose the eight questions. You will select one question from a list of five options to ask about each of the eight Acme Behaviors.
- You will rate candidate responses to each question and input question rating scores into an interview guide spreadsheet. Based on these scores, the spreadsheet's formula will calculate an overall score of fit with Acme Organization.

Structured interview – high procedural autonomy – high evaluation autonomy

- As the hiring manager, you get to choose the eight questions. You will select one question from a list of five options to ask about each of the eight Acme Behaviors.
- You will rate candidate responses to each question and input question rating scores into an interview guide spreadsheet. Based on these scores, you will calculate an overall score of fit with Acme Organization using your judgment.