Retaining the Hiring Professional in Employee Selection through a Mechanical Synthesis Approach

Franziska Helena Linn

Student number: 2706154

Supervisor: Dr. Marvin Neumann

Second Assessor: Dr. Wendy Andrews

Master's Thesis - M.Sc. Work & Organizational Psychology

July, 2023

Vrije Universiteit Amsterdam

Abstract

Hiring managers combine different pieces of information to predict applicants' future job performance. Existing research shows that combining information mechanically (i.e., by an algorithm) yields more valid predictions than combining information holistically (i.e., in the mind of the decision maker). Nevertheless, mechanical decision-making is underutilized in practice, potentially because it limits the fulfillment of decision-makers' autonomy and competence needs. In this study, mechanical synthesis (MS) is proposed as a solution to bridge the gap between predictive validity and subjective needs fulfillment. In MS, holistic judgement is integrated into an algorithm to generate final predictions, aiming to balance predictive validity and subjective needs fulfillment. In a pre-registered between-subjects experiment ($N = 209$), holistic, algorithmic, and MS decision-making approaches were compared regarding their predictive validity and participants' use intentions, perceived autonomy, and perceived competence. MS approaches showed significantly higher predictive validity than holistic approaches. Furthermore, MS approaches significantly outperformed algorithmic approaches but did not differ significantly from holistic approaches regarding perceived autonomy. For use intentions and perceived competence no significant differences were found. The findings imply that MS improves predictive validity compared to holistic prediction without reducing decision-makers' autonomy.

Keywords: mechanical synthesis, holistic prediction, algorithmic prediction, autonomy, competence, self-determination-theory, employee selection

Retaining the Hiring Professional in Employee Selection through a Mechanical Synthesis Approach

Making effective hiring decisions is highly relevant for organizations as human resources are acknowledged as a crucial driver of organizational performance (Liu et al., 2007). Thus, a primary aim of hiring professionals is to identify and select those applicants who will show the highest job performance in the future. To make such performance predictions, information is collected through various procedures such as tests and interviews, and hiring professionals then combine this information to make a hiring decision (Neumann, Niessen, Linde, et al., 2023). Research shows that most hiring decisions are made based on holistic judgement (i.e., applicant information is combined in the decision maker's mind, often intuitively; also referred to as intuitive, or clinical judgement) rather than based on mechanical judgement (i.e., combining information by applying a pre-specified algorithm; also referred to as algorithmic, statistical, or actuarial judgement, Highhouse, 2008; Meijer et al., 2020; Neumann, Niessen, Linde, et al., 2023; Neumann et al., 2021, 2023; Nolan, 2012; Ryan et al., 2015). This is problematic because previous findings show that mechanical judgement significantly outperforms holistic judgement in terms of predictive validity, indicating that performance predictions are less accurate when hiring professionals rely on their intuition and expertise rather than on a pre-specified algorithm (Dawes et al., 1989; Grove et al., 2000; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meehl, 1954; Sawyer, 1966). According to Kahneman et al. (2016; 2021) this is due to the fact that holistic decisions are noisy; individuals do not apply the same predictor weights across different applicants.

Given this shortcoming of holistic judgement, the question arises why hiring professionals nevertheless continue to make decisions holistically rather than mechanically. According to existing literature, one reason for decision makers' reluctance to apply algorithms is that algorithms restrict the fulfillment of subjective needs (Dietvorst et al., 2018; Neumann et al., 2021; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan & Highhouse, 2014). More specifically, based on self-determination-theory (Deci & Ryan, 2000), scholars argue that hiring professionals perceive their sense of autonomy and competence to be threatened by algorithms, resulting in low intentions to use them in future decision processes (Neumann et al., 2021; Nolan, 2012; Nolan & Highhouse, 2014). Conversely, hiring professionals may be more likely to use mechanical decision-making processes when their autonomy and competence is retained. Thus, designing decision-making procedures that combine mechanical and holistic judgements could be a suitable compromise in terms of both predictive validity and practitioners' intentions to utilize algorithms.

One such approach that has so far received little attention is the concept of mechanical synthesis (MS) as defined by Sawyer (1966). MS entails a sequential approach in which decision makers first make a holistic prediction, which is then integrated as a variable into an algorithm to predict a desired outcome variable such as job performance. To configure the final algorithm, weights are attached to each criterion predictor, including the holistic prediction of the decision maker. The weights for the predictors other than the holistic prediction are often chosen based on meta-analytic findings regarding the validity of each predictor for the outcome criterion (Kuncel, 2018). For the holistic prediction, either a fixed weight can be attached (which we call "fixed MS") or practitioners can choose this weight themselves (i.e., "self-designed MS"). In any case, the final algorithm performs a consistently weighted combination of all data presented to the decision maker, as well as the prediction of the decision maker. Thus, mechanical synthesis integrates mechanical and holistic judgement whilst reducing noise in decision-making (Sawyer, 1966). Based on the hybrid nature of MS, it can be expected that this method bridges the gap between retaining the hiring professional and achieving sufficiently valid results compared to purely holistic or purely mechanical selection methods.

To empirically test this expectation, this study aims to compare holistic, algorithmic, and MS approaches of employee selection with regards to their predictive validity, practitioners' use intentions, and practitioners' perceived levels of autonomy and competence. In a between-subjects experiment, participants apply either a holistic, algorithmic, or MS approach to predict the future job performance of applicants based on real applicant data as provided by Kausel et al. (2016). By including fixed and self-designed MS approaches, this study responds to previous calls for investigations of alternative decision-making approaches that have the potential to achieve high levels of both predictive validity and participants' intentions to use this approach (Kuncel et al., 2013; Meijer et al., 2020; Neumann et al., 2021; Nolan, 2012; Nolan & Highhouse, 2014). By doing so, we hope to contribute to solving the problem which has been framed as one of the greatest challenges of organizational psychology: convincing organizations to apply empirically-based selection practices that have been designed to improve the quality of hiring decisions (Highhouse, 2008; Nolan & Highhouse, 2014).

Theoretical Framework

Noise in Decision-Making

Decision-making has received considerable scholarly attention as our daily decisions drive our actions and thus have impact on outcomes in both private and professional domains. Specifically in the domain of employee selection, the decisions of hiring professionals have direct impact on both applicants (i.e., determining whether they get the job or not) and the organization (i.e., improving organizational performance by selecting the best-fitting applicant, Yu & Kuncel, 2022). Thus, it is highly relevant to ensure that hiring professionals make valid selection decisions.

Previous literature on decision science emphasizes the need to distinguish between data collection and data combination as two different steps of decision-making (Meehl, 1954; Nolan, 2012; Nolan & Highhouse, 2014; Sawyer, 1966). Whilst data collection is concerned with accumulating information (e.g., test results or interview scores from job applicants), data combination refers to the process of weighting and integrating this information to form a decision or make a prediction (e.g., regarding applicants' future job performance) (Highhouse & Brooks, 2023; Meehl, 1954; Neumann, Niessen, Linde, et al., 2023; Nolan, 2012; Sawyer, 1966). One of the most prominent findings in the recent literature on decision-making is that human decision-making is noisy (Kahneman et al., 2016; Kahneman et al., 2021). Kahneman et al. (2016) established the term noise, which is referred to as having unwanted variability in judgements of the same data. Noise, or inconsistency, can be problematic because it frequently results in costly judgement errors, such as selecting an applicant for a job whose performance is inferior to other applicants' performance, thereby not achieving the best possible organizational performance (Kahneman et al., 2016; Kahneman et al., 2021). Thus, it should be the goal of both practitioners and scientists to reduce noise in decision-making.

Inconsistency and Inaccuracy of Predictor Weights

Two main drivers have been identified that affect predictive validity of decision-making approaches, namely inaccuracy and inconsistency in weighting criterion predictors (Dawes et al., 1989; Karelaia & Hogarth, 2008; Kausel et al., 2016; Meehl, 1954; Neumann, Niessen, Tendeiro, & Meijer, 2022; Yu, 2018; Yu & Kuncel, 2020). Being inconsistent in attaching weights to predictors reduces comparability between applicants and thus complicates the process of identifying the objectively best-performing applicant (Neumann, Niessen, Tendeiro, & Meijer, 2022; Yu & Kuncel, 2020). Similarly, attaching empirically inaccurate weights to criterion predictors can negatively affect the predictive validity of decisions compared to empirically optimal weighting schemes, as decision makers might over- or underestimate the relevance of single predictors when combining data (Kausel et al., 2016; Neumann, Niessen, Tendeiro, & Meijer, 2022; Yu & Kuncel, 2020). However, research shows that even attaching random or equal positive weights to the predictors outperforms holistic judgement, if applied consistently (Dawes, 1979; Yu, 2018; Yu & Kuncel, 2020). As summarized by Yu and Kuncel (2020), "*mindless consistency* is enough to result in more accuracy than expert [i.e., holistic] judgement" (p. 7). Given that such findings should be less pronounced when predictors differ strongly in validity (Yu & Kuncel, 2020), they indicate that not accurate predictor weights alone but the combination of accurate and consistent weights drive predictive validity of decisions, thereby highlighting the important role of noise in decision-making. On a different note, these findings also show that, in contrast to what people frequently believe, algorithms do not have to be complex but can be very simple yet valid decision-making approaches (Dawes et al., 1989).

Overall, the implications of existing findings on noise are twofold. On the one hand, they suggest that noise is a serious problem threatening the accuracy of hiring decisions and thereby the outcomes for both applicants and organizations. On the other hand, these findings imply that, looking forward, predictive validity of decision can be improved by eliminating inaccuracy and, even more so, inconsistency of predictor weighting in decision-making approaches.

Holistic versus Mechanical Decision-Making Approaches

In decision science, scholars differentiate between holistic and mechanical decision-making approaches (Dawes et al., 1989; Grove et al., 2000; Highhouse & Brooks, 2023;

Kuncel et al., 2013; Meehl, 1954; Meijer et al., 2020; Neumann, Niessen, Linde, et al., 2023; Neumann et al., 2021; Nolan, 2012; Sawyer, 1966). Importantly, the differentiation between holistic and mechanical decision-making depends on the data combination approach, not the data collection approach that is deployed in the decision-making process (Meijer et al., 2020).

Decision-making is considered to be holistic when the decision maker combines information intuitively in their mind, by thinking about the data (Meehl, 1954; Sawyer, 1966). In contrast, decision-making is considered mechanical when a formal rule, such as an algorithm is applied to combine the data following a pre-determined pattern (Meehl, 1954; Sawyer, 1966).

Predictive Validity of Holistic versus Mechanical Approaches

Returning to the question of how to reduce noise in decision-making, Kahneman et al. (2016) state: "The most radical solution to a severe noise problem is to replace human judgment with algorithms" (p. 41). This statement is rooted in the empirically validated assumption that algorithms can be designed in a way that ensures both accuracy and consistency in applying predictor weights (Dawes et al., 1989; Highhouse & Brooks, 2023; Neumann, Niessen, Linde, et al., 2023; Yu & Kuncel, 2020).

The superiority of mechanical decision-making approaches over holistic approaches in terms of predictive validity of the outcome criterion is confirmed by numerous meta-analyses (Dawes et al., 1989; Grove et al., 2000; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meehl, 1954; Sawyer, 1966) and prominently discussed in virtually all studies published on the topic of holistic versus mechanical decision-making (e.g., Dietvorst et al., 2015, 2018; Highhouse & Brooks, 2023; Meehl, 1986; Meijer et al., 2020; Neumann, Niessen, Hurks, & Meijer, 2022; Neumann, Niessen, Linde, et al., 2023; Neumann et al., 2021, 2023; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan & Highhouse, 2014; Yu & Kuncel, 2020). Highhouse and Brooks (2023) summarize the findings of existing research as follows: “Algorithms are good. Algorithms reduce noise. Algorithms result in better judgements” (p. 529). It can thus be concluded that in existing research, the benefits of mechanical over holistic decision-making approaches in terms of predictive validity are communicated clearly and unambiguously.

Holistic and Mechanical Approaches in Practice

Regardless of the variety of studies confirming the benefits of mechanical decision-making approaches, holistic hiring approaches remain by far the dominant method applied by human resources professionals (Highhouse, 2008; Meijer et al., 2020; Neumann, Niessen, Linde, et al., 2023; Neumann et al., 2021; Ryan et al., 2015). This implies that organizations have unfulfilled potential with regard to effectively hiring the best applicant. In existing research, this phenomenon is called “the science-practice gap” (Neumann et al., 2021, p. 207).

Various factors are related to decision makers’ intentions (not) to use algorithms. These include overconfidence in holistic judgment (Dawes et al., 1989; Kahneman et al., 2016; Kahneman et al., 2021; Kausel et al., 2016; Meijer et al., 2020; Nolan, 2012; Yu & Kuncel, 2022), lack of education, or disbelief regarding scientific findings showing the validity of mechanical approaches (Neumann, Niessen, Hurks, & Meijer, 2022), worries that external stakeholders will evaluate one’s algorithm use negatively (Diab et al., 2011; Neumann, Niessen, Hurks, & Meijer, 2022; Neumann, Niessen, Linde, et al., 2023; Nolan et al., 2016), social incentivization not to apply algorithms (Burton et al., 2020), fear of being outperformed or replaced by algorithms (Meehl, 1954, 1986; Nolan et al., 2016), or the perception of algorithms as a threat to one’s self-concept as an expert judge (Meehl, 1986; Yu & Kuncel, 2022).

Although all of these theoretical argumentations seem plausible, there is a growing body of research indicating that the (lack of) subjective needs fulfilment resulting from the use of algorithms might be the key factor influencing decision makers' willingness to use mechanical approaches (Nolan, 2012; Nolan & Highhouse, 2014). In their review of existing literature on evidence-based selection tools, Neumann et al. (2021) suggest that research on subjective psychological needs fulfillment might be the most promising route to develop interventions aimed at reducing user resistance against mechanical selection approaches. Therefore, this stream of research is discussed in detail in the next section.

Self-Determination Theory, Subjective Needs Fulfillment and Intentions to Use Mechanical Decision-Making Approaches

Self-determination-theory (SDT) postulates that individuals' intrinsic motivation to engage in a task or an action depends on the degree to which it contributes to the fulfillment of three subjective psychological needs: competence, autonomy, and relatedness (Deci & Ryan, 2000). The need for competence entails the need to feel effective and skillful, the need for autonomy refers to the need to be in control of one's actions, and the need for relatedness can be described as the need to have meaningful relationships with others (Deci & Ryan, 2000; Nolan, 2012). Importantly, the three needs are neither substitutable nor hierarchical in nature, but rather, they need to be fulfilled independently to fuel intrinsic motivation (Deci & Ryan, 2000; Nolan, 2012).

In the context of employee selection decisions, scholars have shown that the use of algorithms reduces decision makers' perceived fulfillment of subjective needs, especially the needs of autonomy (Dietvorst et al., 2018; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan & Highhouse, 2014) and, to lesser extent, competence (Nolan, 2012). Depending on how algorithms are designed, they may restrict autonomy as they take away decision makers' choices such as which decision strategy to use, which performance indicators to consider, how to weigh them, and when to deviate from the algorithm when making a prediction. Similarly, decision makers might feel less competent when using a pre-specified algorithm as they can no longer utilize their expert knowledge in the process of data combination.

Based on the motivational assumptions underlying SDT, this implies that decision makers are less willing to apply mechanical than holistic decision-making approaches in practice, regardless of the fact that algorithms have been proven to provide more accurate, consistent, and valid decisions than human decision makers (Dawes et al., 1989; Grove et al., 2000; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meehl, 1954; Sawyer, 1966). This leads to a trade-off which Neumann, Niessen, Tendeiro, and Meijer (2022) call the "autonomy-validity dilemma". It describes the quest to design a decision approach that retains decision makers' subjective needs (such as but not limited to autonomy) to increase their use intentions whilst achieving empirically accurate, consistent, and valid decisions. Different approaches for dissolving this dilemma are described in the next section.

Autonomy-Enhancing Approaches for Employee Selection

To find a selection approach that is empirically valid, yet accepted by decision makers, Neumann, Niessen, Tendeiro, and Meijer (2022) tested the effect of a range of autonomy-enhancing features on decision makers' perceived needs fulfillment, use intentions regarding the selection method as well as the validity of their predictions. In their study, participants were given applicant data and were asked to predict the applicants' academic performance using multiple approaches: a holistic approach, a purely mechanical approach, a self-designed algorithm, or an adjustment method where participants were given a mechanical prediction which they could then adjust holistically. The latter two approaches can be considered autonomy-enhancing algorithmic procedures (AEAPs) as they are based on algorithmic data combination whilst allowing the individual to exert some influence (i.e., indicative of autonomy) over the decision strategy (Neumann, Niessen, Tendeiro, & Meijer, 2022). The results showed that participants reported significantly higher perceived autonomy when using AEAPs compared to purely mechanical methods. However, the study's results were also surprising in two ways: first, use intentions were even higher for some AEAPs than for purely holistic methods, although perceived autonomy was highest for holistic decisions (Neumann, Niessen, Tendeiro, & Meijer, 2022). According to Neumann, Niessen, Tendeiro, and Meijer (2022) this indicates that factors other than autonomy – such as competence – should be investigated to understand when and why decision makers are willing to use algorithms. Second, the study showed mixed results regarding the predictive validity of AEAPs, potentially because the investigated AEAPs did not fully ensure consistency in applying predictor weights. Thus, more research is needed to understand how prediction methods that are designed to enhance needs fulfillment affect predictive validity, decision makers' use intentions, and perceived fulfillment of subjective needs such as but not limited to autonomy.

Mechanical Synthesis

A mechanical prediction method which might bridge the gap between predictive validity and subjective needs fulfillment is the method of MS. In MS, the holistic prediction of the decision maker is integrated into an algorithm (Sawyer, 1966). Since the final data combination happens algorithmically, MS is considered a mechanical prediction method (Neumann, Niessen, Hurks, & Meijer, 2022). Nevertheless, implementing MS is not possible without human involvement. Several benefits arise from this sequential approach. Asking the decision maker to make a holistic prediction as part of MS might appeal to their perceived levels of autonomy and competence and might make it more likely for decision makers to accept MS as a prediction method compared to a purely mechanical approach (Kuncel, 2018). At the same time, since MS operates based on an pre-specified algorithm, weights will be applied consistently to performance predictors, thereby reducing noise and increasing predictive validity compared to a purely holistic prediction (Kahneman et al., 2016; Kahneman et al., 2021; Kuncel, 2018; Sawyer, 1966).

One could also argue that in MS not only the consistency but also the accuracy of predictor weights will be higher than in a holistic approach, given that optimal weights for each predictor could be derived from previous research (Kuncel, 2018). Whilst it is true that meta-analytic weights can be attached to performance predictors such as test or interview data (Kuncel, 2018), it is unclear which weight, if any, should be attached to the holistic prediction as part of the final algorithm. This is partly because it is empirically questionable whether including the holistic prediction can improve the final prediction at all (Murphy, 2019; Sackett et al., 2017). Nevertheless, for practical feasibility and acceptance reasons, it is highly relevant to include the holistic prediction, and to understand how the extent of decision makers' involvement in setting the holistic weight affects their use intentions. Thus, we developed two alternative approaches to determine how to incorporate the holistic prediction as a variable into the MS algorithm. These approaches are referred to as *fixed MS* and *self-designed MS*.

Fixed MS. Rather than setting an empirically optimal weight, the fixed MS approach attempts to set a practically feasible weight for the holistic prediction to be integrated into the MS algorithm. This weight is fixed in the sense that it is pre-determined in line with the organization's selection strategy; decision makers cannot change it. Considering practical feasibility is highly relevant to make sure that MS is actually accepted and applied by practitioners. According to advice-taking literature, decision makers tend to have an egocentric bias, meaning that they consider their own opinion more important than external advice (Bonaccio & Dalal, 2006; Krueger, 2003). Thus, when presenting decision makers with a fixed algorithm that includes their own prediction, we consider it unlikely for practitioners to accept a weight of less than fifty percent attached to their holistic prediction. A fixed MS approach ensures consistency of applying predictor weights. However, when letting the holistic prediction account for half the outcome of the prediction, accuracy of the predictor weights is only ensured for the fifty percent of the prediction that is performed by the algorithm, as those weights are derived from meta-analytic results. Nevertheless, we consider the approach of attaching a fixed weight to the holistic prediction promising as it provides some, but not full control to the decision maker, thereby balancing noise reduction and subjective needs fulfillment.

Self-designed MS. In contrast to the fixed MS approach, in the self-designed MS approach decision makers decide themselves how much weight they attach to their holistic prediction. Allowing decision makers to freely choose this weight could come at the cost of lower accuracy of predictor weights, as decision makers are free to choose any weight for their holistic prediction without empirical basis. This could result in a significant reduction of the predictive validity of the decision compared to a fully mechanical approach if decision makers assigned their own holistic judgment substantial weight (Kahneman et al., 2016; Kahneman et al., 2021). Nevertheless, giving decision makers this choice could increase their perceived autonomy and competence as well as their intentions to use this approach.

Research Question and Hypotheses Development

The choice of whether to apply a fixed MS, self-designed MS, fully holistic, or fully mechanical decision-making approach is at the core of the dilemma between predictive validity and subjective needs-fulfillment in decision-making. However, in both MS approaches there is potential to bridge the gap between science and practice as they allow the final decision to emerge from the combination of human expertise and algorithmic consistency. Thus, MS could result in higher predictive validity than purely holistic methods, whilst also leading to higher use intentions and perceived levels of decision makers' autonomy and competence compared to purely mechanical methods. Still, there is a lack of research regarding MS as an approach in employee selection. Therefore, this study aims to answer the following research question:

How do fixed MS and self-designed MS compare to holistic prediction and a prescribed algorithm in terms of predictive validity of applicants' job performance, practitioners' use intentions, and practitioners' levels of perceived autonomy and competence?

Predictive validity and practitioners' use intentions were chosen as outcome constructs because it can be assumed that only a decision-making approach that produces highly accurate predictions and that practitioners are willing to implement can replace holistic judgement in practice. Perceived levels of autonomy and competence were selected as additional outcome constructs to test previous assumptions linking subjective needs fulfillment and practitioners' algorithm aversion. Although SDT references relatedness as a third, relevant subjective need (Deci & Ryan, 2000), relatedness was not included as a variable in this study. This is because feelings of relatedness can be expected to be relevant during information collection (e.g., during selection interviews), or during information combination in a group setting, both of which is not the core of this study.

Hypotheses

As previous research suggests, purely holistic methods have high potential to fulfill decision makers subjective needs (Dietvorst et al., 2018; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan & Highhouse, 2014) and are widely accepted and applied by decision makers (Dietvorst et al., 2015; Highhouse, 2008; Meijer et al., 2020; Neumann, Niessen, Hurks, & Meijer, 2022; Nolan et al., 2016; Ryan et al., 2015) but lack predictive validity (Dawes et al., 1989; Grove et al., 2000; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meehl, 1954; Sawyer, 1966). On the contrary, purely mechanical methods have been shown to be highly predictive (Dawes et al., 1989; Grove et al., 2000; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meehl, 1954; Sawyer, 1966) but to suppress decision makers' sense of autonomy and competence and thus are met with resistance (Dietvorst et al., 2015, 2018; Neumann, Niessen, Hurks, & Meijer, 2022; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan et al., 2016; Nolan & Highhouse, 2014). In contrast, MS approaches entail the consistency of an algorithm whilst requiring the holistic prediction of the decision maker (Sawyer, 1966). Hence, we expect MS approaches to result in higher predictive validity than purely holistic methods and in higher levels of use intentions, perceived autonomy, and perceived competence than purely mechanical methods. Based on these considerations, we formulate the following hypotheses:

Hypothesis 1a: Fixed MS will result in more valid job performance predictions than holistic prediction.

Hypothesis 1b: Self-designed MS will result in more valid job performance predictions than holistic prediction.

Hypothesis 2a: Use intentions will be higher for fixed MS approaches than for algorithmic decision-making when making performance predictions.

Hypothesis 2b: Use intentions will be higher for self-designed MS approaches than for algorithmic decision-making when making performance predictions.

Hypothesis 3a: Perceived autonomy will be higher for fixed MS approaches than for algorithmic decision-making when making performance predictions.

Hypothesis 3b: Perceived autonomy will be higher for self-designed MS approaches than for algorithmic decision-making when making performance predictions.

Hypothesis 4a: Perceived competence will be higher for fixed MS approaches than for algorithmic decision-making when making performance predictions.

Hypothesis 4b: Perceived competence will be higher for self-designed MS approaches than for algorithmic decision-making when making performance predictions.

Methods

Participants

As pre-registered, the target sample size for this study was $n = 51$ participants per group, or $N = 204$ based on a power analysis conducted in G*power. The power analysis was conducted for a one-sided, independent two-sample t-test, assuming a medium effect size ($d = 0.5$) based on previous research (Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan & Highhouse, 2014) as well as the standard input parameters ($\alpha = 0.05$, $1\text{-}\beta = 0.8$) as recommended by Perugini et al. (2018).

Participants were recruited through personal networks as well as through the SONA Systems sampling tool. They were required to be at least 18 years of age to take part in the study. Participants did not receive any monetary reward for their participation, although those taking part via SONA gained study credits for completing the study. Data was excluded from the study if the respective participant did not finish the study or failed both of the two comprehension checks, or the attention check at the end of the study.

The final sample consisted of 209 participants (71% female, 26% male, 2% other, 1% not specified) who were between 18 and 61 years of age (*M* = 29.72, *SD* = 13.05). They were primarily of white/Caucasian ethnicity (90%). Most participants were German (54%) or Dutch (22%). The majority of participants (52%) held a Bachelor's or Master's degree. In terms of participants' occupational status, 43% worked fulltime, 39% worked part-time, 17% were unemployed, and 1% had retired. On average, participants had 9.58 years of work experience (*SD* = 11.48). Approximately two thirds of all participants (66%) indicated that they had never made a hiring decision. The remaining participants had on average 6.41 years of experience in making hiring decisions (*SD* = 7.02, min = 0, max = 25, mode = 1) and made on average 11.51 hiring decisions per year (*SD* = 34.28 , min = 0, max = 200, mode = 1). Due to the wide ranges and high standard deviations for the number of hiring decisions per year and the years of hiring experience in the sample, considering the modes rather than mean values might be more insightful to examine participants' overall experience with making hiring decisions. Based on modes, it can be concluded that most participants of this study have either none or very little experience in making hiring decisions.

Materials

This study relies on existing applicant data which was used as input data for the applicant profiles presented to participants. The applicant data was originally collected by Kausel et al. (2016). The original dataset consists of 236 real applicants and their scores on a general mental ability (GMA) test, a conscientiousness questionnaire, an unstructured interview, as well as a supervisor-rating of their actual job performance three months after being hired. In the original dataset, GMA test scores were provided as percentages, whilst conscientiousness and interview scores were presented on a scale of one to five. For interview scores, the range of the original data set was two to five, as applicants with an interview rating lower than two were not hired.

To avoid misinterpretation of original data due to different scale formats, applicants' GMA and conscientiousness scores were transformed into percentiles. Since it was not feasible to let participants make 236 performance predictions, a subsample of 40 applicants was extracted from the original data set. This subsample is representative of the original sample with regard to its correlation matrix of the applicant data, so that correlations on all off-diagonals do not deviate by more than a self-chosen threshold value of 0.015 from the original correlation matrix. To implement this extraction mechanism, the pre-existing, publicly available algorithm by Neumann, Niessen, Linde, et al. (2023) in the software R was used.

Design and Procedure

A between-subjects online experiment was carried out via the platform QUALTRICS. The experimental design entailed one factor (decision-making approach to make a performance prediction) with four levels (*holistic* decision-making, *algorithmic* decision-making, *self-designed MS* decision-making, and *fixed MS* decision-making). Both randomization and blinding were ensured as participants were randomly but evenly assigned to one of the four conditions, and neither the participants nor the researchers became aware of the condition that each participant was assigned to.

In the beginning, participants received general information about the purpose of the study and the use of their data and provided informed consent. Then, they were introduced to the setting of the study (i.e., that an airline was hiring new ticket agents and that each applicant completed a GMA test, a conscientiousness questionnaire, and an unstructured interview during the application process). Participants were informed that their task would be to predict applicants' job performance based on the applicants' assessment scores. Afterwards, participants were provided with an example of two applicant profiles (including GMA test, conscientiousness questionnaire, and unstructured interview scores) to familiarize themselves with the interpretation of applicant scores based on the underlying scoring scales. After completing a comprehension check, participants were randomly assigned to one of the four experimental conditions to start the prediction task according to the condition-specific instructions. In each condition except for the algorithmic condition, participants predicted the performance of the same 40 applicants in randomized order. Predictions were made one at a time, i.e., participants only considered one applicant profile in isolation rather than comparing multiple applicant profiles. Condition-specific instructions are provided in Table 1. After finishing the prediction task, participants responded to items regarding the dependent measures as well as demographic data. Scales of dependent measures and items within each scale were displayed in randomized order.

Holistic Decision-Making

In the *holistic* condition, participants were instructed to predict the performance of 40 applicants based on the applicant data (i.e., the GMA tests core, conscientiousness questionnaire score, and their interview rating) by using their intuition and expertise.

Algorithmic Decision-Making

In the *algorithmic* condition, participants did not make performance predictions themselves. Rather, they were informed that an algorithm which had been developed based on empirical findings would predict each applicant's performance based on the applicant's information (i.e., their test and interview scores). Participants received information about the predictors, weights, and mathematical equations by which the algorithm operated, and they were educated about the meaning of weights in such an algorithm. Furthermore, it was stated that applying an algorithm would empirically result in more valid performance predictions than using one's intuition and expertise. To understand how the algorithm operates, participants were shown ten applicant profiles and the corresponding performance predictions made by the algorithm. Participants could not adjust these performance predictions.

The algorithm used in this condition was designed based on existing findings by Cortina et al. (2000) regarding the predictive validity of GMA, conscientiousness, and unstructured interviews of job performance, which were used as performance predictors in this study. Following a procedure as performed by Neumann, Niessen, Linde, et al. (2023), optimized meta-analytic regression weights for each predictor were derived from the correlations as presented in the top panel of Figure 3 (p. 339) from Cortina et al. (2000). The GMA, conscientiousness, and interview scores of each applicant were then multiplied with the respective regression weight and summed up to arrive at an algorithmic performance prediction per applicant. Based on this method, the following meta-analytic regression weights were determined for the final algorithm: GMA 53 per cent, conscientiousness 28 per cent, unstructured interview 19 per cent.

Mechanical Synthesis Decision-Making

Two MS conditions were included in the study, namely *self-designed MS decision-making* and *fixed MS decision-making*. In both MS conditions, participants were introduced to a sequential decision-making approach following the logic of MS (Sawyer, 1966). They were informed that their task would be to make performance predictions of 40 applicants by using their intuition and expertise, and that their performance prediction would then be factored into an algorithm to derive a final performance prediction. As in the other conditions, participants received information about the algorithm's predictors, weights, mathematical operations, and empirical validity. Importantly, the two MS conditions differed in the way how the so-called holistic weight was determined (i.e., the weight that was attached to the holistic prediction).

Self-Designed MS Decision-Making. In the *self-designed MS* condition, participants were asked to freely determine a weight "X" between 0.00 and 1.00 with which their performance predictions should be factored into the final algorithm. The weight was chosen by participants once and was applied to all 40 predictions. Depending on the self-chosen weight, the weights for the other three performance predictors were calculated using the optimal weights distribution explained above, accounting together for 1-X. After setting a holistic weight, participants were presented with 40 applicant profiles to make performance predictions. Every time after making a holistic performance prediction for an applicant, participants were shown the performance prediction determined by the algorithm, taking into account their self-chosen holistic weight for their holistic prediction. It was not possible to adjust these final performance predictions.

Fixed MS Decision-Making. In the *fixed MS* condition, participants were not able to determine the weight of their own performance prediction. Rather, they were informed that their holistic prediction was multiplied by a fixed weight of 50 per cent, which was set by the researchers. This value was chosen based on assumptions of advice-taking literature, which implies that it might be unfeasible to introduce instruments in practice that give less responsibility to the hiring professional than to the algorithm (Bonaccio & Dalal, 2006). Thus, participants made 40 holistic performance predictions and each prediction was factored into the pre-specified algorithm by 50 per cent. Accordingly, based on the optimal weight distribution explained above, the following weights resulted for the remaining predictors: GMA 27 per cent, conscientiousness 14 per cent, unstructured interview 9 per cent. After each prediction, participants were shown the final performance prediction, as determined by the algorithm using these weights. They could not adjust these final performance predictions.

Measures

Manipulated Variable

The decision-making approach through which participants were asked to make performance predictions was manipulated, resulting in four conditions: *holistic* decision-making, *algorithmic* decision-making, decision-making based on a *self-designed MS* approach, and decision-making based on a *fixed MS* approach.

Outcome Variables

Predictive validity, practitioners' use intentions, perceived level of autonomy, and perceived level of competence were measured. Unless stated otherwise, participants responded on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). In this study, the outcome variables use intentions, perceived autonomy, and perceived competence will be referred to as attitudinal measures.

Predictive Validity. Predictive validity was operationalized as the correlation between a participant's performance prediction and the applicants' real performance rating (as provided in Kausel et al. (2016). Both ratings were measured on a 5-point, one-decimal scale (1 = very bad, 5 = very good).

Use Intentions. A three-item scale by Nolan (2012) was used to measure participants' intentions to apply a specific decision-making approach to make recruitment decisions in the future ($\alpha = 0.86$). Participants responded to items such as "If I were in charge, I would use this approach to make hiring decisions" (Nolan, 2012, p. 102).

Perceived Autonomy. The six-item scale by Nolan (2012) was used to measure participants' perceived autonomy, as derived from SDT (Deci & Ryan, 2000), when applying a certain decision-making approach in hiring scenarios ($\alpha = 0.84$). Items were formulated as follows: "Hiring [employees] in this way would give me a sense of… control / choice / free will /…" (Nolan, 2012, p. 103).

Perceived Competence. A six-item scale by Nolan (2012) was applied, measuring feelings of competence as derived from SDT (Deci & Ryan, 2000) in settings of employee selection ($\alpha = 0.86$). Participants responded to items such as "Using this approach to hire [employees] would make me feel… effective / capable / useful /…" (Nolan, 2012, p. 103).

Data Analysis

The data was analyzed using R 4.3.0. Given the comparative nature of the hypotheses, one-way, between-subjects ANOVAs and pairwise independent post-hoc t-tests were used as the main tools of analysis. As pre-registered, post-hoc tests were carried out following Tukey's HSD method. In comparison to other methods, Tukey's post-hoc test has been recommended for pairwise multiple comparison procedures in between-subject designs based on its statistical power, simplicity, and the availability of confidence interval boundaries for all comparisons (Jaccard et al., 1984). To ensure that assumptions for this analytical approach were met, homogeneity of variance was tested for each dependent variable across conditions using the Levene's test. Furthermore, histograms and the Shapiro-Wilk test were used to check if each dependent variable was normally distributed in each condition, or alternatively, if the residuals of the respective ANOVA were normally distributed. For the attitudinal measures the reliability of the respective scales was calculated by computing Cronbach's alpha.

To test hypotheses 1a and 1b, the performance predictions of each participant were correlated with the real performance ratings of each applicant, resulting in one Pearson's correlation coefficient per participant. This correlation coefficient was then transformed into Fisher's *z*-score to reduce bias when averaging correlation coefficients (Silver & Dunlap, 1987). The Fisher's *z*-scores per participant were used as the unit of analysis for a subsequent ANOVA and post-hoc test.

To test hypotheses 2a to 4b, the mean score of all items making up the respective dependent variable was formed, which was subsequently used as the unit of analysis for ANOVAs and post-hoc Tukey's HSD tests. To classify the magnitude of effect sizes, the standards as proposed by Cohen (1992) were applied.

Results

Holistic weight

In the self-designed MS condition, the holistic weight that participants assigned to their holistic prediction had a mean of 0.42 (SD = 0.21, min = 0, max = 1), or 42 per cent. Based on this weight and the meta-analytic weights distribution derived by Cortina et al. (2000), the following weights were attached on average to the other performance predictors: GMA 31 per cent, conscientiousness 16 per cent, unstructured interview 11 per cent.

Predictive Validity

Prior to conducting an ANOVA, a significant result was found for the Levene's test for the Fisher's *z*-scores, indicating that homogeneity of variance cannot be assumed for the dependent variable. However, this can be neglected as group sizes are approximately equal. Histograms and the output of the Shapiro-Wilks test indicated that the Fisher's *z*-scores are not normally distributed in the self-designed MS condition. However, the histogram of the residuals of the ANOVA for this condition showed a normal distribution, indicating that this variable can be analyzed by means of an ANOVA, as intended.

Table 2 shows the means and standard deviations for predictive validity per condition, each for the non-transformed Pearson's correlation coefficient and the Fisher's *z*-scores. The mean validity coefficient per condition is displayed in Figure 1. These means suggest that holistic judgement (r = 0.28) is outperformed by both fixed MS and self-designed MS approaches (both r = 0.34). In the next step, these assumptions were tested statistically.

The ANOVA returned a significant result, implying that there are statistically significant differences in predictive validity across conditions ($F_{(3, 205)} = 35.65$, $\eta^2_{partial} = 0.34$, $p < 0.001$). A Tukey's HSD post-hoc test and a subsequent calculation of Cohen's d showed that the holistic condition differed significantly from the fixed MS condition ($M_{diff} = -0.067$, 95% CI [-0.09, -0.04], $p < 0.001$, $d = 1.14$) and from the self-designed MS condition ($M_{diff} = -0.068$, 95% CI [-0.09, -0.04], $p < 0.001$, $d = 1.09$) in terms of predictive validity. Thus, Hypotheses 1a and Hypothesis 1b are supported and large effects can be reported for both comparisons.

Attitudinal Measures

Descriptive statistics of all attitudinal measures can be found in Table 3 and Table 4. The means of the attitudinal measures per condition appear to support the hypotheses, as both MS conditions yield higher values than the algorithmic condition for autonomy, competence, and use intentions (although differences are noticeably small for the latter). However, no conclusions can be drawn from these values alone as the differences between groups need to be tested for statistical significance.

Use Intentions

The Levene's test for the variable use intentions was not significant, meaning there was not enough evidence that the assumption of homogeneity of variance was violated. However, based on histograms and Shapiro-Wilk tests of both the dependent variable across conditions and the residuals of the ANOVA, there is no evidence the criterion of normal distribution is met. Thus, in addition to an ANOVA, a Kruskal-Wallis test was carried out to account for the possibility that the variable of interest is not normally distributed.

Both the ANOVA ($F_{(3, 205)} = 2.22$, $\eta^2_{partial} = 0.03$, $p = 0.09$) and the Kruskal-Wallis test ($H_{(3)} = 7.04$, $p = 0.07$) returned non-significant results. Based on this evidence, there is no support for hypothesis 2a and hypothesis 2b. A subsequent Tukey's HSD test ($q_{(3, 205)} = 3.66$) revealed that none of the comparisons between conditions were significant. However, it should be pointed out that the lower boundary of the 95% confidence interval was very close to zero for the comparison of the fixed MS to the holistic condition ($M_{diff} = 0.4$, 95% CI [-0.03, 0.83], $p = 0.08$, $d = 0.49$). Thus, this result should be treated with caution.

The difference between use intentions in the fixed MS compared to the algorithmic condition was small ($M_{diff} = 0.34$, 95% CI [-0.1, 0.77], $p = 0.2$, $d = 0.37$) whilst the difference between use intentions in the self-designed MS versus algorithmic condition was negligible ($M_{diff} = 0.11$, 95% CI [-0.33, 0.55], $p = 0.92$, $d = 0.12$). Strikingly, the difference between the holistic and algorithmic condition in terms of use intentions was also negligible ($M_{diff} = -0.07$, 95% CI [-0.51, 0.37], $p = 0.98$, $d = 0.08$), whilst a comparison of holistic versus the two MS conditions yielded small effect sizes.

Perceived Autonomy

For perceived autonomy, all prerequisites regarding homogeneity of variance and normal distribution were met. Thus, an ANOVA was carried out, which returned significant results ($F_{(3, 205)} = 7.57$, $\eta^2_{partial} = 0.10$, $p < 0.001$). The subsequent Tukey's HSD test ($q_{(3, 205)} =$

3.66) showed support for hypothesis 3a and hypothesis 3b as both the fixed and the self-designed MS condition differed significantly from the algorithmic condition in terms of perceived autonomy. Both effects can be classified as moderate, as indicated by the effect sizes of $d = 0.6$ (fixed MS vs. algorithmic condition, $M_{diff} = 0.50$, 95% CI [0.13, 0.88], $p < 0.01$) and $d = 0.75$ (self-designed MS vs. algorithmic condition, $M_{diff} = 0.53$, 95% CI [0.15, 0.9], $p < 0.01$).

Although not covered in the hypotheses, a large, significant effect can furthermore be reported for the comparison of holistic versus algorithmic condition with respect to perceived autonomy ($M_{diff} = 0.63$, 95% CI [0.26, 1.01], $p < 0.001$, $d = 0.88$). In contrast, differences in perceived autonomy between the holistic and either of the MS conditions were non-significant and negligible in size.

Perceived Competence

Based on a Levene's test, there was not enough evidence that the assumption of homogeneity of variance for the variable of competence was violated. However, similar to the variable of use intentions, the competence variable did not seem to be normally distributed based on histograms and results of Shapiro-Wilk tests. Thus, both an ANOVA and a Kruskal-Wallis test were conducted to account for a scenario where the data is in fact not normally distributed.

Neither the ANOVA ($F_{(3, 205)} = 2.37$, $\eta^2_{partial} = 0.03$, $p = 0.07$) nor the Kruskal-Wallis ($H_{(3)} = 6.93$, $p = 0.07$) test returned significant results. Therefore, hypotheses 4a and 4b cannot be confirmed based on this analysis. An exploratory Tukey's HSD post-hoc test ($q_{(3, 205)} = 3.66$) revealed that none the comparisons showed significant differences. However, a closer look at the 95% confidence interval boundaries for the comparison of fixed MS to algorithmic ($M_{diff} = 0.30$, 95% CI [-0.07, 0.67], $p = 0.16$, $d = 0.4$), and self-designed MS to algorithmic condition (M_{diff} = 0.35, 95% CI [-0.02, 0.72], p = 0.07, d = 0.48) showed that lower confidence interval boundaries were close to zero.

Discussion

Although algorithms have higher predictive validity than holistic judgement, they are underutilized in practice. Hence, it was the aim of this study to understand how mechanical decision-making approaches should be designed so that they enhance predictive validity compared to holistic decision-making whilst increasing decision makers' use intentions compared to mechanical approaches. By presenting these results, we respond to the call of other scholars to investigate alternative decision-making approaches that are empirically valid yet accepted by decision makers (Kuncel et al., 2013; Meijer et al., 2020; Neumann, Niessen, Hurks, & Meijer, 2022; Nolan, 2012). The results suggest that MS is a promising intervention to improve decision-making as both fixed and self-designed MS are generally accepted by decision makers and result in better predictions than holistic decision-making. The findings further indicate that subjective needs fulfillment is a relevant factor in explaining decision makers' resistance to use algorithms as MS approaches resulted in autonomy-enhancement compared to mechanical approaches.

In existing literature regarding the predictive validity of decision-making approaches, scholars have discussed whether inaccurate predictor weights, inconsistent use of such weights, or both are the main drivers of low predictive validity (Dawes et al., 1989; Kahneman et al., 2016; Kahneman et al., 2021; Yu & Kuncel, 2020). The findings of the present study add to this dialogue. Several scholars emphasize that increased consistency in applying predictor weights is the main reason why MS is likely to outperform holistic judgement, as its sequential nature ensures that holistic judgement is integrated into an algorithm following a pre-determined weighting scheme (Kuncel, 2018; Sawyer, 1966). In line with that reasoning, researchers have shown that self-designed algorithms, and algorithms consisting of equal or even random weights can outperform holistic judgement if valid predictors are chosen (Neumann, Niessen, Tendeiro, & Meijer, 2022; Yu & Kuncel, 2020). Thus, these findings point to consistency as the most relevant factor for achieving predictive validity. In contrast, Kausel et al. (2016) found that inaccuracy in setting predictor weights, rather than inconsistency in applying such weights was the main reason why self-designed algorithms were less predictive than fully mechanical approaches. The results of the present study might resolve this contradiction. In this study, both MS approaches account for consistency in applying predictor weights. In contrast, the holistic approach does not. In terms of accuracy of predictor weights, this contrast is less clear. Whilst in holistic approaches, accuracy of weights is likely low, weights chosen for the MS approaches are not fully accurate either, as MS approaches still attach a weight to the holistic judgement, although it is likely barely valid. Nevertheless, both MS approaches outperform holistic judgement in terms of predictive validity. Based on this reasoning, our findings support the notion that consistency rather than accuracy of predictor weights seem to be the most relevant driver of predictive validity in decision-making.

With regard to perceived autonomy, the results of this study match previous findings by showing that participants perceive higher autonomy when using a decision-making approach that combines human and algorithmic judgement rather than a fully algorithmic approach (Dietvorst et al., 2018; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan & Highhouse, 2014). Also, our data supports the notion of Nolan and Highhouse (2014) that not simply the possibility to experience autonomy but the degree to which a decision-making approach allows for acting autonomously is relevant. In our findings, this is evident as the comparison of the self-designed MS condition to the algorithmic condition yields a greater difference and a larger effect size in terms of autonomy than the comparison of the fixed MS condition to the algorithmic condition. Furthermore, we add to existing literature by pointing out that neither of the MS conditions differed significantly from holistic decision-making in terms of perceived autonomy. Combined with our findings on predictive validity, this implies that replacing holistic judgement with MS approaches has the potential to enhance predictive validity of decisions without taking away decision makers' sense of autonomy.

Nevertheless, we were not able to report significant differences between holistic, algorithmic, and MS approaches with regards to participants' use intentions or perceived level of competence. This is somewhat surprising considering that previous studies found such effects, even though these studies did not study MS approaches (Dietvorst et al., 2018; Highhouse & Brooks, 2023; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan et al., 2016). Although prior findings state that use intentions might result from enhanced autonomy (Nolan, 2012), these findings could not be replicated in our study as differences were found in perceived autonomy but not in participants' use intentions across conditions. Regarding decision makers' perceived level of competence in different decision-making approaches, research is scarce (Neumann, Niessen, Tendeiro, & Meijer, 2022). However, our results suggest that the fulfillment of competence needs might play less of a role than expected for explaining decision makers' propensity to apply different decision-making approaches. A potential explanation could be that competence, unlike autonomy, might be perceived as a domain-specific construct, implying that participants' lack of experience in making hiring decisions might have limited their competence perceptions.

Theoretical Implications

Several implications for theory and practice can be derived from our findings. First, with regard to investigating drivers of predictive validity in decision-making, our findings imply that consistency rather than accuracy of predictor weights should be the focus of further studies in this stream of research. Thus, it might be most promising to centre research efforts around decision-making approaches such as self-designed algorithms or MS variations, as these approaches entail a pre-determined algorithmic pattern that is consistently applied in each prediction. Consequently, research within this domain might move away from investigating ways to educate decision makers about the accuracy of performance indicators and shift towards investigating how decision makers can be aided in consistently applying predictor weights.

A second theoretical contribution is that SDT as a theoretical framework might only partially explain decision makers' propensity to apply certain decision-making approaches. Whilst many studies, including the present one, found significant differences in autonomy perceptions between different decision-making approaches (Dietvorst et al., 2018; Neumann, Niessen, Tendeiro, & Meijer, 2022; Nolan, 2012; Nolan & Highhouse, 2014) there is a lack of similar findings for competence needs. Although we do not doubt that the fulfillment of subjective needs plays a role in determining decision makers' attitudes towards decision-making approaches, it is questionable whether competence should be considered equally important as autonomy in that matter. That being said, the results suggest that autonomy alone might not be sufficient to affect decision makers' use intentions, implying that additional factors outside the theoretical framework of SDT might be of relevance.

Additionally, our study contributes to research on alternative decision-making approaches by providing new insights on the weight that participants assign to their holistic prediction before it is integrated into an algorithm. Such data has so far not been collected in comparable studies. Given that participants attached an average weight of 42 per cent to their holistic judgement, making it the performance predictor with the highest weight, the results imply that egocentric bias, as discussed in advice-taking literature (Bonaccio & Dalal, 2006) may play a role in explaining decision makers' attitudes towards decision-making approaches. This finding provides a starting point for future research on the question of how predictor weights should be distributed between holistic and algorithmic judgement in combined decision-making approaches to ensure both predictive validity and practical feasibility.

Limitations and Directions for Future Research

This study is not without limitations. A larger sample might have increased the chance to find significant effects for the variables use intentions and perceived competence as small effects are more likely to be found when the sample is large (Perugini et al., 2018). Given that this study is one of the first ones to investigate MS approaches as alternatives to algorithmic and holistic approaches, it also would have been interesting to differentiate between participants with and without hiring experience in the sample, as hiring professionals might generally be more aware of their competence in the domain of making hiring decisions. In the present study, a differentiation based on hiring experience was not viable as participants had none or very little hiring experience. Thus, we call upon future researchers to compare lay people and hiring professionals regarding their attitudes towards MS approaches.

Apart from the sample size, the operationalization of the use intentions variable might have prevented us from finding significant results. Although it is common practice to measure participants' use intentions using self-reported measures, relying on actual behaviour of participants might be the more informative. For future studies, we encourage scholars to apply a multiple-stage approach to assess use intentions, where participants decide based on prior experience with different decision-making approaches which approach to use in the next stage (see Dietvorst et al., 2015, 2018) .

The presentation of performance indicators might have been a further limitation of this study. Using two different scales to present three performance indicators could have confused participants. Also, a scale of two to five, on which the interview scores were presented, is much less differentiated than providing percentiles, as it was done for GMA and conscientiousness scores. For future studies, this implies that presenting all performance indicators on the same scale might be most desirable to avoid misinterpretations.

Finally, as this study aimed to compare different decision-making approaches regarding subjective needs fulfilment as proposed by SDT, including not only two but all three subjective needs as outcome variables could have led to even more conclusive findings. Considering the need of relatedness in addition to autonomy and competence might allow future scholars to analyse on a more holistic level to what extent SDT as a theoretical framework provides explanations for decision maker's tendencies to choose certain decision-making approaches over others.

Practical Recommendations

For practice, this study yields a relatively simple yet important implication: Based on the findings, we encourage organizations to implement MS rather than holistic approaches when making hiring decisions. Our results suggest that by doing so, organizations will be able to improve the predictive validity of decisions without taking away autonomy from decision makers. Contingent on our assumption and previous scholars' findings that increased levels of autonomy can result in higher use intentions (Dietvorst et al., 2018; Nolan, 2012; Nolan & Highhouse, 2014), MS might be an approach that is, on the one hand, more empirically valid than holistic decision-making, and on the other hand, more socially accepted than algorithmic approaches.

Conclusion

In this study, a holistic, an algorithmic, a fixed MS, and a self-designed MS decision-making approach were compared regarding predictive validity, decision makers' use intentions, perceived levels of autonomy, and perceived levels of competence. MS approaches significantly outperformed holistic decision-making in terms of predictive validity and led to higher ratings of perceived autonomy than the algorithmic approach. Additionally, MS approaches did not significantly differ from a holistic approach in terms of autonomy. Across all approaches, no significant differences were found regarding decision makers' use intentions and perceived competence. These findings imply that inconsistency in applying predictor weights might be more relevant than inaccuracy in setting predictor weights to explain (lack of) predictive validity in decision-making. Furthermore, they raise the question whether SDT as a theoretical framework is sufficient to explain decision makers' attitudes towards particular decision-making approaches. In sum, we consider MS approaches a promising compromise for organizations to retain the hiring professional whilst reducing noise in hiring decisions.

References

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Burton, J. W., Stein, M.-K., & 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

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/14805-018

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, *53*(2), 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

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

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical Versus Actuarial Judgment. *Science*, *243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

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. S., & 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. S., & 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

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

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. (2023). Improving Workplace Judgments by Reducing Noise: Lessons Learned from a Century of Selection Research. *Annual Review of Organizational Psychology and Organizational Behaviour*, *10*(1), 519-533. https://doi.org/10.1146/annurev-orgpsych-120920-050708

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise Multiple Comparison Procedures: A Review. *Psychological Bulletin*, *96*(3), 589–596. https://doi.org/10.1037/0033-2909.96.3.589

Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making: Algorithmic judgement is more efficient than the human variety. *Harvard Business Review*, *94*(10), 38–46. https://hbr.org/2016/10/noise

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment*. HarperCollinsPublishers.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27–44. https://doi.org/10.1016/j.obhdp.2016.07.005

Krueger, J. I. (2003). Return of the ego-self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review*, *110*(3), 585–590. https://doi.org/10.1037/0033-295x.110.3.585

Kuncel, N. R. (2018). Judgment and Decision Making in Staffing Research and Practice. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology and Employee Performance* (pp. 474–487). SAGE Publications Ltd. https://doi.org/10.4135/9781473914940.n17

Kuncel, N. R., Klieger, D. M., Conelly, 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

Liu, Y., Combs, J. G., Ketchen, D. J., & Ireland, R. D. (2007). The value of human resource management for organizational performance. *Business Horizons*, *50*(6), 503–511. https://doi.org/10.1016/j.bushor.2007.07.002

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. https://doi.org/10.1037/11281-000

Meehl, P. E. (1986). Causes and Effects of My Disturbing Little Book. *Journal of Personality Assessment*, *50*(3), 370–375. https://doi.org/10.1207/s15327752jpa5003_6

Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A Tutorial on Mechanical Decision Making for Personnel and Educational Selection. *Frontiers in Psychology*, Article 3002. https://doi.org/10.3389/fpsyg.2019.03002

Murphy, K. R. (2019). Understanding how and why adding valid predictors can decrease the validity of selection composites: A generalization of Sackett, Dahlke, Shewach, and Kuncel (2017). *International Journal of Selection and Assessment*, *27*(3), 249–255. https://doi.org/10.1111/ijsa.12253

Neumann, M., Niessen, A. S. M., Hurks, P. P. M., & Meijer, R. R. (2022). Holistic and Mechanical Combination in Psychological Assessment: Why Algorithms are Underutilized and What is Needed to Increase their Use.

Neumann, M., Niessen, A. S. M., Linde, M., Tendeiro, J. N., & Meijer, R. R. (2023). “Adding an Egg” in Algorithmic Decision Making: Improving Stakeholder and User Perceptions, and Predictive Validity by Enhancing Autonomy. *PsyArxiv*. https://osf.io/brc9p/?view_only=211357b3a78446078715d47a633d8219

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., & Meijer, R. R. (2023). Predicting decision-makers’ algorithm use. *Computers in Human Behavior*, *145*, 107759. https://doi.org/10.1016/j.chb.2023.107759

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), Article e2270. https://doi.org/10.1002/bdm.2270

Nolan, K. P. (2012). *Basic psychological need fulfillment and user resistance to objective and analytical decision-making practices in employee selection.* Bowling Green State University. http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1343479006

Nolan, K. P., Carter, N., & Dalal, D. (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

Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology*, *31*(1), Article 20. https://doi.org/10.5334/irsp.181

Ryan, A. M., Inceoglu, I., Bartram, D., Golubovich, J., Reeder, M., Derous, E., Nikolaou, I., & Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou & J. K. Oostrom (Eds.), *Current issues in work and organizational psychology. Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). Psychology Press, Taylor & Francis Group.

Sackett, P. R., Dahlke, J. A., Shewach, O. R., & Kuncel, N. R. (2017). Effects of Predictor Weighting Methods on Incremental Validity. *Journal of Applied Psychology*, *102*(10), 1421–1434. https://doi.org/10.1037/apl0000235

Sawyer, J. (1966). Measurement and Prediction Clinical and Statistical. *Psychological Bulletin*, *66*(3), 178–200. https://doi.org/10.1037/h0023624

Silver, N. C., & Dunlap, W. P. (1987). Averaging Correlation Coefficients: Should Fisher's z Transformation Be Used? *Journal of Applied Psychology*, *72*(1), 146–148. https://doi.org/10.1037/0021-9010.72.1.146

Yu, M. C. (2018). *Viewing Expert Judgement in Individual Assessment through the Lens Model* [Doctoral Dissertation]. University of Minnesota, Minnesota. https://conservancy.umn.edu/bitstream/handle/11299/199018/Yu_umn_0130E_19181.pdf?sequence=1

Yu, M. C., & Kuncel, N. R. (2020). Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments. *Personnel Assessment and Decisions*, *6*(2), Article 2, 1–10. https://doi.org/10.25035/pad.2020.02.002

Yu, M. C., & Kuncel, N. R. (2022). Testing the value of expert insight: Comparing local versus general expert judgment models. *International Journal of Selection and Assessment*, *30*(2), 202–215. https://doi.org/10.1111/ijsa.12356

Table 1

Instructions to Participants per Condition.

Condition	Instructions
Holistic	Please review the scores of the applicants and predict how the applicants will perform on the job as an airline ticketing agent. To do so, please use the following approach: Make the decision based on your intuition and expertise, by thinking about the applicants' information.
Algorithmic	Using this approach, you will not make performance predictions yourself. Instead, your performance predictions will be made by an algorithm. The algorithm takes into account the applicant's general mental ability score, conscientiousness score, and interview score. The final prediction is reached by attaching a weight to each piece of information. The weight distribution is determined according to the following, empirically-validated pattern that is based on multiple scientific studies: 53% General mental ability score 28% Conscientiousness score 19% Interview score The weights reflect the importance assigned to each piece of information for the performance prediction. Based on empirical findings, the algorithm above assigns the greatest importance to the general mental ability test score, followed by the conscientiousness score and the interview score. Mathematically, the algorithm corresponds to the following equation: Algorithm prediction = General mental ability score * 0.53 + Conscientiousness score * 0.28 + Interview score * 0.19

The higher the algorithm prediction on a scale of 1.0 to 5.0, the more likely it is that the applicant shows very good job performance. Although the algorithm will probably not result in perfect performance predictions, research shows that using such an algorithm results in more accurate performance predictions than using one’s intuition and expertise alone.

The algorithm predicts the job performance of all 40 applicants. To demonstrate how the algorithm operates, we will show you 10 performance predictions made by the algorithm. You cannot adjust these predictions.

Fixed MS	Please review the scores of the applicants and predict how the applicants will perform on the job as an airline ticketing agent. To do so, please use the following approach: First, make the decision based on your intuition and expertise, by thinking about the applicants' information. Afterwards, your prediction will be incorporated into an algorithm. The algorithm takes into account your performance prediction as well as the applicant's general mental ability score, conscientiousness score, and interview score. The final prediction is reached by attaching a weight to each piece of information. Excluding your performance prediction, the weight distribution for the other three components was determined according to the following, empirically-validated pattern that is based on multiple scientific studies: 50% Your performance prediction 27% General mental ability score 14% Conscientiousness score 9% Interview Score The weights reflect the importance assigned to each piece of information for the performance prediction. Based on empirical findings, when only considering the general mental ability test score, conscientiousness score, and interview score, the algorithm above assigns the greatest importance to the general mental ability test score, followed by the conscientiousness score and the interview score.

	Mathematically, the algorithm corresponds to the following equation: Algorithm prediction = Your performance prediction * 0.5 + General mental ability score * 0.27 + Conscientiousness score * 0.14 + Interview score * 0.09 The higher the algorithm prediction on a scale of 1.0 to 5.0, the more likely it is that the applicant shows very good job performance. Although the algorithm will probably not result in perfect performance predictions, research shows that using such an algorithm results in more accurate performance predictions than using one's intuition and expertise alone. In the next step, you will see applicant data of 40 applicants and predict the job performance of each applicant. Every time after giving a performance prediction, you will be shown the final prediction as determined by the algorithm based on your weighted prediction. You cannot adjust this final prediction.
Self-designed MS	Please review the scores of the applicants and predict how the applicants will perform on the job as an airline ticketing agent. To do so, please use the following approach: First, make the decision based on your intuition and expertise, by thinking about the applicants' information. Afterwards, your prediction will be incorporated into an algorithm. The algorithm takes into account your performance prediction as well as the applicant's general mental ability score, conscientiousness score, and interview score. The final prediction is reached by attaching a weight to each piece of information. You will decide yourself how much weight (x) will be attached to your performance prediction. You can choose a weight between 0 and 1 (for example: 0.30 or 0.77). In total, the weights of all 4 components (your prediction, applicant's general mental ability score, conscientiousness score, and interview score) will add up to 1.0. After you have chosen a weight, the weight distribution for the other three components will be determined according to the following, empirically-validated pattern that is based on multiple scientific studies:

53% General mental ability score
28% Conscientiousness score
19% Interview score

Based on empirical findings, when only considering the general mental ability test score, conscientiousness score, and interview score, the algorithm above assigns the greatest importance to the general mental ability test score, followed by the conscientiousness score and the interview score.

Mathematically, the algorithm corresponds to the following equation:

Algorithm prediction = Your performance prediction * x + General mental ability score * (0.53 * (1-x)) + Conscientiousness score * (0.28 * (1-x)) + Interview score * (0.19 * (1-x))

The weights reflect the importance of each predictor for the final performance prediction. For example, if you believe that your performance prediction is very important for the final outcome compared to the other predictors, you should choose a weight that is closer to 1. Conversely, if you consider your performance prediction not very important for the final outcome compared to the other predictors, you should choose a weight that is closer to 0.

You will decide on the weight of your prediction (x) once. This weight will be used for the calculation of all 40 performance predictions.

The higher the algorithm prediction on a scale of 1.0 to 5.0, the more likely it is that the applicant shows very good job performance. Although the algorithm will probably not result in perfect performance predictions, research shows that using such an algorithm results in more accurate performance predictions than using one’s intuition and expertise alone.

In the next step, you will be asked to choose a weight for your performance prediction. Afterwards, you will see applicant data of 40 applicants and predict the job performance of each applicant. Every time after giving a performance prediction, you will be shown the final prediction as determined by the algorithm based on your weighted prediction. You cannot adjust this final prediction.

Table 2

Sample Size, Means and Standard Deviations for Predictive Validity per Condition and Correlation Coefficient.

Condition	Sample size	Predictive validity			
		Pearson's *r*		Fisher's *z*	
	n	*M*	*SD*	*M*	*SD*
Holistic	53	0.28	0.07	0.28	0.08
Algorithmic	50	0.36	0.00	0.37	0.00
Fixed MS	54	0.34	0.03	0.35	0.03
Self-designed MS	52	0.34	0.04	0.35	0.04

Table 3

Descriptive Statistics and Correlation between Attitudinal Measures.

Measure	1.	2.	3.	*M*	*SD*
1. Use Intentions	-	0.36	0.60	2.82	0.87
2. Perceived Autonomy		-	0.49	2.90	0.77
3. Perceived Competence			-	3.13	0.73

Table 4

Sample Size, Means and Standard Deviations per Condition and Attitudinal Measure.

		Use Intentions		Perceived Autonomy		Perceived Competence	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Holistic	53	2.67	0.78	3.11	0.63	3.09	0.71
Algorithmic	50	2.73	0.95	2.48	0.80	2.92	0.76
Fixed MS	54	3.06	0.87	2.98	0.87	3.22	0.75
Self-designed MS	52	2.83	0.84	3.01	0.59	3.27	0.68

Figure 1

Mean Predictive Validity per Condition.