Holistic and mechanical combination in psychological assessment

Neumann, Marvin; Niessen, A. Susan M.; Hurks, Petra P. M.; Meijer, Rob R.

Holistic and Mechanical Combination in Psychological Assessment: Why Algorithms are Underutilized and What is Needed to Increase their Use

Marvin Neumann[1], A. Susan M. Niessen[1], Petra P. M. Hurks[2], & Rob R. Meijer[1]

[1]Department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, University of Groningen

[2]Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, University of Maastricht

Author note

Marvin Neumann, https://orcid.org/0000-0003-0193-8159

A. Susan M. Niessen, https://orcid.org/0000-0001-8249-9295

Petra P. M. Hurks, https://orcid.org/0000-0002-4366-3707

Rob R. Meijer, https://orcid.org/0000-0001-5368-992X

Correspondence concerning this article should be addressed to Marvin Neumann, Department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. Email: m.neumann@rug.nl.

Submitted for publication. This paper has not been peer-reviewed yet.

Abstract

Although mechanical combination results in more valid judgments and decisions than holistic combination, existing publications suggest that mechanical combination is rarely used in practice. Yet, these publications are either descriptions of anecdotal experiences or outdated surveys. Therefore, in several Western countries, we conducted two surveys (total *N* = 323) and two focus groups to investigate (1) how decision makers in psychological and HR practice combine information, (2) why they do (not) use mechanical combination, and (3) what may be needed to increase its use in practice. Many participants reported mostly using holistic combination, usually in teams. The most common reasons for not using mechanical combination were that algorithms are unavailable in practice and that stakeholders do not accept their use. Furthermore, decision makers do not quantify information, do not believe in research findings on evidence-based decision making, and think that combining holistic and mechanical combination results in the best decisions. The most important reason why mechanical combination is used was to increase predictive validity. To stimulate the use of mechanical combination in practice, our results suggest that decision makers should receive more training on evidence-based decision making, and decision aids supporting the use of mechanical combination should be developed.

Keywords: algorithm aversion, personnel selection, holistic prediction, mechanical prediction, science-practice gap, decision aid

Data availability statement

The data and R scripts that support the findings of our studies are openly available on OSF, https://osf.io/v3thf/.

Acknowledgements

We thank the members of the recruitingrebels e.V. for filling in and distributing the survey from Study 1, and we thank Kienbaum Consultants International GmbH for funding Study 2. We also thank the NIP for letting us recruit participants via a mailing list for Study 3. Lastly, we thank the participants in Study 3 for their fruitful insights on decision making in practice.

Practitioner points

- Combining information with an algorithm (mechanical combination) results in more valid judgments and decisions than combining information in the mind (holistic combination). Yet, decision makers rarely use mechanical combination in practice. To improve predictive validity, transparency, and the opportunity for learning, an algorithm should be used.
- Reasons reported by decision makers why they rarely use mechanical combination are that they do not and cannot quantify all available information, do not believe in research findings on evidence-based decision making, and think that a combination of mechanical and holistic combination results in the best judgments and decisions. Furthermore, decision makers fear negative stakeholder evaluations when they would use algorithms.
- Decision makers showed many misunderstandings regarding holistic and mechanical combination, even after reading an elaborate explanation of the two methods.
- To improve decision making in practice, decision makers should be (1) trained in evidence-based decision making (2) supported in designing evidence-based algorithms, and (3) encouraged to consult the academic literature on evidence-based decision making more regularly.

Holistic and Mechanical Combination in Psychological Assessment: Why Algorithms are Underutilized and What is Needed to Increase their Use

Mechanical (or statistical, actuarial, algorithmic) combination results in more valid judgments and decisions than holistic (or clinical, impressionistic, intuitive, informal) combination (Dawes et al., 1989; Grove et al., 2000; Kuncel et al., 2013; Meehl, 1954). Yet, results from (some relatively old) surveys suggest that mechanical combination is rarely used in psychological practice (Ryan et al., 2015; Ryan & Sackett, 1987; Vrieze & Grove, 2009), which is also referred to as "algorithm aversion" (Dietvorst et al., 2015, p. 114). Algorithm aversion is problematic[1], because it results in suboptimal and untransparent judgments and decisions, which hinders the evaluation and improvement of the decision process (Meijer et al., 2020). Therefore, researchers called for (qualitative) investigations of why decision makers underutilize algorithms (Burton et al., 2020; Dietvorst et al., 2015; Neumann, Niessen, & Meijer, 2021).

The aims of this article were to update existing surveys, to get a clearer picture of how judgments and decisions are made in practice, and, mainly, to investigate why decision makers underutilize algorithms. Underutilization was expected based on findings from existing surveys on the use of mechanical combination. Another aim was to identify factors that could promote the use of algorithms. To achieve these aims, we conducted two online surveys among decision makers involved in personnel selection (reported below in Study 1 and 2, respectively) and organized two focus groups with decision makers from fields such as Industrial-Organizational (I-O)-, educational-, and clinical psychology (Study 3). We organized focus groups with decision makers with different backgrounds since the superiority of mechanical over holistic methods and algorithm aversion has been observed in virtually all social science fields (Dawes et al., 1989; Grove et al., 2000; Kuncel et al., 2013). Yet, we

[1] Our aim was not to "accuse" decision makers who practice holistic combination but to stimulate relevant research on algorithm aversion, and to improve decision making in practice.

tailored our surveys to personnel selection because validity differences between holistic and mechanical combination are especially large in this field (Kuncel et al., 2013; Rynes et al., 2002) and because of the recent discussions and research on this topic in this area (Kuncel, 2018). This emphasis on selection is also reflected in the examples and literature we discuss below. Furthermore, designing a survey that would be comprehensible across fields seemed impractical. Our goal with these studies was to identify existing phenomena (Eronen & Bringmann, 2021) using recent insights of practitioners about practical problems, which is an essential, but often overlooked first step to theory development (Berkman & Wilson, 2021; Campbell & Wilmot, 2018; Ployhart & Bartunek, 2019).

Holistic and Mechanical Prediction

When the aim of combining information is prediction, holistic and mechanical combination is also referred to as holistic and mechanical prediction. In holistic prediction, information is combined in the mind, typically inconsistently, while employing an algorithm results in using the same information in the same way in all cases (Kuncel et al., 2013). An example of a simple algorithm is assigning equal weights (i.e., unit weights, Bobko et al., 2007) to the quantified predictors, and add up the resulting scores to decide, for example, which candidates to hire. Yet, weights can also differ and can be determined in other ways (for a taxonomy of combination methods, see Kuncel, 2018). For example, weights can be based on regression analysis of primary data, meta-analytic estimates, or even expert judgment (Kuncel, 2018). Alternatively, weights may be obtained by regressing decision-makers' holistic predictions on the predictors, which is also called “model of man” (Goldberg, 1970) or judgmental bootstrapping (Armstrong, 2001).

Although often argued by practitioners (Grove & Meehl, 1996), holistic and mechanical prediction cannot be used simultaneously for the same prediction. As Grove and Meehl (1996, p. 300) put it: “If an equation predicts that Jones will do well in dental school, and the dean's committee, looking at the same set of facts, predicts that Jones will do poorly, it would be absurd to say, “The methods don't compete, we use both of them.” One cannot decide both to admit and to reject the applicant; one is forced by the pragmatic context to do one or the other”. Yet, holistic and mechanical prediction can be used *sequentially* for the same prediction (Sawyer, 1966). In *clinical synthesis*, the decision maker receives the collected information (e.g., test scores and interview ratings) plus an algorithm’s prediction, which are then used to make a holistic prediction. In *mechanical synthesis*, the decision maker considers the collected information and makes a holistic prediction, which is subsequently added (with a fixed weight) to an algorithm and combined with all other information. Since the final prediction is made holistically in clinical synthesis, it classifies as holistic prediction. Correspondingly, mechanical synthesis classifies as mechanical prediction, because information is combined consistently according to an algorithm.

The Prevalence of Holistic and Mechanical Prediction in Practice

A couple of publications used anecdotal experiences to suggest that holistic combination is predominantly used in practice (Arkes, 2008; Grove & Meehl, 1996; Highhouse, 2008). Yet, to the best of our knowledge, there exist only a few surveys on the use of holistic and mechanical methods. In the context of personnel selection, Ryan and Sackett (1987) surveyed 163 members of the Society for Industrial and Organizational Psychology about their individual assessment practices and found that only a small minority (2.5%) used mechanical prediction. Other respondents reported combining information purely holistically (55.7%) or using clinical synthesis (41.8%). More recently, Ryan et al. (2015) surveyed 1197 human resource (HR) professionals from several countries on their test use and found that 43% indicated combining test scores and interview ratings in a standardized manner, although it is unclear how ‘standardized’ was defined or interpreted. In line with these findings, clinical psychologists and admission officers also seem to very rarely apply mechanical prediction (Conrad et al., 2016; Vrieze & Grove, 2009).

So, although there exist robust research findings on evidence-based selection and decision making (e.g., Kuncel et al., 2013), our knowledge on how information is *combined* in practice is scarce and outdated or primarily anecdotal (Grove & Meehl, 1996; Ryan & Sackett, 1987). This is surprising, given that surveys on what assessment instruments are used to *collect* information are abundant (Furnham, 2008; Jackson et al., 2018; König et al., 2010; Mann & Chowhan, 2011; Piotrowski & Armstrong, 2006; Risavy et al., 2019; Taylor et al., 2002; Zibarras & Woods, 2010).

Research question 1: How do decision makers combine information in practice?

Factors Related to Algorithm Aversion

Various factors related to algorithm aversion in selection have been investigated (for an overview, see Neumann, Niessen, & Meijer, 2021). A couple of studies based on self-determination theory found that decision makers were more likely to use a self-designed rather than a prescribed algorithm (Neumann, Niessen, Tendeiro, & Meijer, 2021; Nolan & Highhouse, 2014). Yet, Neumann, Niessen, Tendeiro, and Meijer (2021) found mixed evidence that predictions from a self-designed algorithm were more valid than holistic predictions. Relatedly, some studies showed that, compared to strictly using predictions from a prescribed algorithm, decision makers were more likely to use an algorithm when they could (restrictedly) adjust its predictions (clinical synthesis), which resulted in more valid predictions than pure holistic prediction (Dietvorst et al., 2018; Neumann, Niessen, Tendeiro, & Meijer, 2021).

Nolan et al. (2016) drew on attribution theory and found that using standardized hiring practices increased decision-makers' concerns about negative stakeholder perceptions, which increased their fear of losing professional status. This, in turn, decreased their intention to use standardized hiring practices. Some studies also investigated the effects of outcome feedback on performance predictions and mostly showed that outcome feedback *decreased* prediction accuracy (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019; Thiele et al., 2020). Moreover, some research showed that algorithms are considered less useful and more unprofessional, impersonal, and insufficient, compared to holistic prediction (Diab et al., 2011). Yet, another study showed that watching a short educational video on holistic and mechanical prediction increased algorithm use and hence prediction accuracy (Neumann, Hengeveld et al., 2021).

In a sample of HR professionals, Lodato et al. (2011) found that an experiential decision-making style (i.e., a preference for making judgments based on feelings) correlated positively and strongly with a preference for holistic prediction. In contrast, experience in HR management (in years), the possession of a certification as a senior HR professional (SPHR), and organization size were negatively and weakly correlated with a preference for holistic prediction. Furthermore, the frequency of reading work-related academic journals had a negligible but negative relationship with a preference for holistic prediction. Yet, Lodato et al. (2011) noted that the negligible relationship may have resulted from not specifying the term "academic journals", as participants may have believed that trade magazines would fall under this type of literature.

In summary, various factors that may be related to algorithm aversion have been identified, such as autonomy in decision making, outcome feedback, stakeholder perceptions, and individual differences, mostly using experimental study designs. Yet, the perspective of practitioners on algorithm aversion has been rarely entertained in earlier studies, although this is also important for identifying factors related to algorithm aversion (Ployhart & Bartunek, 2019).

Algorithm Appreciation

Compared to algorithm aversion, we know little about when and why decision makers appreciate algorithms (but see Logg et al., 2019). Research on assessment instrument choices showed that validity/effectiveness, efficiency, fairness, and face validity were the most important factors why organizations employ (online) tests in selection (Ryan et al., 2015). We expect that these factors also relate to (not) using algorithms. So, given that practitioner-insights regarding algorithm aversion and appreciation are largely lacking, we had the following research question:

Research question 2: Why do decision makers (not) use algorithms in practice?

Increasing Algorithm Use in Practice

Based on existing surveys, we expected that algorithms are underutilized. Therefore, we conducted focus groups to explore what decision makers need to apply algorithms in practice more often.

Research question 3: What do decision makers need to apply algorithms more often?

Aim and Overview of the Present Studies

We conducted two online surveys and organized two focus groups to investigate (1) how information is combined in practice, (2) why algorithms are (not) used, and (3) what is needed to apply algorithms more often. Since Lodato et al. (2011) found that experience, reading the academic literature, and possessing a certification were weakly related to *holistic* prediction, we also aimed to investigate if these characteristics relate to algorithm use. We expected that obtaining a license and reading up on the superiority of algorithms in the academic literature should increase knowledge and hence algorithm use (Neumann, Hengeveld et al., 2021). Moreover, we also explored the relation between experience and algorithm use since some evidence suggests that it is negatively related to algorithm use and prediction accuracy (Arkes et al., 1986; Logg et al., 2019), likely due to overconfidence of experienced decision makers (Arkes et al., 1986; Dawes, 1994). We measured experience in making hiring decisions in years and amount of decisions made because these measures were inconsistently used in research on the adoption of structured interviews, with sometimes conflicting results (Lievens & De Paepe, 2004; Roulin et al., 2019).

The preregistrations, raw data, codebooks, R scripts, and study materials for all studies are available on https://osf.io/v3thf/?view_only=db9ed843dc3b43e6877c37341204dbb5.

Study 1

Participants

We obtained a convenience sample via LinkedIn, which is typical for survey research among practitioners involved in selection decisions (Jackson et al., 2018; Ryan et al., 2015). The first author distributed the online questionnaire in a German organization's private LinkedIn group whose mission is to bridge the gap between science and practice, and whose members are interested in evidence-based assessment and personnel selection. Furthermore, we employed snowball sampling by asking group members to share the survey link with other eligible decision makers in their network.

We only included participants who reported that they were involved in at least five hiring decisions in the last two years, through making (part of) the hiring decision or providing consultation to others who made the decision. One participant who indicated to be younger than 18 was excluded. In total, we obtained usable data from *N* = 93[2] participants

[2] We had pre-registered to collect data from minimally *N* = 171 participants who report not always using mechanical prediction, to achieve a desired margin of error for observed proportions. Furthermore, the results

(54% female) who ranged in age from 25 to 60 (M = 38.2, SD = 9.4). The sample was primarily German (95%) and 54% of the participants were part of the organization's private LinkedIn group. The other participants had other European nationalities. Participants' mean organizational tenure was 6.5 (SD = 7.9) and role tenure was 4.1 (SD = 5.1). The mean number of years participants were involved in making hiring decisions was 9.8 (SD = 7.6). Other demographic information is displayed in Table 1. The median time it took to complete the survey was 13 minutes.

Procedure

After consenting to take part in the study and reporting their experience, participants indicated whether they use more than one source (i.e., assessment instrument) to obtain information about applicants. Participants who indicated using only one source were asked why, and what source they use. Participants who indicated using multiple sources reported how frequently (1 = *never*, 5 = *always*) they use several methods to combine information (see Table 2), and which method they use most often.

We asked participants who indicated using multiple sources whether they quantify information obtained from *each* source, because mechanical prediction requires quantified information. Participants who indicated using algorithms to some extent (i.e., more often than "never") selected pre-listed approaches of how they construct their algorithm. The list was composed of approaches to data combination presented in Kuncel's (2018) taxonomy[3] and other approaches that we considered relevant.

from an a priori power analysis showed that, for a regression model given five predictors, α = .05, 1-β = .80, and R^2 = .09, N = 136 participants would be needed. After a prolonged time of no additional survey completions, we had to stop the data collection. A post-hoc power analysis showed that power was .67 (given the observed effect size $R^2_{adj.}$ = .10, α = .05, and N = 92; one participant reported only using one source to obtain information about applicants).

[3] We did not include a combination method based on bootstrapped weights because we expected that participants would not understand this answer option without adding an extensive explanation that would have substantially lengthened the survey.

Furthermore, we asked participants to select pre-listed reasons why they do not always use an algorithm. This list was derived from theories that have been tested in the context of algorithm aversion in selection (Neumann, Niessen, & Meijer, 2021), reasons that have been mentioned in position papers (Arkes, 2008; Grove & Meehl, 1996; Highhouse, 2008), and other reasons that we encountered during earlier conversations with decision makers. Participants could also report additional reasons in a text field via an “other” option. Participants who reported always using algorithms were asked to select pre-listed reasons why they do so. This list was based on prior research on factors that explain the use of data collection methods (König et al., 2010; Ryan et al., 2015), but that have not yet been related to data combination methods. Moreover, all participants indicated how frequently (1 = *never or rarely*, 5 = *almost daily*) they consult several sources (e.g., other professionals, academic- or professional literature) to obtain information about how to best make hiring decisions, and whether they possess a license for assessment professionals (German DIN 33430).

In the final part of the survey, we explained the distinction between holistic and mechanical prediction to participants. Then, we presented three fictitious hiring scenarios. For each scenario, the participants indicated whether information was combined holistically, mechanically, or whether they could not tell without further information. These “knowledge check” items were administered to investigate whether participants understood the distinction between holistic and mechanical prediction. The items were administered at the end of the survey to avoid potentially socially desirable answers on questions regarding how decision makers typically combine information.

Results

All but one participant reported that they use multiple sources to obtain information about applicants. Table 2 shows the mean frequency ratings and the number of participants who indicated using a given method most often, for each combination method. We collapsed these methods into the two holistic and mechanical meta-categories by averaging participants' frequency responses to combination method 1, 2, 3, 5, and 7, and 4, 6, and 8, respectively (see Table 2 for the numbering). This showed that holistic methods are by far most often used (87%). Similarly, the mean frequency rating for holistic methods was much higher (M = 2.38, SD = 0.58) than for mechanical methods (M = 1.51, SD = 0.76, d = 1.29, 95% CI [0.89, 1.70]). Furthermore, from the participants who reported using some form of holistic prediction most often (n = 80), 55% reported making holistic predictions in a team discussion. In contrast, 26% reported making holistic predictions on their own after considering the importance of the available information. Moreover, 14% reported using clinical synthesis (combination method 5 and 7), mostly in the form of multi-stage selection. That is, first selecting applicants who pass pre-determined cutoffs and then selecting among the final candidates based on one's own judgment. Among participants who used some form of mechanical prediction most often (n = 12), 75% reported that they first select applicants who pass pre-determined cutoffs and then use a pre-determined algorithm to make final hiring decisions.

Furthermore, most participants (70%) indicated that they do not quantify all obtained information. We also asked participants who reported not always using some form of mechanical prediction (n = 86) why (see Table 3). The most frequently reported reason was that participants cannot or do not want to quantify all information (35%). Other frequently reported reasons were that there are no algorithms available (30%), the use of algorithms is not accepted by stakeholders (e.g., supervisors and colleagues, 24%), and the belief that using an algorithm together with one's own judgment would result in the most valid decisions (21%). Moreover, 14% indicated not knowing that mechanical prediction is more valid than holistic prediction. Less frequently mentioned reasons were that using algorithms would reduce one's perceived competence (5%) and professional status (3%). Moreover, none of the participants were concerned that using an algorithm would be against the law. Other mentioned reasons not pre-listed in the survey were that other decision makers can veto decisions in a group discussion (*n* = 4), and that algorithms, or selection in general, would be unnecessary due to a lack of applicants (*n* = 5). To investigate when decision makers appreciate algorithms, we asked those who always use an algorithm why. Yet, only six participants reported always using an algorithm. Therefore, no specific results on that subsample are reported.

Participants who indicated using some form of mechanical prediction at all (*n* = 50) were also asked how they construct their algorithm (see Table 4). The most common way was by means of discussion with experts/professionals (42%). Other common strategies were to design an algorithm based on scientific research (meta-analyses or primary scientific studies, 34%) or based on one's own knowledge and expertise, without consulting the scientific literature or others (32%). Some participants also indicated that they use an algorithm as prescribed by others (e.g., by their organization or professional standards, 28%) or that they base their algorithm on statistical analyses of data from their organization, or other relevant data (24%). Unit weighting (i.e., weighting all information evenly) was the least common way (10%).

Information Sources

Participants reported that they most frequently obtain information on how to best make selection decisions by consulting other HR professionals (see Table 5). Similarly, blogs, videos, websites, and popular magazines such as Harvard Business Review were frequently consulted. Compared to these sources, participants consulted the academic literature (e.g., Journal of Applied Psychology) less frequently. Moreover, scientists and external consultants were rarely consulted.

Knowledge Check

After reading the explanation of holistic and mechanical prediction, 32% out of all participants who obtained a college degree (*n* = 87) indicated that holistic and mechanical prediction was discussed in their studies. Afterwards, we asked participants to indicate for three fictitious hiring scenarios whether the decision maker combined information holistically, mechanically, or whether they cannot tell without further information. Out of all participants, 43% answered all three “knowledge check” items correctly, while 77% answered at least two items correctly. At least one item was answered correctly by 95% of all participants.

Individual Characteristics

As expected based on Lodato's et al. (2011) results, we also found that reading the academic literature and possessing an assessment license were positively and weakly to moderately related to using mechanical prediction methods (see Supplement S1 in the supplementary material). In contrast, experience in making hiring decisions and organization size showed negligible relationships with using mechanical methods. Results from relative importance analyses (Grömping, 2006) showed that reading the academic literature and possessing an assessment license were the most important predictors.

Study 2

The aim of the second study was to replicate the results from Study 1 using a larger sample that was collected via Amazon Mechanical Turk. We displayed the different combination methods in random order, to prevent order effects. Furthermore, in Study 1, we only asked participants who reported always using algorithms to select reasons why, which resulted in a subsample that was too small for meaningful interpretation. Therefore, in Study 2, all participants who indicated using algorithms at all selected reasons why. Since we expected participants to be mostly U.S. citizens, we slightly revised the sources that participants may consult to obtain information about how to best make hiring decisions. Furthermore, we asked participants if they possess a SHRM license instead of the German

DIN 33430 license. Except for these changes and additional exclusion criteria presented below, the survey was the same as in Study 1.

Participants

The study was introduced as a general study on decision making, to avoid that participants guessed the exclusion criteria. After providing informed consent, participants indicated up to three (from a list of twelve, presented in random order) work activities that they most frequently engage in at work (see Supplement S2). Only participants who selected “staffing organizational units”, “making decisions and solving problems”, or “judging the qualities of things, services, and people” were included. As in Study 1, we excluded participants who reported that they were involved in less than five hiring decisions in the last two years, meaning making (part of) the hiring decision or providing consultation to others who made the decision. Furthermore, participants who failed at least one of two attention checks (see Supplement S3) were excluded. In addition to these pre-registered exclusion criteria, we excluded participants who indicated being younger than 18, or who gave impossible responses to some questions (e.g., role tenure > organizational tenure). We retained usable data from N = 230 participants (55% male) who ranged in age from 21 to 67 (M = 36.4, SD = 10.3). The sample consisted primarily of U.S. citizens (89%). Other participants had other non-European (10%) or European nationalities (< 0.01%). Participants’ mean organizational tenure was 6.9 years (SD = 4.3) and role tenure was 4.5 (SD = 2.8). The mean number of years participants were involved in making hiring decisions was 5.6 (SD = 4.7). On average, participants had made 42.1 (SD = 103.0) hiring decisions in their life. Other demographic information is displayed in Table 6. The median time it took to complete the survey was 12 minutes.

Results

Most of the results found in Study 1 were similar in Study 2. Yet, we also found some differences.

Similarities

All but two participants reported using multiple sources to obtain information about applicants. Also, holistic methods were used most often (82%, see Table 7 for mean frequency ratings per method and the number of participants who indicated using a given method most often). Furthermore, the results from participants who reported using some form of holistic prediction most often (n = 188) showed that making holistic predictions as a team was prevalent (41%), and somewhat more prevalent than making holistic predictions individually (36%). Moreover, 20% reported using clinical synthesis.

Although the reasons for not using algorithms (see Table 8) were also similar, relatively more participants indicated that they feel their status (23%), autonomy (23%), and personal contact with other decision makers (21%) is reduced when using an algorithm. Relatively less participants than in Study 1 indicated that they do not use algorithms because algorithms are unavailable (16%) or because they cannot or do not want to quantify information (18%).

Participants indicated that they frequently consult blogs, videos, websites, other (HR) professionals, and the professional (HR) literature to obtain information on how to best make selection decisions (see Table 9). They consulted the academic literature somewhat less frequently and rarely consulted scientists or external consultants. Moreover, reading the academic literature and possessing an assessment license were the most important predictors in explaining the use of mechanical prediction methods (see Supplement S4).

Differences

In contrast to Study 1, only a small minority indicated that they do not quantify all information (11% versus 70% in Study 1). Furthermore, the mean frequency rating for holistic methods (M = 3.43, SD = 0.59) was similar to the mean frequency rating for mechanical methods (M = 3.37, SD = 0.83, d = 0.09, 95% CI [-0.03, 0.20]), whereas in Study 1 the mean frequency for holistic methods was much higher than for mechanical methods. Moreover, in Study 2, many more participants who obtained a college degree (n = 216) indicated that holistic and mechanical prediction was discussed in their studies (90% versus 32 % in Study 1). Yet, only 7% (versus 43% in Study 1) out of all participants answered all three “knowledge check” items correctly, while 42% (versus 77%) answered at least two items correctly. Out of all participants, 70% (versus 95%) answered at least one item correctly[4].

Participants who indicated that they at least rarely use some form of mechanical prediction (n = 221) were also asked how they construcct their algorithm (see Table 10). In contrast to Study 1, the most common way was to construct the algorithm based on statistical analyses of data from one’s organization, or other relevant data (54%). Furthermore, unit weighting was common (44%), and relatively more common than in Study 1. Other ways of determining algorithms were similarly often reported as in Study 1.

Algorithm Appreciation

We also asked participants who indicated using algorithms at least to some extent (n = 219) why (see Table 11). Commonly reported reasons were that algorithms are more valid (45%), easier to use (37%), yield more valuable information (34%), and are fairer (32%) than holistic prediction. Relatively uncommon reasons were that algorithms are legally safer (22%) and reinforce the employer brand more effectively (18%) than holistic prediction. Only four participants indicated that algorithms would be cheaper than holistic prediction (2%).

Discussion Study 1 and Study 2

[4] As a robustness check, we also inspected the results when only considering participants who answered at least two items correctly. The results of this subsample (see Supplement S5) were closely aligned with the results of the full sample.

The results from Study 1 and 2 showed that most selection decisions are made holistically, and usually in teams. Frequently mentioned reasons why algorithms are not (always) used were that stakeholders do not accept the use of algorithms and that decision makers believe a combination of holistic and mechanical prediction results in the best decisions. Other reasons in Study 1 were that decision makers cannot or do not want to quantify information, and that algorithms are unavailable. Other reasons in Study 2 were that decision makers consider it their professional duty to use holistic prediction and feel that their status and autonomy is reduced when using algorithms. Decision makers who use some form of mechanical prediction mainly do so because of its validity advantages over holistic prediction, and they often base their algorithms on scientific research findings, a discussion with experts (Study 1), and unit weights (Study 2). Despite presenting participants with an elaborate explanation on holistic and mechanical prediction, only a (small) minority answered all knowledge items correctly, suggesting that many misunderstandings regarding evidence-based decision making still exist.

Study 3

The aim of Study 3 was to supplement the quantitative surveys with qualitative data from focus groups, and to investigate what is needed to increase algorithm use in practice.

Method

Participants and Procedure

Participants were members of the Dutch Association of Psychology (Dutch: Nederlands Instituut voor Psychologen, NIP) who indicated their willingness to participate in future research on test use and decision making in an earlier survey on psychological testing, conducted in collaboration with the European Federation of Psychologists' Associations. We invited these members via email to take part in one of three online focus group sessions on decision making in practice. Due to a technical error, the third focus group session was not recorded. Therefore, we only report the results of the first two sessions. Decision makers (mostly psychologists) with backgrounds in I-O, clinical, developmental, forensic, and educational psychology participated in the focus groups. There were nine and 12 participants in focus group 1 and 2, respectively. We conducted focus groups (instead of individual interviews) because we wanted participants to build upon each other’s responses, to obtain a more accurate picture of the relevance of certain reasons why algorithms are underutilized. Participants were asked to read an elaborate explanation on the distinction between holistic and mechanical prediction before signing up for a focus group session. They received a presentation on the distinction again at the beginning of the session. After the presentation, participants responded to the following two questions: (1) “What is the reason you think decision makers do not (or seldom) apply mechanical methods in decision making?” and (2) “What is needed to apply mechanical methods more often?”

We applied thematic analysis to analyze our data (Braun & Clarke, 2006). First, one of the authors transcribed the audio material of the two focus group sessions. Then, the authors worked in two pairs. First, each member of each pair read the text from both focus groups and discussed which parts of the text should be classified as a unit with their partner. Differences in unit classification were resolved by discussion among the authors. Seven themes were suggested based on inspection of the different units (see Table 12). Second, two newly formed author pairs assigned the units to the themes. One pair coded the first focus group and the other pair coded the second focus group. The pairs could choose a second theme when in doubt which theme to choose. Differences in assigned units to themes were resolved by discussion within a pair of authors. Eventually, each pair received all codes (see OSF) and each pair independently wrote a results section. The first author integrated the two results sections as presented in the paper. Per theme, we present an exemplary quote (see Table 13 for the full list of quotes).

Results

Table 12 shows that most units were assigned to primary theme 6 (questions about how to use mechanical methods), 2 (disadvantages of mechanical methods), 7 (reasons why decision makers do not use mechanical methods), and 5.1 (what is needed to increase the use of mechanical methods: information and guidelines). When also considering secondary themes, most units were assigned to theme 7.

Advantages of mechanical methods (theme 1)

The most often mentioned advantage of mechanical methods was transparency. Participants largely agreed that the use of an explicit algorithm would provide insight into how decisions are made (i.e., how information is weighted). Many participants also indicated that mechanical prediction would elicit critical thinking and raise awareness about the validity of predictors. They also mentioned that algorithms constructed based on the available literature would serve as an "anchor", heuristic, or decision aid. Decision makers would be confronted with the algorithm and could learn from it, but also adjust it to their needs, with an argumentation why they would deviate from the default algorithm. Interestingly, several participants who were unfamiliar with mechanical methods were open to use algorithms or to make any implicitly used algorithms more explicit. The general trend was that awareness of how decisions are made was a big advantage of mechanical prediction.

"*I do not know whether you have to impose it on people or something like that. For me it is more about creating awareness. That's where a decision rule can help. This would be the added value for me.*"

Disadvantages of mechanical methods (theme 2)

One mentioned drawback of the mechanical method was that the criterion of interest changes constantly. In personnel selection, for example, it would be impossible to construct an algorithm for each position. That would be inefficient if a position was only filled once or twice a year. Another mentioned drawback was that it is difficult to choose predictor weights. Relatedly, it was mentioned that the mechanical method requires more resources (e.g., time) than the holistic method, which are rarely available for decision makers. Decision makers would not have the time and motivation to consult the literature for predictor validities and weights. Furthermore, it was mentioned that the mechanical method results in a point estimate, but that uncertainty around this estimate would need to be considered. It was also mentioned that some existing algorithms (e.g., in recidivism risk assessment) include controversial predictors such as a person's zip code.

"*I think that decision rules in I-O psychology, of which I am part of, are rarely used because the criterion that you select on changes frequently. And you cannot design a decision rule for a specific function in an organization because such a selection takes place only one or two times a year, maybe even less often. Maybe that is different in clinical psychology. Diagnosis with the DSM. There you have to make a diagnosis because you have to give an indication about a therapy. But it heavily differs per criterion. How often does something occur in an organization? And I think, if you have designed a decision rule, you actually also have to evaluate it. You also have to see whether it actually works.*"

Advantages of holistic methods (theme 3)

Only one statement was related to the advantages of the holistic method. It was mentioned that holistic prediction can (also) be made fully transparent.

"*But then I do not really understand what the difference is between the holistic and mechanical method, because you can also make a holistic prediction fully transparent. You can exactly tell how you do it.*"

Disadvantages of holistic methods (theme 4)

Participants mentioned that research showed that decision makers are bad at combining information holistically, and it was mentioned that this seems to be due to inaccurately and inconsistently weighting information.

"*But it is also true that extensive research has been done into how people integrate different data sources. That is a typical thing that no one is really good at. The same holds for psychologists. They are very bad at that. This was also shown in the research. It also showed that our psychologists put the most weight on the things they had seen themselves. They thought the cognitive ability test was less important.*"

Questions about how to use mechanical methods (theme 6)

A first observation was that for many participants it was unclear what exactly an algorithm was and how to use it, although we had asked participants to read our explanation of holistic and mechanical methods before the session. One participant asked whether algorithms would need to be communicated to everyone involved in the decision-making process, or whether you could also construct individual algorithms. Relatedly, it was asked whether algorithms necessarily have to be evidence-based. Some participants found it remarkable that you could also choose predictor weights yourself. Furthermore, it was asked whether the data combination method would still matter if valid, evidence-based predictors are used. Moreover, some participants were confused about whether the scoring of a test can also be distinguished with regard to the holistic and mechanical method. Lastly, it was asked what to discuss in group discussions when one wants to implement the mechanical method.

"*May I ask something? You just said that it is essential that you quantify things, but that does not necessarily have to be evidence-based?*"

Reasons why decision makers do not use mechanical methods (theme 7)

Participants mentioned different reasons why they do not use algorithms. (1) Holistic and mechanical prediction is, according to some, not sufficiently taught at the university. For many participants, the distinction between holistic and mechanical prediction was unclear, and there were many misunderstandings. (2) Participants thought that the lack of valid instruments in practice would prevent mechanical prediction as they believed that it requires well-measured variables. (3) Participants thought the situations in which they have to make decisions are too complex to apply algorithms. It was mentioned that algorithms ignore the individual, and that information should be differently weighted in different situations. It was also mentioned that it would be the role of the psychologist to identify situations in which information should be weighted differently. Relatedly, (4) participants mentioned that psychologists are "stubborn" and (think) they have a lot of knowledge and can predict events better than an algorithm by using their gut feeling. (5) Stakeholders in the field (e.g., colleagues) who may also be involved in the process do not accept the use of algorithms. Relatedly, some participants suggested that organizational culture could foster or impede the application of algorithms. Organizations with a stronger focus on quantitative than on qualitative information would appreciate algorithms more. (6) Information on how to construct algorithms is not easily available. Furthermore, it would be too time-consuming to find and read the literature on how to construct algorithms. Lastly, (7) algorithms, and in particular the use of cutoff scores, would create an illusion of certainty. Things either do or do not happen. Many participants felt that this is not how the world works.

“*But I also think that it is rarely used because it is not really accepted in the field.*”

What is needed to increase the use of mechanical methods? (theme 5)

We also asked participants what is needed to increase the use of mechanical methods among decision makers. We distinguished three different subthemes (see Table 12).

Providing information and guidelines (theme 5.1)

Participants primarily mentioned the need for discussing and explaining mechanical prediction in guidelines in more detail. Furthermore, it was mentioned that specific training in the application of an algorithm would be very helpful. Moreover, participants desired more information from test commissions on test reliability and validity, which would make it easier to determine predictor weights. It was also mentioned that access to information on algorithms would need to be improved. Ideally, participants would like to have access to a database in which possible algorithms per sub-discipline would be available. Lastly, some members mentioned that evaluating the validity of algorithms and comparing it to current holistic predictions in practice would stimulate its use.

“*If I look at myself, what is needed is training in applying such a decision rule. I would like to see how I could approach such a rule in a manner to avoid bias, or to use more and more the same hobbyhorse*.”

Where should information about mechanical methods be exchanged? (theme 5.2)

Some participants argued that there should be more attention to the use of algorithms in their regular professional training and education (e.g., in the bachelor, master, and post-master in Psychology), and also specific courses should be devoted to mechanical prediction. It was also suggested to use a bottom-up procedure, where for example within- or across companies or institutes, decision makers first discuss how they decide in particular cases, and then based on this information suggest specific algorithms.

"*And this is perhaps something that can be gradually developed, and you can build a community that says these rule work in our case. That you upload this then in a system and that others can also make use of it again.*"

Who should be responsible for encouraging the use of mechanical methods? (theme 5.3)

Participants largely agreed that universities are among others in charge to accommodate decision making more in their curriculum. Furthermore, testing commissions and other diagnostics institutes were expected to cover decision making in more detail. Lastly, in the field of personnel selection, some participants argued that consultancies should collect data on applicants and later performance that may be used to design algorithms.

"*Maybe the big HR consultancies also have to collaborate. I have worked for a big HR consultancy where thousands of applicants are assessed per year. And all this data has been saved. You need data on which the decision rule is based. And I find that we throw away lots of data. The applicants should be followed to see whether you got it right. You have to follow the applicants then, also those we reject, to see whether they end up somewhere where they do a really good job and what we may have done wrong with the decision rule. You have to get organizations so far that they collect data. Back in the days, we did that and then we also* *sometimes adjusted the weighting. We never looked how they scored once they were on the job. The organization can do a lot about it.*"

General Discussion

In line with earlier research (Ryan et al., 2015; Ryan & Sackett, 1987), we consistently found that most decisions are made holistically in practice, usually in teams. Across all three studies, we found that decision makers rarely use algorithms because they lack knowledge and fear negative evaluations from stakeholders when using algorithms. Moreover, participants in Study 1 indicated that they cannot or do not want to quantify information. Also, participants mentioned that algorithms (Study 1) and information to construct them (Study 3) are often unavailable. Participants in Study 2 also reported that their autonomy and status is reduced when using algorithms, and that they consider it their duty to use holistic prediction (also mentioned in Study 3). Furthermore, Study 1 and 2 showed that reading the academic literature and being licensed were positively related to algorithm use. Decision makers who use algorithms do so mainly because of its validity advantages over holistic prediction. The results from Study 1 and 2 suggest that algorithms are often based on research findings (e.g., meta-analyses). While participants in Study 1 indicated that algorithms are often based on a discussion with experts and rarely on unit weights, participants in Study 2 indicated that algorithms are primarily based on data from their organization and unit weights. Lastly, our results from Study 3 suggest that more training and support in using algorithms is needed to increase its use in practice. Furthermore, licensed decision makers should be re-evaluated on evidence-based decision making to retain their license.

Although we supplemented two quantitative surveys with qualitative research to achieve methodological triangulation (Campbell & Fiske, 1959), our studies have several limitations. We used a small convenience sample in Study 1. Therefore, the results may not generalize to other decision makers involved in selection decisions. Despite a larger sample in Study 2, concerns exist regarding the quality of the data, as less participants answered all knowledge check items correctly compared to Study 1. It could be that participants in Study 1 were more intrinsically motivated to learn about evidence-based decision making and hence read the explanation on mechanical and holistic methods more carefully. Yet, in Study 2, the results from a subsample of participants who answered at least two knowledge check items correctly were very similar to the results from the full sample, and most results were in line with the results found in the other studies. Still, given the concerns regarding data quality in Study 2, we primarily interpreted the results from Study 1 and 3. Moreover, we cannot rule out that some participants mistakenly reported using mechanical prediction, given the many misunderstandings we observed. Lastly, due to our cross-sectional, correlational designs, we cannot infer causality.

Our results have several implications for advancing research on algorithm aversion and improving decision making in practice. First, we found that decisions are often made holistically in teams. Yet, existing research on algorithm aversion has mostly focused on individual decision makers (Neumann, Niessen, & Meijer, 2021). Therefore, research on the role of mechanical methods in team decision making would be valuable. Future research may investigate whether algorithms designed in a team are more valid than individually designed algorithms, and to what extent individuals deviate from predictions that stem from collectively vs. individually designed algorithms.

Second, given that knowledge on evidence-based decision making seems to increase algorithm use and predictive accuracy (Neumann, Hengeveld, et al., 2021), decision making should receive much more attention in licenses and the education of practitioners, for example, in on-going organizational trainings. Also, textbooks on psychological and educational testing and test guidelines such as those of the APA/NCME and of the ITC currently do not discuss the benefits of mechanical methods (Meijer et al., 2022). In our opinion, this should change. However, this may not be enough. In line with the broader literature on the science-practice gap (Rynes et al., 2002; Sanders et al., 2008), our results showed that decision makers more often consult the professional than the academic literature to obtain information on how to best make hiring decisions. Yet, the professional literature contains much misinformation, and evidence-based decision making is hardly discussed (Neumann, Niessen, & Meijer, 2021). Therefore, to bring the (I-O) science to the public (Rogelberg et al., 2021; Rynes et al., 2018), information on evidence-based decision making should be disseminated in other outlets next to academic journals, such as trade books, professional journals, social media platforms (e.g., LinkedIn), and podcasts that practitioners frequently consume. Additionally, more active steps should be taken. For example, a (virtual) drop-in center may be established where practitioners can ask experts questions regarding the implementation of algorithms. The resulting input from practitioners may also meaningfully inform theory development on algorithm aversion.

Third, given that many participants mentioned not using mechanical methods because algorithms are unavailable suggests that decision makers think designing algorithms must be complex, as was already noted by Grove and Meehl (1996). Yet, research has shown that when predictors do not substantially differ in validity, simple algorithms with unit (Sackett et al., 2017) or even random weights (with the correct sign, Dawes, 1979; Yu & Kuncel, 2020) outperform holistic prediction. To increase the availability of algorithms and relevant information, testing commissions and consultancies may create decision aids that document evidence-based algorithms for different predictor sets, or support decision makers in developing their own algorithms, such as the ShinyApps by Failenschmid et al. (2021) and Song et al. (2017).

Fourth, we consistently found that decision makers fear negative evaluations from stakeholders when using algorithms, which is in line with existing research (Nolan et al., 2016). Therefore, future research on improving stakeholder perceptions and decision-makers' beliefs about stakeholder perceptions seems valuable. Furthermore, the results from Study 1 and 2 showed that many participants believed a combination of holistic and mechanical methods to result in the best decisions. We found that decision makers mostly combined these two methods in the form of clinical rather than mechanical synthesis. This is in line with research showing that decision makers use an algorithm more often if they can (vs. cannot) adjust an algorithm's prediction (clinical synthesis, Dietvorst et al., 2018). Yet, it remains unknown how mechanical synthesis compares to clinical synthesis in terms of usage and predictive validity. Therefore, future research may compare these two methods.

Conclusion

Although mechanical methods are superior to holistic methods, they are still rarely used, suggesting that there is much room for improving decision making in practice. Many reasons for this underutilization relate to a lack of knowledge of evidence-based decision making. Therefore, proper and elaborate education and providing support in evidence-based decision making seems to be the first step.

References

Arkes, H. R. (2008). Being an advocate for linear models of judgment is not an easy life. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes.* (pp. 47–70). Psychology Press.

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

Armstrong, J. S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 171–192). Springer. https://doi.org/https://doi.org/10.1007/978-0-306-47630-3_9

Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. *Perspectives on Psychological Science*, *16*(4), 864–874. https://doi.org/10.1177/1745691620969650

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, *10*(4), 689–709. https://doi.org/10.1177/1094428106294734

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

https://doi.org/10.1037/h0046016

Campbell, J. P., & Wilmot, M. P. (2018). The functioning of theory in industrial, work and organizational psychology (IWOP). 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., Vol. 1, 2nd ed.* (pp. 3–38). Sage Reference.

Conrad, S. S., Addams, A. N., & Young, G. H. (2016). Holistic review in medical school admissions and selection: A strategic, mission-driven response to shifting societal needs. *Academic Medicine*, *91*(11), 1472–1474. https://journals.lww.com/academicmedicine/Fulltext/2016/11000/Holistic_Review_in_Medical_School_Admissions_and.9.aspx

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

Dawes, R. M. (1994). Prediction and diagnosis: More myths of expertise. In *House of Cards* (pp. 75–105).

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

Diab, D. L., Pui, S., 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

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science, 16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Failenschmid, J. I., Neumann, M., Meijer, R. R., & Niessen, A. S. M. (2021). *The predictive validity of weighted combinations of predictors and criteria [Shiny App]*. https://failenschmid-neumann-meijer-niessen-2021.shinyapps.io/publication/

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

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, *73*(6), 422–432. https://doi.org/10.1037/h0029230

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, *17*, 1–27. https://doi.org/10.18637/jss.v017.i01

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The Clinical-Statistical Controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293

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: Perspectives on Science and Practice*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Jackson, A. T., Young, M. E., Howes, S. S., Knight, P. A., & Reichin, S. L. (2019). Examining factors influencing use of a decision aid in personnel selection. *Personnel Assessment and Decisions*, *5*(1), 1–36. https://doi.org/10.25035/pad.2019.01.001

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

König, C. J., Klehe, U., 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/http://dx.doi.org/10.1111/j.1468-2389.2010.00485.x

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* (2nd ed., pp. 474–487). SAGE Publications Ltd. https://doi.org/10.4135/9781473914940

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

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*, *25*(1), 29–46. https://doi.org/10.1002/job.246

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

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Mann, S. L., & Chowhan, J. (2011). Selection practices in canadian firms: An empirical investigation. *International Journal of Selection and Assessment*, *19*(4), 435–437. https://doi.org/10.1111/j.1468-2389.2011.00571.x

Meehl, P. E. (1954). Empirical comparisons of clinical and actuarial prediction. In P. E. Meehl (Ed.), *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence* (pp. 83–128). University of Minnesota Press. https://doi.org/doi:10.1037/11281-008

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*, *10*, Article 3002. https://doi.org/10.3389/fpsyg.2019.03002

Meijer, R. R., Niessen, A. S. M., & Neumann, M. (2022). Psychological and educational testing and decision making: The lack of knowledge dissemination in textbooks and test guidelines. In L. A. van de Ark, W. H. M. Emons, & R. R. Meijer (Eds.), *Practical measurement: Essays on contemporary psychometrics*.

Neumann, M., Hengeveld, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2021). Education increases decision-rule use: An investigation of education and incentives to improve decision making. *Journal of Experimental Psychology: Applied*. Advance online publication. https://doi.org/10.1037/xap0000372

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. (2021). The autonomy-validity dilemma in mechanical prediction procedures: The quest for a compromise. *Journal of Behavioral Decision Making*. Advance online publication. https://doi.org/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

Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of fortune 1000 firms. *North American Journal of Psychology*, *8*(3), 489–496.

Ployhart, R. E., & Bartunek, J. M. (2019). Editors' comments: There is nothing so theoretical as good practice — A call for phenomenal theory. *Academy of Management Review*, *44*, 493–497. https://doi.org/10.5465/amr.2019.0087

Risavy, S. D., Fisher, P. A., Robie, C. R., & König, C. J. (2019). Selection tool use: A focus on personality testing in Canada, the United States, and Germany. *Personnel Assessment and Decisions*, *5*(1), 62–72. https://doi.org/10.25035/pad.2019.01.004

Rogelberg, S. G., King, E. B., & Alonso, A. (2021). How we can bring I-O psychology science and evidence-based practices to the public. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *15*(2).

Roulin, N., Bourdage, J. S., & Wingate, T. G. (2019). Who is conducting "better" employment interviews? Antecedents of structured interview components use. *Personnel Assessment and Decisions*, *5*(1), 37–48.

https://doi.org/10.25035/pad.2019.01.002

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

Ryan, A. M., & Sackett, P. R. (1987). A survey of individual assessment practices by I/O psychologists. *Personnel Psychology*, *40*(3), 455–488. https://doi.org/http://dx.doi.org/10.1111/j.1744-6570.1987.tb00610.x

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., Colbert, A. E., O'Boyle, E. H., & O'Boyle, E. H. (2018). When the "best available evidence" doesn't win: How doubts about science and scientists threaten the future of evidence-based management. *Journal of Management*, *44*(8), 2995–3010. https://doi.org/10.1177/0149206318796934

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

Sanders, K., van Riemsdijk, M., & Groen, B. (2008). The gap between research and practice: A replication study on the HR professionals' beliefs about effective human resource practices. *International Journal of Human Resource Management*, *19*(10), 1976–1988. https://doi.org/10.1080/09585190802324304

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

Song, Q. C., Wee, S., & Newman, D. A. (2017). Diversity shrinkage: Cross-validating Paretooptimal weights to enhance diversity via hiring practices. *Journal of Applied Psychology, 102*(12), 1636–1657. https://doi.org/10.1037/apl0000240

Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organisations and recruitment firms. *New Zealand Journal of Psychology*, *31*(1), 8–18. https://link.gale.com/apps/doc/A90304040/ITOF?u=groning&sid=ITOF&xid=8fcc27a0

Thiele, A., Jackson, A. T., Stremic, S. M., & Howes, S. S. (2020). Does feedback increase decision aid use among hiring professionals? *Personnel Assessment and Decisions*, *6*(2), 20–30.

Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and Practice*, *40*(5), 525–531. https://doi.org/10.1037/a0014693

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), 1–10. https://doi.org/https://scholarworks.bgsu.edu/pad/vol6/iss2/2

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/10.1348/096317909X425203

Table 1

Frequencies of Sample Demographics in Study 1

Demographic Information	Frequency
Employment status	
Employed	81
Self-employed	12
Unemployed	0
Hours of work	
Full-time (35 hours or more)	81
Part-time	12
Education	
Lower than high school	2
High school	1
Apprenticeship	3
Bachelor degree	10
Master degree	72
MBA	1
PhD	4
Degree program	
Industrial- and organizational psychology	31
Some other psychology discipline	8
Human resource management	6
Business administration	19
Economics	2
Law	2
(Business) Informatics	2
Other (mostly sociology)	17
Organization size (number of employees)	
Less than 100	25
100-499	20
500-1999	14
2000-4999	6
More than 5000	28

Supervisor role	
Yes	38
No	55
Number of subordinates	
1-5	15
6-10	6
11-20	7
21-50	8
51 or more	2
Job title	
Industrial- and organizational psychologist	6
Recruiter	17
HR manager	8
HR business partner	7
HR generalist	2
HR specialist	6
HR director	2
HR executive	0
HR consultant	7
Hiring manager	1
Other	37
Assessment license (DIN 33430) possession	
Yes	14
No	79
Assessment license (DIN 33430) type	
License E (for assessment professionals)	13
License B (for observers)	1
Number of hiring decisions involved in (across employers)	
1-49	29
50-99	20
100-199	9
200-499	20

500 or more	15
Sector of employing/own organization	
Manufacturing	8
Finance	1
Retail	2
Health care	5
Telecommunications	1
Transportation	1
Construction	0
IT	20
Utilities	1
Insurance	0
Educational services	1
Hospitality	0
Business consulting	35
Chemical	2
Pharmaceutical	1
Other	15

Table 2

Combination Methods Presented to Participants and Results from Study 1

Combination Method	Combination Method Number	Frequency of Combination Method most often Used	*M*	*SD*
First, I consider how important the information obtained from these sources is, and then I reach a decision via group consensus, by discussing the information with colleagues/other people involved in the hiring process.	3	44	3.60	1.20
First, I consider how important the information obtained from these sources is, and then I reach a decision by integrating the information using my own judgment.	2	21	3.12	1.18
First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use my own judgment.	7	9	1.90	1.28

First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use a pre-determined formula or rule.	8	9	1.71	1.22
I do not consider upfront how important the information obtained from these sources is. I reach a decision by integrating the information using my own judgment.	1	4	2.14	1.12
I apply a pre-defined formula or rule to the information obtained from these sources and I hire the applicant(s) with the highest overall score(s). Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2	4	2	1.53	1.00
First, I apply a pre-defined formula or rule to the information obtained from these sources. Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2. Then I may adjust this overall score using my own judgment in case I think this is needed.	5	2	1.16	0.50
I apply a pre-defined formula or rule to the information obtained from these sources, but I also explicitly add my own rating with a fixed weight in this rule. Then I hire the	6	1	1.28	0.75

applicant(s) with the highest overall score(s).Exemplary rule: Overall score = Test score*0.4 + Interview rating*0.2 + CV rating*0.1 + own rating*0.3

Note. Participants reported how frequently they use the presented combination methods and responded on a five-point scale (1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *usually*, 5 = *always*). N = 92.

Table 3

Reasons Not to Use an Algorithm – Results from Study 1

Reasons Not to Use an Algorithm	Proportion and 95% CI
I use qualitative information that I cannot or do not want to quantify.	.35 [.25, .45]
There are no rules available.	.30 [.21, .40]
The use of a rule is not accepted by other people around me (e.g., colleagues, supervisors).	.24 [.15, .33]
I think that using a rule together with my own judgment results in the best (i.e., most valid) decisions.	.21 [.12, .30]
Other	.21 [.12, .30]
I do not *believe* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.20 [.11, .28]
I think that using a rule will result in a less diverse workforce than combining information through my own judgment.	.20 [.11, .28]
I think it is my duty as a professional to use my own judgment rather than a rule.	.16 [.08, .24]
I (have to) select personnel based on criteria that are difficult to make explicit (e.g., personal liking, acquaintance).	.15 [.08, .23]
I am not *aware* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.14 [.07, .21]

I do not know how to construct such a rule.	.14 [.07, .21]
I do not have the time to read the academic literature on the use of rules.	.12 [.05, .18]
I feel my autonomy in the hiring process is restricted when using a rule.	.09 [.03, .15]
I do not have access to the academic literature on the use of rules.	.07 [.02, .12]
I feel my personal contact with other decision makers is reduced when using a rule.	.06 [.01, .11]
I feel I cannot demonstrate my competence when using a rule.	.05 [.00, .09]
I feel my professional status is lowered when using a rule.	.03 [.00, .07]
I am concerned that using a rule will be against the law.	0

Note. *N* = 86. Multiple answers possible.

Table 4

Ways to Construct an Algorithm – Results from Study 1

Ways to Construct an Algorithm	Proportion and 95% CI
I determine the rule based on a discussion with experts/professionals.	.42 [.28, .56]
I determine the rule based on scientific research (meta-analyses or primary scientific studies).	.34 [.21, .47]
I determine the rule based on my own knowledge and experience, without consulting the scientific literature or others.	.32 [.19, .45]
I use a rule as prescribed or suggested by others (e.g., by the organization you are working for or by professional standards).	.28 [.16, .40]
I determine the rule based on statistical analyses of data that come from my organization, or other relevant data.	.24 [.12, .36]
I weight all relevant information evenly (the same weights).	.10 [.02, .18]

Note. N = 50. Multiple answers possible.

Table 5

Information Sources in Study 1

Information Sources	*M*	*SD*
Other (HR) professionals	3.10	1.11
(HR) blogs, videos, websites	2.86	1.27
Professional HR literature (e.g., Harvard Business Review, Human Resources Manager, HR Performance, Personalwirtschaft, Wirtschaftspsychologie heute)	2.76	1.09
Scientific literature (Academic journals, e.g., Journal of Applied Psychology, International Journal of Selection and Assessment, Human Resource Management)	2.34	1.15
Popular magazines (e.g., Psychologie heute, Management Wissen, Handelsblatt)	2.09	1.11
External consultants	1.73	0.96
Scientists (university staff)	1.71	0.98

Note. We used the same five-point scale as Rynes et al. (2002), where 1 = *never or rarely*, 2 = *a few times per year*, 3 = *about once a month*, 4 = *several times per month*, 5 = *almost daily*.

Table 6

Frequencies of Sample Demographics in Study 2

Demographic Information	Frequency
Employment status	
Employed	210
Self-employed	17
Unemployed	3
Hours of work	
Full-time (35 hours or more)	223
Part-time	4
Education	
Lower than high school	0
High school	9
Vocational degree	5
Bachelor degree	148
Master degree	65
MBA	2
PhD	1
Degree program	
Industrial- and organizational psychology	17
Some other psychology discipline	8
Human resource management	52
Business administration	85
Economics	14
Law	0
(Business) Informatics	16
Other (mostly sociology)	24
Organization size (number of employees)	
Less than 100	15
100-499	95
500-1999	85
2000-4999	19
More than 5000	16

Supervisor role	
Yes	207
No	20
Number of subordinates	
1-5	27
6-10	78
11-20	75
21-50	22
51 or more	5
Job title	
Industrial- and organizational psychologist	6
Recruiter	23
HR manager	94
HR business partner	17
HR generalist	5
HR specialist	13
HR director	4
HR executive	13
HR consultant	11
Hiring manager	32
Other	12
Assessment license (SHRM) possession	
Yes	159
No	71
Assessment license (SHRM) type	
CP	101
SCP	58
Sector of employing/own organization	
Manufacturing	42
Finance	41
Retail	15
Health care	12
Telecommunications	1

Transportation	0
Construction	8
IT	87
Utilities	0
Insurance	2
Educational services	9
Hospitality	4
Business consulting	5
Chemical	0
Pharmaceutical	0
Other	4

Table 7

Combination Methods Presented to Participants and Results from Study 2

Combination Method	Combination Method Number	Frequency of Combination Method most often Used	*M*	*SD*
First, I consider how important the information obtained from these sources is, and then I reach a decision via group consensus, by discussing the information with colleagues/other people involved in the hiring process.	3	77	3.67	0.91
First, I consider how important the information obtained from these sources is, and then I reach a decision by integrating the information using my own judgment.	2	67	3.57	0.87
First, I apply a pre-defined formula or rule to the information obtained from these sources. Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2. Then I may adjust this overall score using my own judgment in case I think this is needed.	5	19	3.37	1.03
First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile	7	18	3.47	1.00

score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use my own judgment.				
I apply a pre-defined formula or rule to the information obtained from these sources and I hire the applicant(s) with the highest overall score(s). Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2	4	16	3.38	1.09
I apply a pre-defined formula or rule to the information obtained from these sources, but I also explicitly add my own rating with a fixed weight in this rule. Then I hire the applicant(s) with the highest overall score(s). Exemplary rule: Overall score = Test score*0.4 + Interview rating*0.2 + CV rating*0.1 + own rating*0.3	6	15	3.25	1.07
First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use a pre-determined formula or rule.	8	9	3.48	0.95
I do not consider upfront how important the information obtained from these sources is. I reach a decision by integrating the information using my own judgment.	1	7	3.07	1.22

Note. Participants reported how frequently they use the presented combination methods and responded on a five-point scale (1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *usually*, 5 = *always*). *N* = 228.

Table 8

Reasons Not to Use an Algorithm – Results from Study 2

Reasons Not to Use an Algorithm	Proportion and 95% CI
The use of a rule is not accepted by other people around me (e.g., colleagues, supervisors).	.28 [.21, .35]
I do not *believe* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.25 [.18, .32]
I think it is my duty as a professional to use my own judgment rather than a rule.	.23 [.17, .30]
I feel my professional status is lowered when using a rule.	.23 [.16, .29]
I feel my autonomy in the hiring process is restricted when using a rule.	.23 [.16, .29]
I think that using a rule together with my own judgment results in the best (i.e., most valid) decisions.	.23 [.16, .29]
I feel my personal contact with other decision makers is reduced when using a rule.	.21 [.15, .28]
I am not *aware* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.20 [.14, .26]
I use qualitative information that I cannot or do not want to quantify.	.18 [.12, .24]
There are no rules available.	.16 [.10, .22]

I think that using a rule will result in a less diverse workforce than combining information through my own judgment.	.13 [.08, .19]
I (have to) select personnel based on criteria that are difficult to make explicit (e.g., personal liking, acquaintance).	.11 [.06, .16]
I do not know how to construct such a rule.	.08 [.04, .12]
I feel I cannot demonstrate my competence when using a rule.	.08 [.04, .12]
I am concerned that using a rule will be against the law.	.07 [.03, .11]
I do not have access to the academic literature on the use of rules.	.05 [.02, .08]
I do not have the time to read the academic literature on the use of rules.	.04 [.01, .07]
Other	0

Note. *N* = 163. Multiple answers possible.

Table 9

Information Sources in Study 2

Information Sources	*M*	*SD*
(HR) blogs, videos, websites	3.56	1.16
Other (HR) professionals	3.47	0.99
Professional HR literature (e.g., Harvard Business Review, MIT Sloan Management Review, HR Magazine)	3.35	1.13
Scientific literature (Academic journals, e.g., Journal of Applied Psychology, International Journal of Selection and Assessment, Human Resource Management)	3.23	1.25
Popular magazines (e.g., Forbes, Fortune, Fast Company, Inc., Workforce)	3.23	1.21
External consultants	3.03	1.11
Scientists (university staff)	3.01	1.34

Note. We used the same five-point scale as Rynes et al. (2002), where 1 = *never or rarely*, 2 = *a few times per year*, 3 = *about once a month*, 4 = *several times per month*, 5 = *almost daily*.

Table 10

Ways to Construct an Algorithm – Results from Study 2

Ways to Construct an Algorithm	Proportion and 95% CI
I determine the rule based on statistical analyses of data that come from my organization, or other relevant data.	.54 [.48, .61]
I weight all relevant information evenly (the same weights).	.44 [.37, .50]
I determine the rule based on scientific research (meta-analyses or primary scientific studies).	.43 [.36, .49]
I determine the rule based on my own knowledge and experience, without consulting the scientific literature or others.	.33 [.26, .39]
I determine the rule based on a discussion with experts/professionals.	.32 [.26, .38]
I use a rule as prescribed or suggested by others (e.g., by the organization you are working for or by professional standards).	.24 [.19, .30]

Note. *N* = 221. Multiple answers possible.

Table 11

Reasons to Use an Algorithm – Results from Study 2

Reasons to Use an Algorithm	Proportion and 95% CI
Higher predictive validity compared to combining information through my own judgment	.45 [.38, .51]
Easier to use than combining information through my own judgment	.37 [.31, .44]
Yields more valuable information than combining information through my own judgment	.34 [.28, .40]
Fairer than combining information through my own judgment	.32 [.26, .38]
Prior positive experience with using a rule	.26 [.21, .32]
Reduces time to hire more effectively than combining information through my own judgment	.26 [.20, .31]
Reduces time required of the professional compared to combining information through my own judgment	.24 [.19, .30]
It feels legally safer than combining information through my own judgment	.22 [.17, .28]
More transparent than combining information through my own judgment	.22 [.17, .28]
Reinforces employer brand more effectively than combining information through my own judgment	.18 [.13, .23]
Cheaper than combining information through my own judgment	.02 [.00, .04]

Note. *N* = 219. Multiple answers possible.

Table 12

Frequencies of the Units across both Focus Groups per Theme

Theme	Choice 1	Choice 2
1. Advantages of mechanical methods	23	0
2. Disadvantages of mechanical methods	32	1
3. Advantages of holistic methods	0	1
4. Disadvantages of holistic methods	4	0
5. What is needed to increase the use of mechanical methods?		
5.1. Providing information and guidelines	29	2
5.2. Where should information about mechanical methods be exchanged?	19	0
5.3 Who should be responsible for encouraging the use of mechanical methods?	11	2
6. Questions about how to use mechanical methods	41	7
7. Reasons why decision makers do not use mechanical methods	29	22
8. Did not fall under a theme	3	0
Total	191	35

Note. Choice 1 reflects the raters' first coding choice. When in doubt, a second theme could be assigned (Choice 2).

Table 13

Full List of Quotes from Study 3

Theme	Quote
1. Advantages of mechanical methods	"*I do not know whether you have to impose it on people or something like that. For me it is more about creating awareness. That's where a decision rule can help. This would be the added value for me.*"
	"*I think that if you have a decision rule as a sort of guiding principle, it can be useful, because you force people to formulate their thinking explicitly, as it were. Then we say: look, this is known, about what works well. Then you can check with yourself what/how I deviate from what I read here. What do I use then? So as a learning tool, to optimize your own decision rule or strategy, I think it can be useful. Look, we're all professionals, so you can't keep doing something you don't know what you're doing, against your better judgement. Then you are not working professionally. I think someone can file a complaint then. So I see it more as a kind of heuristic maybe, because this is an advice. That's how it works globally. Think about what you do in your own practice and how you can learn from it. So as a kind of intervision tool, just to increase the quality. To make people aware of their thinking.*"

2. Disadvantages of mechanical methods	"*I think that decision rules in I-O psychology, of which I am part of, are rarely used because the criterion that you select on changes frequently. And you cannot design a decision rule for a specific function in an organization because such a selection takes place only one or two times a year, maybe even less often. Maybe that is different in clinical psychology. Diagnosis with the DSM. There you have to make a diagnosis because you have to give an indication about a therapy. But it heavily differs per criterion. How often does something occur in an organization? And I think, if you have designed a decision rule, you actually also have to evaluate it. You also have to see whether it actually works.*" "*The best example is the WISC-5, where we always gave point estimates and then there was a decision rule for IQ scores with regard to educational selection. And then you looked at the didactic scores. If that was okay, you could put a bow on it. But what we see and hear now is, it still works this way because it is the law. But if you talk to people in the field, then you increasingly hear that you should let point estimates loose and give confidence intervals instead. But then using such a decision rule is immediately more difficult.*"
3. Advantages of holistic methods	"*But then I do not really understand what the difference is between the holistic and mechanical method, because you can also make a holistic prediction fully transparent. You can exactly tell how you do it.*"

4. Disadvantages of holistic methods	"*But it is also true that extensive research has been done into how people integrate different data sources. That is a typical thing that no one is really good at. The same holds for psychologists. They are very bad at that. This was also shown in the research. It also showed that our psychologists put the most weight on the things they had seen themselves. They thought the cognitive ability test was less important.*" "*I am also less sharp on one day than on another. If I have accidentally read a book, then that is really important for me. We are not so protocolled in our practice. But a protocol may likely work better. So I sometimes also catch myself going into one direction – or I see my colleague going into one direction – we should formulate this more explicitly.*"
6. Questions about how to use mechanical methods	"*May I ask something? You just said that it is essential that you quantify things, but that does not necessarily have to be evidence-based?*" "*I think very simple and say, just look in the tables from Schmidt and Hunter from 2016. There you can indicate weights, that is based on something and that is evidence-based. But if everyone can simply put something into the formula they like, what predicts what, then it does not matter whether you call that holistic or mechanical, does it?*" "*But I also think that it is rarely used because it is not really accepted in the field.*"

7. Reasons why decision makers do not use mechanical methods	"*But I also tend to say that one time the interview is weighted more heavily than another time. And sometimes an IQ test is just not crucial. This assessment remains in my opinion the core of the psychologist's field.*"
5. What is needed to increase the use of mechanical methods?	
5.1 Providing information and guidelines	"*If I look at myself, what is needed is training in applying such a decision rule. I would like to see how I could approach such a rule in a manner to avoid bias, or to use more and more the same hobbyhorse*." "*I think it would be nice to have examples in the form of a database, where you could search and that it is specific for certain content areas or specializations. We see an enormous diversity in different areas in psychology, where different people use different models on the basis of which they make decisions.*"
5.2 Where should information about mechanical methods be exchanged?	"*And this is perhaps something that can be gradually developed, and you can build a community that says these rule work in our case. That you upload this then in a system and that others can also make use of it again.*" *"This could be exchanged in study programs and courses"*

5.3 Who should be responsible for encouraging the use of mechanical methods?	“*Maybe the big HR consultancies also have to collaborate. I have worked for a big HR consultancy where thousands of applicants are assessed per year. And all this data has been saved. You need data on which the decision rule is based. And I find that we throw away lots of data. The applicants should be followed to see whether you got it right. You have to follow the applicants then, also those we reject, to see whether they end up somewhere where they do a really good job and what we may have done wrong with the decision rule. You have to get organizations so far that they collect data. Back in the days, we did that and then we also sometimes adjusted the weighting. We never looked how they scored once they were on the job. The organization can do a lot about it.*” *“But the COTAN* is the authority for people working in practice and in universities, in the field of testing and test use. So in this sense, I am not sure whether this is the job of the COTAN. I actually think it is not, but maybe they could provide support that a group is getting down to work with this.”*

Note. *the COTAN (in Dutch: Commissie Testaangelegenheden Nederland) is a Dutch committee that evaluates various tests and questionnaires commonly used in the Netherlands with regard to their psychometric properties.

Supplementary Material

Supplement S1

Pre-registered analysis not reported in the main paper

Means, standard deviations, and Spearman's rank correlations among study variables are shown in Table S1. We computed Spearman's rank correlations instead of Pearson's correlation coefficient because the variables "mechanical prediction" and "experience (in years)" were positively skewed. As Table S1 shows, in line with expectations, reading the academic literature and possessing an assessment license were positively and weakly to moderately correlated with the use of mechanical prediction methods. In contrast, organization size and experience (both measured in years and amount of hiring decisions made) had negligible relationships with the use of mechanical prediction methods.

Table S1

Means, standard deviations, and Spearman's rank correlations among study variables in Study 1

Variables	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.
1. Holistic prediction	2.39	0.58	-						
2. Mechanical prediction	1.51	0.76	.08	-					
3. Academic literature	2.34	1.15	-.35*	.26*	-				
4. Assessment license	0.15	0.36	-.24*	.25*	.29*	-			
5. Organization size	2.94	1.60	.02	.02	-.11	-.08	-		
6. Experience (years)	9.75	7.67	-.01	-.17	-.13	-.03	-.18	-	
7. Experience (amount)	2.67	1.49	-.10	.09	.11	-.02	-.01	.52*	-

Note. * $p < .05$, two tailed. Assessment license (DIN 33430) was coded as 0 = No, 1 = Yes. Organization size was measured on a five-point scale (1 = less than 100, 2 = 100 – 499, 3 =

500 – 1999, 4 = 2000 – 4999, 5 = more than 5000). Experience (amount) was measured on a five-point scale (1 = 1 – 49, 2 = 50 – 99, 3 = 100 – 199, 4 = 200 – 499, 5 = 500 or more). *N* = 92-93.

We also fitted a regression model with mechanical prediction as the dependent variable and academic literature, assessment license, organization size, experience (in years), and experience (amount) as independent variables. Together, the predictors explained 15 % of the variance in mechanical prediction ($F(5, 86) = 3.05$, $p = .014$, $R^2_{\text{adj.}} = .10$). As can be seen in Table S2, assessment license ($b = 0.45$, $SE = 0.22$, $t = 2.03$, $p = .045$) and academic literature ($b = 0.12$, $SE = 0.07$, $t = 1.67$, $p = .10$) had the largest regression coefficients.

Table S2

Regression Results from Study 1

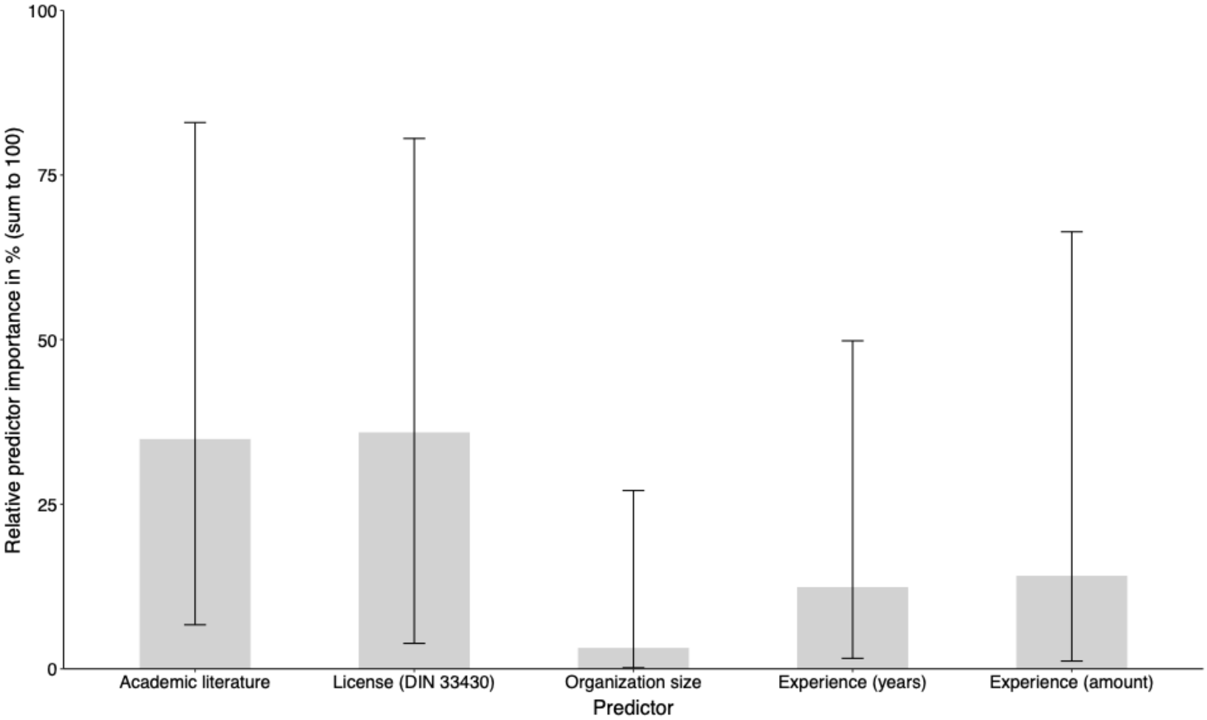
Variable	Estimate (*b*)	SE (*b*)	*t*	*p*
Intercept	0.98	0.28	3.48	< .01
Academic literature	0.12	0.07	1.67	.10
Assessment license	0.45	0.22	2.03	.045
Organization size	0.03	0.05	0.60	.547
Experience (years)	-0.02	0.01	-1.42	.16
Experience (amount)	0.10	0.06	1.59	.115

Note. *N* = 92.

To assess variable importance, we also computed relative importance weights using the *relaimpo* R package (Grömping, 2006). Figure S1 shows relative importance metrics together with 95% bootstrap confidence intervals.

Figure S1

Relative importance predictor weights for predicting the use of mechanical prediction methods in Study 1



Note. Error bars represent 95% bootstrap confidence intervals.

Supplement S2

List of work activities

Staffing organizational units - Recruiting, interviewing, selecting, hiring, and promoting applicants/employees.

Selling or influencing others - Convincing others to buy merchandise/goods.

Guiding, directing, and motivating subordinates - Providing guidance and direction to subordinates, including setting performance standards and monitoring performance.

Controlling machines and processes - Using either control mechanisms or direct physical activity to operate machines or processes (not including computers or vehicles).

Documenting/recording information - Entering, transcribing, recording, storing, or maintaining information in written or electronic form.

Repairing and maintaining equipment - Servicing, repairing, adjusting, and testing machines, devices, moving parts, and equipment that operate on the basis of mechanical or electrical principles.

Scheduling work and activities - Scheduling events, programs, and activities, as well as the work of others.

Making decisions and solving problems - Analyzing information and evaluating results to choose the best solution and solve problems.

Judging the qualities of things, services, or people - Assessing the value, importance, or quality of things or people.

Estimating the quantifiable characteristics of products, events, or information - Estimating sizes, distances, and quantities; or determining time, costs, resources, or materials needed to perform a work activity.

Getting information - Observing, receiving, and otherwise obtaining information from all relevant sources.

Monitoring processes, materials, or surroundings - Monitoring and reviewing information from materials, events, or the environment, to detect or assess problems.

Supplement S3

Attention checks in Study 2

Attention check 1: We included one item in the “Information Source” block in the online survey in which participants were asked to respond with “a few times per year”. Participants who did not choose this answer option were excluded.

Attention check 2: Participants were presented with the statement: “I have 17 fingers”. Those who responded “Yes” were excluded.

Supplement S4

Pre-registered analysis not reported in the main paper

Means, standard deviations, and Spearman's rank correlations among study variables are shown in Table S3. We computed Spearman's rank correlations instead of Pearson's correlation coefficient because the variables "experience (in years)" and "experience (amount)" were positively skewed. In contrast to Study 1, experience (amount) was measured as a continuous variable. As Table S3 shows, in line with expectations, reading the academic literature and possessing an assessment license were positively and moderately correlated with the use of mechanical prediction methods. In contrast, organization size and experience in years showed negligible relationships with the use of mechanical prediction methods. Yet, experience measured as amount of hiring decisions made in one's life showed the expected negative relationship with the use of mechanical prediction methods, although this relationship was small.

Table S3

Means, standard deviations, and Spearman's rank correlations among study variables in Study 2

Variables	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.
1. Holistic prediction	3.43	0.59	-						
2. Mechanical prediction	3.37	0.83	.57*	-					
3. Academic literature	3.25	1.25	.31*	.45*	-				
4. Assessment license	0.69	0.46	.22*	.32*	.39*	-			
5. Organization size	2.68	0.97	-.02	-.05	-.03	.17*	-		
6. Experience (years)	5.65	4.74	.09	-.01	-.17*	-.07	-.08	-	
7. Experience (amount)	42.37	103.4	-.16*	-.15*	-.24*	-.29*	.02	.26*	-

Note. * $p < .05$, two tailed. Assessment license (SHRM) was coded as 0 = No, 1 = Yes. Organization size was measured on a five-point scale (1 = less than 100, 2 = 100 – 499, 3 = 500 – 1999, 4 = 2000 – 4999, 5 = more than 5000). N = 228-230.

As in Study 1, we fitted a regression model with mechanical prediction as the dependent variable and academic literature, assessment license, organization size, experience (in years), and experience (amount) as independent variables. Together, the predictors explained 30 % of the variance in mechanical prediction ($F(5, 222) = 18.58$, $p < .001$, $R^2_{adj.} = .28$). As can be seen in Table S4, assessment license ($b = 0.38$, $SE = 0.11$, $t = 3.41$, $p < .001$) and academic literature ($b = 0.26$, $SE = 0.04$, $t = 6.22$, $p < .001$) had the largest regression coefficients.

Table S4

Regression results from Study 2

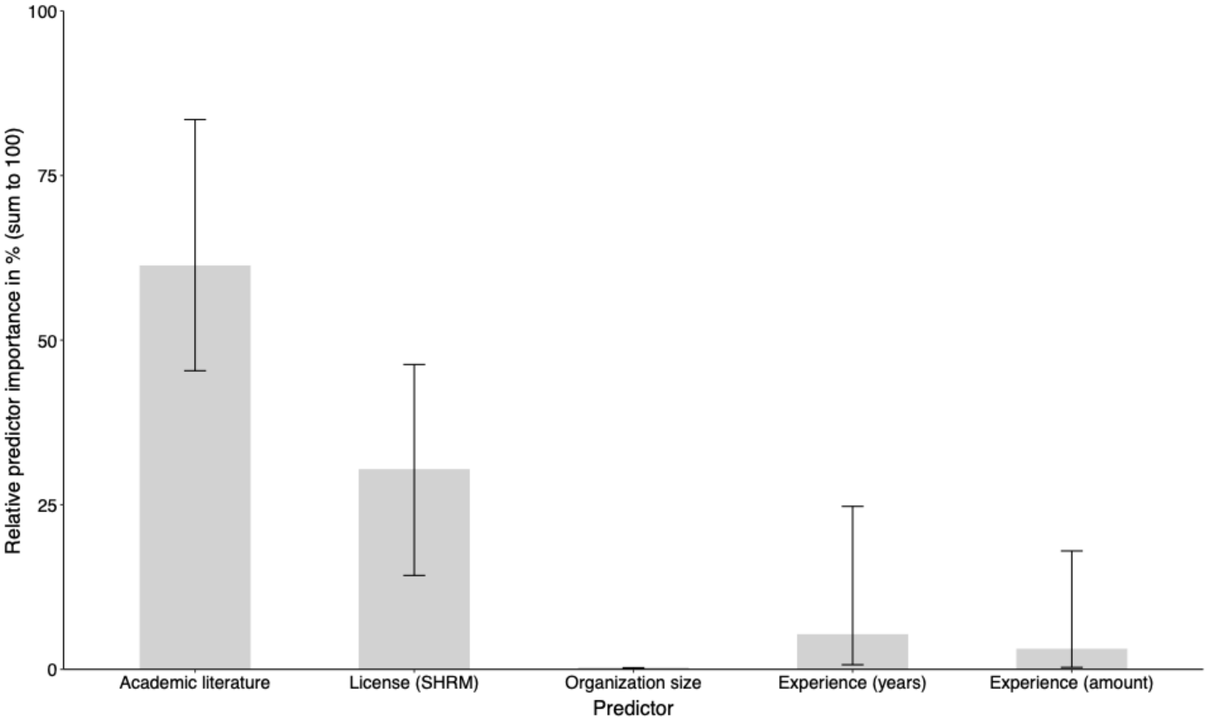
Variable	Estimate (*b*)	SE (*b*)	*t*	*p*
Intercept	2.31	0.21	10.95	< .001
Academic literature	0.26	0.04	6.22	< .001
Assessment license	0.38	0.11	3.41	< .001
Organization size	< 0.01	0.05	0.09	.93
Experience (years)	-0.01	0.01	-0.91	.36
Experience (amount)	< -0.01	< 0.01	-0.68	.50

Note. N = 228.

As in Study 1, we computed relative importance weights using the *relaimpo* R package (Grömping, 2006) to investigate predictor importance. Figure S2 shows relative importance metrics together with 95% bootstrap confidence intervals.

Figure S2

Relative importance predictor weights for predicting the use of mechanical prediction methods in Study 2



Note. Error bars represent 95% bootstrap confidence intervals.

Supplement S5

Compared to Study 1, relatively few participants in Study 2 answered all knowledge check items correctly. Therefore, we also inspected the results of Study 2 when only including participants who answered at least two knowledge items correctly (n = 97). The results of this subsample were closely aligned with the results of the full sample. Very similar to the full sample, 85% of the participants indicated using holistic methods most often. Furthermore, 43% indicated making holistic predictions in teams, while 35% reported making holistic predictions individually. Twenty percent reported using clinical synthesis. Table S5 shows per combination method participants' mean frequency rating and the number of participants who indicated using a given method most often.

All other results were also largely in line with the results from the full sample. We present our results on reasons not to use mechanical prediction and ways participants reported to construct their algorithm in Table S6 and S7, respectively. Furthermore, we present the results on reasons to use mechanical prediction and information sources in Table S8 and S9, respectively.

Table S5

Combination Methods Presented to Participants and Results from the Subsample from Study 2

Combination Method	Combination Method Number	Frequency of Combination Method most often Used	*M*	*SD*
First, I consider how important the information obtained from these sources is, and then I reach a decision via group consensus, by discussing the information with colleagues/other people involved in the hiring process.	3	35	3.71	0.90
First, I consider how important the information obtained from these sources is, and then I reach a decision by integrating the information using my own judgment.	2	29	3.51	0.92
First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use my own judgment.	7	8	3.28	1.12

First, I apply a pre-defined formula or rule to the information obtained from these sources. Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2. Then I may adjust this overall score using my own judgment in case I think this is needed.	5	8	3.22	1.12
I apply a pre-defined formula or rule to the information obtained from these sources, but I also explicitly add my own rating with a fixed weight in this rule. Then I hire the applicant(s) with the highest overall score(s).Exemplary rule: Overall score = Test score*0.4 + Interview rating*0.2 + CV rating*0.1 + own rating*0.3	6	6	3.08	1.18
I apply a pre-defined formula or rule to the information obtained from these sources and I hire the applicant(s) with the highest overall score(s). Exemplary rule: Overall score = Test score*0.5 + Interview rating*0.3 + CV rating*0.2	4	5	3.18	1.20
First, I determine cutoffs, and applicants who fall above these cutoffs progress to the next stage in the hiring process. Exemplary rule: Progress to the next stage = A test percentile score equal to or higher than 50% and an interview rating of 3 or higher (on a five-point scale). To make the final hiring decision, I use a pre-determined formula or rule.	8	4	3.29	1.06

I do not consider upfront how important the information obtained from these sources is. I reach a decision by integrating the information using my own judgment.	1	2	2.93	1.27

Table S6

Reasons Not to Use an Algorithm – Results from the Subsample from Study 2

Reasons not to Use an Algorithm	Proportion and 95% CI
I do not *believe* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.29 [.18, .39]
I feel my autonomy in the hiring process is restricted when using a rule.	.29 [.18, .39]
The use of a rule is not accepted by other people around me (e.g., colleagues, supervisors).	.27 [.17, .38]
I feel my professional status is lowered when using a rule.	.25 [.15, .35]
I use qualitative information that I cannot or do not want to quantify.	.22 [.12, .31]
I am not *aware* that using a rule results in better (i.e., more valid) decisions than combining information through my own judgment.	.21 [.11, .30]
I think that using a rule together with my own judgment results in the best (i.e., most valid) decisions.	.19 [.10, .28]
I think it is my duty as a professional to use my own judgment rather than a rule.	.18 [.09, .27]
I feel my personal contact with other decision makers is reduced when using a rule.	.16 [.08, .25]
There are no rules available.	.14 [.06, .22]

I think that using a rule will result in a less diverse workforce than combining information through my own judgment.	.11 [.04, .18]
I do not know how to construct such a rule.	.08 [.02, .15]
I feel I cannot demonstrate my competence when using a rule.	.08 [.02, .15]
I am concerned that using a rule will be against the law.	.08 [.02, .15]
I (have to) select personnel based on criteria that are difficult to make explicit (e.g., personal liking, acquaintance).	.08 [.02, .15]
I do not have the time to read the academic literature on the use of rules.	.07 [.01, .13]
I do not have access to the academic literature on the use of rules.	.05 [<.01, .10]
Other	0

Note. *N* = 73. Multiple answers possible.

Table S7

Ways to Construct an Algorithm – Results from the Subsample from Study 2

Ways to Construct an Algorithm	Proportion and 95% CI
I determine the rule based on statistical analyses of data that come from my organization, or other relevant data.	.53 [.43, .64]
I determine the rule based on scientific research (meta-analyses or primary scientific studies).	.37 [.27, .47]
I weight all relevant information evenly (the same weights).	.33 [.24, .43]
I determine the rule based on a discussion with experts/professionals.	.32 [.23, .42]
I determine the rule based on my own knowledge and experience, without consulting the scientific literature or others.	.27 [.18, .36]
I use a rule as prescribed or suggested by others (e.g., by the organization you are working for or by professional standards).	.24 [.16, .33]

Note. $N = 90$. Multiple answers possible.

Table S8

Reasons to Use an Algorithm – Results from the Subsample from Study 2

Reasons to Use an Algorithm	Proportion and 95% CI
Higher predictive validity compared to combining information through my own judgment	.40 [.30, .50]
Easier to use than combining information through my own judgment	.33 [.23, .43]
Yields more valuable information than combining information through my own judgment	.32 [.22, .42]
Fairer than combining information through my own judgment	.31 [.21, .40]
Reduces time to hire more effectively than combining information through my own judgment	.28 [.19, .38]
Prior positive experience with using a rule	.27 [.18, .37]
It feels legally safer than combining information through my own judgment	.23 [.14, .31]
Reduces time required of the professional compared to combining information through my own judgment	.22 [.13, .30]
More transparent than combining information through my own judgment	.19 [.11, .28]
Reinforces employer brand more effectively than combining information through my own judgment	.10 [.04, .17]
Cheaper than combining information through my own judgment	.01 [.00, .03]

Note. N = 88. Multiple answers possible.

Table S9

Information Sources in the Subsample from Study 2

Information Sources	*M*	*SD*
Other (HR) professionals	3.49	0.88
(HR) blogs, videos, websites	3.45	1.18
Professional HR literature (e.g., Harvard Business Review, MIT Sloan Management Review, HR Magazine)	3.12	1.13
Popular magazines (e.g., Forbes, Fortune, Fast Company, Inc., Workforce)	3.12	1.30
Scientific literature (Academic journals, e.g., Journal of Applied Psychology, International Journal of Selection and Assessment, Human Resource Management)	3.11	1.27
Scientists (university staff)	2.93	1.34
External consultants	2.83	1.39