Education and Information Technologies
https://doi.org/10.1007/s10639-025-13836-7

Artificial intelligence scoring attitudes: scale development and validation

Erdem Boduroglu[1] · Mahmut Sami Yigiter[2,3]

Received: 6 March 2025 / Accepted: 29 October 2025

Abstract
The increasing role of Artificial Intelligence (AI) applications in education necessitates the development of a valid and reliable measurement tool that can assess students' attitudes towards AI-based scoring systems. The purpose of this study is to develop a scale that measures students' attitudes towards AI-based scoring systems in education and to test the validity and reliability of this scale. In the study, a literature review was conducted and expert opinion was consulted to develop the scale items. The first form of the scale was administered to 416 participants. The construct validity of the scale was examined using exploratory factor analysis (EFA) and rotation procedures. As a result of these procedures, a structure consisting of 12 items and two main factors (AI-SAS positive attitude and AI-SAS negative attitude) was determined. In the next step, a confirmatory factor analysis (CFA) was carried out on the data obtained from 441 participants. The results showed that the scale has robust construct validity. To test concurrent validity, comparisons with the General Attitudes Towards Artificial Intelligence Scale (GAAIS) and the AI Anxiety Scale (AI Anxiety) revealed significant relationships between the AI-SAS and these scales. In addition, measurement invariance was tested to ensure that the scale would measure consistently across different demographic groups. The results showed that the AI-SAS scale has a similar factor structure in different groups according to demographic variables such as gender, type of school, use of artificial intelligence in daily life, and can therefore be used in different subgroups. In conclusion, this study provides a reliable and valid scale to measure students' attitudes towards AI-based scoring systems in education. This scale can be used as a tool for evaluating the impact of using AI in educational practice.

Keywords Artificial intelligence · Attitude · Scale development · AI-based assessment · Psychometric validation · Technology acceptance

Extended author information available on the last page of the article

Published online: 19 November 2025

1 Introduction

Artificial intelligence (AI) technologies have started an important transformation process in many fields, from education to health, from financial services to security (Bahoo et al., 2024; Chen et al., 2020; Holzinger et al., 2019; Radulov, 2019). It is noted that these technologies have recently been widely used and adopted by educational institutions due to the innovations they offer in the field of education (Chen et al., 2020). The use of AI in education offers many benefits such as providing learning paths adapted to the individual needs of students, creating better learning environments, providing feedback by assessing student performance, personalising teaching methods, providing interactive and collaborative learning experiences, and providing career counselling services to students (Ahmad et al., 2021; Hamal et al., 2022; Kamalov et al., 2023; Popenici & Kerr, 2017; Shafique et al., 2023; Tapalova & Zhiyenbayeva, 2022; Wang & Chuang, 2024).

In recent years, AI technologies have enabled the development of innovative approaches in education, particularly in assessment processes. AI has the potential to reduce teacher workload and make assessment processes more efficient by automating assignments such as evaluating and grading student work (Ahmad et al., 2023; Chen et al., 2020; Ramazanoglu & Akın, 2024). These systems offer the possibility of analysing student performance objectively and quickly (Attali & Burstein, 2006). They increase objectivity by requiring less human intervention in the assessment of student performance and support the efficiency and accuracy of assessment processes (Williamson et al., 2006).

Automated scoring systems have some limitations as well as advantages. These systems may have difficulty scoring innovative student responses that do not match the expected answers. This can lead to students focusing on finding the right answer rather than thinking creatively (Hahn et al., 2021). While these tools work well in certain contexts, they may be limited in assessing authentic contexts or writing styles. Further research is needed to improve the reliability of these systems (Xu et al., 2024). In conclusion, AI-based automated scoring systems have the potential to transform assessment processes in education. However, continued research and development is needed to increase the effectiveness of these systems and to minimise potential risks. The opinions, contributions and collaboration of educators, students, policy makers and all other stakeholders are crucial (Owan et al., 2023).

The rapid development of AI technologies has made it important to understand attitudes towards these technologies. Various studies show that understanding the attitudes of individuals and employees towards AI is crucial for the successful integration of these technologies (Del Giudice et al., 2023; Park et al., 2024). Attitudes towards AI are often complex and multidimensional. Attitudes towards AI are assessed with positive and negative sub-dimensions that reflect human similarity, adaptability, quality, fear of use, societal and personal benefits, and concerns (Cao et al., 2021; Schepman & Rodway, 2020). Attitudes towards AI may be influenced by demographic and psychological factors. For example, while men generally have a more positive view of AI, women may be more cautious (Karakuş et al., 2023). In addition, fear and personality traits towards AI are also important factors influencing attitudes. While computer use and knowledge of AI increase positive attitudes, anxiety about learning AI can trigger negative attitudes (Brewer et al., 2022). In this context, how AI is implemented in education will be shaped by the attitudes of teachers and students. The impact of AI in education is directly related to students' attitudes and teachers' awareness of these technologies. Therefore, these factors need to be taken into account for the integration and acceptance of AI (Kaya et al., 2022).

Recently, many scales have been developed to better understand the impact of artificial intelligence technologies in education and individuals' attitudes towards these technologies. Some of these scales are listed in Table 1.

The scales in Table 1 address general attitudes towards AI. It can be seen that there is no comprehensive tool in the literature that can measure students' attitudes towards AI-based scoring systems. Therefore, the development of a unique scale that assesses

Table 1 AI attitudes scales

Authors	Scale Name	Factor Structure	Number of Items
Schepman and Rodway (2020)	General Attitudes Towards Artifcial Intelligence	Negative Attitude Towards Artificial Intelligence, Positive Attitude Towards Artifcial Intelligence	20
Grassini (2023)	AI Attitude Scale (AIAS-4)		4
Park, Woo & Kim (2023)	attitudestowards AI application at work (AAAW)	Humanlikeness, Adaptability, Quality, AI use anxiety, Job insecurity, Personal utility	25
Sindermann et al. (2021)	Attitude Towards Artifcial Intelligence Scale (ATAI)	Acceptance, Fear	5
Yılmaz et al. (2025)	Attitude Scale Towards the Use of Artificial Intelligence Technologies in Nursing (ASUAITIN)	positive attitude, negative attitude	15
Jang et al. (2022)	Attitude Toward the Ethics of Artificial Intelligence (AT-EAI)	Fairness, Transparency, Privacy, Responsibility, Non-maleficence	17
Hadlington et al. (2023)	Attitudes towards AI in Defense (AAID)	Positive Outcomes, Negative Outcomes	15
Marengo et al. (2025)	Generative AI Attitude Scale	Positive Attitude, Negative Attitude	13
Krägeloh et al. (2024)	Artificial Intelligence Attitudes Inventory (AIAI)	Positive attitudes subscale, Negative attitudes subscale	16
Derinalp and Ozyurt (2024)	Student Attitudes toward Artificial Intelligence (SATAI) Scale	Behavioral, Affective, Cognitive	26

attitudes towards AI-based scoring will contribute to a better understanding of both the integration of AI in educational processes and students' attitudes towards these technologies. This need is driven by the increasing role of AI in education and the need to more effectively assess students' interactions with these technologies. Students' positive and negative attitudes towards AI-based scoring systems play a critical role in the adoption, use and development of these systems in educational environments. This scale aims to measure students' attitudes towards AI-based scoring systems in a more accurate and valid way. It is believed that this scale will contribute to increasing students' awareness of the use of artificial intelligence in assessment processes by revealing the positive or negative aspects of students' attitudes.

2 Methodology

2.1 Research design

The aim of the study is to develop a scale to assess students' attitudes to AI-based grading of their exams and assignments. To achieve this, a sequential design model - one of the mixed research methods - was employed (Brannen, 2017). During the scale development process, qualitative analysis was used to generate the item pool, while quantitative analysis was applied during the development and validation phases (Creswell & Plano Clark, 2006). The construction of the Artificial Intelligence Scoring Attitude Scale (AI-SAS) followed the steps shown in Fig. 1 and outlined by DeVellis (2017). To assess the reliability and validity of the scale, exploratory factor analysis (EFA) was conducted on the responses of an initial sample of students using the draft version of the AI-SAS. In the validation phase, the factor structure identified in the EFA was tested with a second student sample, and confirmatory factor analysis (CFA) was conducted on their responses. Concurrent validity was carried out on the first sample's responses to the AI-SAS, GAAIS and AI-Anxiety

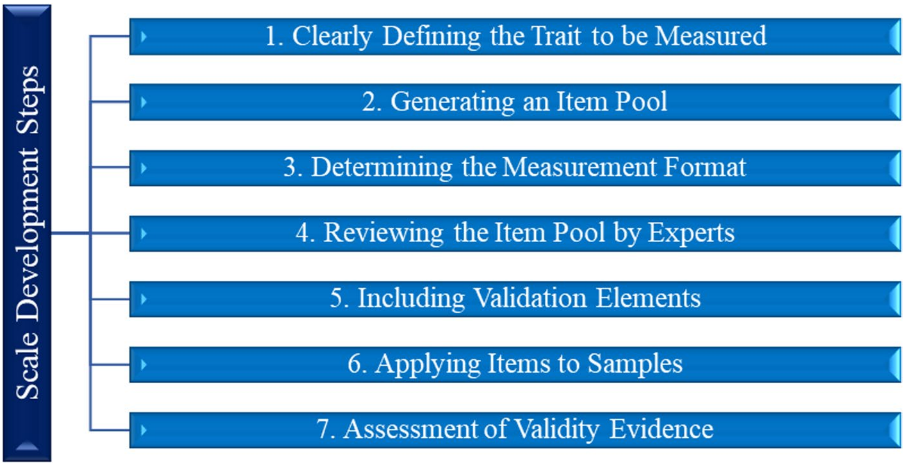


Fig. 1 AI-SAS development steps

scales. Demographic variables obtained from the first sample were used to examine measurement invariance.

2.1.1 Clearly defining the trait to be measured

This study aims to develop a scale to identify and measure students' attitudes towards AI-based scoring systems. With the rapidly increasing use of AI technologies in education, understanding students' attitudes towards these systems is an important step in making the integration of these technologies in education more efficient. The AI-SAS developed in this context focuses on two main dimensions: positive attitudes and negative attitudes. Each dimension examines students' perceptions, emotional reactions and attitudes towards AI-based scoring systems.

The positive attitudes dimension examines students' level of trust in AI-based scoring systems, their belief that the systems will operate in a fair and impartial manner, and their perception of preference. This dimension reflects students' tendency to view AI systems as an accurate and reliable assessment tool. It also measures how they perceive the contribution of AI to the student learning process, the provision of feedback and the impact on student development. Students' confidence in these systems may make them more inclined to use AI as a tool in their educational processes.

The negative attitudes dimension explores students' concerns about AI-based systems, their worries about inaccurate scoring and inconsistent results, and the aspects of the system that they believe may be lacking compared to human assessment. This dimension addresses students' resistance to AI and uncertainties in the feedback they receive from the system. In addition, the lack of the human factor may lead students to find AI-based systems emotionally inadequate. Mistrust and fear of such systems may lead students to develop negative attitudes towards AI-based assessments.

2.1.2 Generating an item pool

In developing the item pool, similar studies in the literature and qualitative analysis techniques were used. First, the items were reviewed from scale development studies in the literature on attitudes towards AI (Derinalp & Ozyurt, 2024; Grassini, 2023; Hadlington et al., 2023; Jang et al., 2022; Krägeloh et al., 2024; Marengo et al., 2025; Park & Kim, 2023; Sindermann et al., 2021; Yılmaz et al., 2025). These studies guided the process of preparing items to be used in measuring attitudes towards AI-based scoring systems. Subsequently, a semi-structured interview form consisting of open-ended questions investigating students' views on the assessment of student answers by AI-based systems was developed and 54 students, 25 high school students and 29 university students, were interviewed. The students' responses to the open-ended questions were analysed by the researchers using content analysis. An inductive approach inspired by Braun and Clarke (2006) was followed to identify meaningful patterns in the data. Two researchers independently coded the responses and reached consensus on the final codes through iterative comparison and discussion. The codes were then grouped into main categories reflecting students' trust, perceived fairness, accuracy concerns, and emotional reactions toward AI-based scoring systems. These categories guided the creation of the item pool. An expert group of 7 people, consisting of experts in the field of information and instructional technologies and measurement and evaluation in education, was formed to create the question pool. The themes from the students' responses were shared with the expert group and the expert group was asked to write a large number of items measuring attitudes towards AI scoring. The expert group wrote a total of 92 items in the first phase. The expert group initially developed 92 items, which were then subjected to detailed descriptive analysis. This analysis focused on assessing the similarity and content of the items. Items with overlapping content were merged, resulting in a draft scale of 32 items (DeVellis, 2017).

2.1.3 Determining the measurement format

The choice of scale development technique depends on the trait being measured. This process should be carried out in parallel with the item generation phase (DeVellis, 2017). As the AI-SAS is an attitude scale, it was developed in Likert-type scale format (Likert, 1932). The number of categories in the Likert-type scale was set at five, as generally accepted in the literature, and scored as '1 = strongly disagree', '2 = disagree', '3 = undecided', '4 = agree' and '5 = strongly agree' (Allen & Seaman, 2007).

2.1.4 Reviewing the item pool by experts

To assess the appropriateness of the 32-item AI-SAS form, an expert opinion form was developed. This form allowed experts to rate each item using the options: 'appropriate', 'should be corrected', 'not appropriate' and 'your opinion' (Berk, 1990). The expert opinion form was first sent to three academicians/teachers specialising in computer education and instructional technology. Their ratings indicated that 12 of the 32 items were inappropriate, while 20 were considered appropriate. Following a consensus among the researchers, the 12 inappropriate items were removed from the scale.In the second phase, the revised 20-item form was sent to four academics/teachers specialising in educational measurement and evaluation. Their feedback indicated that 6 of the remaining 20 items were inappropriate, leaving 14 items as appropriate. In addition, the experts suggested that the terms 'assignment' and 'exam' should be combined as 'assignment/exam' rather than used separately in the items. As a result, a final 14-item AI-SAS form was created to assess students' attitudes towards AI-based grading of homework and exams. In the next stage, this 14-item scale was administered to the participants for further analysis.

After the final 14-item form of the AI-SAS was created, cognitive interviews were conducted with 13 students (3 psychology, 5 law, 2 business and 3 high school) from different departments and levels to determine the appropriateness of the scale items. In the interviews, students were asked what they understood from the items in the 14-item draft form of the AI-SAS and what these items measured. Notes were taken throughout the cognitive interview process. The 14-item first draft of the AI-SAS was finalised by making minor revisions to items that might not have been understood.

2.1.5 Including validation elements

A validation item was included in the AI-SAS. The validation item was included in the 10th row of the scale as "Please mark this AI-related item as 'undecided'.". Participants who did not mark 'undecided' were excluded from the sample. In addition, participants who marked the same option for all items at a high rate were labelled as inattentive responders and excluded from the research sample. As a result, 32 participants in the first sample and 18 participants in the second sample were removed from the data.

2.1.6 Applying items to samples

In order to demonstrate the validity of the AI-SAS, the scale development process was conducted with two separate sample groups. The responses from the first sample were subjected to exploratory factor analysis (EFA) as well as convergent and divergent validity procedures, while the responses from the second sample were subjected to confirmatory factor analysis (CFA). Data collection for both groups took place face-to-face during the autumn term of the 2024–2025 academic year, using convenience sampling with voluntary participation. A critical aspect of scale development is the determination of an appropriate sample size. There are various recommendations in the literature on this issue. Many sources suggest that a sample size between 300 and 500 is sufficient (Comrey & Lee, 1992; Tabachnick & Fidell, 2013). Other researchers suggest that the sample size should be at least five to ten times the number of items in the scale (Bentler & Chou, 1987; Kass & Tinsley, 1979). Some studies even suggest that a sample size between 100 and 200 may be acceptable (Kline, 2014). Furthermore, in terms of statistical power, a minimum sample size of 384 is recommended to ensure representation at a 95% confidence interval (Krejcie & Morgan, 1970). Based on these considerations, the first sample in this study consisted of 416 participants, while the second sample consisted of 441 participants, both of which meet the adequacy criteria for scale validation.

2.1.7 Assessment of validity evidence

In this study, there were two samples to be used in the EFA and CFA analyses. The initial sample sizes were 470 and 465 in the first and second samples respectively. Missing data were analysed first. In the data obtained from the first sample (EFA sample) it was found that the rate of missing data on an item basis was quite low (mean = 1.0%, min = 0.0%, max = 3.7%). Similarly, the rate of missing data for the second sample (CFA sample) was quite low (mean = 0.4%, min = 0.0%, max = 1.3%). Due to the low rate of missing data, the EM algorithm was used to replace the missing data (Dempster et al., 1977). After examining the missing data, a total of 50 inattentive participants who marked one of the different options for the validation item were removed from the sample (Sample1: 32, Sample2: 18). Then, a total of 18 responses that formed response patterns such that they marked the same response to all options (straight-line response) due to the fact that there are two different dimensions in the scale were removed from the sample (Sample1: 12, Sample2: 6) because they reduced data quality (Kim et al., 2019). Following these procedures, the final sample sizes were 416 for the first group and 441 for the second group. In order to assess the normality of the data distribution, the skewness and kurtosis values of the variables were examined. Since all scales had skewness and kurtosis coefficients within the acceptable range of [−1.5, +1.5], the data were considered to be normally distributed (Tabachnick & Fidell, 2013).

Construct validity (EFA, CFA, convergent and divergent validity) and concurrent validity analyses were conducted to demonstrate the validity and reliability of the AI-SAS. Prior to the EFA, the results of the Kaiser Meyer Olkin (KMO) and Bartlett's sphericity tests were analysed. KMO test assesses whether the sample size is suitable for factor analysis and whether the correlations between items are high enough (Pett et al., 2003). If the correlations between items are not high enough, the items cannot be grouped into factors and it should be accepted that the items measure different variables. Bartlett's sphericity test assesses whether the correlations between items are significant (Shrestha, 2021). A KMO value greater than 0.60 and the significance of Bartlett's test ($p<0.01$) indicate that the items can be grouped into factors (Tabachnick & Fidell, 2013). After applying the KMO and Bartlett's sphericity tests, EFA was applied to the data obtained from the first sample. In determining the dimensions of the AI-SAS scale, factors with eigenvalues greater than 1.00 and explaining at least 5% of the total variance were considered (Guttman, 1954; Kaiser, 1960). The scree plot was also analysed (Cattell, 1966). In addition, a parallel analysis was carried out and the number of factors and components suggested by the parallel analysis was also taken into account (Hayton et al., 2004). In EFA, the principal axis factoring (PAF) method, which is recommended for use in social sciences, was used as the factor extraction method (Winter & Dodou, 2012). In exploring the factor structure, EFA was first applied without rotation and then factor loadings were calculated using the 'promax' rotation method, which is recommended when correlations between factors are found. Once the factor structure was obtained, corrected item-total correlations were also calculated. The Average Variance Extracted (AVE) coefficient was calculated to examine convergent validity. AVE is a value that calculates the average of the common variance of the items measuring a construct and shows how much of the variance of the items measuring the construct is explained by the factor. The literature suggests that an AVE of 0.50 and above is sufficient (Hair et al., 2022). The reliability of the factors was assessed using Composite Reliability (CR), Cronbach's Alpha and McDonald's Omega coefficients. CR is a coefficient based on factor loadings that assesses the internal consistency (reliability) of a scale and is expected to be 0.70 and above (Raykov et al., 2016). The interpretation of Cronbach's Alpha and McDonald's Omega was based on Cronbach's (1951) criteria (acceptable ≥ 0.70, good ≥ 0.80 and excellent ≥ 0.90).

In the second stage, first-order (Model I) and second-order (Model II) CFA were applied to the data obtained from the second sample ($n=441$) in order to confirm the construct obtained from the EFA. In evaluating the CFA results, the $\chi2$/sd (chi-squared goodness of fit), CFI (comparative goodness of fit index), TLI (Tucker-Lewis index), RMSEA (root mean square error of approximation) and SRMR (root mean square error of standardised residuals) goodness of fit indices of the model were examined. Acceptable values of model-data goodness of fit indices are as follows $\chi2$/

sd < 5; CFI ≥ 0.90; TLI ≥ 0.90; RMSEA ≤ 0.08; SRMR ≤ 0.08 (Kline, 2014; Bentler & Bonett, 1980; Hu & Bentler, 1999).

In the third step, concurrent validity was examined by calculating Pearson correlation coefficients with two different scales in the literature that are considered to measure a similar and a different construct. In this regard, the General Attitudes towards Artificial Intelligence Scale (GAAIS) (developed by Schepman and Rodway (2020) and adapted to Turkish by Kaya et al. (2022) and the Artificial Intelligence Anxiety Scale (AI-Anxiety) (developed by Wang and Wang (2019) and adapted to Turkish by Akkaya et al. (2021) were used as they contain similar and different items to AI-SAS.

In the fourth stage, the measurement invariance of the scale was examined according to five demographic variables in the data obtained from the first sample ($n = 416$). These variables and the number of participants in their subgroups were as follows Gender (Female = 276, Male = 140), School Type (High School = 195, University = 221), Daily Life AI Use ("Do you use artificial intelligence in your daily life?"; Yes = 326, No = 90), AI Attitude (How would you describe your opinion about artificial intelligence in general?; Positive = 254, Neutral = 140, Negative = 22), Content (Have you ever developed content with artificial intelligence?; Yes = 180, No = 235). Measurement invariance was assessed using multi-group confirmatory factor analysis (MG-CFA) by testing four hierarchical models: configural invariance, metric invariance, scalar invariance, and strict invariance (Vandenberg & Lance, 2000). Changes in comparative fit index (ΔCFI) values were considered to assess model-data fit across these stages (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003; Schumacker & Lomax, 2004). Progression to the next stage was only made when sufficient model fit was confirmed at the previous stage. While some studies suggest testing the significance of chi-squared differences ($\Delta\chi^2$) to determine invariance between stages (Schmitt & Kuljanin, 2008), the chi-squared difference test tends to over-reject the null hypothesis in large samples due to its sensitivity to sample size. As an alternative, Cheung and Rensvold (2002) recommended using changes in CFI (ΔCFI) instead of $\Delta\chi^2$. Following this approach, a ΔCFI value of ≤ 0.01 was adopted as the threshold for establishing measurement invariance across levels in this study (Cheung & Rensvold, 2002).

All analyses in the study were performed in R, using codes written by the researchers. Missing data imputation was performed using 'mvdalab' (Afanador et al., 2022), EFA analyses were performed using 'psych' (Revelle & Revelle, 2015), and CFA and MG-CFA analyses were performed using the 'lavaan' (Rosseel, 2012) and 'SemPlot' (Epskamp, 2015) packages.

3 Results

3.1 Construct validity

Prior to EFA, KMO and Bartlett's tests were examined to test the suitability of the sample size and factorability of the data. KMO value = 0.93 and Bartlett's test ($\chi2$: 2807.19, $P = 0.00$) (Table 2). These findings indicate that the sample size is sufficient and the data can be factorized under the factors.

Table 2 AI-SAS KMO and bartlett's test

Test		Statistic
Bartlett's Test of Sphericity	Approx. Chi-Square	2807.198
	df	91
	p	0.000
KMO		0.93

3.1.1 Exploratory factor analysis (EFA)

First, EFA without factor rotation was performed to explore the factor structure of the scale. The scree plot graph in Fig. 2 shows that there is a two-factor structure as two eigenvalues are greater than 1 (6.64, 1.30, 0.93, 0.74, 0.74, 0.62, …, respectively). Considering that 5% can be accepted as the percentage of variance explained when determining the number of factors, it can be seen that there is a two-factor structure (percentages of variance explained: 44.6%, 6.5%, 2.7%, 1.9%, 1.3%, …) (Guttman, 1954). In addition, the parallel analysis also suggests that the scale has a structure consisting of two factors and two components (Fig. 2). The EFA results show that the first version of the scale, consisting of 14 items, has a structure grouped under two factors.

The EFA results applied to the AI-SAS using the Promax factor rotation method are presented in Table 3.

As a result of EFA with factor rotation, it was decided to remove two items from the scale. While one of these items loaded low on the factors, the other item loaded on two factors (Clark & Watson, 2015; Jordan & Spiess, 2019). The two factors in the AI-SAS were named AI-SAS-Positive Attitude and AI-SAS-Negative Attitude, taking into account the items in the AI-SAS. There was a correlation of 0.66 between AI-SAS-Positive Attitude and AI-SAS-Negative Attitude.

3.1.2 Confirmatory factor analysis (CFA)

As a result of the EFA, the AI-SAS was found to have a structure consisting of two factors and 12 items. To confirm this structure, data were collected a second time from another group and CFA was applied. According to the CFA results, the model-data fit indices were $\chi 2/df = 1.89$, $CFI = 0.999$, $TLI = 0.998$, $RMSEA = 0.045$ and

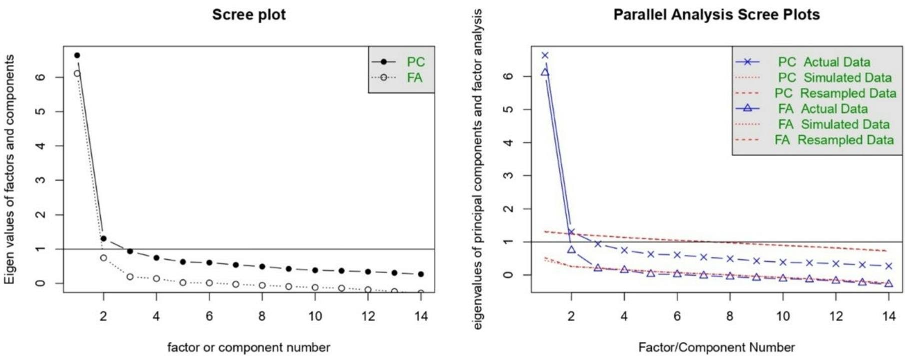


Fig. 2 Scree Plot and Parallel Analysis

Table 3 EFA values of the AI-SAS

	Factors		Mean	Sd	Corrected Item-Total Correlation
Items	1	2			
Item 1	0.850		3.10	1.09	0.696
Item 2	0.871		3.40	1.16	0.675
Item 3	0.745		2.79	1.18	0.621
Item 4	0.801		3.10	1.18	0.638
Item 5*	0.353	0.337	3.48	1.08	Loads on two factors.
Item 6		0.536	3.07	1.13	0.659
Item 7	0.721		2.97	1.03	0.731
Item 8		0.538	2.83	1.10	0.606
Item 9		0.623	2.58	1.10	0.599
Item 10		0.737	2.79	1.13	0.696
Item 11		0.861	2.75	1.18	0.547
Item 12		0.687	2.90	1.12	0.523
Item 13	0.581		3.23	1.12	0.659
Item 14*			3.12	1.60	Factor loadings low.
	12 Items	6 Items	6 Items		
	AI-SAS	AI-SAS-Positive	AI-SAS-Negative		
AI-SAS	1.000				
AI-SAS-Positive	0.917	1.000			
AI-SAS-Negative	0.905	0.659	1.000		

SRMR = 0.033 (Model I). According to these results, the model-data fit is quite good. This shows that the two-factor and 12-item structure of the AI-SAS was confirmed. In addition, a second level CFA was conducted to investigate whether the two-factor structure could be aggregated under a second level factor (Model II). According to the results of the second level CFA, the model-data fit indices were found to be $\chi 2/df = 1.92$, CFI = 0.999, TLI = 0.997, RMSEA = 0.046 and SRMR = 0.033 (Model II). According to the results of the second level CFA, the model-data fit indices are quite good. This shows that the scores obtained from the AI-SAS can be summarized under a single component. (Table 4.)

The factor loadings of Model I and Model II are shown in Fig. 3. It is seen that the factor loadings of the items under the positive attitude factor of AI-SAS (Item 1, Item 2, Item 3, Item 4, Item 7, Item 13) vary between 0.63 and 0.86, and the factor loadings of the items under the negative attitude factor of AI-SAS (Item 6, Item 8, Item 9, Item 10, Item 11, Item 12) vary between 0.67 and 0.83.

3.2 Convergent and divergent validity

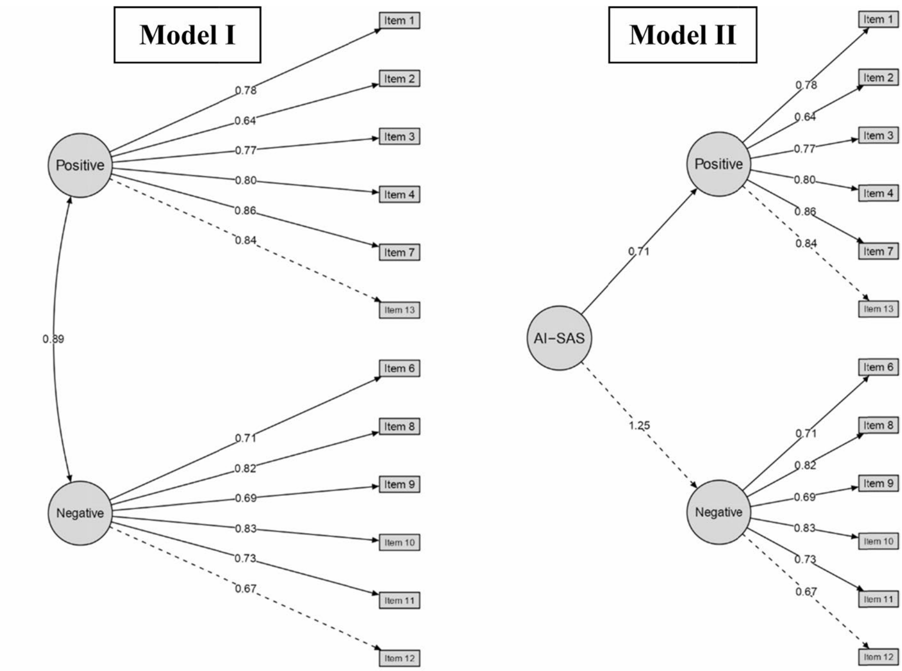
The EFA scores, factor loadings, adjusted item-total correlations, average variance explained (AVE), composite reliability (CR) and reliability coefficients of the

Table 4 CFA model fit indices

Scale	Reference value	Model I	Model II
χ^2		100.3	100.3
p value		53	52
df		0.000	0.000
χ^2/df	<5	1.892	1.929
RMSEA	<0.07	0.045	0.046
SRMR	<0.07	0.033	0.033
GFI	>0.90	0.998	0.998
AGFI	>0.90	0.994	0.993
NFI	>0.90	0.997	0.997
CFI	>0.92	0.999	0.999
TLI	>0.90	0.998	0.997

12-item AI-SAS scale were calculated and convergent and divergent validity were examined (Table 5).

Looking at the convergent analysis values of the AI-SAS scale, the AVE values of the positive and negative factors are 0.63 and 0.53 respectively. The Cronbach Alpha reliability coefficients for the AI-SAS, AI-SAS-Positive and AI-SAS-Negative factors are 0.91, 0.89 and 0.85 respectively, while the McDonald's Omega reliability



Note 1. Model I: Two Factor CFA Model, Model II: Second Level CFA Model.
Note 2. Negative Attitude dimension was reverse coded.

Fig. 3 CFA Path Diagram for Model I and Model II. Note 1. Model I: Two Factor CFA Model, Model II: Second Level CFA Model. Note 2. Negative Attitude dimension was reverse coded

Table 5 AVE, reliability and explained variance percentages of the AI-SAS

Factor	Items	Factor Loadings	Corrected Item-Total Correlation	AVE	CR	Explained Variance Percentages	Cronbach Alpha	McDonald's Omega
AI-SAS Positive Attitude	Item 1	0.833	0.704	0.625	0.895	0.30	0.893	0.930
	Item 2	0.845	0.658					
	Item 3	0.737	0.650					
	Item 4	0.780	0.657					
	Item 7	0.711	0.659					
	Item 13	0.577	0.745					
AI-SAS Negative Attitude	Item 6	0.514	0.631	0.527	0.855	0.23	0.848	0.880
	Item 8	0.532	0.604					
	Item 9	0.620	0.701					
	Item 10	0.724	0.552					
	Item 11	0.850	0.529					
	Item 12	0.667	0.674					
AI-SAS						0.53	0.911	0.928

coefficients are 0.93, 0.93 and 0.88 respectively. It can therefore be concluded that the AI-SAS scale has convergent validity. The factor loadings were above 0.50 in both sub-dimensions and the corrected item-total correlations were at acceptable levels. The total variance explained was 53%, which is considered sufficient for scale development studies in the social sciences. The total explained variance of 53% obtained in this study is considered adequate for self-report scale research (Smedslund et al., 2022). Average variance explained (AVE) values above 0.50 indicate that the scale has convergent validity. The CR values were 0.90 and 0.86 respectively, indicating that the internal consistency of both sub-dimensions was high. The Cronbach's alpha values of the scale and its factors (0.91, 0.89 and 0.85) and the McDonald's omega coefficients (0.93, 0.93 and 0.88) support that the reliability is high. These results suggest that the scale is a valid and reliable instrument for measuring attitudes towards AI-based scoring.

3.3 Concurrent validity

To demonstrate the concurrent validity of the AI-SAS scale, the General Attitudes Towards Artificial Intelligence Scale (GAAIS) consisting of two factors (positive attitudes-negative attitudes) and the Artificial Intelligence Anxiety Scale (AI Anxiety) consisting of four factors (learning, job replacement, socio-technical blindness, AI configuration) were used (N=416) (Table 6). In this study, the Cronbach's alpha values of the GAAIS scale, GAAIS positive attitudes and GAAIS negative attitudes factors were 0.88, 0.90 and 0.82 respectively. The Cronbach's alpha values of the AI fear scale and the learning, job replacement, socio-technical blindness, and AI configuration factors are 0.93, 0.92, 0.82, 0.82, and 0.93, respectively. Looking at the Pearson correlation values (r) between the scales and their sub-dimensions, it can be seen that there is a positive and moderate relationship between AI-SAS and GAAIS (r=0.53). At the same time, there is a moderate and negative relationship between AI-SAS and the AI Anxiety scale (r=−0.34). All these results show the concurrent

Table 6 Pearson correlation values between factors

	1	2	3	4	5	6	7	8	9	10	11
1. AISAS											
2. AISAS-Positive Attitude	0.917										
3. AISAS-Negative Attitude	0.905	0.659									
4. GAIAS	0.526	0.498	0.459								
5. GAAIS-Positive Attitude	0.478	0.506	0.360	0.884							
6. GAAIS-Negative Attitude	−0.369	−0.272	−0.405	−0.735	−0.333						
7. AI Anxiety	−0.335	−0.243	−0.371	−0.569	−0.39	0.581					
8. AI-ANX-Learning	−0.255	−0.192	−0.275	−0.516	−0.410	0.445	0.733				
9. AI-ANX-Job Replacement	−0.323	−0.239	−0.353	−0.459	−0.312	0.473	0.835	0.390			
10. AI-ANX-Sociotechnical Blindness	−0.318	−0.227	−0.356	−0.446	−0.256	0.527	0.859	0.425	0.756		
11.AI-ANX-AI Configuration	−0.195	−0.132	−0.228	−0.418	−0.271	0.450	0.841	0.492	0.615	0.671	

Note AISAS-Negative Attitude variable is reverse coded

validity of the AI-SAS scale. At the same time, the positive attitude factor of the AI-SAS shows small or moderate negative correlations with the factors of the AI-Anxiety scale, while the positive attitude factor of the GAAIS shows positive and moderate correlations. On the other hand, the AI-SAS Negative Attitude factor has a moderate and positive correlation with the GAAIS Negative Attitude and a positive and moderate correlation with the sub-dimensions of the AI-Anxiety scale (as AI-SAS Negative Attitude is reverse coded here, the direction of the relationship should be reversed when interpreting).

3.4 Measurement invariance

Measurement invariance was tested with MG-CFA according to five different demographic variables in the hierarchical order of configural invariance, metric invariance, scalar invariance and strict invariance (Table 7). The value of ΔCFI was used to determine whether invariance was achieved between two hierarchical levels. A ΔCFI<0.01 between levels indicates that measurement invariance is achieved (Cheung & Rensvold, 2002).

The ΔCFI value of the AI-SAS scale is less than 0.01 at all stages according to gender, school type, AI use in daily life, AI attitude and content development with AI variables (see Table 5). Therefore, the AI-SAS scale provides configural, metric, scalar and strict invariance according to five different demographic variables. This finding shows that the AI-SAS scale has a similar factor structure across different subgroups and can therefore be used to measure groups with different characteristics.

Table 7 Measurement invariance by demografic variables

Demografic Variable	Stage	χ^2	sd	χ^2/sd	*RMSEA*	*SRMR*	GFI	AGFI	*TLI*	*CFI*	ΔCFI
Gender	Configural	161.684	106	1.525	0.032	0.040	0.999	0.998	0.994	0.995	-
	Metric	144.083	116	1.242	0.029	0.045	0.998	0.998	0.995	0.996	0.001
	Scalar	163.654	126	1.299	0.032	0.048	0.998	0.997	0.994	0.994	0.002
	Strict	169.961	138	1.232	0.029	0.050	0.998	0.998	0.995	0.995	0.001
School Type	Configural	174.777	106	1.649	0.034	0.041	0.999	0.998	0.993	0.995	-
	Metric	148.032	116	1.276	0.031	0.048	0.998	0.998	0.994	0.995	0.000
	Scalar	171.798	126	1.363	0.036	0.051	0.998	0.997	0.993	0.993	0.002
	Strict	183.952	138	1.333	0.035	0.055	0.998	0.997	0.993	0.993	0.000
Daily Life AI Use	Configural	171.654	106	1.619	0.037	0.043	0.999	0.998	0.992	0.993	-
	Metric	172.256	116	1.485	0.041	0.050	0.998	0.997	0.990	0.991	0.002
	Scalar	182.105	126	1.445	0.039	0.051	0.998	0.997	0.991	0.991	0.000
	Strict	191.748	138	1.389	0.038	0.052	0.998	0.997	0.992	0.991	0.000
AI attitude	Configural	217.297	159	1.367	0.04	0.049	0.998	0.997	0.989	0.992	-
	Metric	205.569	179	1.148	0.034	0.054	0.997	0.996	0.993	0.993	0.001
	Scalar	228.866	199	1.150	0.034	0.057	0.997	0.996	0.992	0.992	0.001
	Strict	247.498	223	1.110	0.03	0.059	0.997	0.996	0.994	0.993	0.001
Content	Configural	172.432	106	1.627	0.033	0.041	0.999	0.998	1.014	1.000	-
	Metric	157.954	116	1.362	0.036	0.050	0.998	0.997	1.009	1.000	0.000
	Scalar	168.071	126	1.334	0.034	0.051	0.998	0.997	1.009	1.000	0.000
	Strict	172.981	138	1.253	0.031	0.052	0.998	0.997	1.010	1.000	0.000

3.5 How to score the scale

The AI-SAS scale consists of 12 items in total, 6 items in the AI-SAS positive attitude dimension and 6 items in the AI-SAS negative attitude dimension. Two items from the original 14-item draft (original Items 5 and 14) were removed after exploratory factor analysis; the remaining items were renumbered for the final 12-item AI-SAS. For cross-reference, see the Table 8 in Appendix-1 below. The items of the scale are given in Table 9 in Appendix-2.

The scores obtained from the items in the dimensions of the scale can be summed and used separately as two dimensions. In addition, the second level CFA results in Table 4 show that the items in the AI-SAS Negative Attitude dimension can be reverse coded and summed with the items in the AI-SAS Positive Attitude dimension and used by obtaining a single AI-SAS score. As a result, the AI-SAS Negative Attitude and AI-SAS Positive Attitude dimensions can be used separately, or the AI-SAS Negative Attitude items can be reverse coded and summed with the AI-SAS Negative Positive items to obtain a single AI-SAS score. It can be interpreted that as the AI-SAS score obtained in this way increases, students' positive attitudes towards AI scoring increase, whereas as the AI-SAS score decreases, students' positive attitudes decrease.

4 Discussion and conclusion

This study was conducted according to the scale development process proposed by DeVellis (2017). An operational plan, including the development steps of the AI-SAS scale, was prepared and the process steps were followed. EFA was applied to the data obtained by applying AI-SAS to the first sample and factor extraction was performed. CFA was applied to the data obtained from the second sample and the factor structure was confirmed. The results of the analysis show that AI-SAS consists of two factors: Positive Attitude, Negative Attitude. The Positive Attitude dimension measures individuals' positive perceptions of AI-based scoring systems and their level of acceptance of these systems. In this dimension, there are items measuring the belief that artificial intelligence will assess exams and assignments in a fair, impartial and reliable way, the confidence that it can produce more consistent and reliable results compared to human assessors, that it will be seen as a facilitating and efficient tool in the assessment process in large-scale exams, and that it will be accepted as a valuable technological innovation in education. The AI-SAS Negative Attitude dimension measures individuals' negative perceptions of AI-based scoring systems and their fear of these systems. This dimension includes items measuring concern about the possibility of making mistakes in AI-based scoring, perceptions of inaccurate or inconsistent scoring, perceptions of inadequacy in correctly understanding students' responses, and general feelings of mistrust and discomfort towards AI-based scoring. The 12-item AI-SAS has adequate reliability and content validity. The AI-SAS was also validated for construct, concurrent, convergent and divergent validity. The results indicate that the AI-SAS has very good psychometric properties and is an adequate instrument for measuring attitudes towards AI-based scoring.

4.1 Theoretical implications

This study develops the Artificial Intelligence Scoring Attitude Scale (AI-SAS) to measure students' attitudes towards AI-based scoring and comprehensively evaluates its psychometric properties. The validated AI-SAS is broadly designed to provide an assessment framework that can be used for a wide range of different AI-based scoring technologies or for comparative assessments. Many scales have already been developed around the concept of AI attitude (Derinalp & Ozyurt, 2024; Grassini, 2023; Hadlington et al., 2023; Jang et al., 2022; Krägeloh et al., 2024; Marengo et al., 2025; Park & Kim, 2023; Sindermann et al., 2021; Yılmaz et al., 2025). The factor structure of six of the ten scales analysed in the literature review section consists of negative and positive dimensions, as in the AI-SAS. While one scale has a single factor, the other scales consist of three, five and six factor structures. AI-SAS had a similar factor structure to the majority of the AI Attitude scales. On the other hand, because AI-SAS measures attitudes towards the scoring of student responses by AI-based systems, it has a different and unique structure from all of these scales, both in terms of theory and items. The constructed AI-SAS consists of two factors: (1) AI-SAS Negative Attitude Factor (i.e., perceptions of fairness, impartiality, trust, and favourability towards AI scoring); (2) AI-SAS Negative Attitude (i.e., perceptions of concern that

AI scoring may be inaccurate, inconsistent scoring, inability to understand answers correctly, and general insecurity and discomfort).

Designed to understand and measure students' attitudes towards Artificial Intelligence (AI)-based scoring systems, AI-SAS can be used as an important research tool, particularly in the fields of educational technology and measurement and evaluation in education. In particular, the process of integrating AI into assessment processes can identify individuals' perceptions, confidence levels and potential resistance points towards the technology and provide important insights into the applicability of AI-based assessment systems in educational settings. On the other hand, the attitudes of stakeholders are very important in the implementation of a new technology in education (Del Giudice et al., 2023; Park et al., 2024). Before using AI-based technologies in nationwide examinations, it is extremely important to obtain the opinions of stakeholders and to organise AI-based scoring systems in accordance with these opinions in terms of the implementation of a new technology (Horowitz et al., 2024; Tverskoi et al., 2022). In this context, AI-SAS can reveal the level of acceptance and reasons for discomfort towards AI assessment by providing deep insights from students in the evaluation of AI-based assessment systems. The data obtained can provide guiding recommendations for educational administrators, policy makers and system developers. AI-SAS can be used to identify individual differences towards AI-based assessment and grading systems. By examining the attitudes of students of different ages, levels of education or departments towards AI scoring, the variables that influence the acceptance or rejection of these systems can be analysed. In addition, AI-SAS can provide suggestions for comparing different AI-based assessment systems. As technology develops, different AI-based assessment systems will be developed and the data obtained from AI-SAS can be used to determine the effectiveness of these systems. When measuring success in education, different measurement methods are used in different industries (e.g. scenario-based questions, open-ended questions, short-answer questions, performance evaluation, interview, etc.). Therefore, students' attitudes towards the evaluation of student responses obtained from different measurement methods with AI-based scoring tools will also be different. AI-SAS can be used to reveal the differences in students' attitudes towards AI scoring according to different measurement methods. Again, AI-SAS can help researchers develop and test theories and models related to AI. With AI-SAS, researchers can conduct modelling studies that explore the relationships between different variables such as students' motivation, anxiety, cognitive engagement, AI literacy, perceived enjoyment of AI (Cengiz & Peker, 2025; Liu & Reinders, 2025; Liu et al., 2025). The findings may provide new insights and understanding for more successful development and application of theories and models related to AI.

4.2 Practical implications

The results of this study show that AI-SAS has satisfactory validity and reliability. AI-SAS can be used to evaluate, rank or compare AI-based evaluation technologies or products. Companies, universities and public organisations are researching and producing AI-based evaluation systems with large budgets. AI-SAS can play an important role in revealing the attitude and satisfaction obtained from these AI-based assessment products. In addition, the students in the sample in this study are at high school and university level, and the age range of the participants varies between 15 and 48 years (mean = 19.3, SD = 3.7). It can therefore be seen that AI-SAS can be used at different levels of education and across a wide age range. In addition, the measurement invariance results in Table 4 show that the AI-SAS has measurement invariance across different groups, so that the results obtained from different groups are comparable. Therefore, the 12-item AI-SAS measurement tool will be able to guide developers and implementers of AI-based assessment systems to improve their systems by providing immediate feedback to users about their systems. In addition, the AI-SAS is designed in two dimensions to measure positive and negative attitudes. This allows the scale to be compared across two different dimensions of AI-based scoring systems.

The AI-SAS can be adapted and used in different studies according to the needs of researchers. In this study, AI-SAS was found to have a positive and moderate correlation with the GAAIS and a negative and moderate correlation with the AI anxiety scale. In this context, AI-SAS can be used in cross-sectional or intervention studies aimed at improving student satisfaction and attitudes towards AI-based assessment systems. In addition, AI-SAS will play an important role in the design of studies aimed at identifying the factors that increase or decrease students' attitudes towards AI-based scoring systems.

In recent years, Artificial Intelligence (AI)-based scoring systems have been increasingly used in education and business, offering innovative solutions to traditional assessment processes. However, the effective adoption of these technologies is directly related to the attitudes of individuals and organisations towards AI-based scoring systems. While users' perceptions of these systems as fair, reliable and useful increase acceptance of the technology, negative attitudes make the adoption process more difficult (Horowitz et al., 2024; Tverskoi et al., 2022). Theoretical approaches to the acceptance of new technologies indicate that individuals' perceptions and attitudes play a critical role in the adoption process. In particular, the Technology Acceptance Model (TAM) developed by Davis (1989) shows that the ease of use and the benefits provided by a technology significantly influence the acceptance process of individuals. In this context, creating positive perceptions about the transparency, reliability and fairness of AI-based scoring systems can facilitate users' adoption of these systems (Chai et al., 2024). In addition to individual attitudes, social and organisational dynamics also shape the acceptance process of AI-based scoring technologies. In particular, factors such as concerns, algorithmic biases and the role of human factors in assessment processes directly influence perceptions of these systems (Starke et al., 2022; Chai et al., 2024). As a result, the effective diffusion of AI-based scoring technologies is directly related to users' attitudes towards these systems. Strengthening the perception that the technology is fair, reliable and useful is one of the main factors that accelerate the adoption process. In this direction, the AI-SAS scale, awareness-raising training, regulations developed within an ethical framework and user-friendly system designs will contribute to determining the level of acceptance of AI-based scoring technologies by the wider masses.

4.3 Limitations and suggestions

In this study, AI-SAS was developed to measure students' attitudes towards AI-based scoring systems. Although all stages of the scale development process were carefully applied in the development of AI-SAS, our study has some limitations. Firstly, probabilistic sampling methods were not used to determine the sample in the development of AI-SAS, but random sampling was used. Therefore, the results of this study may not be generalisable to students in other countries and this needs to be investigated. There may be differences especially in countries with different levels of information and technology. Therefore, this study suggests that future researchers should conduct cross-cultural studies to verify the appropriateness of AI-SAS. Secondly, the sample in this study consists of high school and university students. Primary and secondary school students were not included in the sample due to concerns that they may not have sufficient knowledge, equipment, and attitudes towards AI. Therefore, the fact that the AI-SAS was not investigated among primary and secondary school students is another limitation of this research. Thirdly, the data collected in this study were based on participants' self-reports. Biases such as social desirability, which reflect the tendency to provide responses that present oneself in a favorable light, may have influenced participants' answers. Consequently, positive attitudes may have been overstated, while negative attitudes may have been understated. It is therefore considered valuable to apply the scale to different social or cultural groups in future research. Fourth, the educational context of the participants represents a potential limitation. The participants in this study shared similar educational environments and systems. Therefore, the attitudes of students from similar educational settings may not represent those of individuals in different contexts. Fifth, the cross-sectional design of this study represents another limitation. In future research, it is recommended that the AI-SAS be employed in longitudinal research designs to monitor how students' attitudes toward AI-based scoring evolve over time as AI technologies continue to develop and become more integrated into educational settings. Moreover, experimental studies could be conducted to test the effectiveness of interventions aimed at enhancing students' trust and acceptance of AI-based scoring systems—such as transparency-enhancing feedback mechanisms, AI literacy training, or mixed human–AI evaluation models. Furthermore, integrating the AI-SAS into theoretical frameworks such as the Technology Acceptance Model (TAM) or the Unified Theory of Acceptance and Use of Technology (UTAUT) would provide deeper insights into the cognitive and affective mechanisms underlying the adoption and use of AI-based assessment tools. Such theoretical integration would not only strengthen the construct validity of the AI-SAS but also contribute to advancing theoretical understanding of AI acceptance in educational contexts.

Appendix 1

Table 8 Final item mapping table

Original item number	Final (current) item number	Comment
1	1	unchanged
2	2	unchanged
3	3	unchanged
4	4	unchanged
5	—	Removed after EFA
6	5	renumbered (original 6 → final 5)
7	6	renumbered (original 7 → final 6)
8	7	renumbered
9	8	renumbered
10	9	renumbered
11	10	renumbered
12	11	renumbered
13	12	renumbered (original 13 → final 12)
14	—	Removed after EFA

Note 1.The original item number indicates the numbers in the original form of the scale and those used in the analyses in the article. Two items (original Items 5 and 14) were removed from the original 14-item draft following exploratory factor analysis, and the AI-SAS items were renumbered to obtain the final (current) item number. The final (current) item number corresponds to the item numbers in Appendix 2

Appendix 2

Table 9 AI-SAS items

Factor	Instructions: Dear Participant, This questionnaire is designed to assess your attitudes towards Artificial Intelligence (AI) scoring your assignments/exams. Please read each item carefully and tick the most appropriate option between 1 and 5.	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Positive	(1) I believe the AI will mark assignments/exams fairly and impartially.	1	2	3	4	5
Positive	(2) I am confident that the AI will treat all students equally when scoring assignments/exams.	1	2	3	4	5
Positive	(3) I trust the AI more than human raters to score my assignments/exams.	1	2	3	4	5
Positive	(4) AI should score students' answers in exams with many students instead of humans.	1	2	3	4	5
Negative	(5) AI usually scores my assignments/exams inconsistently.	1	2	3	4	5
Positive	(6) I trust the AI to score my assignments/exams correctly.	1	2	3	4	5
Negative	(7) I think there is a high probability that the AI will score my assignments/exams incorrectly.	1	2	3	4	5
Negative	(8) I think the AI may have difficulty understanding my answers to questions in exams correctly.	1	2	3	4	5
Negative	(9) The idea of AI scoring my assignments/exams makes me uncomfortable.	1	2	3	4	5
Negative	(10) The possibility of making a mistake in AI scoring makes me feel anxious.	1	2	3	4	5
Negative	11. I have difficulty understanding how the AI will score my answers.	1	2	3	4	5
Positive	12. I think AI can be a valuable tool for scoring processes.	1	2	3	4	5

Note 1.The AI-SAS consists of a total of 12 items: six items assess Positive Attitude, and six items assess Negative Attitude (reverse-scored). The items are listed above. Scores for each subscale can be used independently, or a single overall AI-SAS score can be calculated by reverse-scoring the Negative Attitude items (Items 6, 8, 9, 10, 11, and 12) and then summing all items. Items 6, 8, 9, 10, 11, and 12 represent the Negative Attitude dimension and should be reverse scored as follows: 'Strongly Disagree' = 5, 'Disagree' = 4, 'Undecided' = 3, 'Agree' = 2, 'Strongly Agree' = 1

Note 2. The total AI-SAS score is obtained by summing the items from the Positive Attitude dimension and the reverse-scored Negative Attitude dimension. Higher AI-SAS scores indicate a more positive attitude toward AI-based scoring systems, whereas lower AI-SAS scores reflect a more negative attitude toward such systems.

Acknowledgements This study was supported by the Coordination Office for Scientific Research Projects (BAP) of the Social Sciences University of Ankara (ASBÜ) under project code RBB-2025-240.

Author contributions **Dr. Erdem Boduroglu:** Conceptualization, Methodology, Data Collection, Data Analysis, Writing, Review.**Dr. Mahmut Sami Yigiter:** Methodology, Data Collection, Data Analysis, Writing, Visualization, Software.

Funding statement This article is published (or funded) under the open access agreement of the University of Salford.

Data availability The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The author of this study declares that there is no conflict of interest related to the content of this article. The design, data collection, analysis, and interpretation of the research were conducted independently. Moreover, there are no personal, academic, or professional interests that could affect the results of the work.

References

Afanador, N. L., Tran, T., Blanchet, L., & Baumgartner, R. (2022). Mvdalab: Multivariate data analysis laboratory. *Retrieved from Cran R Project* https://cran.r-project.org/web/packages/mvdalab/index.html. *r-project*.

Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M., & Hyder, S. I. (2021). Artificial intelligence and its role in education. *Sustainability, 13*(22), Article 12902.

Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Al-Fuqaha, A. (2023). *Data-driven artificial intelligence in education: A comprehensive review*. IEEE Transactions on Learning Technologies.

Akkaya, B., Özkan, A., & Özkan, H. (2021). Yapay Zeka Kaygı (YZK) Ölçeği: Türkçeye Uyarlama, Geçerlik ve Güvenirlik Çalışması. *Alanya Akademik Bakış, 5*(2), 1125–1146. https://doi.org/10.29023/alanyaakademik.833668

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress, 40*, 64–65.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology Learning and Assessment*, 4(3).

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics*, *4*(2), 23.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berk, R. A. (1990). Importance of expert judgment in content-related validity evidence. *Western Journal of Nursing Research, 12*(5), 659–671. https://doi.org/10.1177/019394599001200507
Brannen, J. (2017). *Mixing methods: Qualitative and quantitative research*. Routledge.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Brewer, P., Bingaman, J., Paintsil, A., Wilson, D., & Dawson, W. (2022). Media use, interpersonal communication, and attitudes toward artificial intelligence. *Science Communication, 44*, 559–592. https://doi.org/10.1177/10755470221130307
Cao, G., Duan, Y., Edwards, J., & Dwivedi, Y. (2021). Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. *Technovation*. https://doi.org/10.1016/J.TECHNOVATION.2021.102312
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
Cengiz, S., & Peker, A. (2025). Generative artificial intelligence acceptance and artificial intelligence anxiety among university students: The sequential mediating role of attitudes toward artificial intelligence and literacy. *Current Psychology*. https://doi.org/10.1007/s12144-025-07433-7
Chai, F., Ma, J., Wang, Y., Zhu, J., & Han, T. (2024). Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2024.1221177
Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access : Practical Innovations, Open Solutions, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5
Clark, L. A., & Watson, D. (2015). Constructing validity: Basic issues in objective scale development. *Methodological issues and strategies in clinical research* (4th ed.)., pp. 187–203). American Psychological Association.
Comrey, A. L., & Lee, H. B. (1992). Interpretation and application of factor analytic results. *Comrey AL, Lee HB. A first course in factor analysis*, *2*, 1992.
Creswell, J. W., & Plano Clark, V. L. (2006). *Designing and conducting mixed methods research*. SAGE.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/bf02310555
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.
Del Giudice, M., Scuotto, V., Orlando, B., & Mustilli, M. (2023). Toward the human– centered approach. A revised model of individual acceptance of AI. *Human Resource Management Review, 33*(1), Article 100856. https://doi.org/10.1016/j.hrmr.2021.100856
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society, Series B: Statistical Methodology, 39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
Derinalp, P., & Ozyurt, M. (2024). Adaptation of the student attitudes toward artificial intelligence scale to the Turkish context: Validity and reliability study. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2024.2352921
DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.
Epskamp, S. (2015). Semplot: Unified visualizations of structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(3), 474–483.
Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1191628
Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*(2), 149–161. https://doi.org/10.1007/bf02289162
Hadlington, L., Binder, J., Gardner, S., Karanika-Murray, M., & Knight, S. (2023). The use of artificial intelligence in a military context: Development of the attitudes toward AI in defense (AAID) scale. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1164810
Hahn, M., Navarro, S., De La Fuente Valentín, L., & Burgos, D. (2021). A systematic review of the effects of automatic scoring and automatic feedback in educational settings. *Ieee Access : Practical Innovations, Open Solutions, 9*, 108190–108198. https://doi.org/10.1109/ACCESS.2021.3100890

Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM) (3rded.). Sage.

Hamal, O., Faddouli, E., Harouni, N. E., M. H. A., & Lu, J. (2022). Artificial intelligent in education. *Sustainability*, *14*(5), 2862.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 9*(4), Article e1312.

Horowitz, M. C., Kahn, L., Macdonald, J., & Schneider, J. (2024). Adopting AI: How familiarity breeds both trust and contempt. *AI & Society*, *39*(4), 1721–1735. https://doi.org/10.1007/s00146-023-01666-5

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jang, Y., Choi, S., & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Education and Information Technologies, 27*(8), 11635–11667. https://doi.org/10.1007/s10639-022-11086-5

Jordan, P., & Spiess, M. (2019). Rethinking the interpretation of item discrimination and factor loadings. *Educational and Psychological Measurement*, *79*(6), 1103–1132. https://doi.org/10.1177/0013164419843164

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kamalov, F., Calonge, S., D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, *15*(16), 12451.

Karakuş, O., Yorulmazlar, M., Cetin, A., & Özsoy, D. (2023). Examining the attitudes of E-Sports players towards artificial intelligence techologies. *International Journal of Recreation and Sports Science*. https://doi.org/10.46463/ijrss.1361388

Kaya, F., Aydın, F., Schepman, A., Rodway, P., Yetişensoy, O., & Kaya, M. (2022). The roles of personality Traits, AI Anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human–Computer Interaction*, *40*, 497–514. https://doi.org/10.1080/10447318.2022.2151730

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review, 37*(2), 214–233. https://doi.org/10.1177/0894439317752406

Kline, R. B. (2014). Convergence of structural equation modeling and multilevel modeling. *The SAGE handbook of innovation in social research methods* (pp. 562–589). SAGE Publications Ltd.

Krägeloh, C. U., Melekhov, V., Alyami, M. M., & Medvedev, O. N. (2024). Artificial Intelligence Attitudes Inventory (AIAI): Development and validation using Rasch methodology. In Research Square. https://doi.org/10.21203/rs.3.rs-4403120/v1

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610.

Likert, R. (1932). A technique for measurement of attitudes. *Archives of Psychology*, *140*, 5–55.

Liu, M., & Reinders, H. (2025). Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. *System*, *128*(103544), 103544. https://doi.org/10.1016/j.system.2024.103544

Liu, Y., Park, Y., & Wang, H. (2025). The mediating effect of user satisfaction and the moderated mediating effect of AI anxiety on the relationship between perceived usefulness and subscription payment intention. *Journal of Retailing and Consumer Services*, *84*(104176), 104176. https://doi.org/10.1016/j.jretconser.2024.104176

Marengo, A., Karaoglan-Yilmaz, F. G., Yılmaz, R., & Ceylan, M. (2025). Development and validation of generative artificial intelligence attitude scale for students. *Frontiers in Computer Science, 7*(1528455). https://doi.org/10.3389/fcomp.2025.1528455

Owan, V., Abang, K., Idika, D., Etta, E., & Bassey, B. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*. https://doi.org/10.29333/ejmste/13428

Park, J., Woo, S. E., & Kim, J. (2024). Attitudes towards artificial intelligence at work: Scale development and validation. *Journal of Occupational and Organizational Psychology, 97*(3), 920–951. https://doi.org/10.1111/joop.12502

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. sage.
Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), Article 22.
Radulov, N. (2019). Artificial intelligence and security. Security 4.0. *Security & Future*, *3*(1), 3–5.
Ramazanoglu, M., & Akın, T. (2024). AI readiness scale for teachers: Development and validation. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13087-y
Raykov, T., Gabler, S., & Dimitrov, D. M. (2016). Maximal reliability and composite reliability: Examining their difference for multicomponent measuring instruments using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(3), 384–391. https://doi.org/10.1080/10705511.2014.966369
Revelle, W., & Revelle, M. W. (2015). Package 'psych.' *The Comprehensive R Archive Network, 337*(338), 161–165.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.
Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports*, *1*, 100014–100014. https://doi.org/10.1016/j.chbr.2020.100014
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.
Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003
Schumacker, R. E., & ve Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Psychology.
Shafique, R., Aljedaani, W., Rustam, F., Lee, E., Mehmood, A., & Choi, G. S. (2023). Role of artificial intelligence in online education: A systematic mapping study. *Ieee Access : Practical Innovations, Open Solutions*, *11*, 52570–52584.
Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics, 9*, 4–11. https://doi.org/10.12691/ajams-9-1-2
Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI - Künstliche Intelligenz, 35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0
Smedslund, G., Arnulf, J. K., & Smedslund, J. (2022). Is psychological science progressing? Explained variance in PsycInfo articles during the period 1956 to 2022. *Frontiers in Psychology, 13*, Article 1089089. https://doi.org/10.3389/fpsyg.2022.1089089
Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221115189
Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th edition). Northridge: Pearson.
Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIed for personalised learning pathways. *Electronic Journal of E-Learning, 20*(5), 639–653.
Tinsley, H. E. A., & Kass, R. A. (1979). The latent structure of the need satisfying properties of leisure activities. *Journal of Leisure Research, 11*(4), 278–291. https://doi.org/10.1080/00222216.1979.11969406
Tverskoi, D., Babu, S., & Gavrilets, S. (2022). The spread of technological innovations: Effects of psychology, culture and policy interventions. *Royal Society Open Science*. https://doi.org/10.1098/rsos.211833
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002
Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies, 29*(4), 4785–4808. https://doi.org/10.1007/s10639-023-12015-w
Wang, Y. Y., & Wang, Y. S. (2019). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments*, *30*(4), 619–634. https://doi.org/10.1080/10494820.2019.1674887

Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (Eds.). (2006). *Automated scoring of complex tasks in computer-based testing*. Lawrence Erlbaum Associates.

Winter, J. D., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*, 695–710.

Xu, W., Mahmud, R., & Hoo, W. (2024). A systematic literature review: Are automated essay scoring systems competent in Real-Life education scenarios? *Ieee Access : Practical Innovations, Open Solutions*, *12*, 77639–77657. https://doi.org/10.1109/ACCESS.2024.3399163

Yılmaz, D., Uzelli, D., & Dikmen, Y. (2025). Psychometrics of the attitude scale towards the use of artificial intelligence technologies in nursing. *BMC Nursing, 24*(1), Article 151. https://doi.org/10.1186/s12912-025-02732-7

Authors and Affiliations

Erdem Boduroglu[1] · **Mahmut Sami Yigiter**[2,3]

✉ Mahmut Sami Yigiter
M.S.Yigiter1@salford.ac.uk; mahmutsamiyigiter@gmail.com

Erdem Boduroglu
erdemboduroglu@gmail.com

1 Niğde Measurement and Evaluation Center, Ministry of National Education, Niğde, Türkiye

2 University of Salford, Post Doctoral Researcher, Salford, United Kingdom

3 Social Sciences University of Ankara, Lecturer, Ankara, Türkiye