

University of Groningen

Education Increases Decision-rule Use

Neumann, Marvin; Hengeveld, Martijn; Niessen, A. Susan M.; Tendeiro, Jorge N.; Meijer, Rob R.

Published in:
Journal of experimental psychology-Applied

DOI:
[10.1037/xap0000372](https://doi.org/10.1037/xap0000372)

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
Final author's version (accepted by publisher, after peer review)

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Neumann, M., Hengeveld, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2022). Education Increases Decision-rule Use: An Investigation of Education and Incentives to Improve Decision Making. *Journal of experimental psychology-Applied*, 28(1), 166-178. <https://doi.org/10.1037/xap0000372>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: [10.1037/xap0000372](https://doi.org/10.1037/xap0000372)

In press, *Journal of Experimental Psychology: Applied*

Education Increases Decision-rule Use: An Investigation of Education and Incentives to Improve Decision Making

Marvin Neumann*¹, Martijn Hengeveld*¹, A. Susan M. Niessen¹, Jorge N. Tendeiro², & Rob R. Meijer¹

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

²Office of Research and Academia-Government-Community Collaboration Education and
Research Center for Artificial Intelligence and Data Innovation Hiroshima University

*Both authors contributed equally to this article.

Author note

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

Martijn Hengeveld, <https://orcid.org/0000-0001-8609-9395>

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

Jorge N. Tendeiro, <https://orcid.org/0000-0003-1660-3642>

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

We have no conflicts of interest to disclose.

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

Abstract

Robust scientific evidence shows that human performance predictions are more valid when information is combined mechanically (with a decision rule) rather than holistically (in the decision-maker's mind). Yet, information is often combined holistically in practice. One reason is that decision makers lack knowledge of evidence-based decision making. In a performance prediction task, we tested whether watching an educational video on evidence-based decision making increased decision-makers' use of a decision rule and their prediction accuracy immediately after the manipulation and a month later. Furthermore, we manipulated whether participants earned incentives for accurate predictions. Existing research showed that incentives decrease decision-rule use and prediction accuracy. We hypothesized that this is the case for decision makers who did not receive educational information about evidence-based decision making, but that incentives increase decision-rule use and prediction accuracy for participants who received educational information. Our results showed that educational information increased decision-rule use. This resulted in increased prediction accuracy, but only immediately after receiving the educational information. In contrast to the existing literature, incentives slightly increased decision-rule use. We did not find evidence that this effect was larger for educated participants. Providing decision makers with educational information may be effective to increase decision-rule use in practice.

Key words: personnel- and educational selection, mechanical- and clinical judgment, education, incentives, decision making

Public significance statement

Combining information with a decision rule results in more valid predictions than combining information holistically in the mind. Yet, decision makers rarely use decision rules in practice. This study suggests that a brief educational intervention can increase decision-makers' use of a decision rule in a human performance prediction task. Consequently, prediction accuracy increased, but only temporarily. Such an educational intervention is easily applicable and may increase evidence-based decision making in practice. But interventions may need to be repeated for a lasting effect.

Education Increases Decision-rule Use: An Investigation of Education and Incentives to Improve Decision Making

Making accurate human performance predictions is important because they reduce costly erroneous decisions such as admitting students that will quit their study program (Kuncel & Hezlett, 2007) or choosing the wrong job candidates (Schmidt & Hunter, 1998). However, decision makers such as hiring managers, assessment psychologists, and admission officers rarely use evidence-based decision-making procedures (Highhouse, 2008; Michel et al., 2019; Ryan & Sackett, 1987; Silzer & Jeanneret, 2011; Slaughter & Kausel, 2014). Hence, an important, yet largely unanswered question is how decision making can be improved in practice (Kuncel, 2018; Milkman et al., 2009; Neumann et al., 2020).

Improving Decision Making

Several methods to improve decision making have been suggested (Milkman et al., 2009). One method is debiasing intuitive judgments (Milkman et al., 2009; Sellier et al., 2019). Most debiasing interventions, such as warning decision makers about and instructing them to avoid bias have not been successful (Fischhoff, 1982; Fischhoff & Broomell, 2020), although encouraging decision makers to consider the opposite of their prediction moderately improves decision making (Mussweiler et al., 2000). Another method is to provide decision makers with outcome feedback on their decisions. Yet, research showed that in probabilistic judgment tasks, outcome feedback decreases judgment consistency and hence prediction accuracy (Arkes et al., 1986; Hammond et al., 1973; Jackson et al., 2019). When multiple sources of information are used, as is the case for most human performance predictions (Clinedinst & Patel, 2018; Morris et al., 2015; Thornton et al., 2010), research based on construal level theory showed that distancing oneself psychologically from the decision (i.e., viewing the decision in an abstract rather than in a context-specific manner) can moderately improve decision making (Fukukura et al., 2013). This is because attention is paid more to

important information and less to salient, but irrelevant information (Trope & Liberman, 2000).

One of the most promising methods is to combine information with a decision rule (Milkman et al., 2009), which can increase prediction accuracy by 50% in the context of human performance prediction (Kuncel et al., 2013). In holistic combination, decision makers use their human judgment to integrate information subjectively in the mind. In mechanical combination, quantified information is combined according to a decision rule or formula in which each piece of information receives an explicit weight (Grove & Meehl, 1996; Meehl, 1954; Meijer et al., 2020). An example of a simple decision rule would be to assign equal weights to a test score, a grade, and an interview rating and to add up the resulting scores. However, weights can also be based on regression analysis of primary data, meta-analyses, or subject matter experts (Bobko et al., 2007; Dawes & Corrigan, 1974; Murphy et al., 2013). In mechanical combination, weights are used consistently across judgments. In contrast, weights are used inconsistently across judgments when information is holistically combined (Hammond & Summers, 1972; Karelaia & Hogarth, 2008; Kuncel et al., 2013; Meijer et al., 2020; Yu & Kuncel, 2020). Importantly, decreased judgment consistency explains the robust finding that mechanical combination results in more accurate predictions than holistic combination (Ægisdóttir et al., 2006; Dawes, 1971; Grove et al., 2000; Kuncel et al., 2013; Meehl, 1954; Sarbin, 1943).

The problem is that, despite these robust research findings, decision makers such as hiring managers, assessment psychologists, and admission officers usually combine information holistically in practice (Highhouse, 2008; Michel et al., 2019; Ryan & Sackett, 1987; Silzer & Jeanneret, 2011; Slaughter & Kausel, 2014) and hold more positive attitudes towards holistic combination (Eastwood et al., 2012; Kirch, 2012). So, in important prediction

contexts, decision makers use suboptimal holistic decision-making procedures, resulting in suboptimal predictions and decisions.

Reasons for decision-makers' preference for holistic- over mechanical information combination are, amongst others, higher autonomy (Nolan & Highhouse, 2014), higher professional status (Nolan et al., 2016), and higher confidence in holistic combination of information (Dietvorst et al., 2015). Another, more fundamental, reason is that many decision makers lack relevant knowledge of evidence-based decision-making practices (Fisher et al., 2020; Jackson et al., 2018; Lawler, 2007; Rynes, 2012; Rynes et al., 2002; Sanders et al., 2008). For example, decision makers believe that they can learn from their prediction errors, can accurately identify incorrect decision-rule predictions that warrant deviation from a rule, and can make more accurate holistic judgments with experience (Dietvorst et al., 2015; Eastwood & Luther, 2016; Highhouse, 2008; Leonard & Swap, 2004). However, in noisy contexts such as human performance prediction, these beliefs are incorrect (Dawes, 1971; Goldberg, 1968; Grove et al., 2000; Jackson et al., 2019; Karelaia & Hogarth, 2008; Yu & Kuncel, 2020).

Knowledge gaps may persist because decision makers lack time to keep up with research findings or have difficulties to interpret them correctly (Majid et al., 2011; Rynes et al., 2002). Furthermore, practitioners tend to read practitioner-oriented journals that rarely cover decision-making research (Neumann et al., 2020), rather than academic journals (Rynes et al., 2007). Moreover, decision-makers' reading of research- or practitioner-oriented journals is at best only very weakly related to knowledge of decision-making research findings (Lodato et al., 2011; Rynes et al., 2002; Sanders et al., 2008). So, decision makers are often unaware of important decision-making research findings (Fisher et al., 2020; Rynes et al., 2002; Sanders et al., 2008), and self-studying the academic- or professional literature seems ineffective and impractical to close knowledge gaps.

Given that knowledge gaps constitute a major reason why decision-making research findings are not translated into practice (Banks et al., 2016; Fisher et al., 2020; Gill, 2018; Terpstra & Rozell, 1997), it is remarkable that there seem to exist only two studies in the context of holistic- vs. mechanical combination in which participants were told that decision-rule predictions are more accurate than holistic predictions. In one of two vignette studies, Eastwood and Luther (2016) found that participants who received information that using a specific decision rule would result in more accurate predictions than holistic predictions reported higher willingness to use such a rule in practice than participants who did not receive such information. In another study, Arkes et al. (1986) found that participants who were told that decision-rule predictions are more valid than holistic predictions made more accurate predictions than participants who were told that holistic predictions are more valid than decision-rule predictions, or participants who were told that these methods are about equally valid. These accuracy differences occurred because participants in the “decision-rule-is-more-valid” condition used an available decision rule more than participants in the other conditions. Importantly, the data in Arkes' et al. (1986) prediction task was simulated such that any additional information not included in the decision rule was random and hence unrelated to the criterion.

Although the existing studies provide valuable insights, Eastwood and Luther (2016) only investigated the effect of educational information on participants' *willingness* to use a decision rule. Hence, no conclusions can be drawn about actual rule use. This also leaves the question unanswered whether educational information can increase decision-rule use enough to also meaningfully increase prediction accuracy. This is an important question to answer because decision makers often deviate from a decision rule, which decreases prediction accuracy (Dawes, 1971; Dietvorst et al., 2018; Guay & Parent, 2018). Although Arkes et al. (1986) also investigated how educational information affects prediction accuracy, their study

was designed in such a way that participants could not beat the decision rule based on any additional information (Arkes et al., 1986, H. Arkes, personal communication, September 15, 2020), which prevents a generalization of their results to real prediction contexts.

We contribute to the existing literature in the following ways. First, by using an experimental prediction task with real data, we investigated whether presenting educational information increased decision-makers' actual decision-rule use, instead of only use intentions (e.g., Eastwood & Luther, 2016) or self-reported use. Furthermore, this allowed us to investigate the effect of educational information on prediction accuracy. Second, compared to simple descriptions of validity differences between mechanical- and holistic combination (Arkes et al., 1986; Eastwood & Luther, 2016), we presented educational information in the form of a short video, which increases ecological validity as videos are often used in organizational trainings (Burgess & Russell, 2003). Third, providing educational information is only practically useful if it has a long-lasting effect on evidence-based decision making. In the context of holistic- vs. mechanical combination, the long-term effectiveness of educational information on evidence-based decision making is unknown since existing evidence is based on cross-sectional study designs (Arkes et al., 1986). However, a study among teachers showed that educational information increased their acceptance of evidence-based teaching practices only temporarily (Ferrero et al., 2020). Therefore, we measured decision-rule deviation and prediction accuracy right after an educational information manipulation, and one month later. Although we do not have a specific hypothesis with regard to a time effect, it seems plausible that decision-rule *deviation* will either be stable or increase over time because knowledge retention in general decreases over time (Arthur et al., 1998; Custers, 2010; Murre & Dros, 2015). Hence, decision-rule deviation should remain stable or increase after a month, depending on whether decision makers sufficiently internalized the educational information.

So, we expect educational information to positively affect decision-makers' attitudes towards decision rules, which should translate into actual decision-rule use (Ajzen, 1991). Attitude changes are likely when the presented arguments and facts are strong, compelling, and falsifiable (Petty & Cacioppo, 1986; Wood, 2000). Furthermore, construal-level theory suggests that educational information is more persuasive when causal (why something is the case) rather than non-causal arguments (that something is the case) are provided, and when arguments are presented in a more general, abstract manner (Reyt et al., 2016; Wiesenfeld et al., 2017). In line with this theoretical framework, we ensured that our educational information provided explanations of why mechanical information combination is superior to holistic combination. Based on the presented theoretical argument and the existing literature, we expect the following:

Hypothesis 1a: Participants who receive educational information on evidence-based decision making will deviate less from a decision rule than participants who do not receive educational information.

Hypothesis 1b: Participants who receive educational information on evidence-based decision making will make more accurate predictions than participants who do not receive educational information.

Incentives and Decision-rule Deviation

Another factor that influences decision making is the presence of incentives. Counterintuitively, research showed that incentivized participants made *less* accurate predictions than participants who were not incentivized for their prediction accuracy, when a decision rule was available (Ashton, 1990; Arkes et al., 1986; Samuels & Whitecotton, 2011); for an explanation see below. Importantly, this effect occurred even when participants were

told that decision-rule predictions are more accurate than holistic predictions (Arkes et al., 1986).

The negative effect of incentives on decision-rule use and prediction accuracy poses a problem because decision makers are often (indirectly and implicitly) incentivized for accurate decision making (Rynes et al., 2005). For example, HR professionals and admission officers may be held accountable for their hiring- and admission decisions. Decision makers may also be motivated to make good decisions because they may be evaluated for such a core task of their job. Moreover, personnel- and educational selection decisions are motivated by increased success ratios (Barrick et al., 1991; Cook, 2016) and the performance gains that partly result from better (i.e., more valid) hiring decisions (Schmidt & Hunter, 1998). To reduce negative incentive effects, we investigated circumstances under which incentives may increase decision-rule use and prediction accuracy. In doing so, we make a theoretical contribution by answering the call for research that may shed light on potential moderators of the incentives – performance relationship (Bonner & Sprinkle, 2002; Camerer & Hogarth, 1999).

In experimental research, monetary incentives are often used to mimic incentives that exist in practice (Bonner et al., 2000; Camerer & Hogarth, 1999). In most tasks, monetary incentives increase effort, which then sometimes translates into increased performance (Bonner et al., 2000; Bonner & Sprinkle, 2002; Garbers & Konradt, 2014; Jenkins et al., 1998; Rynes et al., 2005). However, in judgment and decision-making tasks, such effort can decrease performance (i.e., prediction accuracy, Camerer & Hogarth, 1999). When an imperfect decision rule is present and decision makers do not know that using this rule consistently is a valid judgment strategy, incentives should increase judgment strategy shifts (Arkes et al., 1986) and decision-makers' tendency to add their own judgment to the decision-rule prediction (Camerer & Hogarth, 1999). This, in turn, *increases* decision-makers'

deviation from a decision rule, and hence decreases prediction accuracy. Although valid rule deviations exist, decision makers are unable to identify when such rule deviations are warranted in human performance prediction (Dawes, 1971; Dietvorst et al., 2018). Hence, the best strategy is to follow an existing valid decision rule (Dawes, 1971, 1979; Dawes & Corrigan, 1974; Dietvorst et al., 2018; Guay & Parent, 2018; Sarbin, 1943).

Extending the existing literature, we hypothesize that incentives only increase decision-rule deviation when decision makers are unaware of the most valid judgment strategy. When decision makers know that the best judgment strategy is to follow the decision rule consistently, we expect incentives to *decrease* decision-makers' deviation from the rule and hence increase prediction accuracy. Educated decision makers who are also incentivized and hence would want to make judgment strategy shifts due to increased effort would experience cognitive dissonance (Festinger, 1957). When educational information provides a complete argumentation for the use of decision rules that cannot easily be counter argued, the easiest way to reduce dissonance should be to follow the decision rule consistently. Therefore, we expect educational information and incentives to interact in the following way:

Hypothesis 2a: When no educational information on evidence-based decision making is provided, incentivized participants will deviate *more* from the decision rule than participants who do not receive incentives. When educational information on evidence-based decision making is provided, incentivized participants will deviate *less* from the decision rule than participants who do not receive incentives.

Hypothesis 2b: When no educational information on evidence-based decision making is provided, incentivized participants will make less accurate predictions than participants who do not receive incentives. When educational information on evidence-based decision

making is provided, participants who receive incentives will make more accurate predictions than participants who do not receive incentives.

Method

The study materials, scripts and the dataset used for the analyses are publicly available on <https://osf.io/68qwa/>.

Participants

We conducted a priori power analyses for all relevant effects. The power analysis for a mixed-effects ANOVA between-within interaction resulted in the greatest required sample size ($N = 180$), assuming a medium effect size of $\eta_p^2 = 0.06$, desired power = .80 and $\alpha = .05$. Data was collected until a pre-determined date, given that 180 participants had taken part by that date. The university's research participant pool was used to recruit participants who received a compensation of €9 for their voluntary participation. Participants in this pool are mostly externally employed people and students from multiple Dutch universities and study programs, who are mainly recruited during the authors' university's yearly introduction week. This introduction week is also sometimes attended by students from other universities. The only requirement to enroll in this study was a good comprehension of Dutch, because all materials were in Dutch.

A total of 186 participants took part in the study. Nine participants were excluded based on failing at least one of two attention checks. Furthermore, six participants did not complete the second measurement. The final sample consisted of 171 participants, of whom 68% were currently college students (including students working part-time), 26% employed non-students, and 6% unemployed non-students. Employed, non-student participants mostly held a research university degree (50%) or had completed other types of tertiary education. The mean age was $M = 24.97$ ($SD = 8.49$, range 16-64) and most participants were female

(73%). Furthermore, most participants had the Dutch nationality (90%). Among the other participants, 6% had another European nationality and 4% had a non-European nationality.

This study was approved by the university's ethics committee for psychological research.

Prediction Task

Participants were presented with archival data from a pool of 192 Dutch applicants for the Bachelor Psychology program of the university in 2014. Each participant predicted the first-year GPA of 20 applicants at time 1 and again at time 2, based on three predictors: high school GPA, an admission test score, and a personal statement. We chose these predictors because they are commonly used in admission to higher education (Clinedinst & Patel, 2018; Davis et al., 2018). High school GPA was the mean of all final grades obtained at the end of secondary education (vwo, in Dutch). The admission test was a multiple-choice exam that assessed applicant's knowledge of two chapters from an introductory psychology book that they had to study. Both high school GPA and the admission test were measured or transformed to the Dutch ten-point grading scale and were good predictors of first-year GPA (see Niessen et al., 2018). The personal statement was a document with a maximum of 250 words in which applicants expressed their motivation to study psychology at the university. As participants did not rate the personal statements, we could not calculate its correlation with first-year GPA. However, personal statements have very low predictive validity for GPA (Murphy et al., 2009). Applicants were randomly assigned to participants (without replacement within participants) and were displayed evenly so that each applicant from the pool was judged.

Decision-rule Prediction

Participants also received the predicted first-year GPA for each applicant, based on a regression model including high school GPA and the admission test score as independent variables. This regression model explained 25.3% of the variance in first-year GPA ($F(2, 189)$

= 30.83, $p < .001$). In no condition were participants informed about the predictor validities. Although one may assume that the invalid personal statement urged participants to deviate from the decision rule, this design mimics decision making in practice, as less valid predictors such as personal statements and unstructured interviews are ubiquitous in personnel- and educational selection (Davis et al., 2018; König et al., 2010). Furthermore, the decision maker virtually always has more information than is included in the decision rule (Grove & Meehl, 1996, p. 297). Moreover, a rating of the personal statement would have received a weight of zero in the decision rule because another study based on these data shows that it does not provide any incremental validity over and above high school GPA and the admission test score (Neumann et al., 2021).

Predicted First-year GPA

Participants predicted each applicant's first-year GPA on the Dutch ten-point scale (1 = lowest), up to one decimal.

Observed First-year GPA

Participants were presented with an applicant's observed first-year GPA after making each prediction. Such outcome feedback usually decreases decision-rule use and prediction accuracy (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019). We gave participants outcome feedback to provide a strict test for the educational intervention. Educated participants may be more inclined to deviate from the decision rule when they have to tolerate the decision-rule's prediction errors, compared to no outcome feedback.

Design

We employed a mixed design, with education (yes/no) and incentives (yes/no) as between-subjects factors and time (T1, T2) as a within-subjects factor. Educational information was only presented at T1. Incentives were obtained at both T1 and T2.

Educational Information

We recorded a ten-minute educational video (available on <https://osf.io/68qwa/>) in which characteristics and validity differences of mechanical- and holistic information combination were discussed. Furthermore, participants were informed that decision rules are imperfect, but still result in better predictions than holistic judgments, and that attempting to adjust decision-rule predictions decreases prediction accuracy (Dawes, 1971; Dietvorst et al., 2018). Moreover, mechanisms were discussed that explain why decision rule predictions are more valid than holistic predictions (Dawes & Corrigan, 1974; Kausel et al., 2016). More details are presented in the supplementary material.

Incentives

Participants could earn a monetary incentive per prediction, depending on the absolute deviation between their predicted first-year GPA and that applicant's observed first-year GPA, with a maximum of €5 in total (€2.50 at each time point). Specifically, per prediction, participants could earn 12.5 cents if their prediction was off by 0.5 points or less, 7.5 cents if their prediction was off by 0.7 points or less, and 2.5 cents if their prediction was off by 1.0 or less. So, the more accurate participants' predictions were, the more money they earned. The total incentive was the sum of incentives over all predictions at both time points. The exact incentive scheme is reproduced in Table 1. This incentive scheme adhered to the university's ethical guidelines on using incentives in experimental research.

- Insert Table 1 about here -

Procedure

Both parts of the study were completed online via Qualtrics survey software. All participants were instructed that their task was to predict as accurately as possible applicants' first-year GPA based on their high school GPA, admission test score, and personal statement. Then, participants read that they were free to use a decision rule that was based on high

school GPA and the admission test score, and explained 25.3% of the variance of first-year GPA. All participants received this information before the prediction task at T1 and again before the second prediction task at T2. After they received this information, the educated groups were asked to watch the educational video, after which they answered two attention checks. The incentivized groups were informed about the chance to obtain an incentive up to €5 and were shown the incentive scheme as depicted in Table 1. Participants in the control group who did not watch the educational video and could not earn incentives did not receive any additional information. Finally, all participants started the prediction task. For each of the 20 predictions that participants made at a point of time, they saw the applicant's high school GPA, admission test score, personal statement, and the decision-rule prediction. Then, participants made their prediction and were shown the applicant's observed GPA on the next screen. After they had made all predictions, participants also reported to what extent they used the decision-rule predictions. One month later, participants were invited via email to complete a second set of 20 predictions. After the second measurement, participants' total incentive and their compensation for participation was transferred to their bank account that they had indicated at the end of the first measurement.

Measures¹

Decision-rule Deviation

Decision-rule deviation was operationalized as the mean absolute deviation between participants' predicted first-year GPA (P) and the decision-rule prediction (D) of the 20 predictions ($i = 1, \dots, 20$) made at each time point.

$$\text{Decision rule deviation} = \frac{\sum_{i=1}^{20} |P_i - D_i|}{20}.$$

So, higher scores indicate larger deviations from the decision rule.

¹ This article originated from a graduate student's research project. Therefore, the online survey on OSF includes three additional items that were only used for the student's research project.

Self-reported Decision-rule Use

Based on Arkes et al. (1986), participants indicated on a 7-point scale to what extent they used the decision rule for their predictions (1 = *I never used the mechanical rule*, 7 = *I always used the mechanical rule*).

Prediction Accuracy

Prediction accuracy was operationalized as prediction deviation: the mean absolute deviation between participants' predicted first-year GPA and an applicant's observed first-year GPA (O) of the 20 predictions ($i = 1, \dots, 20$) made at each time point.

$$\text{Prediction deviation} = \frac{\sum_{i=1}^{20} |P_i - O_i|}{20}.$$

So, higher scores indicate larger deviations from applicants' observed first-year GPA.

Results

Correlations between all studied variables are shown in Table 2.

- Insert Table 2 about here -

To investigate the effect of education, incentives, and time, we conducted a mixed-effects ANOVA for each dependent measure (decision-rule deviation, self-reported decision-rule use, and prediction accuracy), with education and incentives as between-subjects factors and time as a within-subjects factor. As substantive significance was most important in this study, we focused on effect sizes rather than p values (Kirk, 1996). We interpreted effect sizes according to the guidelines presented in Cohen (1988).

Decision-rule Deviation

Figure 1 shows the mean decision-rule deviation in each condition at both time points. As hypothesized, we found that educated participants showed less rule-deviation than non-educated participants ($F(1,167) = 19.19, p < .001, \eta_p^2 = .10$). Furthermore, although we did not have a specific hypothesis about a time effect, we assumed that decision-rule deviation

would either stay stable or increase over time. We found a small to moderate interaction between education and time ($F(1,167) = 5.16, p = .024, \eta_p^2 = .03$). The descriptive statistics are shown in Table 3. Educated participants deviated significantly less from the decision rule than non-educated participants, both at T1 ($t(244) = -4.93, p < .001$) and at T2 ($t(244) = -2.92, p = .004$). However, the difference in decision-rule deviation between educated participants and non-educated participants was larger at T1 ($d = -0.72$) compared to T2 ($d = -0.44$). So, educated participants consistently deviated less from the decision rule at both time points (support for hypothesis 1a). Yet, unexpectedly, the smaller difference in rule-deviation between educated and non-educated participants at T2 did not result from increased rule-deviation of educated participants over time, but from non-educated participants showing less rule-deviation at T2, compared to T1 ($t(167) = 3.91, p < .001, d = 0.35$). We did not find evidence for a difference in decision-rule deviation between T1 and T2 for educated participants ($t(167) = 0.44, p = .660, d = 0.04$).

We further hypothesized that incentives decrease decision-rule deviation for educated participants, and increase decision-rule deviation for non-educated participants (hypothesis 2a). However, the interaction effect between education and incentives was negligible ($F(1,167) = 1.35, p = .247, \eta_p^2 = .01$). Contrary to previous findings, we found that, overall, incentivized participants deviated *less* from the decision rule than non-incentivized participants ($F(1,167) = 4.26, p = .041, \eta_p^2 = .03$), by a small to moderate amount.

- Insert Table 3 about here -

Self-reported Decision-Rule Use

We also investigated decision-rule use with a self-report measure. Given that these two measures are supposed to measure the same construct, it should be noted that the correlation between decision-rule deviation and self-reported decision-rule use was only moderate ($r = -.40$ at T1 and $r = -.48$ at T2).

Figure 2 shows the mean self-reported decision-rule use in each condition at both time points. As the Figure suggests, educated participants reported using the decision rule more than non-educated participants ($F(1,167) = 21.26, p < .001, \eta_p^2 = .11$). Furthermore, we found a small to moderate interaction between education and incentives ($F(1,167) = 4.85, p = .029, \eta_p^2 = .03$). The descriptive statistics are shown in Table 4. In general, the results showed that incentivized participants reported to use the rule much more often when they received educational information, compared to incentivized participants who did not receive educational information ($t(167) = 4.61, p < .001, d = 0.93$). For non-incentivized participants, we did not find evidence that educated participants reported to use the rule more often than non-educated participants ($t(167) = 1.79, p = .075, d = 0.32$). Educated, incentivized participants reported somewhat higher levels of rule-use than educated, non-incentivized participants ($t(167) = 1.93, p = .055, d = 0.42$). However, this difference was not statistically significant. Lastly, we did not find evidence that non-educated, incentivized participants reported to use the rule less than non-educated, non-incentivized participants ($t(167) = -1.15, p = .252, d = -0.20$).

In sum, educated participants reported more decision-rule use than non-educated participants, which is in line with our expectations and the results for decision-rule deviation. In contrast to the results for decision-rule deviation, the education effect seemed stronger when participants could also earn incentives.

- Insert Table 4 about here -

Prediction Accuracy

Figure 3 shows the mean prediction deviation in each condition at both time points. In support of hypothesis 1b, we found that educated participants made predictions that deviated moderately less from observed first-year GPA than predictions made by non-educated participants ($F(1,167) = 9.86, p = .002, \eta_p^2 = .06$). Similar to the decision-rule deviation results, we also found a small to moderate interaction between education and time for prediction accuracy ($F(1,167) = 4.60, p = .033, \eta_p^2 = .03$). The descriptive statistics are shown in Table 5. At T1, educated participants made predictions that deviated significantly and moderately less from observed first-year GPA than predictions made by non-educated participants ($t(331) = -3.77, p < .001, d = -0.53$). Yet, at T2, we did not find evidence for this difference, which was negligible in size ($t(331) = -0.88, p = 0.379, d = -0.14$). Furthermore, educated participants' predictions did not deviate significantly more from observed first-year GPA at T2, compared to T1 ($t(167) = -1.34, p = .184, d = -0.23$). Although Figure 3 suggests a trend that non-educated participants made predictions that deviated slightly *less* from observed first-year GPA at T2, compared to T1, this change over time was not statistically significant and negligible in size ($t(167) = 1.73, p = .086, d = 0.23$). Therefore, the results did not support an effect of time on prediction accuracy.

Hypothesis 2b stated that incentives *increase* prediction accuracy for educated participants, but *decrease* prediction accuracy for non-educated participants. However, we did not find evidence for this interaction ($F(1,167) = 0.16, p = .691, \eta_p^2 < .001$).

- Insert Table 5 about here -

Correlations

Consistently deviating from a decision rule (e.g., always increasing the decision-rule prediction by 0.5 points) results in large deviation scores. However, it does not change the rank-order of applicants, although the rank-order is relevant in selection contexts (Dawes, 1979). A measure that reflects ranking well is the correlation coefficient (Dawes, 1979). Therefore, we also calculated correlations between participants' predictions, the decision-rule predictions, and the observed first-year GPA per condition. These correlations are shown in Table 6. Although we were interested in how much the correlation between participants' predictions and observed first-year GPA differed per condition, we could not directly compare them, because the correlation between the "optimal" rule predictions and observed-first year GPA varied slightly per condition due to random allocation of applicants. In other words, the predictability of first-year GPA differed slightly per condition. Therefore, to compare prediction accuracy across conditions, we calculated per condition the difference between two correlations; 1) the correlation between participants' predictions and the observed first-year GPA and 2) the correlation between decision-rule predictions and observed first-year GPA.

To calculate these differences between correlations, we first applied Fisher's z transformation to all correlations between participants' predictions and observed first-year GPA, and all correlations between the rule predictions and observed-first year GPA. Next, we averaged the transformed correlations over time. The difference between these mean correlations was transformed back with the inverse Fisher's z transformation. The resulting differences are shown in Table 6 (column 3). We used one-sided z -tests to compare these differences between conditions, and to compare the validity of participants' predictions with the validity of the decision-rule predictions within conditions. One-sided z -tests were conducted because these aligned with our directional hypotheses, and because there exists substantial evidence that decision-rule predictions are more valid than human predictions

(Kuncel et al., 2013; Meehl, 1954). Two observations stand out. First, when comparing the differences between conditions, educated and incentivized participants made predictions that were more accurate than predictions made by incentivized participants ($z = 1.88, p = .03$) and participants in the control group ($z = 1.92, p = .03$), but not significantly more accurate than predictions made by educated participants ($z = 1.18, p = .12$). Second, in each condition, participants' predictions were less valid than the decision-rule predictions (education and incentives: $\bar{r}_{\text{diff}} = -.10, z = -1.82, p = .03$; education only: $\bar{r}_{\text{diff}} = -.16, z = -3.27, p < .001$; incentives only: $\bar{r}_{\text{diff}} = -.19, z = -4.09, p < .001$; control: $\bar{r}_{\text{diff}} = -.19, z = -4.38, p < .001$).

- Insert Table 6 about here -

Exploratory Analysis

Since we provided outcome feedback, an alternative explanation for some of our findings may be that participants learned to use the decision rule more over the course of a session, which should increase prediction accuracy. Therefore, for both decision-rule deviation and prediction accuracy, we additionally fitted for each time point linear mixed-effects models using the lme4 package (Version 1.1-23, Bates et al., 2015) in R. We compared a model that included education and incentives as fixed effects and participants and prediction trial as random effects with a model for which the random effect of prediction trial was removed. Figure S1 in the supplementary material shows the mean decision-rule deviation per condition and time point for each prediction trial. We did not find evidence that participants learned to use the decision rule more over the course of a session, neither at T1 ($\chi^2(1) = 0.01, p = .93$) nor at T2 ($\chi^2(1) = 0.73, p = .39$). Figure S2 in the supplementary material shows the mean prediction deviation per condition and time point for each prediction trial. We did not find evidence for a learning effect, neither at T1 ($\chi^2(1) = 0.00, p = 1.00$) nor

at T2 ($\chi^2(1) = 0.54, p = .46$). So, we did not find evidence that participants used the decision rule more, or that they made more accurate predictions over the course of a session.

Discussion

Over the last decades, the effects of various methods to improve decision making have been investigated (Milkman et al., 2009). One of the most promising methods is to combine information with a decision rule because this results in more valid predictions than holistic judgment (Kuncel et al., 2013). However, decision rules are underutilized in practice (Michel et al., 2019; Morris et al., 2015; Ryan & Sackett, 1987), partly because decision makers lack knowledge about evidence-based decision making (Rynes, 2012; Vrieze & Grove, 2009). Encouraging certain behaviors to increase decision-makers' knowledge of evidence-based decision making may have no effect, such as reading the scientific- or practitioner-oriented literature (Lodato et al., 2011; Rynes et al., 2002; Sanders et al., 2008), or may even decrease decision-rule use (e.g., learning from outcome feedback; Dietvorst et al., 2015; Jackson et al., 2019). Therefore, the main aim of the present study was to investigate whether providing decision makers with educational information on evidence-based decision making would increase their use of a decision rule and hence prediction accuracy.

In sum, decision-rule use and prediction accuracy increased immediately after educational information was provided, but these effects decreased or disappeared a month later. Unexpectedly, *non-educated* participants used the decision rule more often at a second measurement after one month, compared to the first measurement. A possible explanation is that participants learned from the outcome feedback that was provided. However, we did not find evidence for learning effects, which is in line with existing research (Jackson et al., 2019). Furthermore, we did not find support for our expectation that incentives would

increase decision-rule use and prediction accuracy only when educational information on evidence-based decision making is provided (hypotheses 2a and 2b). However, we found an interaction between education and incentives for the self-report measure, although it slightly deviated from what we expected. One explanation why the results were slightly different for the behavioral- and the self-report measure of rule deviation may be that the self-report measure is prone to socially desirable answers and demand characteristics. This highlights the importance of measuring actual behavior when testing interventions that may improve decision making. Also, we could not replicate earlier findings that incentives decrease decision-rule use and prediction accuracy (Arkes et al., 1986; Ashton, 1990). Rather, our results suggest that incentives *increased* decision-rule use regardless of whether educational information was provided or not, although this effect was rather small. In general, we cannot provide a reasonable explanation of why, overall, incentives increased decision-rule use in our study, and why we did not find evidence for an interaction between education and incentives. On the basis of earlier research we expected that incentives increase effort, which leads decision makers to add their own judgment and in turn decreases decision-rule use (Arkes et al., 1986; Camerer & Hogarth, 1999). Yet, this may depend on individual difference variables (Bonner & Sprinkle, 2002). For example, highly confident decision makers may want to add their own judgment, while less confident decision makers may appreciate the opportunity to use a decision rule in decision tasks where stakes are higher, that is, where incentives can be earned, like in the present study. However, we caution to overinterpret the results and suggest that replication studies may shed more light on incentive effects and potential moderators of the incentives-performance relationship.

Existing research showed that providing information in the form of outcome feedback decreases decision-rule use and prediction accuracy (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019). So, more research is needed to investigate whether decision makers can

learn from outcome feedback interventions. An alternative is to present decision makers with educational information in which the importance of consistent decision-rule use is explained. In line with existing research (Arkes et al., 1986), our results suggest that this form of information seems effective. However, extending existing research, our results showed that providing educational information on evidence-based decision making may increase prediction accuracy only temporary, which suggests that such information may need to be provided regularly in practice.

In line with existing research (Dawes, 1971; Dietvorst et al., 2018; Sarbin, 1943), our results also showed that participants' predictions were less accurate than the actual rule predictions in all conditions. So, deviating from the decision rule decreased prediction accuracy. Although this finding is not new, it shows that more research is needed in which interventions are tested that may increase decision-makers' consistent use of a decision rule (Neumann et al., 2020).

Finally, although educated participants used the decision rule more than non-educated participants, they still deviated by a considerable amount from the decision-rule predictions. This illustrates that there are factors beyond knowledge that contribute to the underutilization of decision rules (Highhouse, 2008; Rynes, 2012). Indeed, research showed that decision makers are more likely to use a decision rule when they retain autonomy in the decision-making process, for example by designing the decision rule themselves (Nolan & Highhouse, 2014) or by adjusting the outcome of a decision rule (Dietvorst et al., 2018). However, the implementation of such alternatives implies that decision makers understand why decision rules are needed in the first place. Therefore, a first step may be to inform decision makers about evidence-based decision making.

Limitations and Future Directions

A first limitation of this study was that the sample consisted primarily of students, which limits the generalizability to decision makers in practice such as admission officers and HR professionals. It is possible that experienced decision makers are less likely to adapt their behavior based on such an educational video, because they have grown overly confident in their own judgments based on their experience (Arkes et al., 1986) and often do not believe that a simple decision rule can outperform their judgment (Arkes, 2008; Dawes, 1976).

A second limitation of this study was that the maximum incentive participants could earn was small (€5), although it was not smaller than the amount used in related research (Arkes et al., 1986; Dietvorst et al., 2018). Furthermore, participants were only incentivized based on their absolute deviation from the observed-first year GPA (i.e., the criterion) to allow comparisons with existing research (Arkes et al., 1986; Ashton, 1990). In future research, decision-rule use could additionally be incentivized, and other incentives may be used. For example, stronger incentives may be procedures in which decision makers are held accountable for their decisions or are required to work with the hired person. Thus, future research is needed to test the effect of educational information on professionals' decision-rule use in high-stakes selection procedures.

A third limitation concerns the choice of predictors. Although we used predictors that are also commonly used in selection- and admission procedures, the personal statement was the only source of information that participants were presented with that was not used to build the decision rule. Another commonly used poor predictor for academic success and job performance is the unstructured interview (König et al., 2010; Lievens & De Paepe, 2004; Michel et al., 2019). It may be that a real, unstructured interview would have had more strongly influenced the deviation from the decision rule for non-educated participants. However, it remains an open question whether watching the educational video would have helped participants to resist deviating from the rule in the presence of a strong, but likely less

valid interview impression. In future research, other common predictors such as the interview and the resumé could be used.

Future research may also investigate whether educational information can increase the perceptions of other stakeholders, such as an organization's employees, managers, and applicants. Mechanical information combination is partly underutilized because decision makers recognize that peers ascribe less credit to their hiring decision outcomes when information is mechanically combined (Nolan et al., 2016). However, informing peers that a competent decision maker is someone who uses evidence-based decision rules and who is aware of the limitations of expert judgment may change peers' perceptions. Since applicant reactions constitute an important part of selection procedures (König et al., 2010; Sackett & Lievens, 2008) and people generally hold negative attitudes towards mechanical combination (Diab et al., 2011; Eastwood et al., 2012), future research could investigate whether educational information can increase applicant reactions towards mechanical combination. Previous research already showed that an educational video increased police officer's fairness perceptions of a test in a real selection procedure (Truxillo et al., 2002). Therefore, educational information may also increase attitudes towards mechanical combination. Organizational justice theory poses a useful theoretical framework for such future research (Gilliland, 1993).

Furthermore, future research could focus on the effectiveness and the underlying mechanisms of different educational interventions. Specifically, it may be investigated whether educational interventions that present causal arguments for the use of decision rules as we did result in more decision-rule use and greater prediction accuracy than simple descriptive statements (Eastwood & Luther, 2016) or instructions (Arkes et al., 1986).

Research that sheds light on potential moderating effects of outcome feedback is also needed. We provided outcome feedback to provide a strict test for an education effect, as

previous research has shown that outcome feedback decreases decision-rule use and prediction accuracy (Arkes et al., 1986; Dietvorst et al., 2015; Jackson et al., 2019). Decision makers who receive educational information would still have to tolerate feedback that shows they almost always make errors to some extent. Hence, it could be that the education effect is stronger when no outcome feedback is provided, as is also more representative of decision making in practice.

Lastly, we measured decision-rule use again one month after participants received educational information. This may be considered a rather short period. Therefore, longitudinal studies are needed in which the effect of an educational intervention is tested after a longer period of time.

Practical Implications

Providing educational information constitutes a feasible and inexpensive intervention that can increase decision-makers' knowledge of evidence-based decision making in practice. As a result, decision rules may be used more often, which can translate into increased prediction accuracy, although this effect may only be temporary. Therefore, organizations could introduce training sessions on evidence-based decision making. Similarly, hiring managers and admission officers could be automatically sent information on evidence-based decision making as reminders when they publish a vacancy in an application system. Although experienced decision makers may resist or ignore such information if it is routinely offered, in other professional fields such as aviation, construction, and chemical production, it is commonplace for experienced professionals to receive such information in the form of occupational safety trainings to reduce consequential errors (Burke et al., 2006; Grote, 2012; Kaplan & Tetrick, 2011). Another way to transfer knowledge on evidence-based decision-making to practitioners could be via more attention for this topic from professional societies and in test guidelines.

Conclusion

With regard to the general superiority of mechanical combination over holistic combination, Meehl (1986) already claimed more than thirty years ago that “there is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one” (p. 373). Yet, many decision makers in practice are still unaware of these robust findings. With this study, we provided a first test of a simple educational intervention to increase decision-makers’ use of mechanical combination.

References

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382. <https://doi.org/10.1177/0011000005285875>
- 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](https://doi.org/10.1016/0749-5978(91)90020-T)
- Arkes, H. R. (2008). Being an advocate for linear models of judgment is not an easy life. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes*. (pp. 47–70). Psychology Press.
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 37*, 93–110. [https://doi.org/10.1016/0749-5978\(86\)90046-4](https://doi.org/10.1016/0749-5978(86)90046-4)
- Arthur, W. J., Bennett, W. J., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57–101. https://doi.org/10.1207/s15327043hup1101_3
- Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research, 28*, 148–180. <https://doi.org/10.2307/2491253>
- Banks, G. C., Pollack, J. M., Bochantin, J. E., Kirkman, B. L., Whelpley, C. E., & O’Boyle, E. H. (2016). Management’s science–practice gap: A grand challenge for all stakeholders. *Academy of Management Journal, 59*, 2205–2231. <https://doi.org/10.5465/amj.2015.0728>
- Barrick, M. R., Day, D. V., Lord, R. G., & Alexander, R. A. (1991). Assessing the utility of

executive leadership. *The Leadership Quarterly*, 2, 9–22. [https://doi.org/10.1016/1048-9843\(91\)90004-L](https://doi.org/10.1016/1048-9843(91)90004-L)

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>

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>

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, 12, 19–64. <https://doi.org/10.2308/jmar.2000.12.1.19>

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, 27, 303–345. [https://doi.org/https://doi.org/10.1016/S0361-3682\(01\)00052-6](https://doi.org/https://doi.org/10.1016/S0361-3682(01)00052-6)

Burgess, J. R. D., & Russell, J. E. A. (2003). The effectiveness of distance learning initiatives in organizations. *Journal of Vocational Behavior*, 63, 289–303. [https://doi.org/https://doi.org/10.1016/S0001-8791\(03\)00045-9](https://doi.org/https://doi.org/10.1016/S0001-8791(03)00045-9)

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health*, 96, 315–324. <https://doi.org/10.2105/AJPH.2004.059840>

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–

42. <https://doi.org/https://doi.org/10.1023/A:1007850605129>

Clinedinst, M. E., & Patel, P. (2018). *State of College Admission 2018*.

https://www.nacacnet.org/globalassets/documents/publications/research/2018_soca/soca18.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

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

Custers, E. J. F. M. (2010). Long-term retention of basic science knowledge: A review study.

Advances in Health Sciences Education, 15, 109–128. <https://doi.org/10.1007/s10459-008-9101-y>

Davis, K. M., Doll, J. F., & Sterner, W. R. (2018). The importance of personal statements in counselor education and psychology doctoral program applications. *Teaching of Psychology, 45*, 256–263. <https://doi.org/10.1177/0098628318779273>

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

<https://doi.org/10.1037/h0030868>

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3–11). Lawrence Erlbaum.

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>

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>
- 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>
- Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*, 458–468. <https://doi.org/10.1002/bdm.741>
- Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020). The effectiveness of refutation texts to correct misconceptions among educators. *Journal of Experimental Psychology: Applied*, *26*, 411–421. <https://doi.org/10.1037/xap0000258>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Fischhoff, B. (1982). Debiasing. In A. Tversky, D. Kahneman, & P. Slovic (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 422–444). Cambridge University Press. [https://doi.org/DOI: 10.1017/CBO9780511809477.032](https://doi.org/DOI:10.1017/CBO9780511809477.032)
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and Decision Making. *Annual Review of Psychology*, *71*, 331–355. <https://doi.org/10.1146/annurev-psych-010419-050747>
- 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>

- Fukukura, J., Ferguson, M. J., & Fujita, K. (2013). Psychological distance can improve decision making under information overload via gist memory. *Journal of Experimental Psychology: General*, *142*, 658–665. <https://doi.org/10.1037/a0030730>
- Garbers, Y., & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational & Organizational Psychology*, *87*, 102–137. <https://doi.org/http://dx.doi.org/10.1111/joop.12039>
- Gill, C. (2018). Don't know, don't care: An exploration of evidence based knowledge and practice in human resource management. *Human Resource Management Review*, *28*, 103–115. <https://doi.org/10.1016/j.hrmr.2017.06.001>
- Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *The American Psychologist*, *23*, 483–496. <https://doi.org/10.1037/h0026206>
- Grote, G. (2012). Safety management in different high-risk domains – All the same? *Safety Science*, *50*, 1983–1992. <https://doi.org/https://doi.org/10.1016/j.ssci.2011.07.017>
- 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>
- Guay, J. P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior*, *45*, 82–100. <https://doi.org/10.1177/0093854817719482>
- Hammond, K. R., & Summers, D. A. (1972). Cognitive control. *Psychological Review*, *79*, 58–

67. <https://doi.org/10.1037/h0031851>

- Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *9*, 30–34. [https://doi.org/10.1016/0030-5073\(73\)90034-2](https://doi.org/10.1016/0030-5073(73)90034-2)
- 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>
- 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>
- Jenkins, G. D. J., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, *83*, 777–787. <https://doi.org/10.1037/0021-9010.83.5.777>
- Kaplan, S., & Tetrick, L. E. (2011). Workplace safety and accidents: An industrial and organizational psychology perspective. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization*. (pp. 455–472). American Psychological Association. <https://doi.org/10.1037/12169-014>
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*, 404–426. <https://doi.org/10.1037/0033-2909.134.3.404>
- Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection:

When and why unstructured interview information can hurt hiring decisions.

Organizational Behavior and Human Decision Processes, 137, 27–44.

<https://doi.org/10.1016/j.obhdp.2016.07.005>

Kirch, D. G. (2012). Transforming admissions: The gateway to medicine. *JAMA: Journal of the American Medical Association*, 308, 2250–2251.

<https://doi.org/10.1001/jama.2012.74126>

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746–759.

<https://doi.org/10.1177/0013164496056005002>

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/http://dx.doi.org/10.1111/j.1468-](https://doi.org/http://dx.doi.org/10.1111/j.1468-2389.2010.00485.x)

[2389.2010.00485.x](https://doi.org/http://dx.doi.org/10.1111/j.1468-2389.2010.00485.x)

Kuncel, N. R. (2018). Judgment and decision making in staffing research and practice. In D. S.

Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The sage handbook of industrial, work and organizational psychology* (2nd ed., pp. 474–487). SAGE

Publications Ltd. <https://doi.org/10.4135/9781473914940>

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success.

Science, 315, 1080–1081. <https://doi.org/10.1126/science.1136618>

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>

<https://doi.org/10.1037/a0034156>

Lawler, 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>

<https://doi.org/10.5465/AMJ.2007.27155013>

Leonard, D., & Swap, W. (2004). Deep smarts. *Harvard Business Review*, 82, 88–97.

- 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>
- 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>
- Majid, S., Foo, S., Luyt, B., Zhang, X., Theng, Y.-L., Chang, Y.-K., & Mokhtar, I. A. (2011). Adopting evidence-based practice in clinical decision making: nurses' perceptions, knowledge, and barriers. *Journal of the Medical Library Association*, *99*, 229–236. <https://doi.org/10.3163/1536-5050.99.3.010>
- Meehl, P. E. (1954). 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/doi:10.1037/11281-008>
- 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>
- 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>
- Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, *4*, 379–383. <https://doi.org/10.1111/j.1745-6924.2009.01142.x>
- 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>

Murphy, K. R., Deckert, P. J., Kinney, T. B., & Kung, M.-C. (2013). Subject matter expert judgments regarding the relative importance of competencies are not useful for choosing the test batteries that best predict performance. *International Journal of Selection and Assessment*, *21*, 419–429. <https://doi.org/10.1111/ijsa.12051>

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.

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE*, *10*, e0120644. <https://doi.org/10.1371/journal.pone.0120644>

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*, 1142–1150. <https://doi.org/10.1177/01461672002611010>

Neumann, M., Niessen, A. S. M., & Meijer, R. R. (2020). Implementing evidence-based assessment and selection in organizations: A review and an agenda for future research. *Organizational Psychology Review*. Advance online publication. <https://doi.org/10.1177/2041386620983419>

Neumann, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2021). The autonomy-validity dilemma in mechanical judgment procedures: The quest for a compromise [Manuscript submitted for publication]. Department of Psychometrics and Statistics, University of Groningen.

Niessen, A.S.M., Meijer, R. R., & Tendeiro, J. N. (2018). Admission testing for higher education: A multi-cohort study on the validity of high-fidelity curriculum-sampling tests. *PLoS ONE*, *13*, 1–22. <https://doi.org/10.1371/journal.pone.0198746>

- 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., & 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>
- Petty, R., & Cacioppo, J. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Reyt, J.-N., Wiesenfeld, B. M., & Trope, Y. (2016). Big picture is better: The social implications of construal level for advice taking. *Organizational Behavior and Human Decision Processes*, 135, 22–31. <https://doi.org/10.1016/j.obhdp.2016.05.004>
- 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/http://dx.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., Gerhart, B., & Parks, L. (2005). Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology*, 56, 571–600. <https://doi.org/10.1146/annurev.psych.56.091103.070254>
- 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. (2008). Personnel selection. *Annual Review of Psychology*, 59, 419–450. <https://doi.org/10.1146/annurev.psych.59.103006.093716>

Samuels, J. A., & Whitecotton, S. M. (2011). An effort based analysis of the paradoxical effects of incentives on decision-aided performance. *Journal of Behavioral Decision Making*, 24, 345–360. <https://doi.org/10.1002/bdm.691>

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>

Sarbin, T. R. (1943). A contribution to the study of actuarial and individual methods of prediction. *American Journal of Sociology*, 48, 593–602. <https://doi.org/10.1086/219248>

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>

Sellier, A.-L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, 30, 1371–1379. <https://doi.org/10.1177/0956797619861429>

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.

- Terpstra, D. E., & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*, 483–495.
<https://doi.org/10.1177/009102609702600405>
- 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.
- Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*, 876–889.
<https://doi.org/10.1037/0022-3514.79.6.876>
- Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and Practice, 40*, 525–531. <https://doi.org/10.1037/a0014693>
- Wiesenfeld, B. M., Reyt, J.-N., Brockner, J., & Trope, Y. (2017). Construal level theory in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 367–400. <https://doi.org/10.1146/annurev-orgpsych-032516-113115>
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology, 51*, 539–570. <https://doi.org/10.1146/annurev.psych.51.1.539>
- 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*. <https://doi.org/https://scholarworks.bgsu.edu/pad/vol6/iss2/2>

Table 1

Incentive scheme with obtained incentive per prediction based on the absolute deviation between the predicted- and observed first-year GPA.

Absolute deviation between predicted- and observed first-year GPA	Incentive
0.5 or less	€12.5 cents
0.7 or less	€7.5 cents
1.0 or less	€2.5 cents

Table 2*Correlations between studied variables.*

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Education	—								
2. Incentives	-.02	—							
3. Decision-rule deviation T1	-.34*	-.11	—						
4. Decision-rule deviation T2	-.21*	-.14	.64*	—					
5. Self-reported decision-rule use T1	.34*	.05	-.40*	-.41*	—				
6. Self-reported decision-rule use T2	.23*	.00	-.38*	-.48*	.58*	—			
7. Prediction deviation T1	-.26*	.00	.63*	.35*	-.21*	-.23*	—		
8. Prediction deviation T2	-.07	-.05	.32*	.54*	-.24*	-.26*	.11	—	
9. Age	-.21*	.05	.31*	.27*	-.04	-.23*	.26*	.18*	—

Note. Education and Incentives were coded as 1 = Yes and 0 = No. * $p < .05$, two-tailed. $N = 171$.

Table 3*Descriptive statistics for decision-rule deviation.*

Conditions	<i>n</i>	T1 <i>M (SD)</i>	T2 <i>M (SD)</i>
Education yes	76	0.39 (0.17)	0.38 (0.19)
Education no	95	0.52 (0.19)	0.45 (0.16)

Table 4*Descriptive statistics for self-reported decision-rule use.*

Condition	<i>n</i>	<i>M (SD)</i>
Control	102	4.12 (1.50)
Incentives only	88	3.81 (1.65)
Education only	84	4.61 (1.51)
Education and incentives	68	5.19 (1.22)

Table 5

Descriptive statistics for prediction accuracy.

Conditions	<i>n</i>	T1 <i>M (SD)</i>	T2 <i>M (SD)</i>
Education yes	76	0.76 (0.21)	0.81 (0.21)
Education no	95	0.89 (0.26)	0.83 (0.19)

Note. Higher scores indicate less prediction accuracy.

Table 6

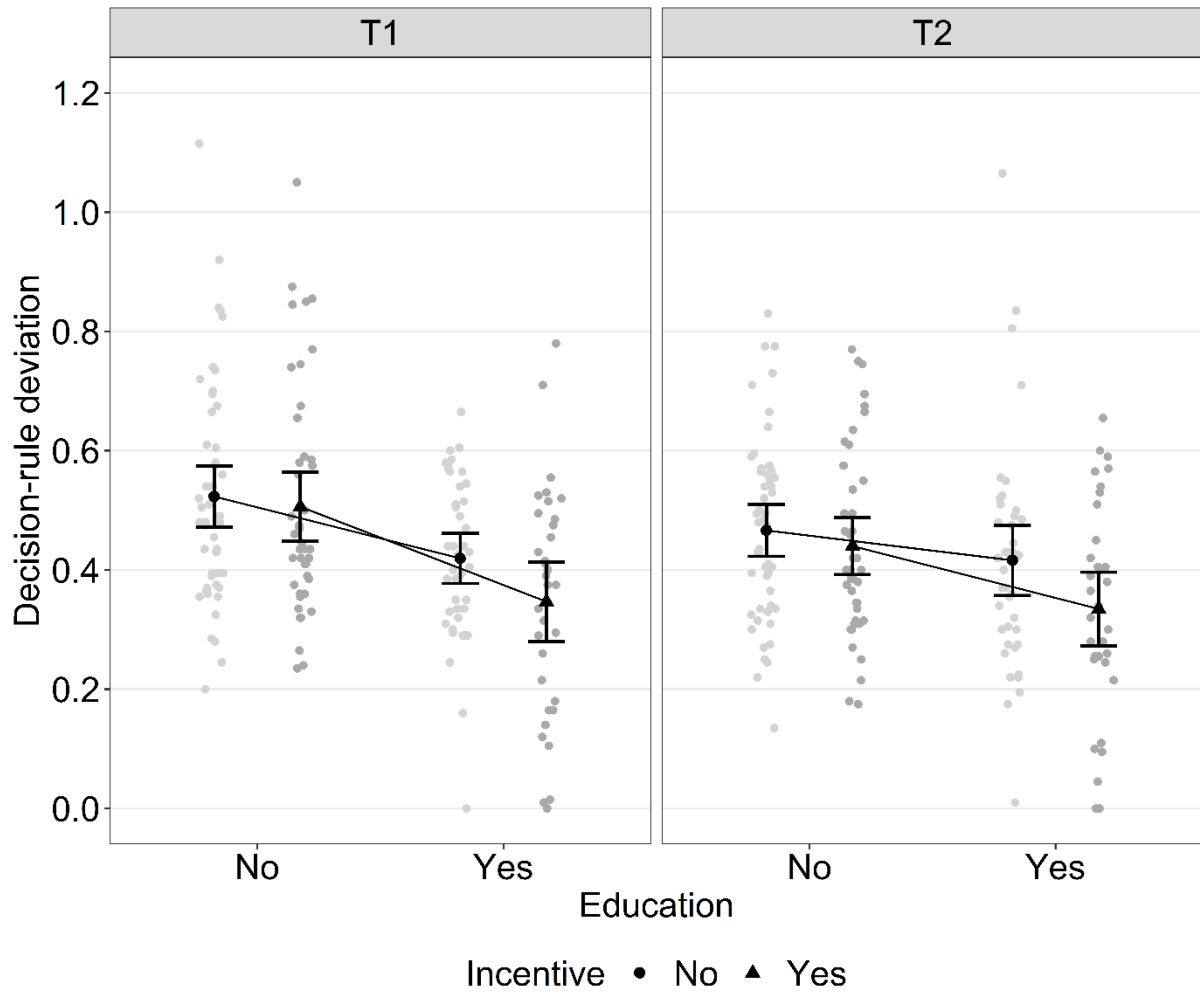
Correlations between participants' predictions and observed GPA, and between rule predictions and observed GPA per condition.

Condition	Correlation between participants' predictions and observed GPA		Correlation between rule predictions and observed GPA		Mean difference between correlations shown in column 1 and 2
	T1	T2	T1	T2	\bar{r}_{diff}
Education + Incentives	.43 [.36, .49]	.39 [.32, .45]	.53 [.47, .58]	.44 [.38, .50]	-.10 [-.17, -.02]
Education + Incentives	.42 [.36, .47]	.38 [.32, .43]	.54 [.49, .59]	.51 [.45, .55]	-.16 [-.22, -.09]
Education - Incentives	.29 [.23, .35]	.38 [.32, .44]	.49 [.44, .54]	.50 [.45, .55]	-.19 [-.26, -.13]
Control	.31 [.26, .37]	.38 [.32, .43]	.45 [.40, .50]	.55 [.51, .59]	-.19 [-.25, -.13]

Note. 95% confidence intervals are presented in brackets.

Figure 1

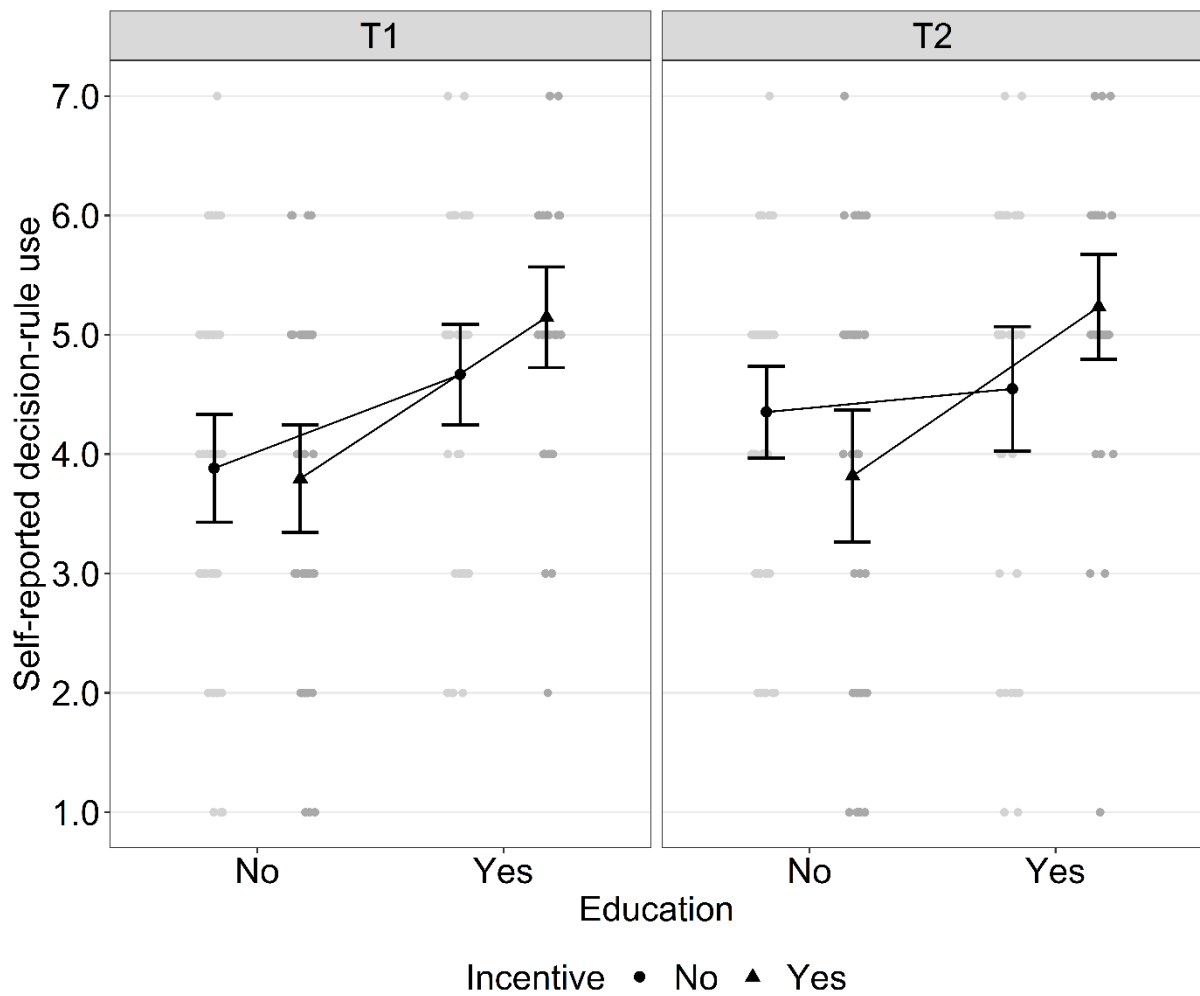
Decision-rule deviation means per condition and time.



Note. Higher values indicate larger deviation from the decision-rule prediction. Error bars represent 95% confidence intervals.

Figure 2

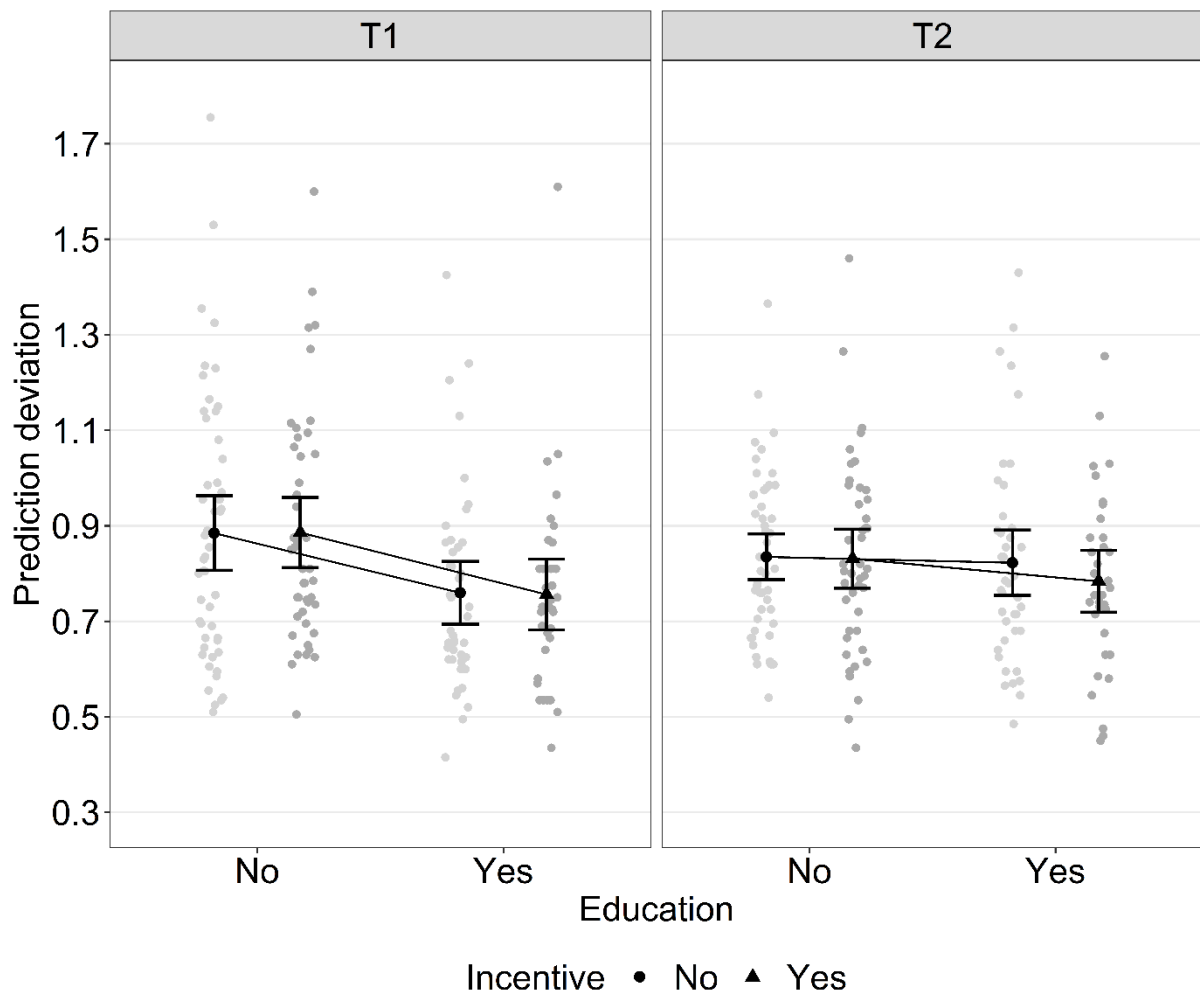
Self-reported decision-rule use means per condition and time.



Note. Higher values indicate more self-reported decision-rule use. Error bars represent 95% confidence intervals.

Figure 3

Prediction deviation means per condition and time.



Note. Higher values indicate larger deviation from observed first-year GPA. Error bars represent 95% confidence intervals.