Development and Validation of an Extended UTAUT-Based Instrument for Measuring Lecturers' Acceptance and Behavioral Intention Toward AI-Based Educational Assessment in Saudi Higher Education

Fahdah Mansour Alzeer[1], Prof. Madya Dr. Adibah Binti Abdul Latif[2]*

[1]PhD Candidate, Faculty of Educational Sciences and Technology, Universiti Technology Malaysia, Malaysia

[2]Faculty of Educational Sciences and Technology, Universiti Technology Malaysia, Malaysia

DOI: https://dx.doi.org/10.47772/IJRISS.2026.100500125

Received: 01 May 2026; Accepted: 06 May 2026; Published: 25 May 2026

ABSTRACT

The increasing integration of artificial intelligence (AI) into educational assessment has created a need for valid, context-sensitive instruments to measure lecturers' acceptance of and behavioural intention toward AI-based assessment practices. Although existing technology acceptance models provide useful theoretical foundations, many available instruments were developed for general educational technologies and may not adequately capture the professional, ethical, and psychological concerns associated with AI-based assessment. This study aimed to develop and validate an extended Unified Theory of Acceptance and Use of Technology (UTAUT)-based instrument for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment in Saudi higher education. The instrument was developed based on UTAUT and extended by incorporating three AI-related constructs: Perceived Trust in AI, Perceived Threat to Professional Autonomy, and Job Security Concern. A quantitative instrument-development design was employed. Items were generated through literature review, construct mapping, and expert validation. The instrument comprised sections on demographic profile, technology acceptance, extended AI-related constructs, and behavioural intention. Rasch measurement analysis was used to examine item fit, person and item reliability, separation indices, dimensionality, rating scale functioning, and item hierarchy. The preliminary findings indicated that the instrument demonstrated acceptable psychometric properties, with most items fitting the Rasch model and representing the intended constructs. The validated instrument is expected to provide researchers, university leaders, and policymakers with a diagnostic tool for assessing lecturers' readiness to adopt AI-based educational assessment. This study contributes to educational measurement by providing a theoretically grounded, contextually relevant instrument for future research on AI adoption in higher-education assessment.

Keywords: AI-based educational assessment; behavioural intention; instrument validation; Rasch measurement model; UTAUT

INTRODUCTION

Artificial intelligence (AI) is increasingly reshaping the landscape of higher education by transforming how teaching, learning, feedback, and assessment are designed and implemented. In educational assessment, AI-based systems offer new possibilities for automated scoring, adaptive testing, intelligent feedback, learning analytics, and data-informed decision-making (Chen et al., 2020; Holmes et al., 2019; Zawacki-Richter et al., 2019). These developments have created opportunities to enhance the efficiency, consistency, scalability, and personalization of assessment. In higher education, where assessment serves not only as a mechanism for grading but also as a means of supporting learning, quality assurance, and academic decision-making, AI-based educational assessment has the potential to support more responsive and evidence-based evaluation practices (Crompton & Burke, 2023; Miao et al., 2021; Wang et al., 2024).

Despite these opportunities, the adoption of AI-based educational assessment is not merely a technical matter. Its successful implementation depends heavily on the acceptance, readiness, and behavioural intention of

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

lecturers who are directly responsible for designing assessment tasks, interpreting student performance, making evaluative judgments, and ensuring the validity and fairness of assessment decisions. Previous studies on educational technology adoption have consistently shown that users' beliefs, perceived usefulness, ease of use, institutional support, and intention to use technology are central to successful implementation (Davis, 1989; Teo, 2011; Venkatesh et al., 2003). Unlike general digital tools used for communication or content delivery, AI-based assessment systems intervene in one of the most sensitive areas of academic practice: the evaluation of student learning. As such, lecturers' acceptance of AI-based assessment may be shaped not only by perceived usefulness and ease of use, but also by deeper professional concerns related to trust, transparency, autonomy, fairness, and job relevance (European Commission, 2022; Holmes et al., 2019; Kizilcec & Lee, 2022; Miao & Holmes, 2023; Shin, 2021; U.S. Department of Education, 2023).

Technology acceptance models, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT), have been widely used to explain users' intentions to adopt technological innovations in educational settings. TAM proposes that perceived usefulness and perceived ease of use influence users' attitudes and behavioural intention toward technology (Davis, 1989). UTAUT further extends this understanding by identifying performance expectancy, effort expectancy, social influence, and facilitating conditions as key determinants of behavioural intention and technology use (Venkatesh et al., 2003; Venkatesh et al., 2016). While these constructs remain useful for understanding general technology adoption, AI-based educational assessment presents additional complexities that may not be fully captured by traditional acceptance frameworks.

This issue is particularly relevant in the context of Saudi higher education, where national digital transformation agendas have accelerated the integration of emerging technologies into university systems. Saudi Arabia's educational reforms and AI-related strategic initiatives have encouraged higher education institutions to modernize teaching, learning, and assessment practices. However, policy support and technological infrastructure alone do not guarantee meaningful adoption among lecturers. Research on technology implementation in higher education indicates that adoption is more likely to occur when users perceive the technology as useful, manageable, institutionally supported, and aligned with their professional responsibilities (Teo, 2011; Venkatesh et al., 2016). If lecturers are not adequately prepared, convinced, or supported, AI-based assessment tools may remain underutilized or may be implemented in ways that do not fully align with sound educational measurement principles.

A major challenge in this area is the limited availability of instruments specifically designed and validated to measure lecturers' acceptance of and behavioural intention toward AI-based educational assessment. Many existing instruments are adapted from general technology acceptance studies and were originally developed for learning management systems, e-learning platforms, or other broad digital technologies (Davis, 1989; Teo, 2011). Although such instruments offer a useful foundation, they may not adequately address the specific pedagogical, ethical, and professional concerns associated with AI in assessment. In particular, constructs such as perceived trust in AI, perceived threat to professional autonomy, and job security concern are highly relevant when AI is used to support or automate assessment-related decisions (Nazaretsky et al., 2022; Shin, 2021). Without incorporating these constructs, measurement instruments may provide an incomplete representation of lecturers' acceptance and readiness.

From the perspective of educational measurement and evaluation, the development of a valid and reliable instrument is essential before meaningful conclusions can be drawn about lecturers' acceptance of AI-based assessment. Instrument validity is not established simply by adapting items from previous studies; rather, it requires systematic construct definition, item development, expert review, empirical testing, and psychometric validation (Boateng et al., 2018; DeVellis & Thorpe, 2021). In this regard, the Rasch measurement model offers a rigorous approach for examining the quality of survey instruments. Through Rasch analysis, researchers can evaluate item fit, person and item reliability, separation indices, dimensionality, rating scale functioning, and item hierarchy (Bond et al., 2020; Boone et al., 2014).

Accordingly, this study focuses on the development and validation of an extended UTAUT-based instrument for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment in Saudi higher education. The proposed instrument is grounded in established technology acceptance theory and extended to include AI-specific concerns that are particularly relevant to assessment contexts. The instrument includes core constructs of technology acceptance, namely performance expectancy, effort expectancy, social influence, and facilitating conditions, as well as extended constructs comprising perceived trust in AI, perceived threat to professional autonomy, and job security concern. Behavioural intention is also measured to capture lecturers' willingness to use, continue using, and recommend AI-based assessment tools.

This study contributes to the field in three main ways. First, it provides a theoretically grounded instrument tailored to the specific context of AI-based educational assessment rather than general educational technology use. Second, it extends existing technology acceptance models by incorporating constructs that reflect lecturers' professional and ethical concerns toward AI-mediated assessment. Third, it applies Rasch measurement analysis to provide psychometric evidence of the instrument's validity and reliability. The resulting instrument is expected to support future research, institutional diagnostics, professional development planning, and policy evaluation related to AI adoption in higher education assessment.

Research Objectives

1. To develop an extended UTAUT-based instrument for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment.
2. To establish the content validity of the instrument through expert review.
3. To examine the psychometric properties of the instrument using Rasch measurement analysis.

Research Questions

1. What constructs and items constitute the extended UTAUT-based instrument for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment?
2. To what extent does the instrument demonstrate content validity?
3. What are the psychometric properties of the instrument based on Rasch measurement analysis?

LITERATURE REVIEW

AI-Based Educational Assessment in Higher Education

Artificial intelligence has increasingly become an important component of educational innovation, particularly in higher education, where institutions are expected to improve teaching quality, assessment efficiency, and evidence-based decision-making. In the context of educational assessment, AI-based systems can support automated scoring, adaptive testing, intelligent feedback, predictive analytics, and learner performance monitoring (Chen et al., 2020; Holmes et al., 2019; Huggins-Manley et al., 2022; Ouyang et al., 2023; Zawacki-Richter et al., 2019). These systems are often viewed as having the potential to improve the speed, consistency, and scalability of assessment, especially in large classes or technology-enhanced learning environments.

However, the use of AI in assessment also raises important concerns related to fairness, validity, transparency, accountability, and professional judgment (Huggins-Manley et al., 2022; Kizilcec & Lee, 2022). Assessment is a core academic responsibility, and decisions derived from assessment outcomes may influence student grades, progression, feedback, and institutional quality assurance. Therefore, the use of AI in assessment cannot be treated as a purely technical innovation. It requires careful attention to educational measurement principles, ethical decision-making, and lecturers' roles as professional evaluators (Holmes et al., 2019; Miao et al., 2021).

Technology Acceptance and Behavioural Intention

Technology acceptance research provides a useful foundation for understanding why individuals adopt or reject new technologies. The Technology Acceptance Model explains technology use through perceived usefulness and perceived ease of use, which influence users' attitudes and behavioural intention (Davis, 1989). In educational contexts, these constructs have been widely applied to examine teachers' and lecturers' adoption of digital learning tools (Acosta-Enriquez et al., 2024; Teo, 2011).

The Unified Theory of Acceptance and Use of Technology further expands this perspective by identifying four major determinants of behavioural intention and technology use: performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2003). These four constructs are highly relevant to AI-based educational assessment. Lecturers may be more willing to use AI-based assessment tools if they believe that these tools can improve assessment efficiency, enhance feedback quality, reduce workload, and support better evaluation of student performance.

Extending UTAUT for AI-Based Educational Assessment

Although UTAUT provides a strong foundation for measuring technology acceptance, AI-based educational assessment introduces additional concerns that may not be fully captured by traditional technology acceptance constructs. AI systems differ from many other educational technologies because they may influence evaluative decisions, automate feedback, generate predictions, or support grading-related processes. These functions are directly linked to lecturers' professional authority and students' academic outcomes.

Perceived trust in AI is an important construct because lecturers may hesitate to rely on AI-generated scores, feedback, or analytics if they are uncertain about the system's accuracy, fairness, or transparency (Nazaretsky et al., 2022). Trust is particularly important in assessment because lecturers must be able to justify assessment decisions and ensure that evaluation processes are fair to students (Nazaretsky et al., 2022; Shin, 2021). The perceived threat to professional autonomy is also relevant because assessment is closely tied to academic judgment and professional discretion. Job security concerns are another important AI-related construct, as lecturers may worry that automation could devalue their professional roles (Kasneci et al., 2023; Miao & Holmes, 2023).

Instrument Development in Educational Measurement

Instrument development is a systematic process that requires clear construct definition, item generation, content validation, pilot testing, and psychometric analysis. In educational and social science research, a questionnaire should not be considered valid merely because it is adapted from previous studies. Instead, evidence must be gathered to demonstrate that the items represent the intended constructs and function appropriately with the target population (Boateng et al., 2018; DeVellis & Thorpe, 2021).

For the present study, instrument development is particularly important because lecturers' acceptance of AI-based educational assessment is a developing area of research. Existing technology acceptance instruments may not sufficiently reflect the context of AI-mediated assessment or the professional concerns of lecturers in Saudi higher education. Therefore, the development of a new or extended instrument requires careful construct mapping and empirical validation.

Rasch Measurement Model for Instrument Validation

The Rasch measurement model provides a rigorous psychometric approach for validating instruments in educational measurement and social science research. Unlike classical test theory, which focuses mainly on total scores and internal consistency, Rasch analysis examines the interaction between person measures and item difficulty on a common measurement scale (Bond et al., 2020; Boone et al., 2014). This allows researchers to evaluate whether items function consistently and contribute meaningfully to measuring the intended construct.

Several Rasch indicators are relevant for instrument validation. Item fit statistics provide evidence of whether each item fits the expected measurement pattern. Person and item reliability indicate the consistency of person responses and item ordering, while separation indices show how well the instrument distinguishes between different levels of respondents and item difficulty. Dimensionality analysis helps determine whether the instrument measures a coherent construct, while rating scale analysis evaluates whether response categories function as intended.

Conceptual Basis of the Present Instrument

Based on the reviewed literature, the instrument developed in this study is grounded in an extended UTAUT framework. The core technology acceptance constructs are performance expectancy, effort expectancy, social influence, and facilitating conditions. The extended AI-related constructs are perceived trust in AI, perceived threat to professional autonomy, and job security concern. Behavioural intention is included as the outcome construct and refers to lecturers' willingness to use, continue using, and recommend AI-based assessment tools.

Therefore, the proposed instrument is designed to measure both general technology acceptance and AI-specific professional concerns within the context of educational assessment. By validating this instrument using Rasch analysis, the study aims to provide a psychometrically sound tool for assessing lecturers' readiness and behavioural intention toward AI-based educational assessment in Saudi higher education.

METHODOLOGY

Research Design

This study employed a quantitative instrument development design to develop and validate an extended UTAUT-based instrument to measure lecturers' technology acceptance and behavioural intention toward AI-based educational assessment in Saudi higher education. This design was considered appropriate because the study's main purpose was to construct, refine, and validate a measurement instrument rather than to test causal effects or intervention outcomes. The study adopted a cross-sectional survey approach, in which data were collected from lecturers at a single point in time to examine the instrument's psychometric functioning.

The instrument-development process was guided by established principles of scale development and educational measurement, including construct identification, item generation, expert review, pilot testing, and psychometric validation (Boateng et al., 2018; DeVellis & Thorpe, 2021). The Rasch measurement model was used as the main validation approach because it enables detailed item-level analysis, including item fit, reliability, separation, dimensionality, and rating scale functioning (Bond et al., 2020; Boone et al., 2014).

Instrument Development Procedure

The instrument was developed through four main stages: construct identification, item generation, content validation, and empirical validation (Huggins-Manley et al., 2022). In the first stage, the core constructs were identified based on the Unified Theory of Acceptance and Use of Technology. These constructs comprised performance expectancy, effort expectancy, social influence, and facilitating conditions. These dimensions were selected because they reflect lecturers' perceptions of usefulness, ease of use, social encouragement, and institutional support regarding the adoption of AI-based educational assessment.

In the second stage, the instrument was extended by incorporating three AI-specific constructs: perceived trust in AI, perceived threat to professional autonomy, and job security concern. These constructs were included because AI-based educational assessment involves more complex professional and ethical considerations than general educational technologies. In the third stage, behavioural intention was included as the outcome construct, referring to lecturers' willingness to use, continue using, and recommend AI-based educational assessment tools.

Instrument Structure

The initial version of the instrument consisted of 100 items distributed across 10 constructs. The technology acceptance section consisted of four constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions. The extended AI-related section consisted of perceived trust in AI, perceived threat to professional autonomy, and job security concern. The behavioural intention section consisted of intention to use, intention to continue use, and intention to recommend or advocate the use of AI-based educational assessment tools.

All items were written specifically for AI-based educational assessment, including automated scoring, intelligent feedback systems, adaptive testing platforms, and analytics dashboards. The items were measured using a four-point Likert scale ranging from 1 = Strongly Disagree to 4 = Strongly Agree. A four-point scale was used to

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

encourage respondents to express a clear level of agreement or disagreement without relying on a neutral midpoint.

Table 1. Instrument Structure

Section	Construct	Number of Items	Description
A	Demographic Information	9	Respondent profile, academic background, AI training, AI use, and institutional support
B	Performance Expectancy	10	Perceived usefulness of AI-based assessment in improving quality, efficiency, accuracy, and feedback
B	Effort Expectancy	10	Perceived ease of learning, operating, and integrating AI-based assessment tools
B	Social Influence	10	Perceived encouragement and support from colleagues, leaders, and the institution
B	Facilitating Conditions	10	Perceived availability of infrastructure, training, guidelines, and technical support
C	Perceived Trust in AI	10	Confidence in the fairness, accuracy, transparency, and ethical use of AI assessment
C	Perceived Threat to Professional Autonomy	10	Concern that AI may reduce lecturers' judgment, control, and academic freedom
C	Job Security Concern	10	Concern that AI may reduce the relevance or security of academic roles
D	Intention to Use	10	Willingness to begin using AI-based assessment tools
D	Intention to Continue Use	10	Motivation to sustain the use of AI-based assessment tools
D	Intention to Recommend or Advocate	10	Willingness to recommend or support AI-based assessment adoption
	Total	100	

Content Validation

Content validation was conducted to ensure that the instrument items were relevant, clear, representative, and appropriate for the intended constructs. The initial instrument was reviewed by five experts in educational measurement, educational technology, assessment, higher education, and AI in education. The experts were selected based on their academic expertise, research experience, and familiarity with instrument development or technology-enhanced assessment.

The expert review focused on four criteria: construct relevance, clarity of item wording, contextual appropriateness, and alignment with AI-based educational assessment. Items that were identified as unclear, repetitive, ambiguous, or weakly aligned with the construct were revised based on expert feedback. The content validation process resulted in the revision of 13 items, but no item was omitted. A total of 100 items were retained for pilot testing. The item-level Content Validity Index ranged from 0.80 to 1.00, while the scale-level Content Validity Index based on average agreement was S-CVI/Ave = 0.91, indicating an acceptable level of content validity for the instrument.

Population and Sampling

The target population of this study consisted of lecturers in Saudi higher education institutions. Lecturers were selected as the target respondents because they are directly involved in designing, implementing, interpreting, and making decisions based on assessment. Their acceptance of AI-based educational assessment is therefore critical to the meaningful and responsible integration of AI into higher education assessment practices.

For the present instrument validation study, the sample consisted of 70 lecturers from three Saudi higher education institutions. The respondents represented different academic disciplines, academic ranks, and years of teaching experience. The study employed stratified sampling to ensure that the respondents reflected the characteristics of lecturers in Saudi higher education.

As this study represents an initial instrument development and pilot validation phase, the final sample of 63 valid respondents was used to examine the preliminary item functioning through Rasch analysis.

Data Collection Procedure

Data were collected using a structured questionnaire administered to lecturers in Saudi higher education institutions. The questionnaire was distributed via email using Google Forms. Respondents were provided with information about the purpose of the study, the voluntary nature of participation, and the confidentiality of their responses.

Prior to data analysis, the data were screened for completeness, missing values, and response irregularities. Responses with substantial missing data or patterned responses were excluded from the final analysis. After data cleaning, a total of 63 valid responses were retained for Rasch measurement analysis.

Data Analysis

Data analysis was conducted using the Rasch measurement model. The Rasch model was selected because it provides a rigorous method for examining the psychometric quality of survey instruments in educational measurement. Unlike classical test theory, Rasch analysis places item difficulty and respondent agreement on the same logit scale, allowing a more precise evaluation of how well the items measure the intended constructs (Bond et al., 2020; Boone et al., 2014).

The analysis examined several indicators of instrument quality. First, item fit was evaluated using infit and outfit mean square statistics. For this study, the acceptable range for infit and outfit mean square values was set at 0.50 to 1.50. Point-measure correlation values were also examined to ensure that each item contributed positively to the overall construct measurement.

Second, person and item reliability and separation indices were examined. Third, dimensionality was examined using principal component analysis of residuals. Finally, rating scale functioning was evaluated to determine whether the four response categories operated as intended. The Rasch analysis was conducted using Winsteps version 5.10.1.0

Ethical Considerations

Ethical considerations were observed throughout the study. Respondents were informed that their participation was voluntary and that they could withdraw from the study at any point without penalty. They were also informed that their responses would remain anonymous and confidential. No personally identifying information was required, except for optional institutional information. The data were analyzed and reported in aggregate form to protect respondents' privacy.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

RESULTS

Respondent Profile

A total of 70 lecturers from Saudi higher education institutions participated in the study. After data screening, 63 valid responses were retained for Rasch measurement analysis. The respondents represented diverse demographic backgrounds across gender, age, academic rank, teaching experience, academic discipline, prior AI-related training, and prior use of AI-based tools.

Table 2. Demographic Profile of Respondents

Demographic Variable	Category	Frequency	Percentage
Gender	Male	31	49.2
	Female	32	50.8
Age	Below 30 years	08	12.7
	31-40 years	24	38.1
	41-50 years	21	33.3
	Above 50 years	10	15.9
Academic Rank	Lecturer	18	28.6
	Assistant Professor	24	38.1
	Associate Professor	14	22.2
	Professor	07	11.1
Teaching Experience	Less than 5 years	10	15.9
	5-10 years	23	36.5
	11-15 years	18	28.6
	More than 15 years	12	19.0
Academic Discipline	Science Social	24	38.1
	Science Technology	17	27.0
	Engineering	22	34.9
AI Training	Yes	27	42.9
	No	36	57.1
Previous Use of AI Tools	Yes	34	54.0
	No	29	46.0

Content Validation Results

The initial instrument was reviewed by five experts in educational measurement, educational technology, assessment, AI in education, and higher education.

Based on expert feedback, 13 items were revised for clarity; no items were merged due to overlap, and no items were removed. The item-level content validity index ranged from [0.8] to [1.0], while the scale-level content validity index was 0.91.

Table 3. Summary of Content Validation Results

Validation Indicator	Result
Number of experts	5
Initial number of items	100
Items revised	13
Items removed	0
Items retained for pilot testing	100
I-CVI range	0.8-1.00
S-CVI/Ave	0.91

Summary of Rasch Model Fit

Rasch measurement analysis was conducted to examine the instrument's psychometric properties. The person's reliability was [0.87], while the item's reliability was [0.95]. The person separation index was [2.59], indicating that the instrument was able to distinguish respondents into approximately [4] levels of acceptance and behavioural intention. The item separation index was [4.36], indicating that the items were sufficiently distributed across endorsement difficulty levels.

Table 4. Rasch Summary Statistics

Rasch Indicator	Person	Item
Reliability	0.87	0.95
Separation Index	2.59	4.36
Mean Measure	0.72	0.00
Standard Deviation	1.18	0.91
Cronbach Alpha/KR-20	0.87	-

Item Fit Analysis

Item fit was examined using infit and outfit mean square statistics. Items with infit and outfit mean square values within the range of [0.50 to 1.50] were considered acceptable.

The results showed that the overall item fit was acceptable, although three items were flagged for closer inspection and were further examined.

Table 5. Summary of Item Fit Statistics

Construct	No of Item	Infit MNSQ Range	Outfit MNSQ Range	PTMEA Corr.	Misfitting Items
Performance Expectancy	10	0.82-1.28	0.79-1.34	0.39-0.63	None
Effort Expectancy	10	0.76-1.31	0.81-1.39	0.36-0.61	None
Social Influence	10	0.84-1.42	0.86-1.45	0.34-0.58	S15 Flagged
Facilitating Conditions	10	0.79-1.36	0.83-1.41	0.37-0.60	Retained

Perceived Trust in AI	10	0.88-1.33	0.85-1.38	0.40-0.64	Retained
Threat to Autonomy	10	0.91-1.46	0.94-1.48	0.32-0.55	RTA8 Flagged
Job Security Concern	10	0.87-1.44	0.90-1.47	0.33-0.56	JSC10 Flagged
Intention to Use	10	0.74-1.25	0.78-1.31	0.42-0.66	Retained
Intention to Continue Use	10	0.81-1.37	0.84-1.43	0.35-0.59	Retained
Intention to Recommend	10	0.85-1.34	0.88-1.40	0.37-0.62	Retained

Although three items were flagged for closer inspection, namely S15 under Social Influence, RTA8 under Perceived Threat to Professional Autonomy, and JSC10 under Job Security Concern, these items were retained after substantive and psychometric review. The decision to retain these items was based on four considerations. First, the flagged items remained within or close to the acceptable Rasch fit range adopted in this study, with construct-level infit and outfit mean square values generally falling between 0.50 and 1.50. Second, all three items showed positive point-measure correlation values, indicating that they contributed in the expected direction to the measurement of the intended construct. Third, the items were theoretically important because they represented context-sensitive aspects of AI-based educational assessment, particularly institutional influence, professional autonomy, and job security concerns. Removing these items at the pilot validation stage could reduce the content coverage of the extended UTAUT framework. Fourth, the overall instrument demonstrated strong item reliability, acceptable person reliability, satisfactory separation indices, and acceptable evidence of dimensionality. Therefore, the flagged items were not removed but were retained with minor wording revisions for further validation in a larger sample.

Dimensionality Analysis

Dimensionality was examined using principal component analysis of Rasch residuals. The raw variance explained by the measures was 46.8%, considered acceptable given the recommended threshold of at least 40%. The unexplained variance in the first contrast was 2.7 eigenvalue units, while the unexplained variance in the second contrast was 2.1 eigenvalue units, suggesting that there was no substantial evidence of a dominant secondary dimension. Therefore, the dimensionality results indicate that the instrument demonstrated acceptable construct validity for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment.

Table 6. Dimensionality Analysis Based on Rasch Residuals

Indicator	**Value**
Raw variance explained by measures	46.8%
Raw unexplained variance	8.5%
Unexplained variance in the first contrast	2.7 (eigenvalue)
Unexplained variance in second contrast	2.1 e(igenvalue)
Interpretation	Acceptable

Rating Scale Functioning

The functioning of the four-point Likert scale was examined to determine whether respondents used the response categories meaningfully. As shown in Table 7, all four categories were used by respondents, with the highest proportion of responses recorded for Category 3, Agree (43.3%), followed by Category 4, Strongly Agree (26.4%).

The average measures increased monotonically from Category 1 (-1.48 logits) to Category 4 (1.72 logits), indicating that higher response categories represented higher levels of agreement. The outfit MNSQ values ranged from 0.88 to 1.12, suggesting acceptable category fit. The thresholds were ordered from -1.12 to 0.94, indicating that the four-point response scale functioned as intended.

Table 7. Rating Scale Functioning

Category	**Label**	**Frequency**	**Percentage**	**Average Measure**	**Outfit MNSQ**	**Threshold**
1	Strongly Disagree	642	10.2	-1.48	1.12	-
2	Disagree	1,268	20.1	-0.54	0.96	-1.12
3	Agree	2,726	43.3	0.68	0.88	0.18
4	Strongly Agree	1,664	26.4	1.72	1.04	0.94

Summary of Instrument Validation

Overall, the Rasch analysis provided evidence that the extended UTAUT-based instrument demonstrated acceptable psychometric properties. Based on the validation results, all 100 items were retained at this developmental validation stage, while three flagged items were revised for clarity. The decision to retain all items was made to preserve the theoretical breadth and content coverage of the extended UTAUT framework, particularly because the instrument was designed to measure both general technology acceptance constructs and AI-specific professional concerns. Nevertheless, the findings indicate that future validation with a larger sample should examine the possibility of item reduction to improve instrument efficiency, reduce respondent burden, and develop a more parsimonious version of the scale without weakening construct validity.

DISCUSSION

The purpose of this study was to develop and validate an extended UTAUT-based instrument for measuring lecturers' technology acceptance and behavioural intention toward AI-based educational assessment in Saudi higher education. The instrument was developed to address a measurement gap in the literature, particularly the limited availability of context-specific tools that measure both general technology acceptance factors and AI-specific professional concerns. Overall, the Rasch analysis provided evidence that the instrument demonstrated acceptable psychometric properties across item fit, reliability, separation, dimensionality, rating scale functioning, and item hierarchy. These findings support the argument that instrument development should be conducted through systematic construct definition, expert validation, and empirical testing rather than relying only on direct adaptation of existing scales (Boateng et al., 2018; DeVellis & Thorpe, 2021).

The results indicate that most items functioned well in measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment. Most items were found to fit the Rasch model within the acceptable infit and outfit mean square range of 0.50-1.50. This suggests that the items contributed meaningfully to measuring the intended constructs and that respondents interpreted them in a relatively consistent manner. In Rasch-based validation, item fit statistics are important because they indicate whether each item behaves as expected within the measurement model (Bond et al., 2020; Boone et al., 2014).

The retention of all 100 items should be interpreted in light of the study's developmental purpose. As this study aimed to develop and validate an extended UTAUT-based instrument rather than to produce a shortened final scale, item retention was guided not only by statistical fit but also by theoretical representativeness and content validity. The three flagged items, S15, RTA8, and JSC10, were examined carefully because they represented constructs central to the adoption of AI-based educational assessment. S15 reflects the role of social influence in the adoption of institutional technology, while RTA8 and JSC10 represent lecturers' concerns about professional autonomy and job security. These concerns are particularly important when AI systems are used in assessment-related decision-making. Although these items showed potential misfit, their fit values were not sufficiently severe to justify deletion at this stage, and their positive point-measure correlations suggested that they continued to function in the intended measurement direction. Therefore, the items were retained but identified for further review, rewording, and revalidation in subsequent studies.

Although all 100 items were retained in the present validation stage, the findings also suggest that future refinement of the instrument may improve its efficiency without compromising construct validity. A 100-item instrument is useful at the developmental stage because it provides broad construct coverage across technology acceptance, AI-specific professional concerns, and behavioural intention. However, for practical use in large-scale institutional diagnostics, a shorter version of the instrument may be more feasible and may reduce respondent fatigue. Therefore, item reduction may be considered in future studies after further evidence is obtained from a larger and more diverse sample. Items may be considered for revision or removal if they consistently demonstrate misfit, low point-measure correlation, redundancy with other items, or weak contribution to construct representation. In this study, item reduction was not conducted immediately because the three flagged items represented theoretically important constructs and did not show severe psychometric weakness. Instead, they were retained with wording refinement and marked for further empirical examination.

A closer construct-level interpretation provides more specific insights into the instrument's measurement functioning. Among the ten constructs, the behavioural intention construct of Intention to Use demonstrated the strongest measurement performance, with infit values ranging from 0.74 to 1.25, outfit values ranging from 0.78 to 1.31, and point-measure correlations ranging from 0.42 to 0.66. These results suggest that the items measuring lecturers' willingness to begin using AI-based assessment tools were relatively coherent, well-targeted, and consistently interpreted by respondents. This may indicate that intention to use is a more direct and familiar construct for lecturers compared with more complex AI-related concerns. Similarly, Perceived Trust in AI showed strong functioning, with positive point-measure correlations ranging from 0.40 to 0.64 and no serious misfitting items. This finding underscores the importance of trust as a measurable construct in AI-based educational assessment, particularly since lecturers must have confidence in the fairness, accuracy, and transparency of AI-generated assessment outcomes.

The core UTAUT constructs also demonstrated acceptable measurement properties. Performance Expectancy showed stable item functioning, with infit values ranging from 0.82 to 1.28, outfit values ranging from 0.79 to 1.34, and point-measure correlations ranging from 0.39 to 0.63. This suggests that lecturers were able to interpret items related to the usefulness of AI-based assessment in a relatively consistent manner. Effort Expectancy also performed acceptably, with infit values ranging from 0.76 to 1.31 and outfit values ranging from 0.81 to 1.39. These results indicate that items related to ease of use, learnability, and integration of AI-based assessment tools were generally functioning as intended. Facilitating Conditions also showed acceptable fit, suggesting that items related to infrastructure, training, technical support, and institutional readiness were meaningful indicators of lecturers' perceived support for AI adoption.

In contrast, several constructs appeared to require closer attention in future validation. Social Influence, Threat to Autonomy, and Job Security Concern each contained one flagged item, namely S15, RTA8, and JSC10. Although the overall fit ranges for these constructs remained within the acceptable Rasch criterion of 0.50 to 1.50, the flagged items suggest that respondents may have interpreted some items less consistently. For Social Influence, the flagged item may reflect variation in how lecturers perceive institutional pressure, peer expectations, or leadership encouragement to use AI-based assessment. In higher education contexts, especially where technology adoption may be shaped by both formal policy and informal academic culture, social influence may not operate uniformly across all lecturers. For Threat to Autonomy, the flagged item suggests that concerns about loss of professional judgment may be more complex and sensitive than general technology acceptance beliefs. Similarly, the flagged item in Job Security Concern indicates that lecturers may differ in how they interpret the relationship between AI automation and the security of their professional roles. These findings suggest that AI-specific professional concerns are measurable, but may require more careful wording, contextualization, and construct-level refinement in future studies.

These construct-level findings strengthen the theoretical argument for extending UTAUT in the context of AI-based educational assessment. The stronger performance of Intention to Use, Performance Expectancy, Effort Expectancy, Facilitating Conditions, and Perceived Trust in AI indicates that lecturers were able to respond consistently to items related to usefulness, ease of use, institutional support, trust, and willingness to adopt AI-based assessment. However, the relatively weaker functioning of selected items within Social Influence, Threat to Autonomy, and Job Security Concern suggests that AI adoption in assessment cannot be understood only through conventional technology acceptance constructs. Instead, lecturers' responses appear to be shaped by professional identity, perceived academic control, institutional culture, and uncertainty about the future role of educators in AI-mediated assessment. This supports the inclusion of AI-specific constructs in the instrument while also indicating that these constructs should be further refined through larger-scale validation and possibly qualitative follow-up.

The item reliability value of 0.95 indicates that the hierarchy of item difficulty was stable and likely to be replicable across similar samples. A high item reliability value suggests that the items were sufficiently spread across different levels of endorsement difficulty, supporting the instrument's usefulness for identifying both readily accepted and more sensitive aspects of AI-based educational assessment. The person reliability value of 0.87 and person separation index of 2.59 indicate that the instrument was able to distinguish respondents according to their level of acceptance and behavioural intention. This is valuable for higher education institutions because lecturer acceptance is a key factor in determining whether educational technologies are meaningfully adopted and sustained in practice (Teo, 2011; Venkatesh et al., 2003).

The sample size of 63 valid respondents should be interpreted in relation to the developmental nature of this validation study. Although the sample was adequate for preliminary Rasch-based item calibration and provided useful evidence on item fit, reliability, separation, dimensionality, and rating scale functioning, the ratio of respondents to items remains a methodological consideration. The instrument consisted of 100 items across 10 constructs, which allowed broad theoretical and content coverage at the early stage of instrument development. However, a larger sample would provide more stable item estimates, stronger evidence of construct functioning across respondent subgroups, and greater confidence in the generalizability of the item hierarchy. Therefore, the present findings should be regarded as preliminary psychometric evidence rather than final validation evidence. Further validation with a larger, more diverse sample of lecturers is recommended before the instrument is used for large-scale institutional diagnosis or cross-context comparisons.

The dimensionality analysis also provided evidence of construct validity. The raw variance explained by measures was 46.8%, while the unexplained variance in the first contrast was 2.7 eigenvalue units. These findings suggest that the instrument measured a coherent underlying construct or a set of closely related constructs associated with lecturers' acceptance and behavioural intention toward AI-based educational assessment. Establishing dimensionality is essential in scale validation because it supports the interpretation that the instrument measures the intended construct structure (Bond et al., 2020; Boone et al., 2014).

The findings support the decision to extend the UTAUT framework by adding AI-specific constructs (Acosta-Enriquez et al., 2024; Nazaretsky et al., 2022). Traditional UTAUT constructs such as performance expectancy, effort expectancy, social influence, and facilitating conditions are important for measuring general technology acceptance (Venkatesh et al., 2003, 2016). However, AI-based educational assessment requires additional attention to constructs that reflect the professional and ethical implications of AI use. The inclusion of perceived trust in AI, perceived threat to professional autonomy, and job security concern strengthens the instrument because these constructs capture concerns that are especially relevant when AI is used in assessment-related decision-making.

From a methodological perspective, the study demonstrates the value of Rasch analysis in instrument validation. While classical reliability indices such as Cronbach's alpha can provide general evidence of internal consistency, Rasch analysis offers deeper diagnostic information about item functioning, category structure, construct representation, and person-item targeting. This is especially important in emerging research areas such as AI-based educational assessment, where constructs are still developing and require careful empirical validation (Bond et al., 2020; Boone et al., 2014).

The findings also have practical implications for Saudi higher education institutions. A validated instrument can support institutional needs analysis by identifying lecturers' readiness, confidence, and concerns regarding AI-based assessment (OECD, 2023; U.S. Department of Education, 2023). It can also guide professional development programmes by showing whether lecturers need support in understanding AI usefulness, developing technical skills, building trust in AI systems, or managing concerns related to autonomy and job security. This is aligned with technology acceptance research, which shows that facilitating conditions and institutional support play an important role in encouraging users' intention to adopt technology (Venkatesh et al., 2003, 2016).

Nevertheless, several limitations should be acknowledged. First, the study used a cross-sectional survey design, capturing lecturers' perceptions at a single point in time. Acceptance of AI-based assessment may change over time as lecturers gain more exposure to, training in, and experience with AI tools. Second, the validation was conducted within the Saudi higher education context; therefore, further validation is needed before the instrument is applied in other national or cultural settings. Third, the study relied on self-reported perceptions, which may be influenced by respondents' familiarity with AI, institutional expectations, or personal attitudes toward digital transformation. Future studies may combine survey data with qualitative interviews or actual usage data to provide a more comprehensive understanding of AI acceptance in assessment.

Another limitation is the sample size, particularly in relation to the instrument's length. Although 63 valid responses were sufficient to provide preliminary Rasch-based evidence for item functioning, the instrument contained 100 items, which may limit the stability and generalizability of some item estimates.

The relatively small sample also limits the extent to which construct-level comparisons, subgroup differences, and item hierarchy can be interpreted with strong confidence. Future studies should therefore validate the instrument using a larger sample of lecturers across different Saudi higher education institutions. A larger sample would also allow more advanced validation procedures, including confirmatory factor analysis, measurement invariance testing, and item reduction.

Future refinement of the instrument should be conducted through several systematic steps. First, the instrument should be validated using a larger, more diverse sample to strengthen the stability and generalizability of the Rasch findings. Second, the three flagged items, S15, RTA8, and JSC10, should be reviewed and reworded based on expert feedback and respondent interpretation. Third, confirmatory factor analysis should be conducted to examine the construct structure of the extended UTAUT model.

Fourth, item reduction should be considered by removing or revising items that repeatedly exhibit misfit, redundancy, or a weak contribution to the construct representation. Finally, measurement invariance analysis should be conducted to determine whether the instrument functions consistently across lecturer subgroups. These steps would support the development of a more concise and robust version of the instrument for future large-scale applications.

CONCLUSION

This study developed and validated an extended UTAUT-based instrument for measuring lecturers' technology acceptance and behavioural intention toward AI-based educational assessment in Saudi higher education. The instrument was designed to address the need for a context-specific measurement tool that captures not only general technology acceptance factors, but also AI-specific professional concerns related to trust, professional autonomy, and job security.

The findings from the Rasch measurement analysis provided evidence that the instrument demonstrated acceptable psychometric properties. The results showed satisfactory item fit, person and item reliability, separation indices, dimensionality, rating scale functioning, and item hierarchy. All items were retained at this developmental validation stage, indicating that the instrument items were generally relevant, coherent, and suitable for preliminary measurement of lecturers' acceptance and behavioural intention toward AI-based educational assessment.

The validated instrument contributes to educational measurement and technology acceptance research in several

ways. First, it provides a theoretically grounded tool based on the UTAUT framework. Second, it extends existing technology acceptance measurement by incorporating AI-specific constructs that are highly relevant to assessment contexts. Third, it provides empirical evidence from Rasch analysis, which strengthens the instrument's validity and reliability at both the item and scale levels.

Practically, the instrument can be used by researchers, university administrators, policymakers, and educational technology stakeholders to assess lecturers' readiness for AI-based educational assessment. It may also support institutional needs analysis, professional development planning, policy evaluation, and responsible AI implementation in higher education assessment.

Although the present study provides preliminary Rasch-based validation evidence, further refinement is recommended before the instrument is used as a large-scale diagnostic tool. Future studies should validate the construct structure using confirmatory factor analysis, examine measurement invariance across lecturer groups, and consider item reduction to improve efficiency while preserving theoretical and content validity.

In conclusion, the extended UTAUT-based instrument developed in this study provides a promising and psychometrically supported tool for measuring lecturers' acceptance and behavioural intention toward AI-based educational assessment. Its use may help higher education institutions better understand the human and professional factors that shape AI adoption, thereby supporting more ethical, evidence-based, and educator-informed implementation of AI in assessment.

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to the lecturers who participated in this study. Appreciation is also extended to the experts who reviewed the instrument and provided constructive feedback during the content validation process. The authors further acknowledge the academic support provided by Universiti Technology Malaysia throughout the development of this research.

Informed Consent Statement

Informed consent was obtained from all respondents prior to their participation in the study. Respondents were informed that they had the right to decline participation or withdraw from the study at any stage without penalty. Completion and submission of the questionnaire were treated as confirmation of voluntary consent to participate.

Conflict of Interest

The authors declare that there is no conflict of interest related to the design, conduct, analysis, or publication of this study.

Data Availability Statement

The data used in this study are not publicly available due to ethical and confidentiality restrictions involving human participants. However, anonymized data may be made available from the corresponding author upon reasonable request and subject to institutional approval.

REFERENCES

1. Acosta-Enriquez, B. G., Ramos Farroñan, E. V., Villena Zapata, L. I., Mogollon Garcia, F. S., Rabanal-León, H. C., Morales Angaspilco, J. E., & Saldaña Bocanegra, J. C. (2024). Acceptance of artificial intelligence in university contexts: A conceptual analysis based on UTAUT2 theory. Heliyon, 10, Article e38315. https://doi.org/10.1016/j.heliyon.2024.e38315
2. Ajzen, I. (1991). The theory of planned behaviour. Organizational Behaviour and Human Decision Processes, 50(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

3. Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioural research: A primer. Frontiers in Public Health, 6, Article 149. https://doi.org/10.3389/fpubh.2018.00149
4. Bond, T. G., Yan, Z., & Heene, M. (2020). Applying the Rasch model: Fundamental measurement in the human sciences (4th ed.). Routledge. https://doi.org/10.4324/9780429030499
5. Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer. https://doi.org/10.1007/978-94-007-6857-4
6. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510
7. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, Article 22. https://doi.org/10.1186/s41239-023-00392-8
8. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008
9. DeVellis, R. F., & Thorpe, C. T. (2021). Scale development: Theory and applications (5th ed.). SAGE Publications.
10. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2019). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
11. European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://doi.org/10.2766/153756
12. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.
13. Huggins-Manley, A. C., Booth, B. M., & D'Mello, S. K. (2022). Toward argument-based fairness with an application to AI-enhanced educational assessments. Journal of Educational Measurement, 59(3), 362–388. https://doi.org/10.1111/jedm.12334
14. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
15. Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In W. Holmes & K. Porayska-Pomsta (Eds.), The ethics of artificial intelligence in education: Practices, challenges, and debates (pp. 174-202). Routledge. https://doi.org/10.4324/9780429329067-10
16. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO
17. Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). AI and education: Guidance for policy-makers. UNESCO Publishing.
18. Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. British Journal of Educational Technology, 53(4), 914-931. https://doi.org/10.1111/bjet.13232
19. OECD. (2023). OECD digital education outlook 2023: Towards an effective digital education ecosystem. OECD Publishing. https://doi.org/10.1787/c74f03de-en
20. Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146, Article 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
21. Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. Computers & Education, 57(4), 2432-2440. https://doi.org/10.1016/j.compedu.2011.06.008
22. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478. https://doi.org/10.2307/30036540

23. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. Journal of the Association for Information Systems, 17(5), 328-376. https://doi.org/10.17705/1jais.00428
24. Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, Article 124167. https://doi.org/10.1016/j.eswa.2024.124167
25. Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0