

# An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology

TANYA NAZARETSKY, Weizmann Institute of Science, Israel

MUTLU CUKUROVA, University College London, UK

GIORA ALEXANDRON, Weizmann Institute of Science, Israel

Evidence from various domains underlines the key role that human factors, and especially, trust, play in the adoption of technology by practitioners. In the case of Artificial Intelligence (AI) driven learning analytics tools, the issue is even more complex due to practitioners' AI-specific misconceptions, myths, and fears (i.e., mass unemployment and ethical concerns). In recent years, artificial intelligence has been introduced increasingly into K-12 education. However, little research has been conducted on the trust and attitudes of K-12 teachers regarding the use and adoption of AI-based Educational Technology (EdTech).

The present study introduces a new instrument to measure teachers' trust in AI-based EdTech, provides evidence of its internal structure validity, and uses it to portray secondary-level school teachers' attitudes toward AI. First, we explain the instrument items creation process based on our preliminary research and review of existing tools in other domains. Second, using Exploratory Factor Analysis we analyze the results from 132 teachers' input. The results reveal eight factors influencing teachers' trust in adopting AI-based EdTech: *Perceived Benefits of AI-based EdTech*, *AI-based EdTech's Lack of Human Characteristics*, *AI-based EdTech's Perceived Lack of Transparency*, *Anxieties Related to Using AI-based EdTech*, *Self-efficacy in Using AI-based EdTech*, *Required Shift in Pedagogy to Adopt AI-based EdTech*, *Preferred Means to Increase Trust in AI-based EdTech*, and *AI-based EdTech vs Human Advice/Recommendation*. Finally, we use the instrument to discuss 132 high-school Biology teachers' responses to the survey items and to what extent they align with the findings from the literature in relevant domains.

The contribution of this research is twofold. First, it introduces a reliable instrument to investigate the role of teachers' trust in AI-based EdTech and the factors influencing it. Second, the findings from the teachers' survey can guide creators of teacher professional development courses and policymakers on improving teachers' trust in, and in turn their willingness to adopt, AI-based EdTech in K-12 education.

Additional Key Words and Phrases: AI, Human Factors, Trust, Teachers, K-12 Education, Blended Learning

## ACM Reference Format:

Tanya Nazaretsky, Mutlu Cukurova, and Giora Alexandron. 2021. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. In .... ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Over the last 25 years significant progress has been made in researching and designing Artificial Intelligence(AI)-based educational technologies. Among other things, the research indicated that AI-based systems for adaptive learning can be as effective as one-to-one human tutoring in STEM topics [51], as well as outperform traditional classroom instruction,

PRE-PRINT VERSION

Accepted to The 12th International Learning Analytics and Knowledge Conference (LAK'22).

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

Manuscript submitted to ACM

laboratory or homework assignments, and reading computerized materials [16, 45]. However, the wide-adoption of AI-based tools in real-world educational settings seems to lag behind [42].

Perhaps it is, at least to a certain extent, due to the ‘AI nature’ of the tools and how AI is perceived in media. Among the positive views and attitudes [36], the media and public rhetoric discourse often raise concerns regarding the potentially harmful impact of AI technologies, including arguments on the observed and expected impact of AI on the future of the workforce, and related fears around mass unemployment [5, 18]. These negative connotations have the potential to skew policymakers’, educators’, and learners’ perceptions, and as a result, lead them to avoid adopting AI-based learning analytics solutions. Indeed, there are clear indications of active resistance to using AI-based educational technology in schools<sup>1 2</sup>. Correspondingly in a recent study, Cukurova et al. [13] demonstrated that when educational research evidence is framed within AI research, it is considered as less credible in comparison to when it is framed instead within neuroscience or educational psychology. This effect was still evident when the subjects’ familiarity with the framing discipline was controlled for. Furthermore, the authors showed that their participants perceived AI to be: less helpful in assisting us to understand how children learn, lacking in adherence to scientific methods, and to be less prestigious compared to neuroscience and educational psychology [13].

Another possible reason for the delay in AI-based technology adoption in educational settings is due to the tendency of research in the AIED area to mostly focus on pedagogical and technological aspects of educational solutions, rather than taking a holistic approach [26] aimed to combine human and technology perspectives in human-machine interactions process. As a result, much less attention was paid to the significance of human agency in the adoption of AI in educational settings [12]. Similar concerns and calls for action were raised by researchers from the learning analytics community and a recent special issue edited by Buckingham-Shum, Ferguson, & Martinez-Maldonado on human-centered learning analytics was published in the *Journal of Learning Analytics* [6]. Nonetheless, our understanding of complex human factors that influence teachers’ trust in the adoption of AI-based Educational Technology and how to measure them is scant at best.

## 2 LITERATURE REVIEW

In contrast to learning and education research, the impact of practitioners’ perceptions of AI on the adoption of AI-based tools in their practice is an active research area in applied sciences, finance, and medicine. There are also previous studies that looked at people’s perceptions of AI in general as well as its range of future outcomes for the society [15]. In this section, we will review this literature, as well as any existing work from educational research, to draw parallels that inspired the research instrument developed in this paper.

To start with, Fast and Horvitz [17] revealed that from the year 2009 the discussion on AI has sharply increased and is more optimistic than pessimistic. Gaines-Ross [19] investigated the lay people’s perception of AI and found out that those individuals who do not have a professional background in technology generally have positive perceptions of AI. Manikonda and Kambhampati [32] collected and analyzed over two million AI-related tweets posted by over 40,000 people and showed that (i) the sentiments expressed in the AI discourse are more positive than an average Twitter discourse; (ii) lay public tend to be more positive about AI than expert tweeters; and (iii) women tend to be more positive about AI impacts than men. In addition, Sharan and Romano [44] examined the relation between personality and Locus of Control characteristics on trust in humans and AI; Gaube et al. [20] compared experts and non-experts’ trust in clinical advice from human experts and AI; and Kizilcec [30] demonstrated that transparency of AI has an impact on

<sup>1</sup><https://www.nytimes.com/2019/04/21/technology/silicon-valley-kansas-schools.html>

<sup>2</sup><https://www.fastcompany.com/90266263/brooklyn-students-walk-out-of-school-over-zuckerberg-backed-learning-system>

people's trust in AI decisions and also examined how much information should be provided to users. Gaube et al. [20] and Tschandl et al. [48] examined AI support in clinical practice and found that practitioners with less experience tend to over-rely on AI support, and may be more susceptible to inaccurate recommendations, while experts tend to trust less on AI advice, especially when they are confident. In addition, Gaube et al. [20] found that although experts show algorithmic aversion (they rated the same quality advice as lower quality when it appeared to come from an AI system instead of human), they may still trust AI advice if they are not confident in their decisions. Similarly, Tschandl et al. [48] found that confidence in decisions is a significant confound in whether people trust in AI to change their decisions or not. More specifically, if experts' initial diagnosis was not in agreement with AI, the experts changed their initial diagnosis less often if they were confident (29.8%) and more often if they were not confident (53.9%). The novices tended to accept AI-based support that contradicted their initial diagnosis even if they were confident. Sharan and Romano [44] examined the effects of personality and locus of control (LOC) on how people trust humans and artificial intelligence. The authors found that, as LOC levels increased, people were less likely to follow suggestions from both humans or AI and people chose the AI suggestions more than those from humans, and self-reported that they believed such recommendations more. This result contrasts with Gaube et al. [20], yet expertise differences between the sources of advice may explain this. Sharan and Romano [44] provided human advice through previous participants (decision-making card game) who lack any expertise in the proposed area, so participants might think AI possesses higher expertise than "previous participants". In Gaube et al. [20]'s research, on the other hand, the human advice came from the experts with years of training and experience (radiologists), and their expertise was valued more by participants. Therefore, when comparing practitioners' trust in advice from humans and artificial intelligence, it is important to take into account the level of expertise pictured for the practitioners.

In another context, Jakesch et al. [28]'s study of AI-mediated interpersonal communication examined the impact of AI-generated vs human-generated media content on people's trust. Researchers found that people give AI-generated profiles in media significantly lower trust ratings compared to human-written profiles when those profiles are mixed with AI. Interestingly, this effect was not found when all profiles are written by either AI or humans alone. This may be due to the inconsistency of content sources, which incurs people's over-elaboration and doubt of the credibility of the profile texts. This study, however, did not identify how to mitigate this effect, it only demonstrates its existence. This gap was filled to a certain extent in Kizilcec's study [30]. Using an online peer assessment platform, Kizilcec [30] examined the effect of different levels of transparency on trust. He demonstrated that procedural transparency on how content was created (i.e how much and which parts of the profile are written by AI, or the mechanism of the AI algorithm used to create content, etc. are provided to users) may lead to greater trust among users.

To the best of our knowledge, the issue of educators' trust in AI was not studied within the learning analytics community, and there is no previously established tool for measuring educators' perceptions of AI, as well as the impact of these perceptions on their trust and adoption of AI tools into their teaching practices. The theory of academic resistance [41] treats the general issue of technology resistance in educational settings, the resistance to organizational change theory [37] proposes a multidimensional view on resistance to change in general and suggests that individuals operate in three dimensions (cognitive, affective, and behavioral) and simultaneously may be ambivalent in all three dimensions, and technology acceptance model (TAM) [14] explains the perceived usefulness of technology and its usage intentions in terms of social influence and cognitive instrumental processes. However, none of the above theories focuses on the 'AI nature' of AI-based tools, which, based on evidence from other disciplines [8, 23] and our preliminary research [35], plays a significant and even critical role in forming practitioners' perceptions and attitudes towards adoption AI-based EdTech.

Therefore, this study introduces a new instrument for measuring different aspects of the trust of teachers in AI-based EdTech, provides evidence of its internal structure validity, and use it to discuss 132 high-school Biology teachers' responses to the instrument items and to what extent they align with the findings from the literature in relevant domains reviewed above.

The rest of the paper is organized as follows. In Section 3, we describe the five-stage procedure of the instrument development including its creation and establishment of the instrument internal validity. In Section 4, we use the instrument to analyze 132 Biology teachers' responses and portray teachers' attitude towards the adoption of AI-based tools in their classroom. In Section 5 we discuss our findings, limitations, and directions for future research. Finally, Section 6 presents the summary and conclusions.

### 3 INSTRUMENT DEVELOPMENT

#### 3.1 Overview of development procedure

The development process included the following steps:

- (1) Top-down design of the categories, items, and instrument, based on the literature review of practitioners' trust in AI and bottom-up design of items based on discussions with science teachers and experts.
- (2) Pilot with a small group of teachers, for polishing the initial design.
- (3) Exploratory Factor Analysis for bottom-up analysis of emerged factors.
- (4) Reliability analysis of the emerged factors to verify their internal consistency and factor labeling.
- (5) Finalization of the instrument.

This is a similar procedure to the one taken by [7] for developing an instrument for measuring undergraduate and graduate research trainee development in a similar context – science education. The flow can be seen in Figure 1.

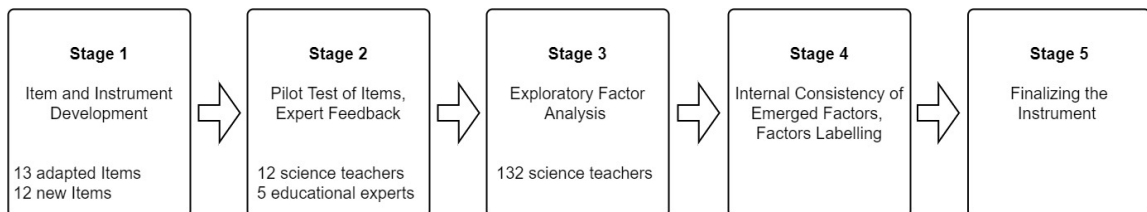


Fig. 1. The development process of the instrument.

#### 3.2 Stage 1: Items/Instrument development

The categories of the instrument, and the items within each category, were developed in a top-down fashion. The initial draft was based on drafting questions relevant to factors studied in existing relevant literature from a wide range of disciplines. These dimensions included questions related to participants' past experience, their knowledge and understanding of AI, perceived credibility associated with AI, perceived relevance of AI to their professional field, their level of expected agency in decision-making processes, expected details in diagnostic justification, and transparency of decision-making processes, type of decisions made in professional life and their perceived importance, perceived risks associated with wrong decisions and accountability, perceived ability of AI systems. In addition, the first draft had

questions related to participants' sociological and demographic information (i.e age, gender, education level) that might influence their trust in AI-based tools.

This initial version was refined and extended by the research team after conducting interviews and group discussions with science teachers and was then presented to a group of five science education and educational technology developers, and revised according to their feedback. Below this process is described in more detail.

**3.2.1 Population.** The research population for this stage consisted of 10 Physics teachers who participated in a year-long professional development (PD) program focusing on training future mentors of in-service teacher communities. Each of the participants had at least two years of experience working with educational technology and applying blended-learning instruction using a digital learning platform for personalized instruction named PeTeL (**P**ersonalized **T**eaching and **L**earning), and at least five years of experience as a science teacher. As our main goal was to focus on beliefs and attitudes towards the adoption of AI-based EdTech tools, and not towards educational technology in general, we chose participants who are frequent users of educational technology and hold positive attitudes towards it but have no prior experience with AI-based tools.

**3.2.2 Qualitative Procedure.** During the PD program, the participants were presented with AI-based educational technology that was being piloted for PeTeL. To make the scenarios presented to teachers concrete, all the interactions were around specific artifacts presented through authentic use-cases. The presented tool was aimed to automatically cluster students into groups of students that exhibit similar response patterns on a multi-dimensional online assessment instrument.

Data were collected via face-to-face interviews and group discussions during the 2019-20 academic year. The collected data were later transcribed and analyzed by the first author. The themes that emerged were discussed with the other authors based on the data and triangulated between the interviews and the group discussions. As a result of this qualitative phase [35], a dozen items were added to the instrument (Q1-Q6, Q16-Q20, and Q23; See the instrument in Appendix A.1).

**3.2.3 Items refinement based on experts feedback.** In order to evaluate the suitability of the instrument for science teachers, it was presented individually to 5 science education and educational technology researchers. The experts were asked to answer the instrument and provide feedback on its comprehensibility and other factors that they find relevant for running it with science teachers. As a result of this stage, the following changes were introduced: (i) The Likert scale of several items was changed to ensure consistency across the instrument; (ii) the wording was clarified, to avoid technical or foreign jargon that teachers may not be familiar with; (iii) examples were added to some of the items to clarify their meaning (e.g., Q1, Q4-Q5, Appendix A.1.1).

The instrument items were written originally in English and were translated to Hebrew by the first author. The translation was validated by two additional researchers, who are experts in AI-based educational technology and science teaching.

### **3.3 Stage 2: A pilot run of the instrument**

**3.3.1 Population.** The experiment was conducted with a different group of 12 high-school science teachers (except one teacher who took part in the Stage 1 experiment too) who participate in a year-long professional development program focusing on co-designing learning analytic features for PeTeL. Each of them had at least 2 years of experience applying blended-learning instruction with PeTeL, and 5 years of experience as a science teacher.

3.3.2 *Procedure.* The experiment took an hour and was conducted remotely due to Covid-19 constraints. The participants were first introduced to the instrument and its purpose and were then requested to fill it individually (during a synchronous Zoom meeting) and provide any feedback that they find relevant.

3.3.3 *Items refinements based on teachers feedback.* As a result of participants' comments the following changes were introduced: (i) the advantages of using AI in education, and on the contrary, reasons for *not* using AI in education, were separated into two separate sections; (ii) the wording of some of the items was modified, and the space for responses to open-ended questions was increased.

The resulting instrument consisted of 25 five-level Likert items (thirteen items adapted from the initially drafted items from the literature review, and twelve new items) to measure different aspects of attitudes towards using AI-based EdTech, and nine items for collecting socio-demographic and qualitative information (see Appendix A.1).

### 3.4 Stage 3: Exploratory Factor Analysis (EFA)

3.4.1 *Population.* The instrument was administered to 132 high-school science teachers (87% females and 13% males; age:  $M = 44$ ,  $SD = 11$ ). The majority of the participants had at least a master's degree in the subject matter or science education (88%), almost all of the rest had a bachelor's degree, while two teachers reported that they have a teaching certificate only. Most teachers (56%) indicated that they had at least six years of experience in using educational technologies in their classroom. All the teachers were participants of a year-long professional development program (during the 2020-21 academic year; online due to the COVID-19 pandemic outbreak).

3.4.2 *Procedure.* Data were collected via an online survey. The participants were presented with a short definition of Artificial Intelligence and were instructed to answer the questions and rate their level of agreement based on their own understandings, opinions, beliefs, and potential attitudes.

3.4.3 *Exploratory Factor Analysis.* To explore the factorial structure of the collected data all 25 Likert items were subjected to an Exploratory Factor Analysis using principal axis (PA) factoring method [46] with oblique rotation (oblimin) [22]. The analysis was done using *fa* component of *psych* R package<sup>3</sup> [38]. There were no missing values in the data as all the corresponding fields were set as required in the online review. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) was used to verify the sampling ( $n=132$ ) adequacy for the analysis ( $KMO = 0.909 > 0.9$ ) [29]. Bartlett's test of sphericity ( $\chi^2 = 1125.64$ ,  $p < 0.001$ ) indicated that correlation structure is adequate for a meaningful factor analysis [50]. The data in the sample violated the normality assumptions. Shapiro Wilk test was statistically significant ( $p < 0.0001$ ) [43], indicating a univariate normality deviation. Multivariate normality was estimated by Mardia's multivariate kurtosis and skewness tests [33]. The null hypothesis was rejected (with  $p$ -value  $< 0.0001$ ), suggesting a violation of multivariate normality of the values in the sample. Due to the violation of normality, factor analysis was performed with a PA estimator, which is suitable for handling ordinal (Likert) scales and non-normal distribution [38].

The Horn's Parallel Analysis [25, 47] suggested a structure of five, six, or seven factors, explaining 42%, 45%, and 48% of the variance, respectively. To estimate the goodness of fit of the proposed models we used multiple fit measures simultaneously [4]. All three models were evaluated by the standards proposed by Hu and Bentler [27], and Bentler and Bonett [3] that are in par with the recommendations of the American Educational Research Association [2]: the root

<sup>3</sup><https://cran.r-project.org/web/packages/psych/index.html>

mean square error of approximations RMSEA ( $\leq 0.06$ , 90% CI  $\leq 0.06$ ), the root mean square of the residuals SRMR ( $\leq 0.08$ ), Tucker Lewis Index of factoring reliability TLI ( $\geq 0.9$ ) and the  $\chi^2/df$  ratio less than 3 [31].

Table 1 contains the fit statistics for all the three models being tested. As can be seen in the table, the 5- and 6-factor models did not meet some of the quality criteria defined above and were thus discarded. The 7-factor model met all the quality criteria: RMSEA = 0.037 with 90% CI = 0.058, TLI = 0.928, SRMR = 0.04 and  $\chi^2/df = 1.29$  (in Table 1 corresponding values are highlighted in bold). Thus, the 7-factor model was chosen as the best model for explaining the response data.

Table 1. Goodness of fit statistics for the three EFA models.

Model	$\chi^2$	df	$\chi^2/df$	TLI	RMSEA	RMSEA 90% CI	SRMR	Explained variance
5-factor model	252*	185	1.36	0.861	0.052	0.035 - 0.069	0.05	42%
6-factor model	212*	165	1.29	0.889	0.046	0.026 - 0.064	0.04	45%
<b>7-factor model</b>	<b>173**</b>	<b>146</b>	<b>1.19</b>	<b>0.928</b>	<b>0.037</b>	<b>0.000 - 0.058</b>	<b>0.04</b>	<b>48%</b>

\* $p < 0.001$ , \*\* $p = 0.059$

Next, for deciding on the mapping of items to factors, we used a threshold of 0.3 [10, 46]. The resulting factor analysis, including factor loadings, are presented in Figure 2 (the full table of factor loadings and proportion of variance explained is presented in Appendix Table 3). We note two exceptions: items with double loading, and items with no loading (not explained by any of the factors). Concerning double loading, items Q2, Q7, Q12, and Q13, had two factors with loading  $> 0.3$ . In that case, we took the more significant factor. On the contrary, items Q4, Q8, and Q20 did not have any factor with loading  $> 0.30$ . The usual approach for such items is to examine them qualitatively to decide if they should be retained, removed, or whether belong to an additional factor. The additional analysis of these items is detailed below.

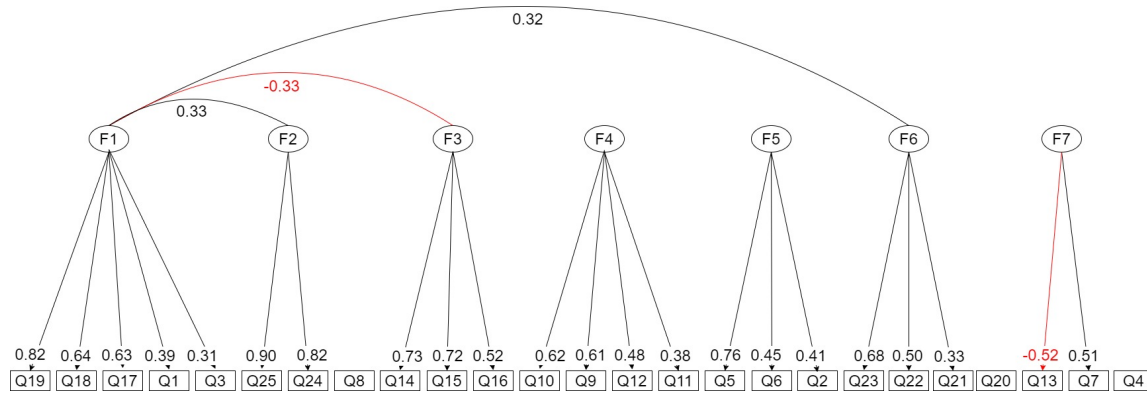


Fig. 2. Path diagram of the 7-factor model with cut-off of 0.30 for factor loading. The items (Q1-Q25) are represented by boxes and the factors (F1-F7) by ellipses. The factors are ordered by the values of the proportion of variance explained by each factor (decreasing order). Factors F2, F3, and F6 are slightly correlated with factor F1. The positive loadings and dependencies are presented in black, negative in red. Q2, Q7, Q12, Q13 have double loadings on factors. For these items, only the connection to the factor with the larger loading is presented in the diagram. Q4, Q8, and Q20 did not have any factor with loading  $> 0.30$  and thus are not connected to any factor.

After deciding on the model, we examined the internal consistency of the assignment of items to factor. This is presented in the next subsection.



### 3.5 Stage 4: Internal Consistency of the Factors and Factors Labeling

The internal consistency of the selected 7-factor model was measured using Cronbach’s  $\alpha$  [11]. As recommended in the Psychometrics literature,  $\alpha$  values in a range of 0.8 to 0.90 are considered as good, and values 0.70 to 0.8 are considered adequate [21], while Kline suggested also to consider values of 0.60 to 0.7 as acceptable for factors that are diverse due their physiological nature [31]. The procedure for evaluating and estimating the obtained factors included computing the  $\alpha$  values for each of the factors separately. In the case of either double loadings or relatively low loadings (0.25 - 0.35), we evaluated models with and without the corresponding items, to evaluate their influence on the  $\alpha$  level. In addition, we tried to combine two or more factors. For example,  $\alpha$  value of factor 7 was very low ( $< 0.3$ ), so we decided to exclude this factor from the analysis.

The resulting values of Cronbach’s  $\alpha$  for all the tested compositions of the factors are presented in Table 2. The factors highlighted in bold ( F1-4, F2-1, F3, F4, F5-3, and F6), with loadings in the range of 0.66 to 0.88, presented the best alpha values, were thus taken as the final mapping of items to factors. Hereafter, we refer to factors F1-4, F2-1, F3, F4, F5-3, and F6 as F1, F2, F3, F4, F5, and F6 correspondingly.

Table 2. A representative sample from the reliability analysis, presenting factor compositions with their corresponding Cronbach  $\alpha$  values. Items with negative correlation (for Factor F7) are presented with a minus sign. The factors that were chosen for further analysis are bolded.

EFA Proposed Factor	Items	Cronbach’s $\alpha$	Chosen Factor	Factor Name
F1-1	Q19, Q18, Q17, Q1, Q3	0.78		
F1-2	Q19, Q18, Q17, Q1, Q2, Q3	0.79		
F1-3	Q19, Q18, Q17, Q1	0.80		
<b>F1-4</b>	<b>Q19, Q18, Q17</b>	<b>0.81</b>	<b>F1</b>	Self-efficacy in Using AI-based EdTech
<b>F2-1</b>	<b>Q25, Q24</b>	<b>0.88</b>	<b>F2</b>	AI-based EdTech vs Human Advice/Recommendation
F2-2	Q25, Q24, Q8	0.71		
<b>F1 + F2</b>	Q19, Q18, Q17, Q25, Q24	0.75		
<b>F3</b>	<b>Q14, Q15, Q16</b>	<b>0.69</b>	<b>F3</b>	Anxieties Related to Using AI-based EdTech
<b>F4</b>	<b>Q10, Q9, Q12, Q11</b>	<b>0.66</b>	<b>F4</b>	AI-based EdTech’s Lack of Human Characteristics
F5-1	Q5, Q6, Q2	0.69		
F5-2	Q5, Q6, Q2, Q7	0.68		
<b>F5-3</b>	<b>Q5, Q6, Q2, Q7, Q1, Q3, Q4</b>	<b>0.75</b>	<b>F5</b>	Perceived Benefits of AI-based EdTec
F5-4	Q5, Q6, Q2, Q13	0.52		
<b>F6</b>	<b>Q23, Q22, Q21</b>	<b>0.67</b>	<b>F6</b>	Preferred Means to Increase Trust in AI-based EdTech
F7-1	-Q13, Q7	0.27		
F7-2	-Q13, Q7, -Q12	0.45		

3.5.1 *Factor Labeling.* Next, the proposed factors were analysed to determine if they indeed represent distinguishable stand-alone properties.

Factor F1 Items Q17-Q19 discuss the aspects of teachers’ confidence in using AI-based technology, so we labeled this factor as “Self-efficacy in Using AI-based EdTech”.

Factor F2 Items Q24 and Q25 are related to acceptance of AI-based recommendation compared to human opinion, so we called this factor “AI-based EdTech vs Human Advice/Recommendation”.

Factor F3 Items Q14 - Q16 discuss possible anxieties related to using AI in teacher pedagogical practice (e.g., the fear to lose autonomy or personal privacy, or being replaced by AI in the future). This factor was labeled “Anxieties Related to Using AI-based EdTech”.



- Factor F4 Items Q9 - Q12 discuss AI-based technology lacking various important values and properties compared to human teachers (e.g. AI has no human intuition and pedagogical expertise, as well as AI does not know the personal history of students outside the online system and does not understand social, emotional and motivational factors driving a student and influencing the learning process). This factor was labeled “AI-based EdTech’s Lack of Human Characteristics”.
- Factor F5 Items Q1-Q7 all describe possible benefits of using AI in education (e.g. diagnosing students’ difficulties, providing automated formative assessment of complex tasks, assisting teachers in preparing lessons, etc). This factor was called “Perceived Benefits of AI-based EdTech”.
- Factor F6 Items Q21 - Q23 are related to possible ways to improve teacher trust in AI (e.g. by increasing the amount of usage and amount of data used for analysis), thus this factor is termed “Preferred Means to Increase Trust in AI-based EdTech”.

### 3.6 Stage 5: Finalizing the instrument

In this stage, items not belonging to any of the factors (Items Q8, Q13, and Q20) were analyzed separately in order to determine whether they represent additional aspects not captured by any other factor, or whether they should be removed from the instrument.

- Item Q8 is slightly correlated with the factor F2 (“Trust in AI vs Human Advice/Recommendation”), however, its removal from the factor significantly increased factor reliability (see Table 2). So, we decided to remove this item from the instrument.
- Item Q13 is related to the lack of transparency in how AI makes decisions. We decided to retain this item in the instrument based on our preliminary research with teachers [35], where the issues of lack of general understanding of how machine learning works and what can and cannot be expected from AI-based system were detected as substantial reasons for not trusting AI-based insights. We called this factor “AI-based EdTech’s Perceived Lack of Transparency”.
- Item Q20 deals with an issue of significant change in ways teachers teach today due to potential AI-based EdTech usage. Based on our previous research [35], we decided to retain this item in the Instrument, as this subject was raised frequently in our interviews with teachers and teacher group discussions related to their experiences with recently developed AI-based EdTech tools (presented in [35]) in their classrooms. We called this factor “Required Shift in Pedagogy to Adopt AI-based EdTech”.

Next, we examined if the proposed factors were aligned with the top-down categories of the instrument:

- The category “Perceived benefits of AI in education” (Appendix A.1.1) is fully aligned with the factor F5.
- The category “Reasons for not trusting AI diagnosis” (Appendix A.1.2) includes the factors F3, F4, and individual item Q13.
- The category “Working alongside AI to improve pedagogy” (Appendix A.1.3) includes the factors F1, F2, F6, and individual item Q20.

To conclude, the final version of the proposed instrument consists of 24 Likert Items divided into eight sub-categories, namely: *Perceived Benefits of AI-based EdTech*, *AI-based EdTech’s Lack of Human Characteristics*, *AI-based EdTech’s Perceived Lack of Transparency*, *Anxieties Related to Using AI-based EdTech*, *Self-efficacy in Using AI-based EdTech*, *Required Shift in Pedagogy to Adopt AI-based EdTech*, *Preferred Means to Increase Trust in AI-based EdTech*, and *AI-based EdTech vs Human Advice/Recommendation*

#### 4 ANALYSIS OF TEACHER RESPONSES

Finally, we use the instrument to measure the distribution of teachers’ location on each of the factors that resulted from the analysis. Per factor, the score of the teacher on that factor was computed by averaging the scores of the items belonging to that factor [9, 46]. The scores of individual items (Q13 and Q20) were taken as they are. The resulting distributions of the scores by the chosen subcategories are presented in Figure 3.

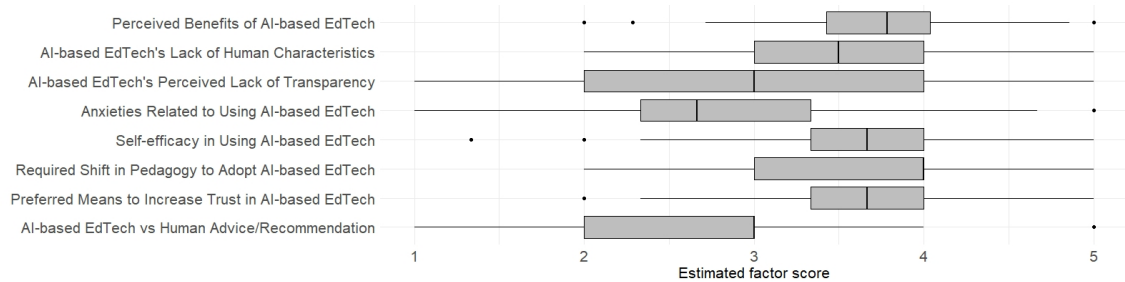


Fig. 3. The distributions of the scores within each subcategory. The scores are on a scale of 1=“Strongly disagree” to 5=“Strongly agree”, with 3 representing a neutral position.

According to the data, teachers perceive AI-based EdTech as beneficial for both students’ self-learning as well as for providing assistance to teachers (Mean = 3.8, Median = 3.8, SD = 0.58). Moreover, teachers’ self-efficacy in their ability to use AI effectively is quite high (Mean = 3.8, Median = 3.7, SD = 0.63), despite the fact that only a minority of the teachers (12%) disagreed with the statement that adoption of AI-based EdTech will require a significant change in their pedagogy. As per possible obstacles, most of the teachers (Mean = 3.5, Median = 3.5, SD = 0.66) considered the lack of human properties (emotions, affect, taking into consideration previous history of students, and pedagogical intuition) as the main reason for not trusting AI-based EdTech diagnostics. Moreover, we analyzed the responses to an open-ended question asking to think about *additional* reasons for not trusting AI in educational settings. The question was defined as optional, so only 36 of the 132 teachers answered it. Among these, more than half (17 out of 36) mentioned issues related to AI’s lack of human abilities to communicate and feel, which emphasizes the importance of this factor even more. Another reason for not using AI-based EdTech was related to teachers’ uncertainty regarding how AI makes decisions (lack of transparency) (Mean = 3.0, Median = 3.0, SD = 1.07). In addition, the results strongly indicated that teachers trust AI-based EdTech less than they trust their peers or experts (Mean = 2.7, Median = 3, SD = 0.8). Interestingly, in contrast to what was expected, teachers were not so afraid to be replaced by AI (Mean = 2.8, Median = 2.7, SD = 0.8). Finally, we find it promising that teachers agreed (Mean = 3.7, Median = 3.7, SD = 0.63) that knowing more about how AI-based EdTech work and how they make decisions would increase their trust in such technologies.

#### 5 DISCUSSION

The analysis of teachers’ responses to the trust instrument reveals eight dimensions influencing teachers’ trust in AI-based EdTech: *Perceived Benefits of AI-based EdTech*, *AI-based EdTech’s Lack of Human Characteristics*, *AI-based EdTech’s Perceived Lack of Transparency*, *Anxieties Related to Using AI-based EdTech*, *Self-efficacy in Using AI-based EdTech*, *Required Shift in Pedagogy to Adopt AI-based EdTech*, *Preferred Means to Increase Trust in AI-based EdTech*, and *AI-based EdTech vs Human Advice/Recommendation*.

Looking at the responses teachers provided on these dimensions, we observe that the sampled teachers have positive perceptions of the usefulness of AI-based EdTech (“Perceived Benefits of AI-based EdTech”) and believe that they will be able to effectively use it in their classroom (“Self-efficacy in Using AI-based EdTech”). However, they are concerned that adopting such technology would require a significant change in their teaching practices (“Required Shift in Pedagogy to Adopt AI-based EdTech”) which might imply additional workload for teachers. The results also show that the teachers appreciate the proposed ways to improve their trust in AI (“Preferred Means to Increase Trust in AI-based EdTech”), and that being replaced by AI is not a major concern of teachers (“Anxieties Related to Using AI-based EdTech”). This result is in contrast to what we would expect based on the research studies that sampled participants from the general public (i.e. [13, 17]) rather than teachers specifically, or the perceptions on AI-related anxieties might be changing towards more positive in recent years after the previous research. In addition, our results strongly indicate that teachers’ have less trust in the advice provided by AI than in one provided by experts or peer teachers (“AI-based EdTech vs Human Advice/Recommendation”), and that they are concerned about the lack of transparency on how AI makes its decisions (“AI-based EdTech’s Perceived Lack of Transparency”). Transparency is dependent on factors like availability of information, conditions of its accessibility, how this information may pragmatically assist in decision making as well as the current knowledge of the users [49]. Here we did not study such nuances, so the referred lack of transparency is entirely in the perception of teachers regardless of how transparent the AI models used in EdTech actually are. Finally, AI’s lack of human characteristics like emotions and intuition (“AI-based EdTech’s Lack of Human Characteristics”) is considered by teachers as one of the biggest limitations of this technology in the context of K-12 education.

Overall, these findings provide useful insights into the factors that affect teachers’ trust in AI-based EdTech, which can inform the design of such technology as well as the design of professional development opportunities for teachers to adopt them in their practice.

First, AI tools should be designed to address the actual needs and to meet the real-world constraints of the K-12 classrooms. This would lead to a decrease in the shift teachers are expected to make in order to adopt the tools in their practice, and thus reduce the barriers to adoption. Previous learning analytics research has shown that this can potentially be achieved by adopting participatory or co-design processes [24, 34] that involve the users (teachers and students) early in the design process. Designing AI-based EdTech “with teachers and for teachers”, would help to address some of the barriers of trust discovered in this research paper at the early stages of conception and design.

Second, similarly to the findings of [30] and [20], our results emphasize the importance of avoiding the presentations of so-called “black-box” solutions and creating instead more transparent systems by providing not only a ‘bottom line’ but also explanations on the reasons that led to a particular decision. Frameworks for explaining AI-based decision making to end-users were shown to be useful in other domains (see for example [39]), and may be embedded into teacher-facing AI-based EdTech. On the other hand, as also evidenced by [30], complete transparency can also potentially lead to information overload for teachers and learners, as sometimes referred to in AI literature as the transparency paradox [40]. Therefore, rather than approaching transparency as a binary concept in AI-based EdTech, to what extent transparency would benefit teachers and learners in a particular context should be studied as a dynamic concept.

Similarly, providing explanations on why AI engines made certain decisions may require revealing some of their internal machinery. While *what* and *how* to reveal depends on the tool, the context, and the conceptual approach, a clear implication of this research is that trust requires some level of transparency. Since that transparency is also dependent on the level of understanding of the underlying machinery, preparing teachers for the ‘AI era’ may also require teachers to improve their knowledge of how AI-based EdTech they intend to use works. Recent research has

also suggested that educating teachers on the internals of AI-based EdTech can increase their awareness of privacy concerns [53], an issue that was raised by the teachers in our research as well (see Q16).

Third, akin to Almada [1] who highlighted the difference between the notions of “automated decision-making” and “human decisions based on automated processing”, our findings stress the significance of allowing teachers to review and override the automated AI decisions in their practice. Providing such functionality can increase teachers’ autonomy and control, reduce the fear of AI-based systems making a painful mistake for which they would be accountable, highlight the importance of a human teacher’s agency in the loop and also reduce teachers’ anxiety of being replaced by AI.

### 5.1 Limitations and Future Research

The presented research provides sufficient initial evidence of the internal consistency of the proposed instrument, however, validating a new instrument is an ongoing process that requires different types of evidence [2, 7]. Since our validation was conducted with a certain population of teachers (high-school science teachers), on the next step we intend to administer the instrument to a diverse population of teachers (in terms of subject matter, technological experience, and other characteristics). These data will be used to further examine the internal validity of the instrument via Confirmatory Factor Analysis (CFA) and can provide more insights on the distribution of teachers responses via Item Factor Analysis (IFA) within the structural equation modeling (SEM) and item response theory (IRT) frameworks [52].

In addition, the proposed instrument can be used to measure trust dimensions in certain groups of teachers (with various general attitudes towards technology and AI, level of expertise in subject matter, LOC characteristics, gender, etc.) to identify the severity of each issue among them. The effectiveness of professional development courses aimed at preparing teachers to work with AI-based EdTech can also be estimated by measuring the shift in the distribution of teachers’ location on each of the proposed dimensions before and after the professional development courses.

## 6 SUMMARY AND CONCLUSIONS

Evidence from various domains confirms that the adoption of AI-based technology by practitioners is not a straightforward process, and users’ willingness to adopt such tools in their practice is likely to be influenced by their trust in them. AI-based educational technology is increasingly entering K-12 education, however the issue of how to measure and support teachers’ trust in the adoption of such technology received only minor attention within the learning analytics research community. To the best of our knowledge, there is no previously established instrument for measuring educators’ perceptions of AI-based EdTech, and the impact of these perceptions on their trust in the adoption of AI tools into their teaching practice.

The goal of the presented research is to introduce a new instrument for measuring teachers’ trust in AI-based EdTech and present our results on its initial implementation with 132 high-school Biology teachers. This paper describes the instrument items creation process and provides a sufficient amount of evidence on the internal structure of the instrument and the internal validity of its components. The results reveal eight factors influencing teachers’ trust in adopting AI-based EdTech: *Perceived Benefits of AI-based EdTech*, *AI-based EdTech’s Lack of Human Characteristics*, *AI-based EdTech’s Perceived Lack of Transparency*, *Anxieties Related to Using AI-based EdTech*, *Self-efficacy in Using AI-based EdTech*, *Required Shift in Pedagogy to Adopt AI-based EdTech*, *Preferred Means to Increase Trust in AI-based EdTech*, and *AI-based EdTech vs Human Advice/Recommendation*. Then we use these emerged factors to analyze teacher responses to the survey items and discuss to what extent they align with the findings from the literature in educational

and other domains. We conclude that the instrument presented is a valuable tool to generate valuable insights into teachers' trust in AI-based EdTech.

The contribution of this research is twofold. First, it presents a novel instrument for measuring teachers' trust in AI-based EdTech. Second, it uses the instrument to portray teachers' trust among a group of science teachers and analyzes the factors influencing it. These extend our theoretical understanding of the human factors affecting acceptance of AI-based EdTech among K-12 educators, and can be used to develop practical means to increase it, for example through more informed co-design of AI-based EdTech and professional development programs for teachers.

## ACKNOWLEDGMENTS

TN is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. The research of GA was supported by the Willner Family Leadership Institute for the Weizmann Institute of Science and the Iancovici-Fallmann Memorial Fund, established by Ruth and Henry Yancovich. The research of TN was partially supported by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center. The authors thank Moriah Ariely, Carmel Bar, Michal Walter, and Asaf Bar-Yosef for leading the professional development programs and for useful discussions.

## REFERENCES

- [1] Marco Almada. 2019. Human intervention in automated decision-making: Toward the construction of contestable systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 2–11.
- [2] American Educational Research Association, American Psychological Association, National Council on Measurement in Education, et al. 2018. *Standards for educational and psychological testing*. American Educational Research Association.
- [3] Peter M Bentler and Douglas G Bonett. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin* 88, 3 (1980), 588.
- [4] Timothy A Brown. 2015. *Confirmatory factor analysis for applied research*. Guilford publications.
- [5] Erik Brynjolfsson and Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.
- [6] Simon Buckingham Shum, Rebecca Ferguson, and Roberto Martinez-Maldonado. 2019. Human-centred learning analytics. *Journal of Learning Analytics* 6, 2 (2019), 1–9.
- [7] Amanda R Butz and Janet L Branchaw. 2020. Entering Research Learning Assessment (ERLA): Validity evidence for an instrument to measure undergraduate and graduate research trainee development. *CBE—Life Sciences Education* 19, 2 (2020), ar18.
- [8] Stephen Cave, Kate Coughlan, and Kanta Dihal. 2019. "Scary Robots": Examining Public Responses to AI (AIES '19). Association for Computing Machinery, New York, NY, USA, 331–337.
- [9] AL Comrey and HB Lee. 1992. Interpretation and application of factor analytic results. *Comrey AL, Lee HB. A first course in factor analysis* 2 (1992), 1992.
- [10] Anna B Costello and Jason Osborne. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation* 10, 1 (2005), 7.
- [11] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.
- [12] Larry Cuban, Heather Kirkpatrick, and Craig Peck. 2001. High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American educational research journal* 38, 4 (2001), 813–834.
- [13] Mutlu Cukurova, Rosemary Luckin, and Carmel Kent. 2020. Impact of an artificial intelligence research frame on the perceived credibility of educational research evidence. *International Journal of Artificial Intelligence in Education* 30, 2 (2020), 205–235.
- [14] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.
- [15] Thomas G Dietterich and Eric J Horvitz. 2015. Rise of concerns about AI: reflections and directions. *Commun. ACM* 58, 10 (2015), 38–40.
- [16] Benedict du Boulay. 2016. Recent meta-reviews and meta-analyses of AIED systems. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 536–537.
- [17] Ethan Fast and Eric Horvitz. 2017. Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.
- [18] Carl Benedikt Frey and Michael Osborne. 2013. The future of employment. (2013).
- [19] L Gaines-Ross. 2016. What do people—not techies, not companies—think about artificial intelligence. *Harvard Business Review* 24 (2016).
- [20] Susanne Gaube, Harini Suresh, Martina Raue, Alexander Merritt, Seth J Berkowitz, Eva Lermer, Joseph F Coughlin, John V Gutttag, Errol Colak, and Marzyeh Ghassemi. 2021. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ digital medicine* 4, 1 (2021), 1–8.

- [21] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, and Ronald Tatham. 2006. *Multivariate data analysis*. Uppersaddle River.
- [22] Harry H Harman. 1976. *Modern factor analysis*. University of Chicago press.
- [23] Monika Hengstler, Ellen Enkel, and Selina Duelli. 2016. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change* 105 (2016), 105–120.
- [24] Kenneth Holstein, Bruce M McLaren, and Vincent Alevan. 2019. Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics* 6, 2 (2019), 27–52.
- [25] John L Horn. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 2 (1965), 179–185.
- [26] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.
- [27] Li-tze Hu and Peter M Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal* 6, 1 (1999), 1–55.
- [28] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [29] Henry F Kaiser. 1974. An index of factorial simplicity. *Psychometrika* 39, 1 (1974), 31–36.
- [30] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2390–2395.
- [31] Paul Kline. 2013. *Handbook of psychological testing*. Routledge.
- [32] Lydia Manikonda and Subbarao Kambhampati. 2018. Tweeting AI: perceptions of lay versus expert twitterati. In *Twelfth International AAAI Conference on Web and Social Media*.
- [33] Kanti V Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530.
- [34] Roberto Martinez-Maldonado, Abelardo Pardo, Negin Mirriahi, Kalina Yacef, Judy Kay, and Andrew Clayphan. 2015. LATUX: An Iterative Workflow for Designing, Validating, and Deploying Learning Analytics Visualizations. *Journal of Learning Analytics* 2, 3 (2015), 9–39.
- [35] Tanya Nazaretsky, Mutlu Cukurova, and Giora Alexandron. 2021. Confirmation bias and trust: Human factors that influence teachers’ adoption of AI-based educational technology. In *Companion Proceedings of the Sixteenth European Conference on Technology Enhanced Learning (AI for Blended-Learning: Empowering Teachers in Real Classrooms Workshop)*.
- [36] Changhoon Oh, Taeyoung Lee, Yoojung Kim, SoHyun Park, Saebom Kwon, and Bongwon Suh. 2017. Us vs. them: Understanding artificial intelligence technophobia over the google deepmind challenge match. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2523–2534.
- [37] Sandy Kristin Piderit. 2000. Rethinking resistance and recognizing ambivalence: A multidimensional view of attitudes toward an organizational change. *Academy of management review* 25, 4 (2000), 783–794.
- [38] William R Revelle. 2017. *psych: Procedures for personality and psychological research*. (2017).
- [39] Aaron Rich, Giora Alexandron, and Reuven Naveh. 2009. An explanation-based constraint debugger. In *Haifa Verification Conference*. Springer, 52–56.
- [40] Neil M Richards and Jonathan H King. 2013. Three paradoxes of big data. *Stan. L. Rev. Online* 66 (2013), 41.
- [41] Bart Rienties. 2014. Understanding academics’ resistance towards (online) student evaluation. *Assessment & Evaluation in Higher Education* 39, 8 (2014), 987–1001.
- [42] Ido Roll and Ruth Wylie. 2016. Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 582–599.
- [43] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.
- [44] Navya Nishith Sharan and Daniela Maria Romano. 2020. The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon* 6, 8 (2020), e04572.
- [45] Saiying Steenbergen-Hu and Harris Cooper. 2014. A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning. *Journal of educational psychology* 106, 2 (2014), 331.
- [46] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using multivariate statistics*. Vol. 5. Pearson Boston, MA.
- [47] Marieke E Timmerman and Urbano Lorenzo-Seva. 2011. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological methods* 16, 2 (2011), 209.
- [48] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, et al. 2020. Human–computer collaboration for skin cancer recognition. *Nature Medicine* 26, 8 (2020), 1229–1234.
- [49] Matteo Turilli and Luciano Floridi. 2009. The ethics of information transparency. *Ethics and Information Technology* 11, 2 (2009), 105–112.
- [50] Cees Van der Eijk and Jonathan Rose. 2015. Risky business: factor analysis of survey data—assessing the probability of incorrect dimensionalisation. *PLoS one* 10, 3 (2015), e0118900.
- [51] Kurt VanLehn. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist* 46, 4 (2011), 197–221.
- [52] RJ Wirth and Michael C Edwards. 2007. Item factor analysis: current approaches and future directions. *Psychological methods* 12, 1 (2007), 58.
- [53] Elad Yacobson, Orly Fuhrman, Sara Hershkovitz, and Giora Alexandron. 2021. De-identification is Insufficient to Protect Student Privacy, or—What Can a Field Trip Reveal? *Journal of Learning Analytics* 8, 2 (2021), 83–92.

## A APPENDIX

### A.1 Instrument

The proposed instrument consists of three main parts: (i) General information related to teachers' knowledge of AI; This part contains a short explanation of AI-term and asks teachers to name a few applications of AI in education and other domains; (ii) 25 5-level scale Likert items related to teachers attitudes and beliefs towards using AI-based tools in educational settings; and (iii) Socio-demographic information (age, gender, years of teaching experience, years of experience of using technology in pedagogy, education level, teaching discipline). The 25 Likert items are presented below. The Likert scale used was: 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly agree.

*A.1.1 Perceived benefits of AI in an educational setting.* Below are some sentences that discuss the views that educators have in the field of adopting artificial intelligence-based tools for personalizing teaching. The sentences address the potential benefits of using artificial intelligence in education. Read each sentence carefully and decide how much it characterizes your opinion and feeling. Circle the number that reflects your position.

- Q1 Artificial Intelligence can assist in formative assessment of complex tasks (e.g., open questions, quality of scientific reasoning), and suggest personalized feedback in real-time (AI makes diagnosis + action).
- Q2 Artificial Intelligence can assist in the development of automated systems for assigning individualized learning paths for students' self-learning.
- Q3 Artificial Intelligence can assist in creating intelligent agents (robots or software) that may serve as learning pals or teaching assistants.
- Q4 Artificial Intelligence can assist in diagnosing students' difficulties to help teachers assign personalized tasks that match individual needs (AI makes the diagnosis, the teacher decides intervention).
- Q5 Artificial Intelligence can assist teachers with in-class management activities, such as identifying students who are off-task (AI focuses on management rather than direct learning).
- Q6 Artificial Intelligence can assist teachers in planning tasks such as preparing lessons and activities before class time.
- Q7 Artificial Intelligence can improve teacher professional training, for example using AI-based avatars to simulate and practice teaching scenarios.

In your opinion, are there other benefits of using AI in the field of education not mentioned above?

*A.1.2 Reasons for not trusting AI diagnosis.* Below are some sentences that discuss the views that educators have in the field of adopting artificial intelligence-based tools for personalizing teaching. The sentences address the potential obstacles related to using artificial intelligence in education. Read each sentence carefully and decide how much it characterizes your opinion and feeling. Circle the number that reflects your position.

- Q8 Discrepancies between the AI diagnosis and the teacher's opinion.
- Q9 AI-developers lack educational experience.
- Q10 AI algorithms do not understand social, emotional, and motivational factors that are very important in teaching.
- Q11 The superiority of the teachers' real-life experience and intuition over a computer's.
- Q12 AI algorithms do not know the history of the student outside the system.
- Q13 Lack of transparency in how AI makes decisions.
- Q14 Removes autonomy and control from teachers to AI.
- Q15 When AI becomes more prevalent, fewer teachers will be required.
- Q16 Use of data raises risks to teacher and learner privacy.

In your opinion, are there other reasons for not trusting AI in the field of education not mentioned above?

*A.1.3 Working alongside AI to improve pedagogy.* To what extent do you agree or disagree with the following statement? Circle the number that reflects your position.



- Q17 I fully trust using AI-based personalized learning tools in my classroom.
- Q18 I believe I will be successful in using AI-based tools for personalization in my classroom when and if they will be available to me.
- Q19 I will use AI-based tools for personalization in my class when and if they will be available to me.
- Q20 I am sure that using AI-based tools for personalization in my classroom will require me to make a significant change in my pedagogy.
- Q21 As they become more prevalent, AI-based tools for personalized learning will gain my full trust.
- Q22 The more I know how an AI-based tool makes decisions, the more I will be able to trust it.
- Q23 The more data is available for AI-based tool, the more I will be able to trust its insights.
- Q24 I rely on the recommendation of an AI-based tool at least as much as I trust a recommendation from a fellow teacher.
- Q25 I rely on the recommendation of an AI-based tool at least as much as I trust a recommendation from an expert (e.g., a researcher from the department of science teaching).

**A.2 Exploratory Factor Analysis**

Table 3. 7-factor model: Standardized loadings (pattern matrix) based upon correlation matrix. The loadings that are greater than 0.30 are highlighted in bold. h2 is the amount of variance in the item explained by the factors. It is the sum of the squared loadings (communality). u2 = 1 - h2 is residual variance (uniqueness). com is item complexity, which measures how much an item reflects a single factor. Below presented the percentage of variance explained by each of the factors.

	F1	F2	F3	F4	F5	F6	F7	h2	u2	com
Q1	<b>0.39</b>	0.00	-0.13	0.13	0.24	0.14	0.1	0.41	0.59	2.92
Q2	<b>0.37</b>	-0.14	-0.03	0.13	<b>0.41</b>	0.11	0.01	0.41	0.59	2.6
Q3	<b>0.31</b>	0.14	-0.02	0.10	0.20	-0.13	0.30	0.34	0.66	3.8
Q4	0.15	0.01	-0.24	0.22	0.24	0.24	0.2	0.41	0.59	5.65
Q5	-0.04	-0.01	-0.06	-0.05	<b>0.76</b>	-0.03	-0.05	0.53	0.47	1.0
Q6	0.09	0.11	0.15	-0.10	<b>0.45</b>	0.13	0.21	0.47	0.53	2.2
Q7	-0.06	0.18	0.24	-0.07	<b>0.31</b>	0.06	<b>0.51</b>	0.59	0.41	2.6
Q8	0.16	-0.30	0.28	0.29	0.03	-0.10	0.11	0.31	0.69	4.1
Q9	-0.19	-0.09	0.10	<b>0.61</b>	-0.01	0.19	0.09	0.50	0.50	1.6
Q10	0.00	0.02	-0.04	<b>0.62</b>	-0.10	-0.17	-0.04	0.42	0.58	1.2
Q11	0.05	-0.16	0.14	<b>0.38</b>	-0.07	0.06	-0.13	0.26	0.74	2.1
Q12	-0.01	0.05	0.08	<b>0.48</b>	0.05	0.09	<b>-0.36</b>	0.42	0.58	2.0
Q13	-0.08	0.03	0.20	0.08	<b>0.34</b>	-0.03	<b>-0.52</b>	0.41	0.59	2.2
Q14	-0.10	0.10	<b>0.73</b>	0.11	0.02	-0.06	-0.12	0.66	0.34	1.2
Q15	0.06	-0.02	<b>0.72</b>	0.00	-0.06	-0.06	0.06	0.51	0.49	1.1
Q16	-0.06	-0.25	<b>0.52</b>	-0.10	0.04	0.23	0.13	0.41	0.59	2.2
Q17	<b>0.63</b>	0.06	-0.11	-0.14	0.05	0.23	0.01	0.71	0.29	1.5
Q18	<b>0.64</b>	-0.09	-0.01	-0.09	-0.08	0.26	-0.10	0.53	0.47	1.5
Q19	<b>0.82</b>	0.15	0.00	0.01	-0.01	-0.15	0.05	0.73	0.27	1.1
Q20	0.01	0.07	0.20	0.12	-0.14	0.24	0.13	0.12	0.88	4.0
Q21	0.18	0.27	0.05	0.07	0.14	<b>0.33</b>	-0.03	0.37	0.63	3.1
Q22	0.08	0.13	-0.02	0.10	-0.02	<b>0.50</b>	0.04	0.33	0.67	1.3
Q23	0.05	0.13	-0.06	-0.01	0.04	<b>0.68</b>	0.00	0.58	0.42	1.1
Q24	0.05	<b>0.82</b>	-0.02	-0.04	-0.01	0.12	-0.06	0.76	0.24	1.1
Q25	0.02	<b>0.90</b>	0.03	0.01	0.00	-0.03	0.07	0.83	0.17	1.0
Percentage of variance	10%	8%	7%	6%	6%	6%	4%			