Implementing evidence-based assessment and selection in organizations

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

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
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

Article

Implementing evidence-based assessment and selection in organizations: A review and an agenda for future research

Marvin Neumann
University of Groningen, Netherlands

A. Susan M. Niessen
University of Groningen, Netherlands

Rob R. Meijer
University of Groningen, Netherlands

Abstract

In personnel- and educational selection, a substantial gap exists between research and practice, since evidence-based assessment instruments and decision-making procedures are underutilized. We provide an overview of studies that investigated interventions to encourage the use of evidence-based assessment methods, or factors related to their use. The most promising studies were grounded in self-determination theory. Training and autonomy in the design of evidence-based assessment methods were positively related to their use, while negative stakeholder perceptions decreased practitioners' intentions to use evidence-based assessment methods. Use of evidence-based decision-making procedures was positively related to access to such procedures, information to use it, and autonomy over the procedure, but negatively related to receiving outcome feedback. A review of the professional selection literature showed that the implementation of evidence-based assessment was hardly discussed. We conclude with an agenda for future research on encouraging evidence-based assessment practice.

Corresponding author:
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

In personnel- and educational selection, practitioners such as human resource managers, organizational psychologists, and admission officers should be interested in using valid, evidence-based assessment because it can result in large performance- and financial gains (Hoffman et al., 2017; Huselid, 1995; Kuncel & Hezlett, 2007; Schmidt & Hunter, 1998; Terpstra & Rozell, 1993). Yet, there is a substantial gap between evidence-based and actual assessment practices (Bolander & Sandberg, 2013; Highhouse, 2008; Morris et al., 2015; Ployhart et al., 2017; Rynes et al., 2002; Sanders et al., 2008). Some have even argued that this gap is widening (Rynes, 2012).

There are different reasons why evidence-based assessment is unterutilized. Reasons are the unawareness of or disbelief in research findings (Fisher et al., 2020; Highhouse, 2008), the restriction of practitioners' autonomy (Nolan & Highhouse, 2014), and the reduction of the credit received from other stakeholders for decisions made (Nolan et al., 2016). Despite substantial progress in research on performance prediction and decision-making, the challenge of increasing the use of evidence-based assessment in selection has not been resolved in the last century (Ryan & Ployhart, 2014, p. 695). This shows that selection still is the _supreme problem_ in applied psychology (Ployhart et al., 2017). To solve this supreme problem, studies on factors associated with or interventions designed to increase practitioners' use and acceptance of evidence-based assessment practices emerged. These studies seem wide in scope, covering applied (e.g., Dietvorst et al., 2018), exploratory (e.g., Roulin et al., 2019), and theory-driven research (Nolan & Highhouse, 2014). Some studies (Nolan & Highhouse, 2014) were based on self-determination theory (Deci & Ryan, 2000) and mostly investigated the effect of the practitioner's autonomy on the use of evidence-based assessment. Other studies were grounded in attribution theory (Kelley, 1973) and investigated specifically how stakeholder perceptions may influence practitioner's use evidence-based assessment (Nolan et al., 2016, 2020). In some cases, studies were not clearly based on theory (e.g., Dietvorst et al., 2018; Roulin et al., 2019). Given this diversity, it remains unclear which theoretical frameworks are most promising for studying interventions that may increase evidence-based assessment in selection practice. Therefore, the first aim of this review was to describe and synthesize the existing research on factors related to and interventions designed to encourage the adoption of evidence-based assessment, and to get an insight into the most promising theoretical frameworks that have been used in this research.

Providing an overview of ways to overcome the science-practice gap in selection also requires insight into how the professional community perceives and discusses the implementation of evidence-based assessment in practice. The scientific discussion of the science-practice gap has been mainly concerned with the underutilization of assessment practices that result in better performance predictions. Yet, practitioners also often try to optimize other factors than performance, such as organizational fit (Barrick & Parks-Leduc, 2019), assessment costs (Klehe, 2004), and diversity (Pyburn et al., 2008). Sometimes, these practical factors have also inspired research, as in the case of the validity-diversity dilemma (Ployhart & Holtz, 2008; Rupp et al., 2020). For these reasons, the second aim was to investigate how evidence-based assessment and the translation of scientific evidence into practice is discussed in professional journals that are read by practitioners, and how this discussion aligns with the discussion in the scientific literature. To answer these questions, we conducted a review of the professional literature as well. Since practitioners care about more aspects than performance prediction (König et al., 2010), we expected that the discussion on the implementation of evidence-based assessment regarding performance prediction would be rather limited. Therefore, we also investigated which other factors related to selection received attention in the professional literature.

Based on both reviews, we (1) give practical recommendations about effective interventions

to establish evidence-based assessment, (2) discuss the most promising theoretical frameworks that were used in these studies, and (3) provide an agenda for future research. We chose to focus on personnel- and educational selection because these fields are both concerned with human performance prediction, often use similar predictors (Michel et al., 2019; Risavy et al., 2019), and findings on evidence-based assessment are comparable across both fields (Kuncel et al., 2013).

The science-practice gap

In designing selection procedures, two choices are of main importance: What information is collected (for example, standardized test scores or interview impressions, Kuncel & Hezlett, 2007; Kuncel et al., 2001; Schmidt & Hunter, 1998) and how that information is combined to make predictions and decisions (Kuncel et al., 2013). In information collection, another important distinction is between the constructs assessed and the instruments used to measure those constructs (Arthur & Villado, 2008). Scientists largely agree that cognitive abilities, and to a lesser extent personality, are the most relevant constructs that explain differences in academic- and job performance (Kuncel et al., 2004; Sackett, Lievens, et al., 2017; Stanek & Ones, 2018). Instead, practitioners primarily consider personality and applied social skills rather than cognitive abilities to be the most important constructs (Fisher et al., 2020; Ryan et al., 2015; Sackett & Walmsley, 2014).With regard to assessment instruments, scientific evidence showed that scores on cognitive ability tests, assessment centers, work sample tests, and structured interviews are valid predictors of job performance (Huffcutt et al., 2014; Ones et al., 2010; Roth et al., 2005; Sackett, Shewach, & Keiser, 2017). However, less valid instruments such as analyses of CV's and cover letters, and unstructured interviews are prevalent in practice (König et al., 2010; Lievens & De Paepe, 2004; Risavy et al., 2019; Zibarras & Woods, 2010).

After information about applicants is collected, it needs to be combined into a judgment or prediction. This can be done holistically, that is through intuitively combining information "in the mind," or mechanically, through combining information by means of a consistently applied rule (Grove & Meehl, 1996; Meehl, 1954a). A large amount of studies showed that mechanical combination of information results in predictions that are often equally or more valid than predictions based on holistic combination of information (Grove et al., 2000; Meehl, 1954a; Sawyer, 1966), especially when predicting human performance (Kuncel et al., 2013). Despite these consistent findings, information is typically combined holistically in practice (Highhouse, 2008; Morris et al., 2015; Prien et al., 2003; Ryan & Sackett, 1987; Silzer & Jeanneret, 2011; Slaughter & Kausel, 2014). So, the gap between evidence-based assessment practices and actual assessment practices exists in both information collection and combination (Highhouse, 2008; Rynes, 2012), and both gaps are considered in this review.

Academic literature

Method

Selection of studies

We conducted a literature search in the databases PsycInfo, Web of Science, Grey Literature Report, and ResearchGate, up to and including 2020. A non-exhaustive list of important search terms included "mechanical combination," "holistic combination," "evidence-based assessment," "decision aid," "standardized selection," "use intention," and "implementation" (the complete list is presented in the Online Appendix). Search terms were chosen to cover the PICO elements (participants, interventions, comparators, and outcomes) of existing studies (Shamseer et al., 2015). Other search terms were based on key words of relevant articles before conducting the systematic search. We identified empirical studies of interventions designed to

increase decision makers' use and acceptance of evidence-based assessment procedures. In addition, we included studies that investigated factors associated with the use and acceptance of evidence-based assessment. So, a study was included if it contained a dependent measure of decision makers' use (intentions) or acceptance of evidence-based assessment. We only included studies that focused on personnel- or educational selection. Therefore, participants in the included studies were adults that were either HR professionals, admission officers, staff involved in selection procedures, (working) adults, or students. Furthermore, we only included book chapters, journal articles, dissertations, and errata/corrections, published in English, German, or Dutch.

Two independent reviewers (authors 1 and 2) initially screened the titles and abstracts of 4060 unique documents for inclusion. A coding scheme with three categories (0 = no inclusion, 1 = inclusion, 2 = no inclusion but inspect reference list) was used in the first round. In the second round, both reviewers read each of the remaining documents and coded them for in- or exclusion. Inter-rater agreement was high in the first- (absolute agreement 98.6%, κ = .705) and second round (absolute agreement 94.5%, κ = .786). Disagreements were resolved through discussion until consensus was reached. The first author investigated the papers coded as inspect reference list and all reference lists of the final papers that were included in the review. Papers from the reference list search and papers that we found or that were published during the writing process were added. Eventually, 21 articles met the inclusion criteria. A flow chart that depicts the literature selection process is shown in Figure 1.

After we identified the final articles, we loosely followed the steps of thematic analysis outlined in Braun and Clarke (2006) to establish common research topics. Then, we revised these research topics based on the reviewers' suggestions and eventually identified five topics that were studied in association with

evidence-based assessment use or acceptance: 1. Practitioner characteristics (six studies), 2. Communication and presentation of scientific evidence (nine studies), 3. Feedforward and outcome feedback (seven studies), 4. Motivational factors (12 studies), and 5. Stakeholder perceptions (four studies). These topics differ in the extent to which the organizational or social context, the prediction and selection context, or practitioner characteristics are related to the adoption of evidence-based assessment practices. Studies on the communication and presentation of scientific evidence are aimed at persuading stakeholders of the advantage of evidence-based assessment. In contrast, in outcome feedback studies, the goal was to let practitioners experience their relatively greater prediction errors compared to evidence-based methods, and to induce learning. Hence, these topics also differ with regard to the relevant theoretical frameworks. We deliberately chose this grouping because it most clearly highlighted topics on which research exists and where future research is needed. To guide the reader through the results, we emphasize to what extent a section discusses information collection, information combination, or both.

Results

Information on the study design, participants, and sample size for each of the included studies from the academic literature search is shown in Table 1. Furthermore, we mention effect sizes as reported in the original studies.

Practitioner characteristics

Investigating whether practitioner characteristics, such as their professional and educational background, are related to their use of evidence-based assessment practices is important because organizations may select, for example, HR professionals based on such individual differences. Other practitioner characteristics that have been investigated include experience and training, and decision-making
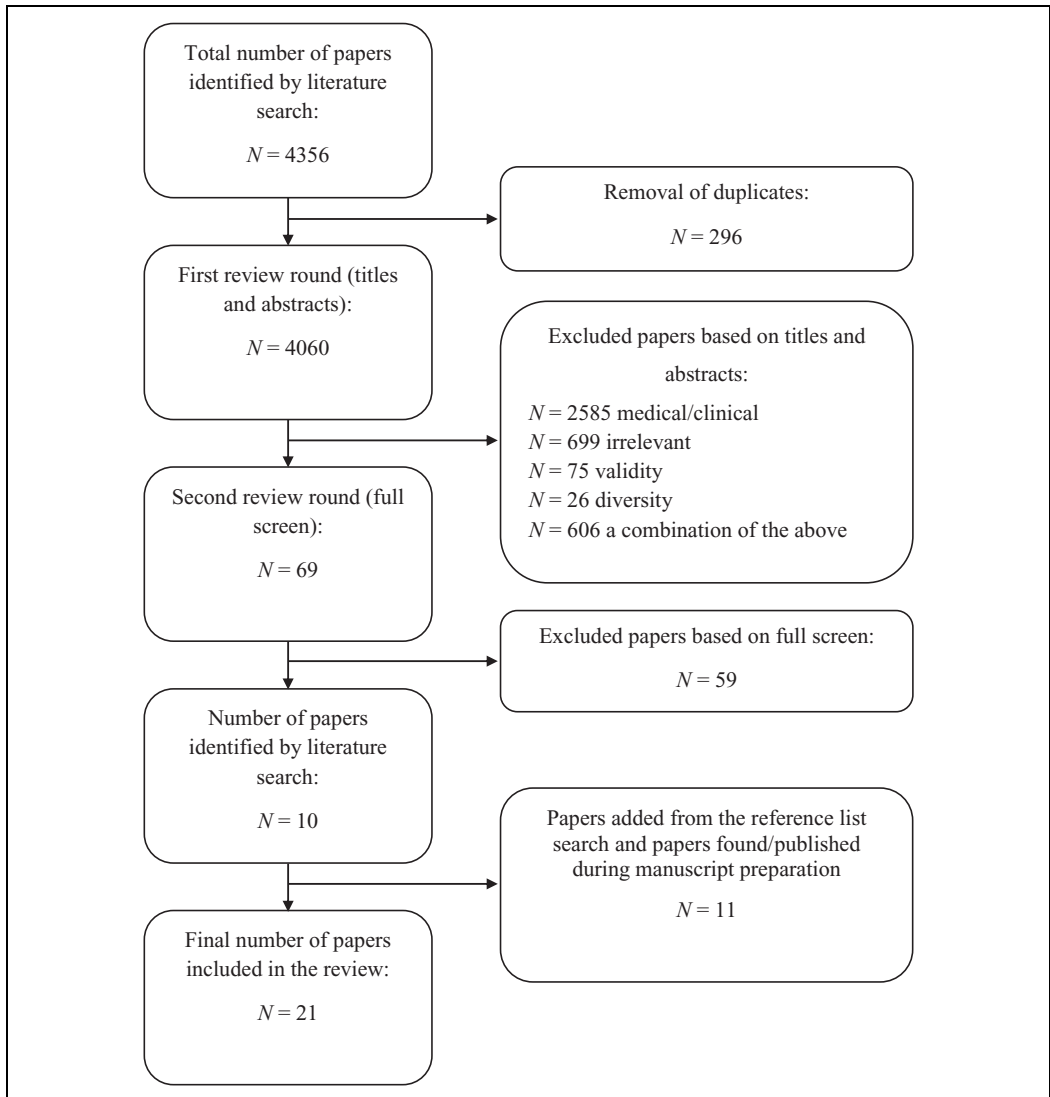
Figure 1. Flow chart of the academic literature selection process.

styles and personality traits. Almost all studies focused on information collection.

Professional and educational background. A practical rather than theoretical explanation for the science-practice gap may be that practitioners have diverse educational backgrounds, including (human resource) management, I/O psychology,

and other academic or professional qualifications, and hence also differ in their knowledge of the selection- and decision-making literature (Jackson et al., 2018). Jackson et al. (2018) compared perceptions of validity to the actual validity of 13 employee assessment methods among U.K. practitioners ($n = 193$) and laypeople ($n = 283$). They found that occupational

Table 1. Study characteristics of the included studies.

Study topic	Article	Study	Study design	Collection or combination	Participants	Sample size (N)
Practitioner characteristics	Jackson et al. (2018)	1	Online survey	Collection	U.K. practitioners and laypeople	476
Practitioner characteristics, Motivational factors	Lievens and De Paepe (2004)	2	Online survey	Collection	Belgian HR professionals	127
Practitioner characteristics	Roulin et al. (2019)	1	Survey	Collection	Canadian experienced interviewers	131
Practitioner characteristics, Motivational factors	Tsai et al. (2016)	1	(Online) survey	Collection	Experienced interviewers (HR professionals and line managers)	327
Practitioner characteristics	Lodato (2008)	1	Online survey	Both	HR professionals mostly from the U.S.	206
Practitioner characteristics	Lodato et al. (2011)	1	Online survey	Both	HR professionals mostly from the U.S.	206
Communication/presentation of scientific evidence	Zhang et al. (2018)	1	Online experiment	Collection	Amazon Mturk adult sample	533
Communication/presentation of scientific evidence	Zhang et al. (2018)	2	Online experiment	Collection	Amazon Mturk full-time employees with management/recruiting duties	156
Communication/presentation of scientific evidence	Zhang et al. (2018)	3	Online experiment	Collection	Amazon Mturk adult sample	306
Communication/presentation of scientific evidence	Voss and Lake (2020)	1	Online experiment	Collection	Psychology students	246
Communication/presentation of scientific evidence					Amazon Mturk sample of working adults without interview experience	245
Communication/presentation of scientific evidence					Amazon Mturk sample of hiring managers	144
Communication/presentation of scientific evidence	Highhouse et al. (2017)	1	Online experiment	Collection	HR professionals	201
Communication/presentation of scientific evidence	Zhang et al. (2019)	1	Online experiment	Collection	U.S. Amazon Mturk adult sample	253
Communication/presentation of scientific evidence	Zhang et al. (2019)	2	Online experiment	Collection	Prolific (online crowdsourcing platform) sample of employees with hiring experience	197
Feedforward and outcome feedback	Arkes et al. (1986)	1	Lab experiment	Combination	U.S. undergraduates	226
Feedforward and outcome feedback	Jackson et al. (2019)	1	Online experiment	Both	Amazon Mturk sample	154

(continued)

Table 1. (continued)

Study topic	Article	Study	Study design	Collection or combination	Participants	Sample size (N)
Feedforward and outcome feedback	Jackson et al. (2019)	2	Online experiment	Combination	Hiring professionals	519
Feedforward and outcome feedback	Thiele et al. (2020)	1	Online experiment	Combination	Hiring professionals	900
Feedforward and outcome feedback	Dietvorst et al. (2015)	1	Online experiment	Combination	Most likely students	361
Motivational factors	Dietvorst et al. (2015)	2	Online experiment	Combination	Most likely students	206
Feedforward and outcome feedback	Dietvorst et al. (2015)	4	Online experiment	Combination	Most likely students	354
Feedforward and outcome feedback	Yankelevich (2010)	1	Online experiment	Combination	U.S. adult full-time job occupants	323
Motivational factors	Dietvorst et al. (2018)	1	Online lab experiment	Combination	Most likely students	288
Motivational factors	Dietvorst et al. (2018)	2	Online experiment	Combination	Amazon Mturk sample	816
Motivational factors	Dietvorst et al. (2018)	3	Online experiment	Combination	Amazon Mturk sample	818
Motivational factors	Nolan and Highhouse (2014)	1	Online experiment	Both	Amazon Mturk job occupants	296
Motivational factors	Nolan and Highhouse (2014)	2	Online experiment	Both	Amazon Mturk job occupants	142
Motivational factors	Nolan (2013)	1	Online experiment	Both	Amazon Mturk job occupants	296
Motivational factors	Nolan (2013)	2	Online experiment	Both	Amazon Mturk job occupants	259
Stakeholder perceptions	Nolan et al. (2016)	1	Online experiment	Both	Amazon Mturk job occupants	468
Stakeholder perceptions	Nolan et al. (2016)	2	Online experiment	Both	Amazon Mturk job occupants with experience in employee selection	150
Stakeholder perceptions	Nolan et al. (2020)	1	Online experiment	Collection	Amazon Mturk job occupants with hiring experience	360
Stakeholder perceptions	Diab et al. (2011)	1	Online experiment	Both	U.S. and non- U.S. adult job occupants	418

psychologists' (OP) estimates matched actual validity more closely than those of practitioners with a human resource management (HRM) degree ($d = 0.61$), and with a professional HR qualification (HRq, $d = 0.75$). No statistically significant differences were found between HRM or HRq practitioners and laypeople ($d = -0.45$ and -0.47, respectively). OP-practitioners also reported the largest discrepancies between their validity perceptions and their perceptions of frequency of use of assessment methods, suggesting that they are most aware of the science-practice gap. In contrast, in a survey of 127 Belgian HR professionals, Lievens and De Paepe (2004) found no significant relationship between the use of high-structure interviews and holding a degree in I/O psychology ($r = .07$).

Experience and training. Experienced practitioners may not use evidence-based assessment and selection practices because they are overly confident in their own judgments (Arkes et al., 1986; Kausel et al., 2016; Kleinmuntz, 1990). Roulin et al. (2019) explored the relationship between interviewer experience and the use of interview structure components without utilizing a specific theory or formulating hypotheses. Components such as question consistency (asking the same questions consistently across candidates), asking sophisticated questions, taking notes, and evaluation standardization increase interview structure, while rapport building and probing reduce structure (Chapman & Zweig, 2005). Results from a survey among experienced Canadian interviewers ($N = 131$) showed that more experienced interviewers engaged in more probing ($r = .18$) and asked less sophisticated questions ($r = -.30$), but also engaged in more note-taking ($r = .27$). Furthermore, having received interviewer training was positively associated with higher question consistency ($r = .33$), more note-taking ($r = .28$), and higher evaluation standardization ($r = .39$, Roulin et al., 2019). Similarly, Lievens and De Paepe (2004) also found a moderately positive relation between interviewer training and the use of

higher interview structure ($r = .37$). However, they did not find a significant positive association for months of interviewing experience ($r = .05$). In contrast, Lodato (2008) surveyed HR professionals ($N = 206$) and found that more years of experience in HRM and being certified as a senior professional in HR was weakly negatively related to a preference for making intuition-based hiring decisions ($r = -.20$ and $-.21$, respectively). However, it was not significantly related to a preference for making mechanical hiring decisions ($r = .14$ and .13, respectively).

Decision-making styles. Practitioners may be aware of the superiority of evidence-based assessment practices, but they may not implement them (Fisher, 2008; Highhouse, 2008; Phillips & Gully, 2008; Rynes et al., 2002) because their preferred decision-making styles are more in line with intuitive assessment approaches (Lodato, 2008). According to cognitive experiential self-theory (CEST, Epstein et al., 1992), people with an experiential decision-making style prefer to rely on initial feelings when making decisions. Furthermore, decisive people tend to make sense of much information quickly, while indecisive people may prefer structure and support when making decisions. Based on CEST, Lodato (2008) found that an experiential decision-making style was strongly positively correlated with preferences for intuition-based hiring ($r = .64$). However, there were no significant relationships between decisiveness and preference for an intuitive-hiring approach ($r = -.02$), or between rational-thinking style and preference for a mechanical hiring approach ($r = .12$). Moreover, work experience and having a senior HR certification did not explain any additional variance in intuition-based hiring preference over decision-making style (Lodato et al., 2011). Notably, the scale for intuitive-hiring preference that served as the criterion in this study may be viewed as a contextualized measure of the experiential decision-making style scale, which could explain the strong relationship.

Personality. Practitioners' personality characteristics may also be related to the use of specific components of the structured interview. Extraverted practitioners may engage more in rapport building and probing, because these behaviors increase their share of the conversation. Indeed, Roulin et al. (2019) found that more extraverted interviewers used more rapport building ($r = .31$) and probing ($r = .24$). Furthermore, conscientious practitioners who are more organized and self-disciplined may engage in note-taking more often and may prefer standardized evaluation and question consistency. Conscientious interviewers were more likely to increase interview structure through using evaluation standardization ($r = .22$, Roulin et al., 2019).

Tsai et al. (2016) also investigated the relationship between personality traits and intentions to use a structured interview in a sample ($N = 327$) of HR professionals and line managers. Agreeableness and conscientiousness were positively related to interviewers' intention to use a structured interview ($r = .26$ and $r = .22$, respectively), and the relationship between conscientiousness and use intentions was stronger when interviewers were accountable for the outcomes of the selection procedure. However, the expected negative relationship between extroversion and intentions to use structured interviewing was not found. Lastly, based on Holland's (1997) RIASEC model, Lievens and De Paepe (2004) showed that interviewers who scored high on the conventional dimension (i.e. methodical, systematic) also used higher interview structure ($b = 0.20$). Unexpectedly, interviewers who scored higher on the social dimension and interviewers who considered themselves to be a good judge of human character did not use significantly less interview structure.

In sum, a variety of practitioner characteristics have been investigated in six different studies. Research on practitioners' educational background produced mixed findings. In one study, professionals with a background in I/O psychology seemed most aware of the validity of different information collection methods. However, another study found no relation between such a background and the degree of structure used in interviews. Furthermore, the results of two studies suggest that training rather than experience is related to the use of higher structured interviews. Moreover, results from one study on decision-making styles suggest that HR professionals who prefer to make intuitive decisions in everyday situations also tend to rely more on their intuition in employee selection. In terms of personality, two studies found that conscientiousness was related to the use of structured interviewing.

Communication and presentation of scientific evidence

The goal of several studies was to investigate the effect of the way validity information is presented on the understanding and perceived advantages of evidence-based assessment. Examples of presentation formats are non-traditional metrics, graphical visual aids, and narrative stories. Theoretical perspectives on why these alternative presentation formats would improve validity communication were, except for two studies (Highhouse et al., 2017; Zhang et al., 2019), not explicitly described. All of these studies communicated the advantage of structured- over unstructured interviews. So, they solely focused on information collection.

Non-traditional metrics. Non-traditional metrics that may yield more successful validity communication than the—often not well understood—correlation coefficient (Huberty, 2002; Voss & Lake, 2020) include, for example, the binomial effect size display (a tabular display that shows the change in success rate that is attributable to an intervention such as the use of particular tests in a 2×2 Table, Rosenthal & Rubin, 1982). Voss and Lake (2020) conducted

an experiment among students ($N = 246$), and found that communicating the validity difference between an unstructured- and structured interview with non-traditional metrics (six variations), compared to traditional metrics (r and R^2), increased participants' self-reported understanding, which in turn increased their perceived usefulness of the structured interview. Yet, these effects were not replicated in a sample of working adults without hiring experience ($N = 245$) and in a sample of hiring managers ($N = 144$).

Visual aids. Validity information may also be communicated with graphical visual aids like icon arrays (a graph consisting of icons that symbolize individuals, Galesic et al., 2009) and expectancy charts (a less sophisticated graphical visual aid, Guion, 2011). However, graph literacy may still be required to correctly interpret the graphs (Galesic et al., 2009). Zhang et al. (2018) conducted an experiment among laypeople ($N = 533$), providing them with validity information about a structured- (small or large validity) and unstructured interview, and random selection. Validity information was presented in the form of a tabular display, a graphical expectancy chart, or a graphical icon array. Participants perceived an icon array as more useful for communicating the validity advantage of a structured interview over an unstructured interview and random selection compared to a tabular display ($d = 0.28$) and an expectancy chart ($d = 0.22$), when the validity of the structured interview was small and large. However, among employees with management or recruiting/interviewing as main work duties ($N = 156$), the results were only replicated for participants with high graph literacy, who perceived the structured interview to be more useful when its validity was presented graphically ($R^2 = .16$). Furthermore, Zhang et al. (2018) found that, controlling for hiring experience, laypeople from the U.S. ($n = 158$) were more willing to use a structured interview when validity information was presented with an expectancy chart ($d = 0.53$) or an icon array ($d = 0.46$), compared to a tabular presentation. However, no such effect was found in a non-U.S. sample of laypeople ($n = 148$).

Evaluability. When practitioners are unfamiliar with correlation coefficients, they may find it difficult to differentiate between the validity of methods and they may underestimate the utility of valid assessment instruments (Brooks et al., 2014; Muchinsky, 2004). General evaluability theory (Hsee & Zhang, 2010) suggests that presenting validity information of different instruments jointly rather than separately provides context for the evaluation of values and therefore aids evaluability. Furthermore, based on the theory of sensory perception (Volkmann, 1951), presenting practically realistic upper validity limits of assessment instruments should help people in evaluating validity information (Highhouse et al., 2017).

Highhouse et al. (2017) used hypothetical hiring scenarios in which HR professionals ($N = 201$) evaluated structured- and unstructured interviews. These instruments and their validities were presented jointly or separately, and half of the participants were also provided with a range of validity coefficients of other commonly used instruments in personnel selection. Participants slightly preferred the unstructured interview over the structured interview when these instruments were presented separately, but preferred the structured interview when presented jointly ($d = 0.61$). In addition, professionals preferred the structured interview regardless of a separate or joint presentation when they also received a range of commonly observed validity estimates of several assessment instruments ($d = 0.39$).

Storytelling. Communicating information about evidence-based assessment in the form of a story rather than advice may increase positive attitudes toward evidence-based assessment. Transportation theory (Green & Brock, 2000) suggests that stories transport the reader's focus to the story's character rather than the

persuasive message. This may reduce the reader's counter arguing and in turn increase positive attitudes toward evidence-based assessment (Zhang et al., 2019).

To investigate these hypotheses, Zhang et al. (2019) presented a sample of adults ($N = 253$) recruited via MTurk with a script that described a structured interview and reasons for its underutilization. Participants either read a story in which an I/O consultant described the development and implementation of a structured interview for a company without mentioning that it was superior, or a script in which structured interviews were clearly recommended based on their superiority compared to unstructured interviews. Participants who read the story counter argued less than participants who received direct advice, and consequently reported more positive attitudes toward the structured interview.

In a second study, experienced hiring professionals ($N = 197$), reported their attitudes toward the structured interview before and after reading a similar script as in study 1 (Zhang et al., 2019). Compared to a pretest, attitudes toward the structured interview increased slightly after reading either of the scripts ($d = 0.22$), and the effect was slightly stronger for participants who read the story (no effect sizes reported). Replicating the results from study 1, participants who read the story were more engaged with its persuasive message and counter argued less, which translated into higher intentions to use the structured interview.

Overall, nine studies showed that the effects of different presentation modes of validity information were small and seemed to depend on individual and demographic differences. Furthermore, a complication of using tabular or visual displays to communicate validity evidence in the form of the expected number of successful hires is that a base rate needs to be assumed, which can substantially differ across job selection contexts (50% was assumed in Zhang et al., 2018). The results also suggest that displaying a range of validity coefficients of possible assessment instruments facilitates decision makers' evaluation of validity information. Lastly, advocating structured interviews in the form of a story compared to direct advice indirectly increased hiring professionals' intentions to use the structured interview.

Feedforward and outcome feedback

The use of evidence-based assessment methods may be increased by informing practitioners what predictors to use and how to combine them before engaging in a prediction task (feedforward). Alternatively, information could be provided in the form of outcome feedback, which could increase the use of evidence-based assessment practices because practitioners recognize their own limitations (Slaughter & Kausel, 2014). For example, practitioners may recognize that test scores predict job performance better than their unstructured interview ratings. Studies that investigated feedforward and outcome feedback have mainly focused on the use of combination methods and were not explicitly based on a theoretical framework.

Feedforward. In an experiment with U.S. undergraduate students ($N = 226$), Arkes et al. (1986) showed that participants who were informed that mechanical rule predictions are more valid than holistic predictions used an available mechanical rule more consistently, and hence made more accurate academic performance predictions, than participants who were informed that holistic predictions are more valid than mechanical rule predictions, or than participants who were told that both methods are equally valid. Similarly, Jackson et al. (2019) presented participants with some hiring experience ($N = 154$) with 10 applicant pairs and their scores on three predictors. For each pair, participants predicted an applicant's performance percentile rank and decided whom to hire. Half of the participants was shown the validity of the predictors and a rule they could use to combine them, while the other half was

not. Various simulated predictor validities were used, which were unrealistically high ($R^2 = .96$ and .50, respectively). They found that when participants had access to the rule and predictor validity was lower, predictions of participants with little hiring experience matched the rule's predictions more closely than those of participants with a lot of experience. In addition, experienced participants seemed to have learned from the outcome feedback provided in this study since their predictions matched the rule's predictions more over time.

Outcome feedback. Five studies investigated the effect of presenting decision makers after each individual prediction with outcome feedback on the use of mechanical rules, of which four studies found negative effects and one study found no effects. In the same study as reported above, Arkes et al. (1986) also found that participants who received outcome feedback after each prediction used the mechanical rule less consistently, and hence made less accurate predictions, than participants who did not receive feedback. Similarly, in a second experiment with hiring professionals ($N = 519$), Jackson et al. (2019) manipulated whether participants received outcome feedback, and also included a realistic validity condition ($R^2 = .20$). Just like Arkes et al. (1986) they found that participants who received outcome feedback made hiring choices (but not performance predictions) that were significantly *less* likely to match the rule. Furthermore, they found multiple complex interactions between feedback, number of predictions, the presence of a decision rule, and predictor validity.

Conversely, Thiele et al. (2020) found no effect of outcome feedback on matching mechanical rule predictions or hiring choices. Using a modified version of Jackson's et al. (2019) prediction task, they asked hiring professionals ($N = 900$) to predict the job performance of 20 applicant pairs based on three predictors and an available mechanical rule. In a 2×2 between-subjects design, they manipulated whether participants received outcome feedback on their own predictions or not, and on the mechanical rule predictions or not. Both feedback types resulted in near-zero effects.

In contrast to the studies mentioned above, Dietvorst et al. (2015) conducted an experiment in which they investigated the effect of outcome feedback on what method (mechanical rule vs. own intuition) participants chose to use for making subsequent incentivized predictions. They found that students ($N = 361$) chose to use the rule predictions less often than their own intuitive predictions when they had previously experienced the rule's prediction performance and—importantly—its error. Interestingly, making their own predictions—and therefore seeing their error—did not diminish their reliance on their own predictions. Most notably, participants that had experienced both their own and the rule's prediction performance were *least* likely to use the rule predictions. Furthermore, results from another experiment ($N = 354$) also showed some, albeit less strong, evidence that students who had experienced the rule's prediction performance chose to use the rule predictions less often than intuitive predictions provided by another person. This suggests that people tolerate error less if made by a rule than by a human.

Practitioners may also resist mechanical judgment because it makes prediction error more salient (Dawes, 1979; Highhouse, 2008) as compared to holistic judgment (Yankelevich, 2010). Mechanical judgment results in a quantified prediction that is visible and hence can be more easily compared to an outcome than an ambiguous holistic judgment. Two aspects of prediction error are the error rate and the error type. Error types can be false positives, that is, selecting an unsuited candidate, or false negatives: rejecting a suited candidate. Yankelevich (2010) hypothesized that participants would evaluate a hypothetical hiring procedure as more useful when the error rate is low, and when errors are false negatives, as false positives are likely perceived as more costly than

false negatives (Martin, 2008). Furthermore, Yankelevich (2010) hypothesized that reporting false positives would affect the usefulness perceptions of mechanical judgment more negatively than perceptions of holistic judgment. U.S. full-time job occupants ($N = 323$) evaluated a hypothetical hiring procedure as more useful when the error rate was low ($\eta^2 = .04$), and when a holistic approach was used ($\eta^2 = .05$). However, the hypothesized interactions between judgment approach and error type, judgment approach and error rate, and the main effect of error type were not found. Furthermore, the hiring procedure was perceived as more legal when holistic combination was used ($\eta^2 = .02$) and when the errors led to false positive rather than false negative decisions ($\eta^2 = .02$).

In conclusion, six studies provide strong support that feedforward increases decision makers' rule use, while feedback that demonstrates a rule's performance—and thus its error—*decreases* decision makers' rule use (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019). Furthermore, the results of one study showed that people find more accurate hiring procedures slightly more useful (Yankelevich, 2010). Notably, some of the reviewed studies used prediction tasks with dubious predictors such as race and socio-economic status (Dietvorst et al., 2015, 2018) and assumed unrealistically high predictor validities (Jackson et al., 2019), which are rarely found in practice.

Motivational factors

Practitioners' motivation to use evidence-based assessment can be determined by internal factors, such as psychological needs, or by external factors, such as incentives, accountability, and the goal to either attract or select applicants. We identified studies investigating the effects of internal and external motivational factors on the use of evidence-based information collection and combination methods.

Internal motivational factors

Fulfillment of Basic Psychological Needs. According to self-determination theory, people strive to satisfy their three main needs of autonomy, competence, and relatedness (Deci & Ryan, 2000). Structured and mechanical information collection and combination methods may offer less potential to satisfy these needs than their unstructured and holistic counterparts. For example, structured interviews and mechanical combination may violate autonomy needs because practitioners are bound to pre-defined questions and mechanical rules, which restricts their expression of idiosyncratic preferences (Dipboye, 1994). Similarly, competence needs may be violated if practitioners cannot demonstrate their ability to come up with spontaneous candidate-tailored questions or to detect assumed mechanical rule exceptions and complex predictor interactions (Meehl, 1954b, p. 24). Lastly, a structured interview may violate relatedness needs because increased structure disrupts communication flow and prevents interaction quality (Dipboye et al., 2012). Hence, practitioners may be less likely to use structured and mechanical selection approaches if these violate basic human needs.

Autonomy. Autonomy needs may be satisfied in two different ways. With regard to mechanical information combination, practitioners could be allowed to adjust mechanical rule predictions. Alternatively, practitioners could design mechanical rules by choosing the predictor weights. Results from an experiment by Dietvorst et al. (2018) showed that students ($N = 288$) chose to use a rule more often if they could adjust its prediction. Participants who could restrictedly adjust the rule's predictions chose the rule much more often (73–76%) than participants who could not change the rule's predictions (32%). Because the participants who could not change the rule's predictions chose to use their own holistic predictions more often, they were less accurate than participants who could adjust the rule's predictions.

Dietvorst et al. (2018) also conducted a second experiment where adults recruited via MTurk ($N = 816$) were assigned to conditions in which they could adjust the rule's predictions by either varying degrees, or not at all. Again, participants chose to use the rule significantly more often when they could adjust its predictions (70% vs. 47%). However, no differences were found across groups with different degrees of allowed adjustment. Lastly, a third experiment with adults recruited via MTurk ($N = 818$) showed that allowing people to adjust a rule's prediction increased the likelihood of choosing the rule without the possibility to adjust its predictions in subsequent predictions. In a first stage of incentivized predictions, participants were randomly assigned to conditions in which they could adjust the rule predictions freely, adjust the rule predictions a little, or could not adjust the rule predictions. In a second stage, participants chose between purely using their own predictions, purely using the rule predictions, and adjusting rule predictions. Participants that could adjust rule predictions in the first stage were more likely to use rule predictions that they could not adjust in the second stage.

Decision makers' autonomy needs may also be satisfied if they have control over the design of evidence-based assessment procedures. In an online experiment, Nolan and Highhouse (2014) asked U.S. job occupants ($N = 296$) to imagine that they had to fill their old position after being promoted. The interview structure and method to combine interview attributes was manipulated. Participants reported perceiving more autonomy when using an unstructured interview and a holistic combination method than when using a structured interview ($d = 1.79$) and a mechanical combination method ($d = 0.32$).

In a second experiment, only scenarios describing a structured interview and mechanical combination were presented, which varied in autonomy potential. U.S. job occupants ($N = 142$) reported higher use intentions for a hiring approach when they could form the mechanical rule by choosing the interview attribute weights (more autonomy), compared to using organization-determined attribute weights (less autonomy, $\eta^2 = .07$). No differences in use intentions were found between participants who could choose the interview attributes and questions (more autonomy), compared to participants who used organization-determined attributes and questions (less autonomy, $\eta^2 = .00$). Nolan and Highhouse (2014) also found an interaction effect: use intentions were lowest when both the information collection and information combination method offered less autonomy, but use intentions were highest when the information combination method offered more autonomy and the information collection method offered less autonomy ($\eta^2 = .04$). Similarly, in their study of 127 Belgian HR professionals, Lievens and De Paepe (2004) found that a desire to retain autonomy was negatively related to the use of higher interview structure ($r = -.19$).

Competence. In two online experiments, Nolan (2013) found that U.S. job occupants ($N = 296$ and $N = 259$) also perceived more potential for competence in an unstructured interview ($\eta^2 = .25$) and in holistic information combination ($\eta^2 = .02$), and that this *perceived* competence potential was strongly related to use intentions ($r = .75$). The samples, designs and manipulations reported in Nolan (2013) were similar as those reported in Nolan and Highhouse (2014).

Relatedness. Nolan (2013) also found that U.S. job occupants perceived more potential for relatedness in an unstructured interview ($\eta^2 = .66$). Furthermore, their perceived relatedness potential was strongly related to use intentions ($r = .69$). In line with these results, Lievens and De Paepe (2004) showed that interviewers' desire to establish personal contact was negatively related to the use of higher interview structure ($r = -.28$).

External motivational factors

Incentives. Although it may not be common to directly incentivize practitioners for their hiring decisions, increased success ratios should serve as incentives for practitioners to use evidence-based assessment methods (Cook, 2016). Hence, incentivizing accurate decision-making may result in increased use of evidence-based assessment. The studies that investigated the effect of incentives on the use of evidence-based assessment methods were not based on any theory. Paradoxically, Arkes et al. (1986) showed that participants who received monetary incentives made fewer correct predictions than participants who did not receive incentives. Incentivized participants performed worse because they used a mechanical rule less consistently after negative outcome feedback, although this effect was not statistically significant (Arkes et al., 1986).

In the second experiment described in Dietvorst et al. (2015), they investigated the effect of the type of incentive on choosing to use a mechanical rule. Students ($N = 206$) were randomly assigned to one of three types of incentives they would receive in a prediction task: $1 when their prediction was within a narrow range of the true criterion score, $1 when their prediction was within a wider range, or an incentive (between $1 and $10) based on average absolute error. Participants that were paid when predictions were within a wider range were less likely to choose the rule (13%) than participants in the narrow range condition (26%) and the absolute average error condition (34%). Thus, participants chose to use their own predictions relatively more often than the rule's predictions when the incentive was easier to achieve.

Accountability. Practitioners who feel more accountable for the selection process and outcome may be more likely to use evidence-based assessment (Brtek & Motowidlo, 2002). With regard to the use of structured interview components, Tsai et al. (2016) hypothesized and found that interviewers who felt being more accountable for the procedure or the outcome of the interviews they usually conduct intended to use structured interviews more often ($r = .13$ and $r = .15$, respectively).

Goal of the Procedure. In a study mentioned above, Roulin et al. (2019) also showed that interviewers who reported that their goal was more to select rather than attract applicants asked more sophisticated questions ($r = .26$), asked questions more consistently ($r = .19$), and engaged in more standardized evaluation ($r = .27$) and note-taking ($r = .20$).

Overall, eight studies showed that internal motivational factors (i.e., psychological needs) play an important role in the use of evidence-based assessment. With regard to information combination, the studies consistently showed moderate to strong evidence that decision makers' rule use increased when they retained some autonomy. Adjusting rule predictions by only a small amount increased rule use and, importantly, decision makers' likelihood to use unadjustable imperfect rules. Furthermore, practitioners showed higher intentions to use a rule when they were involved in its design. With regard to information collection, the results are less consistent but still suggest that psychological need satisfaction affected the use of structured interviews. In contrast, four studies showed that the effect of external motivational factors on the use of evidence-based combination methods is mixed. Incentives decreased mechanical rule use, which decreased prediction accuracy. In addition, decision makers preferred making holistic predictions when incentives were easier to achieve. Regarding information collection, interviewers who feel being more accountable for the interview procedure and the outcome may be slightly more likely to use structured interviews. Lastly, interviewers whose goal is to select rather than attract applicants seem to use slightly more structure in their interviews.

Stakeholder perceptions

Stakeholder perceptions can also influence practitioners' implementation of evidence-based assessment and selection. For example, applicant perceptions majorly influence HR practitioners' choice of assessment instruments (König et al., 2010). Furthermore, stakeholders such as employees may devalue practitioners' status when decisions result from evidence-based assessment. Attribution theory's discounting principle (Kelley, 1973) suggests that people can attribute the cause of an event to internal and external factors. Applied to selection, people may ascribe selection decisions based on, for example, unstructured interviews and holistic combination to the practitioner (internal factor), but not if decisions are based on standardized tests and mechanical rules (external factor). If practitioners receive less credit for their decision outcomes, their professional status may be threatened (Meehl, 1986), which may decrease the use of evidence-based assessment. So, the goal of the reviewed studies was to test how stakeholder perceptions and the organizational context may influence practitioners' use of evidence-based assessment practices. The identified studies both investigated information collection and combination.

Stakeholder perceptions of selection decision aids. In a between-subjects experiment, Diab et al. (2011) investigated job occupants' ($N = 418$) usefulness- and legality perceptions of different hypothetical assessment procedures. These procedures varied according to the information collection method (structured interview vs. paper-and-pencil test), information combination method (holistic vs. mechanical), and job occupant's assumed role (applicant vs. employer). Neither the information collection method nor the assumed role affected usefulness- or legality perceptions ($\eta_p^2 = .00–.02$). However, the holistic combination method was perceived as more useful in a U.S. subsample ($d = 0.60$), but not in a non-U.S. subsample. In

contrast, effect sizes in the non-U.S. subsample suggested that participants preferred holistic combination for interviews ($d = 0.39$), but preferred mechanical combination for test scores ($d = 0.22$, Diab et al., 2011).

Devaluation of professional status. Based on attribution theory's discounting principle (Kelley, 1973), Nolan et al. (2016) hypothesized that stakeholders give less credit to HR managers for outcomes of evidence-based assessment methods. Job occupants ($N = 468$) who evaluated the hiring strategies of their HR manager in a hypothetical hiring scenario attributed the outcomes significantly less to the HR manager when the information collection (unstructured vs. structured interview) and information combination (intuitive vs. computer) methods were standardized ($\eta_p^2 = .14$ and $.15$, respectively). However, employment decision outcomes were also perceived as more stable for the standardized methods ($\eta_p^2 = .09$ and $.07$).

Secondly, Nolan et al. (2016) hypothesized that practitioners' potential awareness of these perceptions may explain their underutilization of evidence-based assessment methods. In an experiment with job occupants with hiring experience ($N = 150$), Nolan et al. (2016) found that practitioners' beliefs about employees' perceptions of their causality/control over the hiring process were lowest when a structured interview was combined with mechanical combination, and highest for the unstandardized counterparts. Furthermore, practitioners believed that employees perceive the hiring outcome to be more stable when the structured interview was used together with the mechanical combination approach. Moreover, results from a mediation analysis showed that practitioners' beliefs about employees' perceptions of their causality/control over the hiring process had a negative relationship with the perceived threat of their professional status as a result of the increasing use of standardized technology, which, in turn, had a negative relationship with use intentions. In a partial replication with job

occupants with hiring experience ($N = 360$), Nolan et al. (2020) showed that practitioners believed that other people in their organization would think they had less control over the hiring process when they would use a structured (vs. an unstructured) interview ($d = 0.94$). This, in turn, increased practitioners' perceived threat of professional status, which decreased intentions to use a structured interview.

In sum, the results of four studies imply that educating practitioners to use evidence-based assessment may not suffice. Even if they are aware of evidence-based assessment and selection practices, they also have good reasons to secure the perceived value of their contribution, which may be at risk when they use evidence-based assessment. However, the results suggest that it may be effective to show stakeholders such as applicants how subjective impressions can be quantified, as they struggle to imagine that subjective methods such as an interview can be objectively scored.

Professional literature

Method

Selection of articles

We conducted a second systematic search (from 2005 to 2020) in the databases PsycInfo and Business Source Premier, using the same search terms as in the academic literature search. We focused on articles published after 2005 to avoid reviewing an outdated discussion on evidence-based assessment and its implementation in the professional literature that may not accurately reflect current practice anymore. Rynes et al. (2002) surveyed professionals to investigate how frequently they read different periodicals. Based on their results, we searched the most frequently read periodicals for this review, including (in descending order of frequency) *HR Magazine, HR Focus, Human Resource Management Journal, Workforce, Fortune, Forbes, Harvard Business Review, Inc.,* and *Fast Company.* We could not access articles published in

Wall Street Journal, Human Resource Executive, Business Week, and *Human Resource Plann-ing Journal.*[1] Furthermore, we included *California Management Review* and *MIT Sloan Management Review*, which are considered 'bridge journals' that transfer evidence-based practices into practice (Rynes et al., 2007). The search resulted in 569 articles.

Given that many assessment practices in selection are not evidence-based, we expected to find a small number of publications that encourage the use of evidence-based assessment. This raised the questions how assessment for selection is discussed and recommended in general, and what the reasons for organizations' current assessment practices are. To answer these questions, the coding scheme differed from the one used for the academic literature search. We included an article if it described (1) an intervention that may encourage evidence-based assessment or factors related to its use in selection, (2) an organization's selection process, (3) constructs or selection methods and their validity, or (4) advice on what selection methods to use or how to make selection decisions. Two independent reviewers (authors 1 and 2) screened and coded the articles for inclusion (0 = exclusion, 1 = inclusion, absolute agreement 80%, $\kappa = .53$). Disagreements were resolved through discussion until consensus was reached. Eventually, 207 articles were included. Articles that described interventions to encourage evidence-based assessment or factors related to its use (inclusion criterion 1) were classified according to the same research topic scheme that was applied to the results of the academic literature. All other identified articles did not fit this scheme because they were not concerned with interventions that encourage evidence-based assessment or factors related to its use. Given the amount and content diversity of these articles, we used the process of thematic analysis as described in Braun and Clarke (2006) to identify other topics discussed in the professional literature.

Results

The identified articles constitute only a very small portion (less than 1%) of all published articles in the reviewed periodicals ($N = 48.220$), which shows that selection is an underrepresented topic. Furthermore, only 10 articles (5% of all articles about selection) discussed interventions to encourage the implementation of evidence-based assessment practices or factors related to its use. These articles could all be classified under the research topics practitioner characteristics (experience and training) and motivational factors (autonomy).

Experience and training

With regard to information collection, the only mentioned intervention was training professionals in asking (behavioral) interview questions (Bortz, 2018; Fernández-Aráoz et al., 2009; Holmes, 2019; Lytle, 2013; Tyler, 2005). With regard to information combination, the most commonly mentioned intervention was developing awareness of unconscious decision-making biases (Bortz, 2018; Fernández-Aráoz, 2007; Segal, 2006; Wright, 2016), although it is questionable whether such unconscious bias trainings are truly evidence-based (FitzGerald et al., 2019; Paluck & Green, 2009; Williamson & Foley, 2018).

Autonomy

Furthermore, increasing autonomy, as also mentioned in the academic literature, was mentioned in one article as a means to increase decision-rule use (Kuncel et al., 2014). Specifically, the authors suggest that mechanical rules can be used to narrow the candidate pool, after which the final selection decision may be determined by holistic judgment. Another article discussed structuring holistic judgments in case mechanical rules are strongly resisted (Kahneman et al., 2016). These authors suggest that practitioners

could independently predict candidates' performance and resolve differences in a group discussion moderated by a group leader who has access to all independent predictions.

Other topics discussed in the professional literature

Rather than discussing the implementation of evidence-based assessment, 85 articles (41%) discussed the constructs that are or should be assessed in selection procedures, thus information collection. Some articles acknowledged the importance of evidence-based constructs such as intelligence (e.g., Menkes, 2005), although it was suggested to assess intelligence in an interview. However, the majority was concerned with constructs such as curiosity, grit, creativity, integrity, empathy, community involvement, humor, passion, and "the tone of the conversation" (e.g., Anders, 2013; Buchanan, 2015; Colvin, 2014; Fisher & Yang, 2009; Gino, 2018; Harnish, 2013; Lee & Duckworth, 2018; Tkaczyk, 2014). Reasons to assess these constructs were to a large extent based on personal beliefs and experience instead of empirical evidence. Although the methods used to assess these constructs were rarely mentioned explicitly, it was often implied that the interview was the predominant instrument. Furthermore, numerous articles suggested that practitioners consider the criterion of "cultural fit" as important as job performance (e.g., Cappelli, 2019; Hennigan & Evans, 2018; Meinert, 2013; Rockwood, 2015). Most articles that discussed interviewing explicitly mentioned or implied the use of unstructured interviews, while only a few advocated structured interviews (e.g., Cappelli, 2019; Holmes, 2019; Moore, 2017). Remarkably, reasons mentioned for interview standardization and standardized testing were often organizational goals such as cost reduction or legal defensibility, rather than increased validity (Bateson et al., 2014; Freeman, 2014; Grensing-Pophal, 2006; Taylor Arnold, 2008).

Lastly, few articles discussed information combination. The articles rarely explicitly recommended a certain combination method, but often implied that information is predominantly combined holistically (Clifford, 2006; Fernández- Aráoz et al., 2009; Greco, 2007, 2009) with one exception that advocated a mechanical approach (Cadrain, 2010). In general, however, the literature search also revealed a considerable amount of misinformation. For example, some articles overstated the validity of cover letters (Fried, 2010), recommended dubious interview questions (Lee, 2017), or suggested to infer sincerity, genuine enthusiasm, and warmth from someone's voice (Buchanan, 2005), although empirical research showed that these beliefs and recommendations are false or unsupported (Levashina et al., 2014; Murphy et al., 2009). Furthermore, most articles provided little argumentations for the described approaches and recommendations, and arguments based on empirical research were rarely presented.

In conclusion, the volume of professional articles about selection in the periodicals we reviewed was small, and articles that discussed evidence-based assessment practices and their implementation were extremely rare. Remarkably, besides interview training, none of the very few specific approaches that practitioners suggested to increase evidence-based assessment emerged from the academic review. The professional literature seems to focus mostly on potentially relevant constructs and the "right" interview questions to ask in an interview, but not much on how to assess those constructs validly, and how to use those assessments to make selection decisions.

Discussion

Decades of selection research have produced valuable knowledge regarding evidence-based practices. Yet, many of these evidence-based instruments and procedures are rarely translated into practice (Highhouse, 2008; Ryan

& Sackett, 1987; Slaughter & Kausel, 2014; Thornton et al., 2010). Therefore, the main aim of this paper was to provide an overview of the scientific literature about factors associated with or interventions designed to promote the implementation of evidence-based assessment in the context of human performance prediction and selection. Two major findings of our academic literature review were that (1) the limited number of studies that have been conducted on this topic covered a wide variety of factors and theoretical frameworks (e.g., Deci & Ryan, 2000; Green & Brock, 2000; Hsee & Zhang, 2010; Kelley, 1973) and (2) there are currently few consistent findings that result in feasible recommendations. Furthermore, there were no studies concerned with factors related to evidence-based assessment practices in educational selection, such as college admission procedures, although some studies did utilize stimulus data from an educational context in experimental research (Arkes et al., 1986; Dietvorst et al., 2015, 2018). Given the importance of selection into educational programs and the many debates on this topic (Niessen & Meijer, 2017) we found this very surprising. Lastly, the review of the professional literature showed that very few articles discussed evidence-based assessment or how to encourage its use.

The academic review identified a couple of factors that seem less promising to increase the use of evidence-based assessment. For example, no conclusive evidence emerged that practitioners with specific educational backgrounds or experience in assessment would be more likely to use evidence-based assessment. Furthermore, providing decision makers with outcome feedback on their predictions, which has been suggested as a more useful intervention (Slaughter & Kausel, 2014), was studied extensively. However, outcome feedback *decreased* the use of evidence-based assessment. These less promising factors were rarely explicitly based on theory, but rather driven by practical explanations. Given their applied focus, it is ironic that these interventions seem

difficult to implement in practice. For example, outcome feedback is usually not available in practice. Hence, the practical value of this research remains unclear. Factors that may also be difficult to intervene on in practice concern differences in decision makers' personality traits or decision-making styles that hinder or aid the adoption of evidence-based assessment methods (Lodato, 2008; Roulin et al., 2019; Tsai et al., 2016). Another intervention that has been researched is to present validity information in formats that may be easier to understand for decision makers. Although this is a feasible intervention, positive effects on the use of evidence-based methods were rarely found in samples of practitioners with hiring experience. Therefore, this research seems most useful for convincing other stakeholders that may be involved in assessment.

A more promising intervention that emerged from the review and that has also been suggested to increase the use of evidence-based assessment (Slaughter & Kausel, 2014) is to provide decision makers with information on how to use and combine predictors (feedforward), which is also easier to implement in practice. Therefore, future research may investigate how such information should be presented such that the use of evidence-based assessment is most effectively increased (Dalal & Bonaccio, 2010). Research on internal motivational factors showed the most promising results to increase the use of evidence-based assessment methods, and hence more so than research on external motivational factors. This is in line with existing findings that supporting people's autonomous behavior is positively related to a variety of organizational outcomes (Deci et al., 2017) and acceptance of organizational change (Gagné et al., 2000). Since the implementation of evidence-based assessment requires change, fostering practitioners' autonomy seems more effective than controlling their use of evidence-based assessment practices via external factors such as incentives or accountability. Research on internal motivational factors was sometimes explicitly

based on self-determination theory (Nolan & Highhouse, 2014). Other researchers did not explicitly mention this theory, although their results also fit this framework (Dietvorst et al., 2018; Lievens & De Paepe, 2004). Self-determination theory seemed to be the most promising theoretical framework, and we encourage researchers who investigate interventions that satisfy psychological needs to use and extend this theory.

Although half of the reviewed studies were not clearly based on any theory, one other useful framework emerged from the review. Attribution theory's discounting principle (Kelley, 1973) was used in research on stakeholder perceptions, which showed that practitioners may underutilize evidence-based assessment methods because it decreases the professional status that stakeholders ascribe to them. Therefore, attribution theory may inform research that focuses on a tradeoff between decision maker contribution and standardization. Another useful theory that did not emerge from this review, but that has been applied in earlier selection research (van der Zee et al., 2002) is the theory of planned behavior (Ajzen, 1991). This theory suggests that one's intention to perform a certain behavior is influenced by three factors; one's attitude toward a certain behavior (attitude), one's perceived ease of performing a behavior (perceived behavioral control), and the expectations of others (subjective norms). Hence, this framework may be useful for future research that jointly investigates the effects of stakeholder perceptions, attitudes, and confidence.

In sum, a few promising interventions to increase evidence-based assessment emerged from the review. However, the relevance of some findings remains unclear because it is unknown to what extent the variables that were manipulated in experimental studies are representative of or implementable in practice, such as incentives (Arkes et al., 1986; Dietvorst et al., 2015) and outcome feedback (e.g., Thiele et al., 2020). Therefore, researchers should conduct some basic prevalence research and become

more engaged with practitioners who may provide valuable input on the feasibility of interventions, and on important boundary conditions that researchers may use when building theory (Campbell & Wilmot, 2018; Ployhart & Bartunek, 2019). Furthermore, the reviewed studies exclusively used cross-sectional designs and very often focused on the hiring interview. Therefore, future research may use longitudinal study designs and include other commonly used or recommended instruments, such as standardized tests, assessment centers, and personality questionnaires, and may also differentiate between assessed constructs (e.g., cognitive ability, personality) and assessment methods (e.g., tests, interviews, see Lievens et al., 2005). Ideally, future research should investigate professionals' behavior in real prediction tasks.

Although the academic review revealed that some interventions seem promising to increase evidence-based assessment in practice, these interventions, and evidence-based assessment in general, were barely discussed in the professional selection literature. Furthermore, articles in which organizations reported on the use of evidence-based assessment procedures were almost absent. This hinders the dissemination of these procedures because organizational assessment practices are also majorly influenced by the assessment practices of other organizations (König et al., 2010). Therefore, based on the theory of diffusion of innovations (Rogers, 2010; Rynes, 2012), a fruitful strategy for researchers may be to collaborate with influential organizations to implement evidence-based assessment. Other organizations may mimic this best practice (see also Holmes, 2019).

The results of the professional review also showed that practitioners discuss primarily the importance of various constructs, but hardly discuss how to assess these constructs validly. In the scientific literature, there exists large agreement on what constructs are important (Sackett et al., 2017). Researchers rather seem concerned with the measurement of those constructs (e.g., personality, Sackett et al., 2017).

Hence, researchers may provide practitioners with an accessible overview of the empirical value of different constructs for different criteria (e.g., cultural fit and job performance), and on how those constructs can be assessed validly. We speculate that one reason why evidence-based instruments such as cognitive ability tests are underutilized is that these instruments do not measure practitioners' constructs of interest and cannot be adjusted to measure those constructs. Practitioners may consider a method such as the unstructured interview to be highly flexible to measure any construct of interest, which would explain its popularity.

Practical recommendations

Training professionals in evidence-based assessment practices, like conducting structured interviews, may be a promising intervention to increase the use of evidence-based assessment (Lievens & De Paepe, 2004; Roulin et al., 2019). The review of the professional literature showed that some organizations already provide interview training, which suggests that they also consider training to be useful. Training may be tailored to characteristics of the setting and the interviewers. Extraverted people, for example, were more inclined to engage in probing (Roulin et al., 2019) and thus may benefit the most from receiving training on this component. However, no research has been conducted on the effect of training on the use of other evidence-based information collection- or combination methods.

Given that practitioner-oriented journals, which rarely discuss evidence-based assessment, are preferred by practitioners over academic journals, researchers may contribute more strongly to the professional literature, and publish tutorials on evidence-based assessment in open-access journals (e.g., Meijer et al., 2020). Moreover, since practitioners' educational background is related to their knowledge of evidence-based assessment methods (Jackson et al., 2018),

it is important that university programs teach evidence-based assessment.

Another promising and feasible intervention may be to allow practitioners some autonomy in designing evidence-based assessments. For example, practitioners who are involved in the design of interviews (e.g., choosing the question order) may use higher structured interviews (Lievens & De Paepe, 2004). Similarly, adjustable or practitioner-determined mechanical rules could be implemented, as this increased decision makers' use (intentions) of a rule (Dietvorst et al., 2018; Nolan, 2013). Furthermore, introducing some autonomy may have the additional advantage that practitioners' professional status is not considerably devalued by other stakeholders. In general, practitioners should be provided with predictor validity information and a mechanical rule, as this already increased rule use and accuracy (Jackson et al., 2019).

The science-practice gap may also be reduced by influencing the perceptions of other stakeholders like applicants and managers, as they seem to affect the adoption of evidence-based assessment practices (Diab et al., 2011; Nolan et al., 2016). This is important because these stakeholders may not be as well informed about evidence-based assessment as assessment professionals and hence do not feel a need for change (Lawler, 2007). The value (e.g., validity) of evidence-based assessment may be communicated to these stakeholders with non-traditional metrics, graphical visual aids, and stories (Voss & Lake, 2020; Zhang et al., 2018, 2019). However, we note that stories could also be used to promote non-evidence-based methods. In fact, storytelling and anecdotal forms of communication without any empirical support were abundant in the professional literature.

The results from the professional literature showed that practitioners may want to predict criteria such as cultural fit in addition to performance. If practitioners have reasons to predict multiple outcomes, they can still use standardized instruments and mechanical rules (Meijer et al., 2020). For example, based on primary data

or meta-analytic findings, practitioners could apply the formula presented in Murphy (2019), to experience how the multivariate validity of a composite changes based on different weighting schemes of predictors and criteria (e.g., cultural fit and job performance). Similarly, when facing the dilemma to maximize both predictive validity and diversity (Pyburn et al., 2008), practitioners could use Pareto-optimization methods for which there exist user-friendly online tools (Rupp et al., 2020).

Agenda for future research

In Table 2, we provide a summary of important research questions to address in future research. Based on the integration of the results of the academic- and professional review, we structured our agenda according to (1) general research questions (2) the aims of assessment and selection (3) the design of the selection procedure (4) the assessment professionals (5) other stakeholders, and (6) the evaluation of the selection procedure. We used a similar procedure as in the academic review to identify future research topics. However, the topics in the academic review and the future research agenda differ because we added to the research questions that derived from the reviews other important research questions that should be addressed. Two general research questions were identified that we think should guide future research on the implementation of evidence-based assessment in selection.

General research questions

One general research question that should be investigated is how information collection and combination are related. Table 1 shows that information collection and combination methods were often studied in isolation, but there is some evidence that they interact with regard to usefulness perceptions and use intentions (Diab et al., 2011; Nolan & Highhouse, 2014). It remains

Table 2. Future research questions related to the use of evidence-based practices in personnel- and educational selection.

Study topic	Research questions
General	1. Do interventions that increase the use of evidence-based assessment instruments also increase the use of evidence-based decision-making procedures and vice versa?
	2. Why are evidence-based assessment practices underutilized?
Aims of assessment and selection	3. Do researchers and practitioners view the aim of selection differently?
	4. (How) does the integration of multiple criteria affect the acceptance of evidence-based assessment for different stakeholders with different aims?
Design of the selection procedure	5. How do rule transparency, rule complexity, and practitioners' understanding of the rule affect rule use?
	6. Do practitioners prefer autonomy in the design or in the outcome of the selection procedure, and how do these different levels of autonomy affect predictive validity?
	7. Do practitioners overestimate the costs of evidence-based assessment methods, compared to traditional assessment methods (e.g., unstructured interviewing and holistic information combination)?
Assessment professionals	8. How can experienced professionals best be encouraged to use evidence-based assessment practices?
	9. How can we increase practitioners' confidence in evidence-based assessment?
	10. How can practitioners' error aversion be reduced?
Other stakeholders	11. What interventions convince client organizations and assessees/applicants to accept or ask for evidence-based assessment practices?
	12. How are organizational differences related to the use of evidence-based collection and combination methods?
Evaluation of the selection procedure	13. What factors determine whether organizations collect information for evaluation purposes?
	14. Can practitioners effectively learn from feedback, and if yes, how?

unclear whether interventions that increase the use of evidence-based instruments are also useful to increase the use of evidence-based decision-making procedures and vice versa (research question (RQ) 1 in Table 2). Importantly, a question that remains unanswered is whether interventions that increase only evidence-based information collection – or combination methods can effectively improve our decisions. For example, using moderately valid assessment instruments may not improve decision making much when the scores on these instruments are holistically combined.

Another general research question that should be explored is why evidence-based assessment practices are underutilized (RQ 2). Exemplary research showed that restricted autonomy and the recognition of negative stakeholder evaluations explain practitioners' intentions to use evidence-based assessment (Nolan et al., 2016; Nolan & Highhouse, 2014). To enhance theory development, qualitative studies that employ, for example, think aloud protocols may explore practitioners' reasoning behind their choice of information collection- and combination methods. Similarly, case studies of organizations that have recently implemented evidence-based assessment may be valuable to investigate antecedents to evidence-based assessment (for an example, see Holmes, 2019).

Aims of assessment and selection

In future research, qualitative studies could be used to shed more light on whether researchers and practitioners view the aims of selection differently (RQ 3). For example, practitioners may not realize that selecting a candidate implies a prediction. As an anonymous reviewer noted: "It is not clear that selection practitioners realize that they are essentially in a 'prediction market'." While researchers' primary aim is to make valid performance predictions, practitioners also use selection practices for other aims, such as improving the organization's brand (Russell & Brannan, 2016).

Assessment professionals and managers may also have different aims when selecting candidates. Nolan et al. (2016) showed that practitioners who did not expect to work together with an applicant considered the fit between the applicant's ability and the job demands more important than the fit between the applicant's values and the organization's characteristics. In contrast, practitioners who expected to work together with an applicant considered the fit between the applicant's values and the organization's characteristics more important. Future research may investigate whether collection and combination methods designed to predict multiple criteria can increase the use of evidence-based assessment of multiple stakeholders who have different aims for selecting candidates (RQ 4). Given that organizations consider fit important (Dipboye, 1994), experiments that focus on use intentions and predictive accuracy of different collection and combination methods may also, in addition to performance, include fit as an outcome variable.

Design of the selection procedure

Practitioners' acceptance of evidence-based assessment and selection practices may depend on the design and complexity of those practices. For example, some practitioners have the implicit belief that accurate predictions require a complex combination of predictors, which is a hallmark of expert judgment (Highhouse, 2008). If practitioners expect rules to be complex, they may be skeptical about *simple* mechanical rules (Bobko et al., 2007; Dawes, 1979) and think that their holistic predictions are more valid. Therefore, future research should investigate how rule characteristics, such as transparency and complexity, but also practitioners' understanding of mechanical rules relate to the use of such rules (RQ 5, see also Shin & Park, 2019).

Although intervention studies in which practitioners retained some autonomy in the selection procedure showed the most promising results to increase the use of evidence-based assessment (Dietvorst et al., 2018; Nolan, 2013), it remains an open question whether practitioners prefer autonomy in the design (Nolan & Highhouse, 2014) or in the outcome of a selection procedure (Dietvorst et al., 2018). In multi-stage selection procedures, practitioners may prefer to design information collection- and combination methods in a first stage, but prefer to adjust prediction outcomes in a second stage, because it balances autonomy and effort (Nolan & Highhouse, 2014). However, autonomy should not substantially decrease predictive validity and decision quality, and should result in substantial improvement compared to holistic procedures. Therefore, future studies should investigate predictive validity in combination with different levels and types of autonomy, in order to find a feasible balance between predictive validity and acceptance of procedures by practitioners (RQ 6, Kuncel, 2008; Kuncel et al., 2013).

Small adjustments of rule predictions and predictor weights should not decrease predictive validity much (Dawes, 1979; Dietvorst et al., 2018). However, when predictors differ substantially in predictive validity, suboptimal practitioner-determined weights may lower predictive validity (Murphy, 2019). Besides autonomy, future research may investigate the role of relatedness needs more closely since practitioners often engage in group discussions to make selection decisions (Bolander & Sandberg,

2013). Existing research showed that the fulfillment of relatedness needs was related to higher intentions to use structured interviews (Nolan, 2013). Relatedness needs may also be investigated in regards to other collection methods than the interview, and with more emphasis on information combination.

When assessment and selection procedures are designed, practitioners' use of evidence-based assessment methods may be influenced by financial costs (Klehe, 2004). Our professional review and existing research suggests that practitioners consider factors like costs at least as important as predictive validity (König et al., 2010). The costs of conventional selection practices (interview scheduling, CV checks) may not be reported in as much detail as the costs of evidence-based information collection methods (psychometric testing). Hence, the costs of evidence-based assessment practices may be overestimated (RQ 7). In future research, costs of evidence-based assessment practices can be contrasted with unstructured interviewing and holistic candidate discussions. Research topics like these are good examples where relevant research arises from the coproduction between academic research and practice.

Assessment professionals

There is some evidence that experienced professionals implement evidence-based assessment practices less often than less experienced professionals (e.g., Arkes et al., 1986; Roulin et al., 2019), because they are overly confident in the quality of their decisions due to their experience (Arkes et al., 1986; Kausel et al., 2016). Therefore, future research may focus more specifically on experienced professionals (RQ 8), who may have a strong influence on current assessment practices. Specifically, it may be investigated how we can increase practitioners' confidence in evidence-based assessment (RQ 9). Practitioners often believe that more information is always better, although adding invalid to valid information can decrease decision makers' prediction

accuracy, but increase their confidence in holistic judgments (Dana et al., 2013; Kausel et al., 2016). Hence, future research may focus on contextual factors that influence practitioners' choice of information collection methods by, for example, identifying interventions that increase confidence in judgments that are based on less, but valid information.

Furthermore, future research could investigate whether interventions that reveal the inconsistency in practitioners' judgment could increase their confidence in evidence-based assessment (Kahneman et al., 2016). Recently, practitioners have discussed whether using mechanical rules could also increase diversity due to consistent judgment, as opposed to biased and inconsistent holistic judgment (Lam, 2015). Practitioners may be more convinced by the effect of consistent judgment on diversity (e.g., ratio of men and women) and fairness (treating all persons equally) outcomes than by performance outcomes, which may also be difficult to observe. Therefore, explaining the importance of consistency in evidence-based assessment may be more effective when focused on diversity and fairness than performance prediction. Ideally, such research would use realistic prediction tasks with qualitative and quantitative predictor information.

Also, in future research it should be investigated how practitioners' error aversion can be effectively reduced (RQ 10) so that they consider evidence-based assessment useful under realistic levels of predictive validity (Einhorn, 1986). Especially in high-stakes contexts such as executive selection, practitioners may be highly motivated and hence particularly averse to prediction errors, which may reduce their use of evidence-based information collection- and combination methods (Arkes et al., 1986).

Other stakeholders

In future research, a broader network of stakeholders should be investigated, such as applicants, assessment professionals, managers, and

client organizations of assessment professionals. Client organizations constitute a stakeholder group that has received little attention thus far but their influence is important as they sometimes demand practices from professionals such as expert judgment that are costly and at odds with evidence-based assessment (Yu & Kuncel, 2020). Future research may identify interventions that would convince clients to accept or ask for evidence-based assessment practices (RQ 11).

Another interesting question is how organizational culture relates to, and provides boundary conditions for the implementation of evidence-based assessment and selection practices (RQ 12, Potworowski & Green, 2012). For example, organizations that value critical thinking and transformations may be more receptive to an adoption of evidence-based assessment practices, while more hierarchical organizations may struggle to overcome organizational inertia.

Evaluation of the selection procedure

A crucial step in assessment is the choice of valid predictors (Dawes & Corrigan, 1974). Predictor validity information that is obtained from valid local research findings may increase acceptance among stakeholders (Kahneman et al., 2016). Yet, this requires that organizations systematically collect relevant information (e.g., interview and supervisor ratings), which is not common practice (Cappelli, 2019). Therefore, more research is needed that focuses on factors that explain the systematic collection of such information (RQ 13).

Although a plausible intervention is to provide practitioners with outcome feedback (Slaughter & Kausel, 2014), studies consistently showed that outcome feedback decreased the use of evidence-based practices (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019). Therefore, it is important to investigate whether practitioners can learn from feedback in selection, and if yes, what kind of feedback would be most effective (RQ 14). For example, the existing studies have

provided feedback after each prediction made (case-level feedback). Decision makers who receive case-level feedback experience that they almost always err, and as a result disregard a mechanical rule (Arkes et al., 1986). Therefore, receiving feedback (e.g., in the form of a correlation coefficient) after a number of predictions have been made (group-level feedback) may have more potential to increase the use of evidence-based assessment. Since allowing practitioners to design the selection procedure increased the use of evidence-based information collection- and combination methods (Lievens & De Paepe, 2004; Nolan & Highhouse, 2014), group-level feedback based on self-designed selection procedures may be most informative. Researchers may also vary the mode of the presented feedback. Existing research has usually provided feedback numerically (e.g., Jackson et al., 2019). However, in practice, if feedback is obtained at all, it is often provided in a narrative form. Furthermore, it is important that feedback interventions sustainably affect attitudes toward evidence-based assessment as the lack of feedback elicits hindsight bias and sense-making, which increases practitioners' confidence (Kahneman & Klein, 2009).

Finally, future studies should investigate if findings on encouraging evidence-based practices from other fields are useful in the context of personnel- and educational selection. For example, Kaplan et al. (2001) showed in an accountancy bond rating task that decision makers were more likely to use a mechanical rule when they could design the rule. Furthermore, informing decision makers in a legal judgment scenario that mechanical rule predictions are more accurate than holistic predictions increased their intention to support the use a mechanical rule in practice (Eastwood & Luther, 2016).

Conclusion and take-home message

With this review, we provided an overview of the research that has investigated factors associated

with and interventions for effective implementation of evidence-based assessment in practice. Albeit small in volume, these studies address a topic that is essential to utilizing the benefits of the knowledge that assessment researchers have produced in the past century. We hope this overview inspires researchers from different disciplines to expand this line of research further, as the reviewed literature suggests that future research requires expertise from fields such as I/O psychology, judgment- and decision-making, psychometrics, management, and science communication. Finally, besides encouraging researchers and practitioners to be outspoken about evidence-based practices in assessment and selection, we especially encourage researchers to employ autonomy supportive principles when trying to implement evidence-based practices, as these principles seem most promising to narrow the science-practice gaps that exist in this field.

ORCID iD

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

Supplemental material

Supplemental material for this article is available online.

Note

1. We used all search engines of our academic institution, alternative engines such as Google Scholar, and also consulted the journals' websites and archives, to retrieve articles published in these journals.

References

References marked with an asterisk indicate studies included in the review.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Anders, G. (2013). Insider trading for talent. *Forbes*, *192*, 1.

*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*, 93–110. https://doi.org/10.1016/0749-5978(86)90046-4

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

Barrick, M. R., & Parks-Leduc, L. (2019). Selection for fit. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 171–193. https://doi.org/10.1146/annurev-orgpsych-012218-015028

Bateson, J., Wirtz, J., Burke, E., Vaughan, C., Jones, G., Pavan Gobburu, V. N., Mathur, A., Hoskins, J., Baron, G., & Gajano, A. M. (2014). When hiring, first test, and then interview: Interaction. *Harvard Business Review*, *92*, 20.

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*, 689–709. https://doi.org/10.1177/1094428106294734

Bolander, P., & Sandberg, J. (2013). How employee selection decisions are made in practice. *Organization Studies*, *34*, 285–311. https://doi.org/10.1177/0170840612464757

Bortz, D. (2018). Hiring blind. *HR Magazine*, *63*, 54–59.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in*

Psychology, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). Are common language effect sizes easier to understand than traditional effect sizes? *Journal of Applied Psychology*, *99*, 332–340. https://doi.org/10.1037/a0034745

Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology*, *87*, 185–191. https://doi.org/10.1037/0021-9010.87.1.185

Buchanan, L. (2005). Knowing what to listen for. *Harvard Business Review*, *83*, 25.

Buchanan, L. (2015). The fine art of people. *Inc.*, *37*, 56–61.

Cadrain, D. (2010). Police, firefighter exams challenge employers. *HR Magazine*, *55*, 14.

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.

Cappelli, P. (2019). Your approach to hiring is all wrong. *Harvard Business Review*, *97*, 48–58.

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

Clifford, S. (2006). The new science of hiring. *Inc.*, *28*, 90–98.

Colvin, G. (2014). Employers are looking for new hires with something extra: Empathy. *Fortune*, *170*, 55.

Cook, M. (2016). *Personnel selection: Adding value through people—A changing picture* (6th ed.). Wiley-Blackwell.

Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes*, *112*, 11–23. https://doi.org/10.1016/j.obhdp.2009.11.007

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making*, *8*, 512–520.

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

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*, 95–106. https://doi.org/10.1037/h0037613

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

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227–268. https://doi.org/10.1207/S15327965PLI1104_01

*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*, 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*, 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*, 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. *Research in Personnel and Human Resource Management*, *12*, 79–123.

Dipboye, R. L., Macan, T., & Shahani-Denning, C. (2012). The selection interview from the interviewer and applicant perspectives: Can't have one without the other. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 323–352). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0015

Eastwood, J., & Luther, K. (2016). What you should want from your professional: The impact of educational information on people's attitudes toward simple actuarial tools. *Professional Psychology: Research and Practice*, *47*, 402–412. https://doi.org/10.1037/pro0000111

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

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *62*, 328–339. https://doi.org/10.1037/0022-3514.62.2.328

Fernández-Aráoz, C. (2007). Making people decisions in the new global environment. *MIT Sloan Management Review*, *49*, 17–20.

Fernández-Aráoz, C., Groysberg, B., & Nohria, N. (2009). The definite guide to recruiting in good times and bad. *Harvard Business Review*, *87*, 74–84.

Fisher, A., & Yang, J. L. (2009). Now hiring. *Fortune*, *159*, 68–77.

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

Fisher, P., Risavy, S., Robie, C., König, C., Christiansen, N., Tett, R., & Simonet, D. (2020). Selection myths: A conceptual replication of HR professionals' beliefs about effective human resource practices in the United States and Canada. *Journal of Personnel Psychology*. Advance online publication. https://doi.org/10.1027/1866-5888/a000263

FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology*, *7*, 29. https://doi.org/10.1186/s40359-019-0299-7

Freeman, C. (2014). Employers use video interviews to connect with candidates, cut costs. *HR Focus*, *91*, 7–9.

Fried, J. (2010). Never read another resumé. *Inc.*, *32*, 36–37.

Gagné, M., Koestner, R., & Zuckerman, M. (2000). Facilitating acceptance of organizational change: The importance of self-determination. *Journal of Applied Social Psychology*, *30*, 1843–1852. https://doi.org/10.1111/j.1559-1816.2000.tb02471.x

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*, 210–216. https://doi.org/10.1037/a0014474

Gino, F. (2018). The business case for curiosity. *Harvard Business Review*, *96*, 48–57.

Greco, S. (2007). When is it safe to hire? *Inc.*, *29*, 52–53.

Greco, S. (2009). What is culture. *HR Magazine*, *54*, 42–44.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Grensing-Pophal, L. (2006). Internal selections. *HR Magazine*, *51*, 75–78.

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*, 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*, 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). Routledge/Taylor & Francis Group.

Harnish, V. (2013). Five steps to find (and keep) young stars. *Fortune*, *168*, 22–1NULL.

Hennigan, M., & Evans, L. (2018). Does hiring for "cultural fit" perpetuate bias? *HR Magazine*, *63*, 28–29.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

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

Hoffman, M., Kahn, L. B., & Li, D. (2017). Discretion in hiring. *The Quarterly Journal of Economics*, *133*, 765–800. https://doi.org/10.1093/qje/qjx042

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Psychological Assessment Resources.

Holmes, D. E. (2019). Expanding the pool. *Harvard Business Review*, *97*, 58–61.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*, 343–355. https://doi.org/10.1177/1745691610374586

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, *62*, 227–240. https://doi.org/10.1177/0013164402062002002

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

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, *38*, 635–672. https://doi.org/10.2307/256741

*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*, 1–36. 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*, 33–56. https://doi.org/10.1111/joop.12187

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

Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*, *94*, 38–46.

Kaplan, S. E., Reneau, J. H., & Whitecotton, S. (2001). The effects of predictive ability information, locus of control, and decision maker involvement on decision aid reliance. *Journal of Behavioral Decision Making*, *14*, 35–50. https://doi.org/10.1002/1099-0771(200101)14:1<35::AID-BDM364>3.0.CO;2-D

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

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *38*, 107–128. https://doi.org/10.1037/h0034225

Klehe, U. C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *International Journal of Selection and Assessment*, *12*, 327–342. https://doi.org/10.1111/j.0965-075X.2004.00288.x

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*, 296–310. https://doi.org/10.1037/0033-2909.107.3.296

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*, 17–27. https://doi.org/10.1111/j.1468-2389.2010.00485.x

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

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081. https://doi.org/10.1126/science.1136618

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 162–181. https://doi.org/10.1037/0033-2909.127.1.162

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161. https://doi.org/10.1037/0022-3514.86.1.148

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*, 1060–1072. https://doi.org/10.1037/a0034156

Kuncel, N. R., Klieger, D. M., & Ones, D. S. (2014). In hiring, algorithms beat instinct. *Harvard Business Review*, *92*, 32.

Lam, B. (2015). *For more workplace diversity, should algorithms make hiring decisions? The Atlantic*. www.theatlantic.com/business/archive/2015/06/algorithm-hiring-diversity-HR/396374/

Lawler, E. E. I. I. I. (2007). Why HR practices are not evidence-based. *Academy of Management Journal*, *50*, 1033–1036. https://doi.org/10.5465/AMJ.2007.27155013

Lee, T. (2017). 10 classic interview questions to ask . . . . . . and the responses you want to hear. *HR Magazine*, *62*, 56–61.

Lee, T. H., & Duckworth, A. L. (2018). Organizational grit. *Harvard Business Review*, *96*, 98–105.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*, 241–293. https://doi.org/10.1111/peps.12052

*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*, 29–46. https://doi.org/10.1002/job.246

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

*Lodato, M. A. (2008). *Going with your gut: An investigation of why managers prefer intuitive employee selection* (Doctoral dissertation). http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1206311034

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

Lytle, T. (2013). Streamline hiring. *HR Magazine*, *58*, 63–65.

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

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

Meehl, P. E. (1954b). The special powers of the clinician. In *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*

(pp. 24–28). University of Minnesota Press. https://doi.org/10.1037/11281-004

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, *50*, 370–375. https://doi.org/10.1207/s15327752jpa5003_6

Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A tutorial on mechanical decision-making for personnel and educational selection. *Frontiers in Psychology*, *10*, 3002. https://doi.org/10.3389/fpsyg.2019.03002

Meinert, D. (2013). Cultural similarities influence hiring decisions. *HR Magazine*, *58*, 18.

Menkes, J. (2005). Hiring for smarts. *Harvard Business Review*, *83*, 100–109.

Michel, R., Belur, V., Naemi, B., & Kell, H. (2019). Graduate admissions practices: A targeted review of the literature. In *ETS research report series*. https://doi.org/10.1002/ets2.12271

Moore, D. A. (2017). How to improve the accuracy and reduce the cost of personnel selection. *California Management Review*, *60*, 8–17. http://10.0.4.153/0008125617725288

Morris, S. B., Daisley, R. L., Wheeler, M., & Boyer, P. (2015). A meta-analysis of the relationship between individual assessments and job performance. *Journal of Applied Psychology*, *100*, 5–20. https://doi.org/10.1037/a0036938

Muchinsky, P. M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology*, *57*, 175–209. https://doi.org/10.1111/j.1744-6570.2004.tb02488.x

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

Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, *84*, 83–86.

Niessen, A. S. M., & Meijer, R. R. (2017). On the use of broadened admission criteria in higher education. *Perspectives on Psychological Science*, *12*, 436–448. https://doi.org/10.1177/1745691616683050

*Nolan, K. P. (2013). Basic psychological need fulfillment and user resistance to objective and analytical decision-making practices in employee selection. (Doctoral dissertation). http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1343479006

*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*, 30–47. https://doi.org/10.25035/pad.2016.004

Nolan, K. P., Dalal, D. K., & Carter, N. (2020). Threat of technological unemployment, use intentions, and the promotion of structured interviews in personnel selection. *Personnel Assessment and Decisions*, *6*, 38–53. https://doi.org/10.25035/pad.2020.02.006

*Nolan, K. P., & Highhouse, S. (2014). Need for autonomy and resistance to standardized employee selection practices. *Human Performance*, *27*, 328–346. https://doi.org/10.1080/08959285.2014.929691

Nolan, K. P., Langhammer, K., & Salter, N. P. (2016). Evaluating fit in employee selection: Beliefs about how, when, and why. *Consulting Psychology Journal*, *68*, 222–251. https://doi.org/10.1037/cpb0000065

Ones, D. S., Dilchert, S., Viswesvaran, C., & Salgado, J. F. (2010). Cognitive abilities. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 255–275). Routledge/Taylor & Francis Group.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607

Phillips, J. M., & Gully, S. M. (2008). The role of perceptions versus reality in managers' choice of selection decision aids. *Industrial and Organizational Psychology: Perspectives on Science*

and Practice, *1*, 361–363. https://doi.org/10.1111/j.1754-9434.2008.00064.x

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

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*, 153–172. https://doi.org/10.1111/j.1744-6570.2008.00109.x

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

Potworowski, G. A., & Green, L. A. (2012). Culture and evidence-based management. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 272–292). Oxford University Press.

Prien, E. P., Schippmann, J. S., & Prien, K. O. (2003). *Individual assessment: As practiced in industry and consulting*. Lawrence Erlbaum Associates Publishers.

Pyburn, K. M. J., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, *61*, 143–151. https://doi.org/10.1111/j.1744-6570.2008.00108.x

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*. https://doi.org/10.25035/pad.2019.01.004

Rockwood, K. (2015). Beyond fitting in. *Inc.*, *37*, 112–113.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166–169. https://doi.org/10.1037/0022-0663.74.2.166

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

*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*, 37–48. https://doi.org/10.25035/pad.2019.01.002

Rupp, D. E., Song, Q. C., & Strah, N. (2020). Addressing the so-called validity–diversity trade-off: Exploring the practicalities and legal defensibility of pareto-optimization for reducing adverse impact within personnel selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *13*, 246–271. https://doi.org/10.1017/iop.2020.19

Russell, S., & Brannan, M. J. (2016). "Getting the right people on the bus": Recruitment, selection and integration for the branded organization. *European Management Journal*, *34*, 114–124. https://doi.org/https://doi.org/10.1016/j.emj.2016.01.001

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 J. K. Oostrom & I. Nikolaou (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). Taylor & Francis.

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, *65*, 693–717. https://doi.org/10.1146/annurev-psych-010213-115134

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

Rynes, S. L. (2012). The research-practice gap in I/O psychology and related fields: Challenges and potential solutions. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (Vol. 1., pp. 409–452). Oxford University Press.

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*, 149–174. https://doi.org/10.1002/hrm.10029

Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, *50*, 987–1008. https://doi.org/10.5465/AMJ.2007.27151939

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, *102*, 254–273. https://doi.org/10.1037/apl0000151.supp

Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology*, *102*, 1435–1447. https://doi.org/10.1037/apl0000236

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science*, *9*, 538–551. https://doi.org/10.1177/1745691614543972

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*, 1976–1988. https://doi.org/10.1080/09585190802324304

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

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

Segal, J. A. (2006). "Good fit" isn't always legit. *HR Magazine*, *51*, 121–126.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-p) 2015: Elaboration and explanation. *BMJ*, *349*, 1–25. https://doi.org/10.1136/bmj.g7647

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Silzer, R., & Jeanneret, R. (2011). Individual psychological assessment: A practice and science in search of common ground. *Industrial and Organizational Psychology*, *4*, 270–296. https://doi.org/10.1111/j.1754-9434.2011.01341.x

Slaughter, J. E., & Kausel, E. E. (2014). Employee selection decisions. In S. Highhouse, R. S. Dalal, & E. Salas (Eds.), *Judgment and decision making at work* (pp. 57–79). Routledge/Taylor & Francis Group.

Stanek, K. C., & Ones, D. S. (2018). Taxonomies and compendia of cognitive ability and personality constructs and measures relevant to industrial, work and organizational psychology. 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. 366–407). SAGE Reference.

Taylor Arnold, J. (2008). Getting facts fast. *HR Magazine*, *53*, 57–62.

Terpstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, *46*, 27–48. https://doi.org/10.1111/j.1744-6570.1993.tb00866.x

*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*, 20–30. https://doi.org/10.25035/pad.2020.02.004

Thornton, G. C. I., Hollenbeck, G. P., & Johnson, S. K. (2010). Selecting leaders: Executives and high potentials. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 823–840). Routledge/Taylor & Francis Group.

Tkaczyk, C. (2014). How Google works. *Fortune*, *170*, 103.

*Tsai, W., Chen, F. H., Chen, H., & Tseng, K. (2016). When will interviewers be willing to use high-structured job interviews? The role of personality. *International Journal of Selection and Assessment*, *24*, 92–105. https://doi.org/10.1111/ijsa.12133

Tyler, K. (2005). Train for smarter hiring. *HR Magazine*, *50*, 89–93.

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

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads; the University of Oklahoma lectures in social psychology* (pp. 273–298). Harper.

*Voss, N. M., & Lake, C. J. (2020). Communicating validity information to differentially experienced audiences: The effects of numeracy and nontraditional metrics. *Personnel Assessment and Decisions*, *6*, 11–19. https://doi.org/https://doi.org/10.25035/pad.2020.02.003

Williamson, S., & Foley, M. (2018). Unconscious bias training: The 'silver bullet' for gender equity? *Australian Journal of Public Administration*, *77*, 355–359. https://doi.org/10.1111/1467-8500.12313

Wright, A. D. (2016). Autism speaks, and employers listen. *HR Magazine*, *61*, 60–64.

*Yankelevich, M. (2010). Attitudes toward holistic and mechanical judgment in employee selection: Role of error rate and false positive and false negative error. (Doctoral dissertation). http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1269269662

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

*Zhang, D. C., Highhouse, S., Brooks, M. E., & Zhang, Y. (2018). Communicating the validity of structured job interviews with graphical visual aids. *International Journal of Selection and Assessment*, *26*, 93–108. https://doi.org/10.1111/ijsa.12220

*Zhang, D. C., Zhu, X., Ritter, K.-J., & Thiele, A. (2019). Telling stories to communicate the value of the pre-employment structured job interview. *International Journal of Selection and Assessment*, *27*, 299–314. https://doi.org/10.1111/ijsa.12264

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*, 499–511. https://doi.org/10.1348/096317909X425203

Author biographies

Marvin Neumann is a PhD candidate in the department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, at the University of Groningen, the Netherlands. His research focuses on personnel- and educational selection, mechanical prediction, and the scientist-practitioner gap in assessment and selection.

A. Susan M. Niessen is an assistant professor in the department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, at the University of Groningen, the Netherlands. She has published in the areas of personnel- and educational selection, test-use and decision making, predictive validity, applicant perceptions, and test bias.

Rob R. Meijer is a full professor in the department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, University of Groningen, the Netherlands. His research focuses on applied psychometrics, educational- and personnel selection, and decision making through tests.