Computers and Education: Artificial Intelligence 2 (2021) 100028

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Algorithm appreciation or aversion? Comparing in-service and pre-service teachers' acceptance of computerized expert models

Esther Kaufmann

Institute of Psychology, University of Konstanz, Fach 31, 78457, Konstanz, Germany

ARTICLE INFO

Keywords:
Artificial intelligence
Digitalization
Algorithm acceptance
Teacher education
Pre-service teachers

ABSTRACT

Although computerized expert models (i.e., algorithms) could improve educational decisions and judgments, initial research has demonstrated that teachers, like other professional groups, tend to be "algorithm averse." In the current study, we use behavioral and questionnaire data to examine the extent to which in-service and pre-service (i.e., students in training to become) teachers accept advice from expert models and investigate how teachers' acceptance of expert models could be improved. Although it is often presumed that younger generations are less algorithm averse, we demonstrate that both in-service and pre-service teachers prefer advice from a human source (school counselor) than from an expert model, to a similar extent. Furthermore, we find that advice acceptance depends on the difficulty of the decision task, but we find no evidence that pre-service teachers' acceptance of computerized advice depends on their numeracy or the Big Five traits of openness and neuroticism. Finally, we find that in-service teachers lacked knowledge of computerized expert models but indicated that advice from expert models would be superior to human advice in certain kinds of tasks. Our results indicate that both in- and pre-service teachers could profit from training about the definition and value of computerized expert models, and we provide suggestions for training and future research.

1. Introduction

Although the education environment has changed since the introduction of computers and the use of Artificial Intelligence and will continue to change, the teachers' task of making a plethora of important decisions as part of their daily work remains. Teachers need to make decisions about students, for instance, on their competency and motivation or how to allocate resources. However, the questions arising from the changing educational environment provide an option to support teachers in their judgment and decision-making. So far, the option of using the potential of computerized expert models remains greatly untapped, and computerized expert models (computerized algorithm advice)—that is, decision-making algorithms based on the statistical analysis of large datasets—could be used to improve teachers' decisions (Dawes et al., 1989; Grove et al., 2000; Kaufmann & Wittmann, 2016; Kuncel et al., 2013; Meehl, 1954). This is surprising because, in recent years, a rich palette of teachers' technology-adoption research has been conducted (Scherer et al., 2019).

Initial research on teachers' acceptance of computerized expert models has demonstrated that teachers, like other people, appear to be rather "algorithm averse" (Kaufmann & Budescu, 2020; Burton et al., 2020). In the current study, we use behavioral, and questionnaire data to (re-) examine in-service and pre-service teachers' acceptance of expert models. We were particularly interested in whether the next generation of teachers is more prepared to accept digitalized advice than the current generation. We also examine (a) whether teachers' acceptance of expert models depends on task difficulty and/or their personal characteristics, and (b) how teachers perceive expert model advice relative to human advice, and (c) how much they know about expert models.

1.1. *Teachers' judgment and decision-making in a digitalized environment*

Given the changing environment of teaching with digitalization, teachers must be able to accurately evaluate their students' competencies to support their students' learning and development optimally. When a teacher overestimates students' competencies, they may create a learning environment that is too challenging, with negative consequences for students' self-concept, learning, and motivation. However, if a teacher underestimates students' competencies, they may create a learning environment that is insufficiently challenging, leading to boredom and performances well below students' potential. Hence, inaccurate teacher judgments can result in conditions that prevent

E-mail address: esther.kaufmann@uni-konstanz.de.

https://doi.org/10.1016/j.caeai.2021.100028
Received 30 March 2021; Received in revised form 16 June 2021; Accepted 19 July 2021
Available online 23 July 2021

students from achieving their full potential and increase inequality between students' learning rates.

Given the consequences of inaccurate judgments, researchers and teacher education programs have recognized the ability to accurately diagnose students' academic competencies as an essential competency for teachers (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004; Shavelson, 1973) and an important facet of teachers' general and subject-specific pedagogic knowledge (Baumert & Kunter, 2006) and adaptive competency (Beck et al., 2008). However, several reviews (Hoge & Coladarci, 1989;[1] Südkamp et al., 2012; Urhahne & Wijnia, 2021) have demonstrated that the full potential of teachers' judgment accuracy is often not reached, and subsequently, there is still potential to improve teachers' judgment accuracy.

Researchers have suggested various ways that teachers' judgment accuracy could be improved (Südkamp & Praetorius, 2017); for example, feedback and interventions can improve teachers' judgment accuracy (Karing & Seidel, 2017). As outlined in the following sections, we argue that teachers can benefit from using artificial intelligence and increase their judgment accuracy from the feedback and advice of expert models produced with algorithms.

1.1.1. The potential of expert models

Expert models are formal decision-making tools. They consist of *algorithms* that are evidence-based forecasting formulas and rules such as statistical models, decision aids, or other mechanical procedures (Dietvorst et al., 2016) developed based on large datasets (Meehl, 1954). Today, expert models are commonly digitalized. One advantage of computerized expert models is that they can incorporate and combine multiple sources of information in a consistent way: given the same input, expert models—unlike human judges—reach the same decision every time. Furthermore, expert models are neither distracted by irrelevant cues nor affected by the many cognitive biases that are known to skew human judgment (Kahneman et al., 1982).

Meehl (1954) was the first to present evidence that statistical models outperform human judges. His evaluation of whether clinical psychologists outperform models has been discussed at length under the topics of bootstrapping (Goldberg, 1976; Karelaia & Hogarth, 2008) and "man versus model of man" (Einhorn, 1974). Overall, expert models tend to outperform and increase the accuracy of human judgments. Since then, several meta-analyses have confirmed this result in various decision domains (Dawes et al., 1989; Grove et al., 2000), including education (Kaufmann & Wittmann, 2016; Kuncel et al., 2013). However, there is also evidence that the superiority of expert models over human judges has been underestimated because past meta-analyses have failed to account for study artifacts like measurement error or sampling bias (Kaufmann & Wittmann, 2016).

The digital transformation of contemporary society, school reforms, and an increase in standardized testing have all led to an increase in digitalized data within schools (Schweizerische Konferenz der kantonalen Erziehungsdirektoren, 2018) and an increase in data-driven decision making in the educational field (Mandinach, 2012; Mandinach & Schildkamp, in press; Marsh et al., 2006; Schildkamp, 2019; Schildkamp et al., 2013). As a result, it is now possible to develop computerized expert models that can inform teachers' decisions about student potential and motivation and which students are most likely to benefit from particular interventions. As an example of how an expert model can improve educational outcomes, Connor et al. (2007) demonstrated that teachers using an algorithm to guide individualized reading instruction to first-grade students resulted in better reading development compared with a control group. The potential for using computerized expert models to improve teachers' decision-making hinges, however, on the teachers' willingness to accept computerized, algorithmic advice.

However, taking a broader scope, several direct or indirect factors (e.g., attitudes toward technology) are associated with teachers' adoption of technology (Straub, 2009). Vannatta and Fordham (2004) suggested a combination of factors (amount of technology training, time spent beyond contractual work week, and openness to change) that best predict the use of classroom technology. However, as we outline in the following section, although there is rich literature on technology adoption, only a few studies have focused explicitly on computerized, algorithmic advice.

1.1.2. Neglection of advice from expert models (algorithm aversion)

People generally tend to prefer advice from humans than algorithms, a phenomenon called "algorithm aversion" (Burton et al., 2020; Dietvorst et al., 2016). So far, only Kaufmann and Budescu (2020) have investigated teachers' acceptance of computerized expert models specifically. In Kaufmann and Budescu (2020), middle- and high-school teachers ($N = 435$) were asked to make judgments that were either difficult or easy based on students' profiles. Before reaching a decision, they were either offered (Studies I and II) or could ask for (Study III) advice from either an expert model, a school counselor, or both. Teachers requested and followed advice from an expert model less frequently than they requested and/or followed advice from a school counselor.

Interestingly, although judgments based on multiple opinions are usually more accurate than judgments made by a single person (cf. the "wisdom of the crowds" Galton, 1907; Kaufmann & Wittmann, 2016; Budescu & Chen, 2015; Karelaia & Hogarth, 2008; Meehl, 1954), teachers rarely opted to receive advice from both a school counselor and an expert model. The results of the study demonstrated that teachers, like other professionals, appear to be algorithm averse. Kaufmann and Budescu (2020) also found no evidence that in-service teachers' acceptance of computerized expert models depended on either their Big Five traits of openness and neuroticism or their objective numeracy competency (i.e., their ability to understand and use probabilistic and mathematical concepts).

However, several studies have indicated that the acceptance of expert models varies widely (Chacon, Kausel, & Reyes, 2021; Dietvorst et al., 2014; Germann & Merkle, 2019; Logg et al., 2019; Longoni et al., 2019; Yeomans et al., 2019; Önkal et al., 2009). Furthermore, there is some evidence that people are more averse to using algorithms to make subjective (e.g., recommending a romantic partner or a gift) as opposed to objective judgments (e.g., diagnosing a disease or piloting a plane; Castelo et al., 2019; Logg et al., 2019). Other research has demonstrated that people—including teachers—are more likely to follow advice from expert models when making difficult versus easy decisions (Kaufmann & Budescu, 2020). Hence, it seems that task difficulty plays a role in advice taking (Gino & Moore, 2007; Schrah et al., 2006). Finally, there are inconsistent results on teachers' judgments of students' achievements within different subject matter (Kolovou et al., 2021; Südkamp et al., 2012).

To date, it is unknown whether algorithm aversion might also depend on the characteristics of the decision-makers. Based on a review of 61 studies of algorithm aversion published between 1950 and 2018, Burton et al. (2020) recommended a combination of domain-specific training and general training in algorithm literacy (e.g., training on statistical concepts like error and uncertainty) to improve algorithm acceptance. However, they also pointed out that members of more recent generations might already have higher algorithm literacy than their older peers. So far, however, there is little evidence regarding whether either domain expertise or (proxies of) algorithm literacy are related to expert model acceptance. Existing studies on expert model acceptance have typically been based on student samples or, in other words, people with little experience in the decision domain (Burton

[1] Kaufmann (2020) provide an up-to-date meta-analysis of Hoge and Coladarci's review and demonstrate that although teachers' judgment accuracy within this review is underestimated, there is still potential to increase it with algorithms.

et al., 2020), and studies that systematically compare expert model acceptance of domain "experts" and "novices" have been rare. Those that have been conducted have demonstrated that experts discount algorithm advice sources (Logg et al., 2019). Additionally, Authors (2020a) only included in-service teachers in their study, and so their results have no information about whether acceptance of expert models varies as a function of expertise in educational decision making.

1.2. Overview of the three studies in the current study

In sum, computerized expert models can improve teachers' decisions, but only if teachers are willing to use them. So far, only one study has investigated the extent to which in-service teachers accept digitalized advice (Kaufmann & Budescu, 2020). In the current study, we contribute to the literature by examining the extent that pre-service teachers accept expert models using a mix of behavioral and questionnaire data. We also contribute to the literature by comparing pre-service and in-service teachers' *perceptions* of advice from expert models compared to a school counselor, and in-service teachers' *knowledge* of computerized expert models and when they think expert models should be used.

We were particularly interested in identifying ways to potentially improve teachers' acceptance of expert models. Comparing experts and novices in educational decision-making can, for instance, indicates whether pre-service teachers are also in need of training or whether training in the decision domain might improve acceptance. Therefore, our general hypothesis was that pre-service teachers would be more inclined to accept expert model advice than in-service teachers and may also prefer the advice by an expert model compare to the advice by a school counselor.

2. Study I: behavioral data on expert model acceptance

In Study I, we use behavioral data to compare pre-service and in-service teachers' acceptance of advice from computerized expert models. Specifically, we examine the extent to which both groups of teachers seek advice from a human expert (school counselor), an expert model, or both, when deciding which of two students should receive additional tutoring. We also examine how participants react to advice (i.e., whether they follow advice and how certain they feel about their judgment with and without advice) and explore advice behavior in difficult and easy judgment tasks.

We expected that both pre-service and in-service teachers would prefer advice from a school counselor than from an expert model. However, we also expected that pre-service teachers would demonstrate more acceptance of expert models than in-service teachers due to their lack of educational decision-making experience and because they have presumably grown up within a more digitalized environment (Burton et al., 2020). Additionally, in line with previous research (Kaufmann & Budescu, 2020), we expected that both in-service and pre-service teachers would be more willing to accept advice from expert models when making difficult judgments and would also want more advice when making difficult judgments.

Finally, we explored whether pre-service teachers' advice behavior was related to their numeracy and the Big Five personality characteristics of openness and neuroticism. We focused exclusively on pre-service teachers because previous research has found no evidence that in-service teachers' acceptance of expert models was related to their openness, neuroticism, or numeracy (Kaufmann & Budescu, 2020) and expected to replicate this result in our current study.

2.1. Method for study I

2.1.1. Sample

The sample consisted of in-service and pre-service teachers in Switzerland. To recruit in-service teachers, we invited selected school leaders in different cantons (regions) to participate in the study. The school leaders forwarded an email containing a link to the online study to teachers at their school. The sample of in-service teachers consisted of the sample teachers from the third study by Authors (2020a; $N = 76$ teachers recruited in May 2018) along with 23 additional teachers recruited in May 2019 ($N = 99$ in-service teachers). We followed a similar strategy to recruit pre-service teachers. In June 2019, we asked the head of research at a university of teacher education to forward an email containing a link to the online study to students pursuing a middle/secondary teaching degree (*Sekundarstufe* 1; $N = 63$ pre-service teachers).

The sample of in-service teachers consisted mostly of experienced teachers; 70.7 % had 10+ years of experience. There were slightly more high-school (44.4 %) than middle/secondary-school (38.4 %) teachers. About one-third of the in-service teachers (37.4 %) were women, and most were between 30 and 60 years old (87.0 %). The pre-service teachers all had some, but fewer than five years, teaching experience (in Switzerland, students pursuing a degree in teaching must complete multiple in-school internships). Most pre-service teachers were female (74.6 %), and almost all (90.5 %) were between 20 and 29 years old.

2.1.2. Procedure

Fig. 1 outlines the procedure. First, the participants read general information about the study, including a definition of computerized expert models. The participants then completed four judgment tasks (Kaufmann & Budescu, 2020). In each task, participants were presented with two student profiles and asked to decide which of the two students should receive additional tutoring. Each profile contained information about the student's biography (e.g., class level and gender) and their grades in different school subjects. An example judgment task is available in the Appendix (see Figure A1).

Two of the judgment tasks were difficult, and two were easy. The *difficult tasks* involved two students who varied in either gender or class level but had similar grades. Hence, it was unclear who was more in need of extra tutoring. In the *easy tasks*, students had different grades in the relevant subjects.

For pre-service teachers, we also manipulated the tutoring subject. Namely, pre-service teachers were asked to complete one difficult and one easy task for tutoring in *language* and one difficult and one easy task for tutoring in *math*. In-service teachers only judged students' needs in *math* tutoring (Kaufmann & Budescu, 2020).

After reading the profiles but before indicating their decision, the participants had to decide (a) whether they wanted to receive advice and, if so, (b) whether they wanted advice from a computerized expert model, a school counselor, or both.

The participants then made their decision. We then checked if advice was requested and if the decision was in line with the advice (= 1) or disagreed with the advice (= 0); this was to measure if the teachers followed the advice requested. After each decision, they also *judged their certainty* ("How certain are you about your judgment?" 1 = not at all certain, 9 = very certain). After completing all four judgment tasks, the participants completed several additional questionnaire items (see Study II). Pre-service teachers completed measures of numeracy (Weller et al., 2013), neuroticism, and openness (Rammstedt & Danner, 2016).

2.1.3. Analytic strategy

We first aggregated the data within each student profile (see Table A1). Depending on the variable of interest, we then aggregated the data either across participants or across tasks. We used t-tests and χ^2-tests on the different subsamples of the participants to compare the frequency of how often the pre-service and in-service teachers requested advice and how frequently they requested advice from one or two sources. Similarly, we used t-tests and χ^2-tests based on different subsamples of tasks to compare how frequently advice was requested for easy and difficult tasks, how frequently advice was requested from one or two sources, and how frequently advice was requested from school

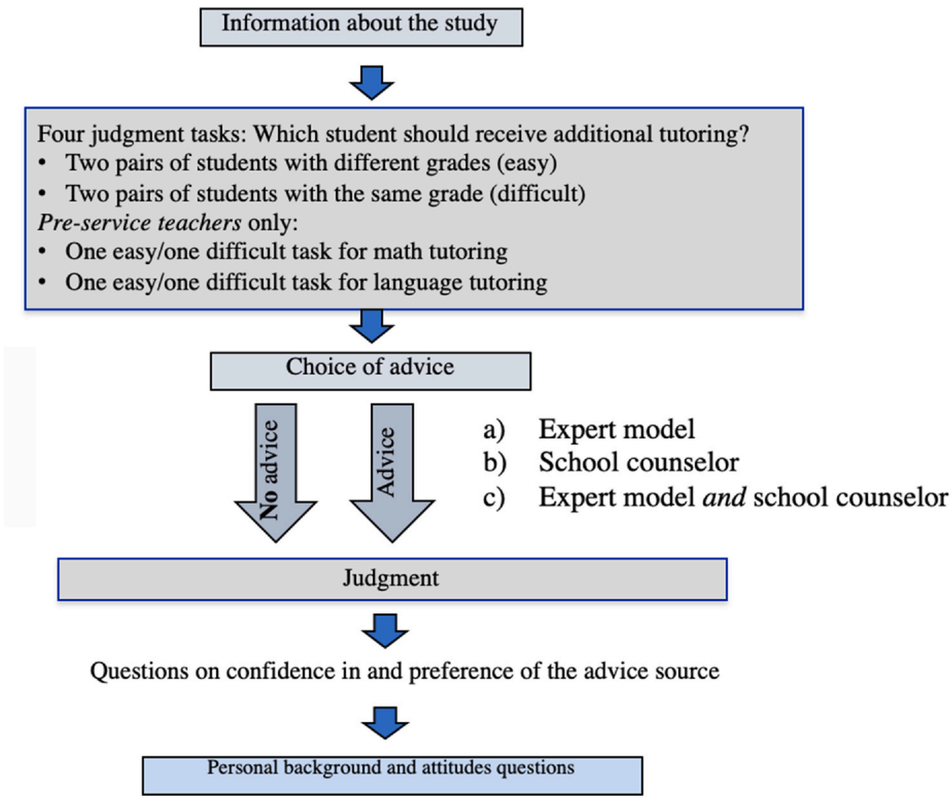

Fig. 1. Study I procedure.

counselors versus expert models. We used Pearson correlations to examine the association between pre-service teachers' numeracy, neuroticism, and openness and their judgment certainty with and without advice, as well as their tendency to follow advice.

2.2. Results

2.2.1. Advice seeking

Fig. 2 displays how frequent pre-service and in-service teachers requested advice. Pre-service teachers (93.4 %, $n = 58/63$) requested advice at least once more frequently, compared (75 %, $n = 74/99$) to in-service teachers ($\chi^2(162) = 7.65$, $p < .01$).

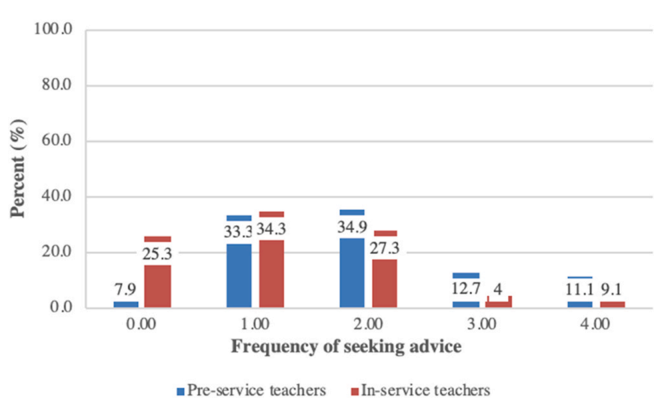

Fig. 2. Frequency of advice seeking, pre-service, and in-service teachers across four tasks.

Both pre-service and in-service teac hers were more likely to request advice in difficult than in easy tasks (in-service teachers: 71.8 % versus 19.2 %, $t(98) = -8.7$, $p < .001$; pre-service teachers: 88.9 % versus 30.1 %, $t(61) = -10.08$, $p < .001$; see Table A1). Among pre-service teachers,[2] the frequency of advice seeking did not depend on whether the tutoring was for math or language ($t(62) = -0.12$, $p = .89$), and it was independent of whether the task was difficult or easy ($t(62) = 0.96$, $p = .34$ and $t(62) = 0.57$, $p = .56$, respectively).

2.2.2. Preferred number of advice sources

Table 1 displays the proportion of tasks in which advice was requested from one or two sources in each subsample and as a function of task difficulty. Here we analyze only the subset of tasks in which advice was demanded (in-service: $k = 136$, 34.4 % of all tasks; pre-service: $k = 117$, 46.6 %).

Overall, advice was requested more or less equally from one (54.2 %, $k = 137$) or two sources (45.8 %, $k = 116$, $\chi^2(1, k = 253) = 1.73$, $p =$

Table 1
Percent of tasks in which advice was requested by the number of subsamples, requested source, and task difficulty (in-service: K = 136 tasks, pre-service: K = 117 tasks).

Sample	Task difficulty	Any advice requested	Number of requested sources	
			1	2
Pre-service	Easy	22.2	7.7	14.5
	Difficult	77.8	42.7	35.0
	Total		50.4	49.6
In -service	Easy	22.1	9.6	12.5
	Difficult	77.9	47.8	30.1
	Total		57.4	42.6

[2] Contrary to pre-service teachers, in-service teachers only judged students' need for *math* tutoring; hence, no subject differences are reported.

.18). There was also no difference with regard to how frequent either in-service or pre-service teachers (i.e., when data were aggregated across the participants as opposed to tasks) requested advice from one or two sources (in-service: $t(73) = 1.31$, $p = .19$; pre-service: $t(57) = 0.07$, $p = .94$).

The number of advice sources depended on the task difficulty ($\chi^2(1, k = 253) = 6.4$, $p = .01$). Specifically, as illustrated in Table 1, teachers preferred advice from just one source in difficult tasks and from two sources in easy tasks. The unexpected preference for advice from just one source in difficult tasks was observed among both in-service and pre-service teachers, although the within-group comparisons just missed significance (in-service teachers: $\chi^2(1, k = 136) = 3.09$, $p = .06$; pre-service teachers: $\chi^2(1, k = 117) = 3.34$, $p = .054$).

Focusing only on difficult tasks in which advice was sought, in-service teachers preferred to receive advice from just one (61 %) over two (39 %) sources ($\chi^2(1, k = 106) = 5.43$, $p = .02$). In contrast, pre-service teachers did not demonstrate a preference for advice from either one or two sources in difficult tasks ($\chi^2(1, k = 91) = 0.89$, $p = .34$). We could not compare the preference for advice from one or two sources in easy tasks due to the very small sample of easy tasks in which advice was sought (pre-service teachers: $k = 9$; in-service teachers: $k = 13$).

2.2.3. Preference for advice from school counselor versus the expert model

We used the subsample of $k = 137$ tasks in which only one advice source was selected to analyze whether participants preferred advice from a school counselor or an expert model. In these cases, advice was more frequently requested from a school counselor ($k = 113$, 82.5 %) than from an expert model ($k = 24$, 17.5 %, $\chi^2(1, k = 137) = 57.8$, $p < .001$), see Table 2. The same pattern of preference was observed among in-service ($\chi^2(1, k = 78) = 34.6$, $p < .001$) and pre-service ($\chi^2(1, k = 59) = 23.2$, $p < .001$)) teachers. Neither in-service nor pre-service teachers' preferences for advice from a school counselor depended on task difficulty (in-service teachers: $\chi^2(1, k = 78) = 0.46$, $p = .37$; pre-service teachers: $\chi^2(1, k = 59) = 1.51$, $p = .22$).

2.2.4. Effect of advice on judgment certainty

We used the subsamples of $n = 65$ in-service and $n = 50$ pre-service teachers who completed at least one task *with* and at least one task *without* advice to examine whether receiving advice was related to teachers' judgment certainty. Receiving advice did not affect in-service teachers' judgment certainty ($t(64) = -1.88$, $p = .06$). In contrast, pre-service teachers' judgment certainty was significantly higher when they received advice ($M = 7.55$ versus $M = 5.99$, $t(49) = 5.42$, $p < .001$; see also Table A2 in the Appendix), independent of subject (math: $M = 7.86$ with versus $M = 6.18$ without advice, $t(36) = 4.64$, $p < .001$; language: $M = 7.5$ with versus $M = 6.01$ without advice, $t(33) = 2.89$, $p = .01$). Compared to in-service teachers who received advice ($n = 58$), pre-service teachers who received advice ($n = 74$) felt more certain about their judgment ($M = 7.6$ versus $M = 5.29$, $t(130) = -5.45$, $p < .001$). Pre-service teachers who received advice felt more certain than in-service teachers who received advice regardless of whether they received advice from a school counselor, expert model, or both (see Table A3 in the Appendix). Interestingly, pre-service teachers felt more certain about their judgment when deciding which student should receive additional tutoring in math ($M = 6.94$, $SD = 1.35$) than in language ($M = 6.35$, $SD = 1.68$; $t(62) = 3.18$, $p = .002$). There was no evidence that pre-service teachers' judgment certainty was related to their numeracy ($r = 0.06$, $p = .62$), openness ($r = -0.05$, $p = .7$) or neuroticism ($r = -0.12$, $p = .31$).

Table 2
Percent of tasks in which advice was requested from a school counselor versus and an expert model by subsample and task difficulty (by pre-service teachers: k = 78 and by in-service teachers: k = 59 tasks in which advice was requested from just one source).

Sample	Task difficulty	Preferred source of advice (%)		
		Counselor	Expert Model	Total
Pre-service	Easy	10.2	5.1	15.3
	Difficult	71.2	13.6	84.7
	Total	81.4	18.6	
In-service	Easy	12.8	3.8	16.7
	Difficult	70.5	12.8	83.3
	Total	83.3	16.7	

2.2.5. Following advice

Overall, the participants almost always followed the advice they received when they requested it (93.6 % of the time; see Table A4 in the Appendix). Pre-service and in-service teachers did not differ in following advice (95.6 % versus 93.9 %), and there were also only minor differences in the tendency to follow advice for easy (pre-service: 100 %, in-service: 90 %) and difficult tasks (pre-service: 92.1 %, in-service: 94.3 %). Additionally, among pre-service teachers, following advice was unrelated to numeracy ($r = -0.3$ to -0.09 across profiles), neuroticism, or openness ($r = -0.3$ to 0.03 across profiles).

2.3. Discussion

The results indicate that, when given the opportunity, pre-service and in-service teachers do request and follow advice, particularly when they have to make difficult judgments. In line with previous studies (Kaufmann & Budescu, 2020) on in-service teachers, our current study also focuses on pre-service teachers demonstrates that pre-service teachers also prefer advice from a human source (school counselor) over advice from an expert model.

Interestingly, we found that teachers tended to prefer two sources of advice for easier tasks, but just one source of advice for difficult tasks. Potentially, teachers may be less concerned about receiving contradictory advice during easy tasks because they are interested only in confirming their own judgment. For difficult tasks, teachers may wish to avoid having to deal with contradictory advice. Contrary to previous research (Kaufmann & Budescu, 2020), there was no significant interaction between task difficulty and the requested number of advice sources.

We found no evidence that pre-service teachers' advice behaviors depended on whether they had to decide whether students should receive tutoring in math or language. However, they felt more certain about their judgments when deciding which student should receive additional tutoring in math than in language. Potentially, pre-service teachers perceive grades as a more objective indicator of math ability than language ability and believe it is easier to gauge a student's math ability than their language ability.

In line with previous results based on in-service teachers (Kaufmann & Budescu, 2020), numeracy, openness, and neuroticism were unrelated to how pre-service teachers reacted to advice regarding their judgment certainty and propensity to follow advice.

Taken together, the results of Study I suggest that both in-service and pre-service teachers prefer to receive advice from a school counselor than from an expert model, independent of either task difficulty or their personal characteristics. They also do not appear to expect that advice from an expert model could provide any "added value" over and above advice from a school counselor. Thus, it does not appear that the next generation of Swiss teachers is more accepting of expert models than the current generation of teachers.

3. Study II: questionnaire study on perceptions of advice from expert models

Study I used behavioral data to examine and compare the extent to which pre-service and in-service teachers accept advice from computerized expert models. In Study II, we use questionnaire data to shed light on the results of Study I. Specifically, we compare in-service and pre-service teachers' perceptions of advice from computerized expert models relative to advice from a human source (school counselor) across six characteristics. We additionally examine whether age, gender, and teaching experience predict in-service and pre-service teachers' perceptions of advice from both sources (Kaufmann & Budescu, 2020).

3.1. Method for study II

3.1.1. Sample and procedure

The sample included a total of $N = 432$ Swiss in-service and pre-service teachers. The sample of $N = 369$ in-service teachers consisted of the sample of in-service teachers from the study by Authors (2020a; $n = 107$ from Study 1, $n = 163$ from Study 2) and the sample of in-service teachers from Study I in the current study ($n = 99$). The sample of $N = 63$ pre-service teachers was the same as in Study I. Just over half (56.4 %, $n = 208$) of the in-service teachers taught at the middle school level, 27.6 % ($n = 102$) taught at the high school level, and 16 % ($n = 59$) taught at both levels. However, the proportion of women was lower in the sample of in-service teachers (42 %, $n = 155$) than in the sample of pre-service teachers (75 %, $n = 47$). Additionally, only a minority of in-service teachers were 30 years or younger (7.9 %, $n = 29$); in comparison, almost all of the pre-service teacher samples were 30 years or younger (96.8 %, $n = 61$).

After completing the judgment task described above (or a very similar task for the in-service teachers from Kaufmann & Budescu, 2020), participants used a 7-point scale (1 = preference for school counselor, 7 = preference for expert model) to indicate the extent to which they perceived advice from an expert model as more or less *objective, reliable, accurate, trustworthy*, *transparent*, and *beneficial for students* (= six ratings) than advice from a school counselor. The participants also reported their age, gender, and teaching experience.

3.1.2. Analytic strategy

We used t-tests to test whether in-service and pre-service teachers' average ratings differed from the scale mid-point (= 4). A significant *t*-test indicates that the participants perceived a difference between advice from school counselors and advice from expert models. We also used t-tests to examine whether pre-service teachers' ratings differed from in-service teachers' ratings. Finally, to verify our reported results, we ran a series of six regressions (i.e., one regression for each item) using data from the combined sample to examine whether teachers demographic variables like age, gender, and teaching experience were associated with each of the item ratings (Kaufmann & Budescu, 2020).

3.2. Results

Table 3 displays how pre-service and in-service teachers perceived advice from expert models relative to advice from school counselors. In-service teachers rated advice from school counselors as more beneficial for students, more transparent, more trustworthy, more accurate, and more reliable than advice from an expert model (all $p < .001$). In-service teachers rated advice from school counselors and advice from expert models as equally objective ($t(365) = 1.27, p = .20$). Pre-service teachers likewise rated advice from school counselors as more beneficial for students, more reliable, more accurate, and more trustworthy (all $p <$ .00). However, pre-service teachers rated advice from school counselors and expert models as equally transparent ($t(62) = -0.14, p = .88$).

Compared to pre-service teachers, in-service teachers perceived less of a difference between advice from school counselors and advice from an expert model with regard to the objectivity of the advice, but more of a difference with regard to how transparent the advice was for students. There were no indications that pre-service and in-service teachers differed about their perceptions of how much more reliable, accurate, trustworthy, or beneficial advice from a school counselor was relative to advice from an expert model. The results of the regression analyses provided no evidence that any of the ratings were related to age, gender, or teaching experience (see Table A5 in the Appendix).

Table 3
Pre-service and in-service teachers' ratings of advice from a school counselor versus an expert model.

Rating	Sample	N	M	SD	In favor of	*t*-test
Objectivity						
	Pre-service	63	5.23	1.43	Expert model	$t(427) = -4.88, p < .001$***
	In-service	366	4.11	1.72	–	
Reliability						
	Pre-service	63	3.11	1.21	School counselor	$t(430) = -.75, p = .45$
	In-service	369	2.97	1.37	School counselor	
Accuracy						
	Pre-service	63	3.53	1.26	School counselor	$t(429) = -.56, p = .57$
	In-service	368	3.43	1.42	School counselor	
Trustworthy						
	Pre-service	63	2.85	1.3	School counselor	$t(428) = .48, p = .63$
	In-service	367	2.95	1.45	School counselor	
Transparency						
	Pre-service	63	3.96	1.75	–	$t(428) = -2.02, p = .04$*
	In-service	367	3.49	1.71	School counselor	
Beneficial for students						
	Pre-service	63	2.69	1.1	School counselor	$t(427) = 2.92, p = .004$**
	In-service	366	3.28	1.51	School counselor	

Note: *$p < .05$. **$p < .01$. ***$p < .001$.

3.3. Discussion

The results of Study II confirm that pre-service and in-service teachers perceive advice from school counselors more favorably than they perceive advice from expert models. The results of Study II by pre-service teachers largely confirmed previous results regarding how in-service teachers perceive advice from expert models relative to advice from school counselors (Kaufmann & Budescu, 2020). This general result also confirms previous research in other decision domains such as medicine (Dietvorst et al., 2014; Önkal et al., 2009) that people prefer advice from a human source. This in line with the results of our Study I.

Interestingly, teachers do not appear to consider the perceived superior objectivity of advice from expert models as particularly advantageous. Thus, interventions that make teachers more aware of cognitive biases, the inconsistency of human judgment, and their negative consequences may help to improve their acceptance of expert models.

4. Study III: knowledge and potential application fields of computerized expert models

Study I used behavioral data, and Study II used questionnaire data to examine pre-service and in-service teachers' acceptance and perceptions of advice from computerized expert models. In Study III, we use an online questionnaire to examine in-service teachers' *knowledge* of expert models. We examine whether their knowledge about, and perceived utility of, expert models is related to their numeracy, age, and gender. Additionally, we asked teachers to list tasks in which they would prefer advice from an expert model and tasks in which they would prefer advice from a human expert. Our general aim was to identify how interventions might best improve teachers' acceptance of expert models.

4.1. Method for study III

4.1.1. Sample

We recruited a new sample of 47 in-service Swiss primary school teachers to complete an online survey. We contacted school leaders who forwarded an email containing a link to the online survey to teachers at their schools. The gender balance of the sample (78.7 % women) was consistent with the gender balance of primary school teachers in Switzerland (84 % women; Bundesamt für Statistic [Office for Statistics], 2016). Approximately half of the teachers were between 20 and 40 years old (57.4 %), the rest were between 40 and 60 years (42.6 %).

4.1.2. Procedure and measures

After clicking on the survey link, participants were informed that their answers were anonymous and that their data was legally protected. After confirming their willingness to participate, participants completed a 15-min survey. First, they were asked to define "expert models" using an open-end format. We considered the definition a complete and accurate definition if it contained "database" or related words (e.g., "dataset") as well as "algorithm" or related words (e.g., "mathematical combination," "equation"). We scored definitions as 0 (neither concept), 1 (either "algorithm" *or* "database"), or 2 (both "algorithm" and "database"). Hence, our coding strategy considers the two essential cornerstones of the term "expert models," namely, data and their aggregation. Additionally, in line with Logg et al. (2019) and to compare our results with theirs, we further analyzed the type of data aggregation (algorithm definition) into four subcategories (1 = Math, Equation, calculation, 2 = step by step procedure, 3 = logic, formula, 4 = others; Table 2).

After providing their answer, the participants were presented with the correct definition ("advice based on an algorithm arising from the analysis of a large data set"). The participants were then asked to estimate the *utility of expert models* and the *utility of human advice* using 5-point rating scales (5 = useful, 1 = useless).

Next, the participants used an open-answer format to describe three tasks in which expert models would provide better advice than a human judge and three tasks in which human advice would be superior. We first grouped the answers according to the decision domain (e.g., education or finance). We further classified situations in the educational domain as either *for assisting students*, *for assisting teachers*, *for preparing teaching lessons*, or *other*. Finally, teachers provided their age, gender, and completed a numeracy measure (Weller et al., 2013).

4.1.3. Analytical strategy

We analyzed the distribution of knowledge scores and used a *t*-test to compare how the participants rated the utility of expert models and human judges. We descriptively analyzed the relationship between knowledge of expert models, numeracy, and the perceived utility of advice from expert models and from human sources.

4.2. Results

4.2.1. Knowledge and perceived utility of advice from expert models

Most participants demonstrated at least some knowledge about expert models: 17 % ($n = 8$) provided a fully correct definition, and 68.1 % ($n = 32$) provided a partially correct definition. A minority (12.8 %, $n = 6$) provided an incorrect definition and 2.1 % ($n = 1$) did not answer. As described above, we then compared our results with the coding suggested by Logg et al. (2019). We also had a group of participants that did not provide any information about data aggregation (21.2 %, $n = 10$); hence, these participants were excluded from our comparison with Logg et al. (2019). Our comparison indicates that the participants mostly use terms such as math/equation/calculation (59.4 %, $n = 22$; Logg et al., 2019, p. 42 %), or a step-by-step procedure (24.3 %, $n = 9$; Logg et al., 2019, p. 26 %) in their definition, but seldom logic/formula related definitions (2.7 %, $n = 1$; Logg et al., 2019, p. 14 %). Some participants were also categorized into the "other" category (13.5 %, $n = 5$; Logg et al., 2019, p. 18 %).

After the participants obtained the definition of an expert model, they rated the utility of advice from expert models ($M = 3.04$, $SD = 0.15$) as significantly lower than the utility of human advice ($M = 3.70$, $SD = 0.12$; $t(46) = -3.55$, $p = .001$). As displayed in Table 4, there were no apparent relationships between the participants' knowledge of expert models and either the perceived utility of human advice or the perceived utility of advice from computerized expert models, with the exception of teachers' knowledge and their numeracy ($r = 0.4$, $p = .01$). We found that teachers with knowledge on expert models clearly perceive a higher utility of expert models than teachers with no knowledge on expert models ($M = 1.6$, $SD = 0.8$ vs. $M = 3.32$, $SD = 1.5$, $t(46) = -3.19$, $p = .001$). This result must be interpreted with caution due to the small number of teachers without knowledge of expert models.

4.2.2. Ideal applications of computerized expert models

Twenty (42.6 %) participants identified three tasks in which a computerized expert model would provide better advice than a human source, 12 (25.5 %) participants indicated two tasks, six (12.8 %) participants indicated one task and nine teachers (19.1 %) did not answer (Total of $K = 90$ tasks). About one-third of the responses were very general (38.9 %; $k = 35$; e.g., "straightforward tasks"). Just under one third were explicitly related to the educational domain (28.9 %; $k = 26$), and about one third (32.1 %; $k = 29$) referred to other domains (20 % finance, 4.4 % traffic, 4.4 % media, and 3.3 % nature). With regard to the educational field, participants indicated that advice from expert models would be better for supporting aspects of their own work (e.g., grading or examination corrections, 53.8 %, $k = 14$), preparing lectures (19.2 %, $k = 5$), promoting the students' possibilities (demonstrating their learning progress, 11.5 %, $k = 3$), or other tasks (15.4 %, $k = 4$; e.g., allocation of student grants).

Twenty-five (53.2 %) teachers indicated three tasks in which human advice would be better, ten (21.3 %) indicated two tasks, three (6.3 %) indicated one task, and nine (19.1 %) did not answer the question (Total $K = 98$ tasks). Many of the responses were very general (43.9 %, $k = 43$; e.g., "emotional decisions" $k = 7$, "new solutions," "when no data are available," "if there is no right or wrong"). About one quarter (23.5 %, $k = 23$) were specific to the educational domain; the rest were in other domains (e.g., 10.2 % personnel and 12.2 % relationship decisions). In the educational domain, the participants mentioned tasks for supporting their daily work (61 %, $k = 14$), when meeting with parents (21.7 %, $k = 5$), or tasks related to school administration (17.4 %, $k = 4$; e.g., allocation of students to a particular class).

4.3. Discussion

The results of Study III confirm the results of Studies I and II; teachers perceive human advice as more useful than advice from expert models. More importantly, the results of Study III suggest that many primary teachers may lack a full understanding of computerized expert models, and people who have greater knowledge of expert models also have higher numeracy competency. This demonstrates that teachers with

Table 4
Teachers' knowledge of expert models, numeracy, and the perceived utility of expert model and human advice.

Knowledge of expert models	Frequency (%)	N	Numeracy		Utility of Advice			
					Expert model		Human	
			M	SD	M	SD	M	SD
None	10.4	6	3.2	1.9	1.6	0.8	3.3	0.5
Partial	62.5	32	3.9	1.4	3.2	1.0	3.8	1.0
Complete	22.9	11	3.9	0.6	3.37	0.7	3.8	0.6
No Answer	4.2	1	–	–	4.0	–	3.0	–
Total	100	47	3.7	1.5	3.0	1.0	3.7	0.9

higher numeracy competency might also be more interested in the topic of expert models.

Our results about teachers' knowledge on expert models—focusing on the data aggregation—demonstrate that the pattern of answer frequency is in the same direction as Logg et al. (2019), and we assume, therefore, that our data are largely comparable and confirm the results by Logg et al. (2019). However, we highlight that in both studies, participants recall their definitions. We, therefore, assume that teachers' knowledge of expert models might be underestimated. Thus, our results might underestimate teachers' knowledge of expert models due to the recall approach used. Future studies should use a recognition approach because expert models might not have been introduced in teachers' daily lives, and their expert model knowledge might not be easily retrievable.

Our results also demonstrate that interestingly, participants mentioned slightly more educational situations in which they perceived that advice from expert models would be better than advice from human sources instead of the other way around. Based on their responses, the participants would be more accepting of using expert models to support their own work tasks (e.g., evaluating students' work) but would prefer advice from humans in more emotional and complex situations. Training programs that first focus on how expert models can streamline teachers' daily work may thus be met with less resistance and help teachers eventually "warm-up" to the idea of using expert models in other sorts of judgment tasks. It is also noteworthy that the results of Study III supported the judgment task used in Study I and in a previous study (Kaufmann & Budescu, 2020), as there was no indication that choosing which student should receive tutoring was a task in which teachers would prefer advice from a human source. We note, however, that the small sample size may limit the generalizability of the results.

5. General discussion

Although computerized expert models can improve human decision-making, people are often reluctant to use them. Moreover, the acceptance of advice from computerized expert models has not been well-studied in the educational domain. This paper used behavioral (Study I) and questionnaires (Studies II and III) to examine teachers' acceptance, perceptions, and knowledge of computerized expert models. Furthermore, to examine whether the next generation was more familiar and comfortable with computerized expert models in Studies I and II, we systematically compared pre- and in-service teachers.

In line with the only other previous study of teachers' acceptance of expert models (Kaufmann & Budescu, 2020), all three studies indicate that teachers generally prefer advice from other people as opposed to a computerized expert model. Even when they could receive advice from both sources, teachers preferred to receive advice from only the human source (Study I). Additionally, neither in-service nor pre-service teachers recognize the value of considering multiple opinions (i.e., the wisdom of the crowd phenomenon; Galton, 1907). The results of Study I confirm that advice seeking depends on task difficulty; specifically, teachers are more likely to seek advice in difficult than in easy tasks (similar to the results in Kaufmann & Budescu, 2020). They also appear to prefer *less* advice in more difficult tasks. In contrast, we found no evidence that pre-service teachers' advice behavior depended on the subject, although they felt more certain about their judgments in math than in language lessons. Future research should investigate how task difficulty and subject affect (teachers') advice behavior in greater detail.

Research on algorithm aversion has been based largely on student (i.e., novice) samples (Germann & Merkle, 2019). Studies have seldom focused on expert judges, and few studies have systematically compared experts and novices (Eining et al., 1997; Logg et al., 2019). We, therefore, added to the literature by systematically comparing experts and novices in education. The results from Study I demonstrate no indication that pre-service teachers had more favorable attitudes toward computerized expert models than in-service teachers, despite (presumably) growing up in a more digitalized environment. It has been demonstrated that attitudes towards technology influence the use of technology (Scherer et al., 2019; Vannatta & Fordham, 2004), and hence, teachers' attitudes are a starting point to change in- and pre-service teachers' algorithm advice behaviors. Therefore, our results suggest that the next generation of teachers is no better adapted to digitalized, data-based educational decision-making than the current generation.

However, our results demonstrate the indication that pre-service teachers had more faith in the advice they received (regardless of source) than in-service teachers. Whereas in-service teachers felt just as certain about their judgments whether or not they received advice, pre-service teachers felt more certain about their judgments when they received advice compared to when they did not. We speculate that in-service teachers may not appreciate advice because it calls their expertise into question (Kleinmuntz, 1990). However, when they requested and received advice, both pre-service and in-service teachers tended to follow it.

Our results provide evidence of how training programs could improve teachers' acceptance of expert models. The results of Study 1 indicate that just being a member of a younger, presumably more digitalized generation does not on its own guarantee higher acceptance of expert models. Instead, the results suggest that *both* younger and older generations of teachers could benefit from interventions that increase their acceptance. In light of the current results, informing teachers about the importance of being objective and the weaknesses in human judgments (based on the results of Study II) and improving knowledge about expert models (based on the results of Study III) might be particularly helpful for increasing their acceptance of expert models by also changing their attitudes. Training programs that first focus on using computerized expert models to support teachers' daily work tasks (e.g., evaluating students' work) may encounter the least resistance. Subsequently, teachers can be trained to use expert models in other tasks (e.g., administration tasks).

Across the three studies, we found no indication that personal characteristics such as openness, neuroticism, or numeracy were related to perceptions or behaviors vis-à-vis expert models. We highlight that openness plays an important role in a broader scope of technology adoption. Research on technology adoption demonstrates that openness to changes is positively related to the use of technology within classrooms (Vannatta & Fordham, 2004). However, our studies demonstrate that openness does not impact the use of algorithm advice; however, future research is required to verify these results. Further studies should also verify if the suggested training on statistical concepts, such as error and uncertainty, could improve algorithm literacy (Burton et al., 2020) and test the relevance of numeracy for algorithm literacy training.

In Study III, the teachers indicated that they would prefer human advice during administrative tasks and tasks involving parents, but expert model advice for supporting their daily work. Studies in the medical domain have demonstrated that patients prefer physicians not to receive advice from a expert model (Promberger & Baron, 2006) and that physicians who do not use diagnostic aids are also perceived as more competent and professional than physicians who use such aids (Arkes et al., 2007). Future research should explore whether the same pattern of findings also applies to the educational domain, and hence explain why teachers prefer not to use advice from computerized expert models in tasks that involve parents (e.g., perhaps because parents do not want them to and because parents would see them as less competent). Future studies should also examine how parents and other stakeholders feel about teachers' use of expert models. Potentially, improving the algorithm acceptance of other stakeholders may also improve teachers' acceptance.

Future research should address some of the limitations of the current study. For example, in Study I, we used a subjective and longer-term judgment task (Castelo et al., 2019; Logg et al., 2019), and our behavioral results may thus not extend to more objective judgment tasks or shorter-term decisions (e.g., preparing a lecture; Himmelstein, Atanasov, & Budescu, 2021). Our small convenience sample also limits our results; future studies with larger and more representative samples of teachers are needed to confirm our results. Finally, in line with Yeomans et al. (2019), we recommend research on other factors that may influence teachers' acceptance of algorithms, such as the control of the algorithm (Dietvorst et al., 2016) and personalization (Longoni et al., 2019). For example, Dietvorst et al. (2014) found that observing an algorithm make a mistake causes people to distrust algorithms even more, even after observing humans make the same mistake. Research investigating whether teachers respond in the same way as other experts in other studies would be useful for designing appropriate interventions specifically for the educational domain. Currently, it is unclear if experts from different domains behave similarly because of the different types of tasks (subjective vs. objective), and we argue that there are differences in how experts judge the tasks due to the judged tasks (Hammond, 1996).

To summarize, our research demonstrates that, like in-service teachers, pre-service teachers are algorithm averse, meaning that they prefer human advice for their judgment and decision-making processes. Therefore, in a broader scope, our study has to be seen as a further piece of research on the acceptance of artificial intelligence advice within the educational field. Furthermore, we highlight that several factors may interplay with teachers' reliance on computerized expert models (e.g., attitudes and type of tasks), and these need further elaboration. Such research is necessary because when teachers can make more accurate assessments with the support of computerized expert models as artificial intelligence, they are better able to create learning environments that best foster students' academic development and may also decrease inequality among students. Therefore, the insights from our research contribute to improving students' potential to thrive and reducing inequality by more accurate teachers' judgments because of the potential of artificial intelligence.

Conflict of interest and authorship conformation form

Please check the following as appropriate:

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
o This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
o The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript
o The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript:

Appendices.

Student A: Mary	Characteristics	Student B: Tom
female	**sex**	male
12 years old	**age**	13 years old
sixth grade (Middle School)	**class level**	seventh grade (Middle School)
good behavior	**working habits**	cooperative
none	**absence**	minimal (2 days per semester)
none	**problems**	none
member of the tennis team	**extracurricular activities**	member of an art and drama group
received no support	**previous student needs**	received no support
	grades	
F	**math**	F
D	**science**	D
C	**english**	C
D	**social studies**	D

Fig. A1. An example judgment task from Study I. This is a difficult task because the students have the same math grade

Table A1
Frequency of pre-service and in-service teachers' judgments in the four judgment tasks when receiving no advice, advice from an expert model, advice from a school counselor, or both (%; k in parentheses)

		Profil I: Easy Task: Math			
		Pre-service (63)		In-service (99)	
Chosen Student		A*	B	A*	B

(continued on next page)

Table A1 (*continued*)

		Profil I: Easy Task: Math			
		Pre-service (63)		In-service (99)	
Chosen Student		A*	B	A*	B
Without Advice		76.2 (48)	1.6 (1)	62.6 (62)	20.2(20)
With Advice	Counselor	7.9 (5)	0 (0)	5.1 (5)	0 (0)
22.2 (14), Pre-service	Expert model	1.6 (1)	0 (0)	0 (0)	1.0 (1)
17.2 (17), In-service	Both	12.7 (8)	0 (0)	10.1 (10)	1.0 (1)
		Profil II: Easy Task: Language/Math			
		Pre-service (63)		In-service (99)	
Chosen Student		A*	B	A*	B
Without Advice		76.2 (48)	4.8 (3)	61.6 (61)	25.3 (25)
With Advice	Counselor	1.6 (1)	0 (0)	5.1 (5)	0 (0)
19.0 (12), Pre-service	Expert model	3.2 (2)	0 (0)	2.0 (2)	0 (0)
13.1 (13), In-service	Both	14.3 (9)	0 (0)	5.1 (5)	1.0 (1)
		Profil III: Hard Task: Math			
		Pre-service (62)		In-service (99)	
Chosen Student		A*	B	A*	B
Without Advice		11.1 (7)	19.0 (12)	33.3 (33)	13.1 (13)
With Advice	Counselor	27 (17)	3.2 (2)	25.3 (25)	2.0 (2)
69.8 (44), Pre-service	Expert model	6.3 (4)	0 (0)	3.0 (3)	1.0 (1)
22.2 (14), In-service	Both	31.7 (20)	1.6 (0)	21.2 (21)	1.0 (1)
		Profil IV: Hard Task: Language/Math			
		Pre-service (62)		In-service (99)	
Chosen Student		A	B*	A*	B
Without Advice		9.5 (6)	14.3 (9)	17.2 (17)	29.3 (29)
With Advice	Counselor	4.8 (3)	30.2 (19)	27.3 (27)	1.0 (1)
74.6 (74), Pre-service	Expert model	1.6 (1)	4.8 (3)	6.1 (6)	0 (0)
53.5 (53), In-service	Both	1.6 (1)	30.2 (19)	18.2 (18)	1.0 (1)

Note. The student recommended by either the expert model or the school counselor is marked by a *.

Table A2
In- and pre-service teachers' judgment certainty (M, SD, N) across tasks as a function of task difficulty and advice

				No advice	Demanding advice	Type of advice (M (SD))		
Pre-service						One advice source		Two advice source
	Task difficulty	Profile(s)	N	M (SD)	M (SD)	School counselor	Expert model	
	Easy	I	63	6.06 (2.12), N = 49	7.50 (1.91), N = 14	8.00 (1.00), N = 5	9.00 (0.00), N = 1	7.00 (2.32), N = 8
		II	63	5.62 (2.17), N = 51	6.66 (2.06), N = 12	5.00 (0.00), N = 1	5.00 (2.82), N = 2	7.22 (1.85), N = 9
		Total	126	5.84 (2.14), N = 100	7.11 (1.98), N = 26	7.5 (1.51), N = 6	6.33 (3.05), N = 3	7.11 (2.02), N = 17
	Hard	III	63	5.58 (1.64), N = 19	8.00 (1.12), N = 44	7.84 (1.39), N = 19	8.0 (1.15), N = 4	8.14, (0.85), N = 21
		IV	63	4.62 (2.06), N = 16	7.63 (1.55), N = 47	7.43 (1.70), N = 23	7.25 (2.87), N = 4	7.95 (1.35), N = 20
		Total	126	5.14 (1.88), N = 35	7.81 (1.43), N = 91	7.62 (1.56), N = 42	7.62 (2.06), N = 8	8.05 (1.12), N = 41
	Overall		252	5.66 (1.92), N = 135	7.65 (1.59), N = 117	7.6 (1.50), N = 48	7.27 (2.28), N = 11	7.77 (1.49), N = 58
In-servicerowhead								
	Easy	I	99	6.32 (1.88), N = 82	5.71 (3.15), N = 17	7.40 (2.07), N = 5	9.00 (0.00), N = 1	4.63 (3.23), N = 11
		II	99	6.43 (1.79), N = 86	5.46 (3.52), N = 13	5.20 (3.63), N = 5	5.00 (5.65), N = 2	5.83 (3.54), N = 6
		Total	198	6.37 (1.83), N = 168	5.60 (3.26), N = 30	6.30 (3.02), N = 10	6.33 (4.61), N = 3	5.05 (3.28), N = 17
	Hard	III	99	6.11 (2.17), N = 46	5.60 (3.18), N = 53	5.51 (3.14), N = 27	8.00 (1.15), N = 4	5.81 (3.52), N = 22
		IV	99	5.69 (1.96), N = 46	5.41 (3.07), N = 53	5.75 (2.95), N = 28	3.50 (2.50), N = 6	5.52 (3.33), N = 19
		Total	198	5.90 (2.07), N = 92	5.50 (3.12), N = 106	5.63 (3.02), N = 55	4.10 (2.23), N = 10	5.68 (3.4), N = 41
	Overall		396	6.20 (1.92), N = 260	5.52 (3.14), N = 136	5.73 (3.00), N = 65	4.61 (2.87), N = 13	5.50 (3.35), N = 58
Overall			648	6.02 (2.00), N = 395	6.51 (2.75), N = 253	6.53 (2.65), N = 113	5.83 (2.89), N = 24	6.63 (2.82), N = 116

Table A3
Pre-service and in-service teachers' judgment certainty after receiving advice from different sources

Advice source	Teachers	M	SD	n	*t*-test
School counselor	Pre-service	5.3	1.9	44	$t(75) = .93, p = .35$
	In-service	5.7	1.8	33	
Expert model	Pre-service	5.4	2.59	9	$t(19) = -1.38, p = .18$
	In-service	4.1	1.33	12	
Both sources	Pre-service	6.4	1.9	32	$t(63) = -1.32, p = .19$
	In-service	5.6	2.59	33	

Table A4
Distribution of in- and pre-service teachers who followed the advice given as a function of task difficulty and type of advice.

In-service	Task difficulty	Profile(s)	N	No advice	Demanding advice	Type of advice (N)			Overall (%)
						One advice source		Two advice source	
				%	%	School counselour	Expert model		
	Easy	I	99	75.6	17.2 (17)	5	(1)	10(1)	88.2
		II	99	61.6	13.1 (13)	5	2	5(1)	92.3
		Total	198	73.2	15.2 (30)	10	2(1)	15(2)	90
	Hard	III	99	71.7	53.5 (53)	25(2)	3(1)	21(1)	92.5
		IV	99	37	53.5 (53)	27(1)	6	18(1)	96.2
		Total	198	54.3	53.5 (106)	53(3)	9(1)	39(2)	94.3
	Overall		396	66.5	34.3(136)	62(3)	11(2)	54(4)	95.6
Pre-service	Easy	I	63	98	22.2 (14)	5	1	8	100
		II	63	94.1	19 (12)	1	2	9	100
		Total	126	96	20.6 (26)	6	3	17	100
	Hard	III	63	36.8	69.8 (44)	17(2)	4	20(0)	95.34
		IV	63	60	76.2 (48)	19(3)	3(1)	19(1)	89.1
		Total	126	47.1	73(92)	36(5)	7(1)	39(1)	92.1
	Overall		252	83.6	46.8(118)	42(5)	10(1)	56(1)	93.9
Overall			648	72.3	39.2(254)	104(8)	21(3)	110(5)	93.6

Note. Numbers in parentheses are the number of teachers who did not follow the advice given for one task.

Table A5
Results of the regression analyses of teaching experience, gender, and age as predictors of the perceived difference between human over expert model advice across six characteristics

	Characteristic					
Independent variables	Objectivity	Reliability	Accuracy	Trustworthy	Transparency	Beneficial for students
Experience	0.15	0.06	0.004	0.00	0.18	−0.18*
Gender	−0.02	0.13	0.006	0.05	0.00	−0.03
Age	−0.16*	0.03	−0.04	0.04	−0.14*	−0.05
R^2 (corrected)	0.06***	0.00	0.00	0.00	0.15*	0.2*
F	10.96	0.36	0.26	0.58	3.11	3.26
N	425	428	427	426	426	425

Note. Displayed are the standardized beta values from the saturated model. Higher values indicate more of a perceived difference between advice from human sources and expert models. *$p < .05$. ***$p < .0$.

References

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

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle kompetenz von Lehrkräften [key word: Professional competence of teachers]. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2.

Beck, E., Baer, M., Guldimann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M., & Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens [Adaptive teaching skills. Analysis and structure, changeability and effect of action-controlling teacher knowledge.* Waxmann.

Budescu, D. V., & Chen, E. (2015). Identifying expertise to extract the wisdom of crowds. *Management Science, 61*, 267–280. https://doi.org/10.1287/mnsc.2014.1909.

Bundesamt für Statistic, Swiss Statistical Report. (2016). Personal von Bildungsinstitutionen (People of education institutions). Available at: https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/personal-bildungsinstitutionen.assetdetail.6446987.html.

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

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(6), 809–825. https://doi.org/10.1177/0022243719851788.

Chacon, A, Kausel, E, E, & Reyes, T (2021). *A longitudinal diary approach for understanding algorithm use. Working paper at Pontificia Universidad Católica de Chile.* Submitted for publication.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years. Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464–465. https://doi.org/10.1126/science.1134513.

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science, 243* (4899), 1668–1674. https://doi.org/10.1126/science.2648573.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). 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. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643.

Einhorn, H. J. (1974). Cue definition and residual judgment. *Organizational Behavior & Human Performance, 12*(1), 30–49. https://doi.org/10.1016/0030-5073(74)90035-X.

Eining, M. M., Jones, D. R., & Loebbecke, J. K. (1997). Reliance on decision aids: An examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice & Theory, 16*(2), 1–19.

Galton, F. (1907). Vox populi. *Nature, 75*, 450–451. https://doi.org/10.1038/075450a0.

Germann, M., & Merkle, C. (2019). *Algorithm aversion in financial investing.* Germany: Working paper University Mannheim.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making, 20*, 21–35. https://doi.org/10.1002/bdm.539.

Goldberg, L. R. (1976). Man versus model of man: Just how conflicting is that evidence? *Organizational Behavior & Human Performance, 16*(1), 13–22. https://doi.org/10.1016/0030-5073(76)90003-9.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19.

Hammond, K, R (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* Oxford, UK: University Press.

Himmelstein, M, Atanasov, P, & Budescu, D, V (2021). Forecasting forecaster accuracy: Contributions of past performance and individual differences. *Judgment and Decision Making*, 323–362.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313. https://doi.org/10.2307/1170184.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

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

Karing, C., & Seidel, T. (2017). Förderung diagnostischer Kompetenz. In A. Südkamp, A.-K. Praetorius, & Hrsg (Eds.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen [Diagnostic competence of teachers: Theoretical and methodological developments]* (pp. 201–202). Waxmann.

Kaufmann, E. (2020). How accurately do teachers' judge students? Re-analysis of Hoge and Coladarci (1989) meta-analysis. *Contemporary Educational Psychology, 63*, Article 101902. https://doi.org/10.1016/j.cedpsych.2020.101902.
Kaufmann, E., & Budescu, D. V. (2020). Do teachers consider advice? On the acceptance of computerized expert models. *Journal of Educational Measurement, 57*(2), 311–342. https://doi.org/10.1111/jedm.12251.
Kaufmann, E., & Wittmann, W. W. (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS ONE, 11*(6), Article e0157914. https://doi.org/10.1371/journal.pone.0157914c.
Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107*(3), 296–310. https://doi.org/10.1037/0033-2909.107.3.296.
Kolovou, D., Naumann, A., Hochweber, J., & Praetorius, A. K. (2021). Content-specificity of teachers' judgment accuracy regarding students' academic achievement. *Teaching and Teacher Education, 100*, 103298.
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, 9*(6), 1060–1072. https://doi.org/10.1037/a0034156.
Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.
Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013.
Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064.
Mandinach, E. B., & Schildkamp, K. (in press). The complexity of data-based decision making: An introduction to the special issue. Studies In Educational Evaluation. https://doi.org/10.1016/j.stueduc.2020.100906.
Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. RAND Corporation.
Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis. MN: University of Minnesota Press.
Önkal, D., Brehm, N., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making, 22*, 390–409. https://doi.org/10.1002/bdm.637.
Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19*(5), 455–468. https://doi.org/10.1002/bdm.542.
Rammstedt, B., & Danner, D. (2016). Die Facettenstruktur des Big Five Inventory (BFI) [The facet structure of the Big Five Inventory (BFI): Validation for the German adaptation of the BFI]. *Diagnostica, 63*(1), 70–84. https://doi.org/10.1026/0012-1924/a000161.
Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (tam): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009.
Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research, 61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716.
Schildkamp, K., Lai, M. K., & Earl, L. (2013). *Data-based decision making in education: Challenges and opportunities*. Springer.
Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making, 19*(1), 43–60. https://doi.org/10.1002/bdm.514.
Schweizerische Konferenz der kantonalen Erziehungsdirektoren. (2018). Digitalisierungsstrategie: Strategie der EDK vom 21. In *Juni 2018 für den Umgang mit Wandel durch Digitalisierung im Bildungswesen*. [Digitization strategy: EDK's strategy of 21 June 2018 for managing change through digitization in education]. Available at: https://edudoc.ch/record/131564/files/pb_digi-strategie_d.pdf.
Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany]. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften* [Standards of teaching education: Educational science, Germany], 16, 2004 *Beschluss der Kultusministerkonferenz vom December*. Bonn.
Shavelson, R. (1973). What is the basic teaching skill? *Journal of Teacher Education, 24*(2), 144–151. https://doi.org/10.1177/002248717302400213.
Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research, 79*(2), 625–649. https://doi.org/10.3102/0034654308325896.
Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–763. https://doi.org/10.1037/a0027627.
Südkamp, A., Praetorius, A.-K., & Hrsg. (2017). Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen. In *[Diagnostic competence of teachers: Theoretical and methodological developments]*. Waxmann.
Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review, 32*. https://doi.org/10.1016/j.edurev.2020.100374. Article 100374.
Vannatta, R. A., & Fordham, N. (2004). Teacher dispositions as predictors of classroom technology use. *Journal of Research on Technology in Education, 36*(3), 253–271. https://doi.org/10.1080/15391523.2004.10782415.
Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A rasch analysis approach. *Journal of Behavioral Decision Making, 26*(2), 198–212. https://doi.org/10.1002/bdm.1751.
Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*, 403–414. https://doi.org/10.1002/bdm.2118.