frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 16 January 2026
DOI 10.3389/feduc.2026.1754136

Check for updates

OPEN ACCESS

EDITED BY
Xinya Liang,
University of Arkansas, United States

REVIEWED BY
Nozimjon Tuygunov,
The University of Hong Kong,
Hong Kong SAR, China
Ji Li,
University of Arkansas for Medical Sciences,
United States

*CORRESPONDENCE
Ghalia Y. Bhadila
✉ gbhadila@kau.edu.sa

RECEIVED 25 November 2025
REVISED 28 December 2025
ACCEPTED 05 January 2026
PUBLISHED 16 January 2026

CITATION
Bhadila GY, Bahdila D, Saber NO and Alyafi DA (2026) Impact of artificial intelligence on task performance and perceived task load: a pragmatic randomized experiment.
Front. Educ. 11:1754136.
doi: 10.3389/feduc.2026.1754136

COPYRIGHT
© 2026 Bhadila, Bahdila, Saber and Alyafi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Impact of artificial intelligence on task performance and perceived task load: a pragmatic randomized experiment

Ghalia Y. Bhadila[1]*, Dania Bahdila[1], Nujud O. Saber[2] and Dana A. Alyafi[3]

[1]Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia, [2]Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia, [3]Department of Orthodontics, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction: This study aimed to assess the impact of artificial intelligence (AI) assistance on immediate task performance and evaluate perceived task load and AI acceptance among dental interns in an educational setting.

Methods: A pragmatic experiment was conducted among 132 dental interns during the 2024–2025 academic year. Participants were randomly allocated to either an AI-assisted group (n = 67) or a baseline knowledge control group (n = 65) to complete a 15-question quiz based on pediatric orthodontic cases. Perceived task load was measured using the National Aeronautics and Space Administration Task Load Index. AI acceptance was assessed using the Technology Acceptance Model (TAM). Task performance (quiz scores), task load, and AI acceptance were analyzed using Wilcoxon rank-sum tests and an adjusted generalized regression model.

Results: The AI-assisted group achieved higher task performance scores (median, 13 vs. 11; $p < 0.0001$) and lower perceived task load scores (median, 21.7 vs. 41.7; $p < 0.0001$) than the control group. The AI-assisted group had 1.67 times higher odds of answering a question correctly compared to controls in the adjusted model. Responses to the TAM demonstrated high levels of perceived usefulness, perceived ease of use, and behavioral intention (Cronbach's α = 0.92–0.95).

Conclusion: The AI-assisted group demonstrated improved immediate task performance and reduced perceived task load compared to the control group. This study serves as a preliminary step toward understanding how AI tools can support clinical learning and decision-making processes in educational settings.

KEYWORDS

artificial intelligence, ChatGPT, dental education, interceptive orthodontics, pediatric dentistry

1 Introduction

The incorporation of artificial intelligence (AI) into dental clinical decision-making represents a major transformation in modern dentistry, fundamentally changing the approaches to diagnosis, treatment planning, and patient care (Semerci and Yardımcı, 2024). The integration of AI into dental practice demonstrates promise for delivering more personalized and comprehensive care, thereby enhancing clinical decision-making (Samaranayake et al., 2025; Tuygunov et al., 2025). For example, ChatGPT has demonstrated a relatively high diagnostic performance (72.2%) in various pediatric dental cases (Bhadila et al., 2025). Additionally, a recent review highlighted the accuracy and efficiency of AI-based systems in orthodontics, particularly in diagnosing skeletal malocclusions, detecting landmarks, indexing images, and guiding treatment planning, ultimately improving time efficiency and reducing the clinician's task load (Gracea et al., 2025).

AI also holds potential for reducing the mental and task load burdens experienced by clinicians during data collection, analysis, record-keeping, and decision-making, which may help decrease burnout and improve patient outcomes (Gandhi et al., 2023). Perceived task load can be reliably measured using the National Aeronautics and Space Administration Task Load Index (NASA-TLX), a validated tool that assesses perceived mental, physical, and temporal demands, as well as performance, effort, and frustration (Hart and Staveland, 1988). The application of the NASA-TLX in dental education has revealed significant relationships between task complexity and perceived task load (Al-Saud, 2023). For example, a study with second-year dental students showed that complex preclinical tasks increase cognitive load, leading to performance decline when mental demand exceeds manageable levels (Al-Saud, 2023).

As with any emerging technology, its effectiveness depends on users' acceptance and perceived ease of use. Those constructs are articulated within the Technology Acceptance Model (TAM), one of the widely applied frameworks for understanding how new technologies are embraced (Ibrahim and Shiring, 2022). The TAM examines perceived usefulness (PU) and perceived ease of use (PEOU) to provide insights into the factors shaping practitioner acceptance of AI-powered tools, with PU emerging as a key predictor of adoption (Lee et al., 2025).

The intersection of AI assistance, perceived task load, and technology acceptance has promising implications for dental education and practice. Understanding how AI adoption influences performance in tasks that require aspects of diagnostic reasoning and cognitive demand can help guide AI integration in educational settings and support the digital transformation of dentistry. However, research examining the combined impact of AI tools on task performance, perceived task load reduction, and user acceptance in dental education remains scarce.

Therefore, this study aimed to evaluate the effect of AI-assisted tools on immediate task performance among dental interns and assess perceived task load and technology acceptance using NASA-TLX and TAM. The hypothesis was that among a sample of dental interns in an educational setting, the AI-assisted group would demonstrate improved immediate task performance and reduced perceived task load compared to the baseline knowledge group. This study serves as a preliminary step toward understanding how AI tools can support learning and decision-making processes in educational settings.

2 Materials and methods

2.1 Ethical considerations

The study was approved by the Ethical Research Committee of the Faculty of Dentistry at King Abdulaziz University Dental Hospital (KAUDH), Jeddah, Saudi Arabia (Protocol Code: 28–02-25). Written informed consent was obtained from all participants prior to the study. Participants were informed that participation was voluntary and that they could withdraw at any time without any negative consequences.

2.2 Experiment design

This study was designed as a pragmatic randomized experiment to examine the impact of AI assistance on immediate task performance (i.e., quiz scores) in pediatric orthodontic clinical decision-making scenarios and perceived task load among dental interns. Pragmatic experiments are designed to evaluate the effectiveness of an intervention in real-life routine practices (Schwartz and Lellouch, 1967; Torgerson and Torgerson, 2007). This approach allowed the evaluation of AI assistance on immediate task performance and perceived task load under a standard classroom quiz experience, where variations in digital literacy and natural user behavior were preserved. This pragmatic design helped provide early practical relevance in natural educational settings. The Consolidated Standards of Reporting Trials (CONSORT) extension for pragmatic trial guidelines was followed in this study (Bennett, 2005; Zwarenstein et al., 2008) (Figure 1). Data were collected between May 2025 and June 2025. Participants were recruited from KAUDH and were eligible if they were dental interns enrolled during the 2024–2025 academic year and agreed to participate. None of the enrolled students had received formal training in AI tools as part of their curriculum. Recruitment was conducted via email invitation with instructions to bring fully charged electronic devices. On the day of data collection, the participants received an overview of the study and provided written informed consent.

In our context, dental interns remain under the supervision of dental school throughout their seventh year of training. After completing six academic years, they rotate across affiliated public and private hospitals, where they deliver supervised clinical care comparable to general dental practice. Although they function in general dental practice settings during this internship year, they receive their bachelor's degree only after successfully completing their internship. Therefore, they were selected as our sample of choice, as this hybrid model allowed us to capture the combined real-world general practice exposure with structured educational supervision.

2.3 Sampling

A priori power analysis was conducted using G*Power (version 3.1) to determine the minimum sample size required to detect a statistically significant difference in mean scores between the two independent groups. Assuming a moderate effect size ($d = 0.5$), significance level of 0.05, and power of 80%, the analysis indicated that 128 participants (64 per group) would be necessary to achieve an adequate statistical power. A convenience sampling approach was employed, followed by random allocation of participants to one of the two groups.

- Group I (AI-assisted): Dental interns were permitted to use AI-assisted tools for task completion.
- Group II (Baseline Knowledge): Dental interns completing the task without AI assistance.

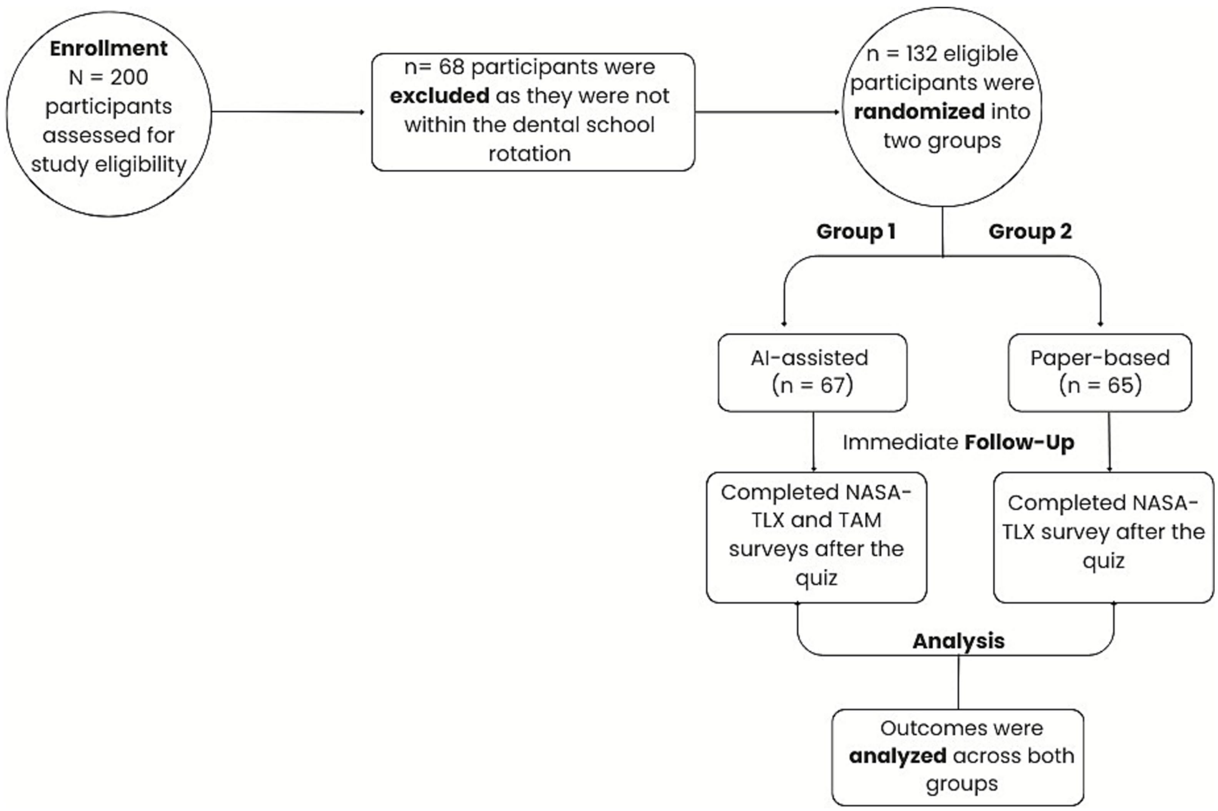

FIGURE 1
CONSORT flowchart of the study participants. NASA-TLX, NASA task load index; TAM, technology acceptance model.

Simple randomization was performed using a 1:1 allocation ratio generated in Microsoft Excel. Equal numbers of printed slips labeled "Group I," or "Group II" were folded and placed in an opaque container. The slips were mixed thoroughly, and at the time of data collection, the participants were asked to randomly draw one slip each, without replacement. This ensured equal probability of allocation and concealment. However, given the nature of the intervention, blinding was not feasible.

The design aimed to mimic the real-world classroom conditions in dental schools. Selection bias was reduced through randomization, ensuring approximately equal group size. Both groups completed the quiz simultaneously in the same, quiet room under identical supervision and time limits to minimize environmental bias. Each participant worked individually on their personal laptop or tablet. Students were instructed to complete the activity individually, without collaboration, and proctoring was performed to ensure compliance. The maximum time allowed to solve the 15-question test was 15 min.

The control group (baseline knowledge group) was chosen to reflect the standard quiz practice at our institution, in which students independently solved quizzes without external digital assistance. Meanwhile, the AI-assisted group models the emerging clinical practice in which AI tools may be used for decision support, supplementing baseline knowledge. Although administering the quiz through Google Forms for both groups could have minimized potential Hawthorne effects, this approach would have introduced a high risk of contamination, as preventing Internet access among control participants was not feasible if Google Forms were to be used. Therefore, the control group completed a paper-based quiz reflecting the standard classroom practice at our institution, whereas the intervention group completed the same quiz on Google Forms with access to AI tools.

To minimize potential performance anxiety, demoralization, or disability bias between groups, we informed all participants prior to the study about the following:

1. The quiz was part of a simple educational experiment and not a graded test.
2. The quiz was not relevant or weighted in their internship evaluation.
3. Individual responses were completely anonymous, as no names or potentially identifying data were collected.
4. Key answers were immediately shared with all participants after the tasks and surveys were completed.
5. The intervention group had access to the Internet during the quiz but did not receive any new AI instructional content or preparatory material beyond that available to the control group.

To minimize excessive variability in tool usage, participants in Group I (AI-assisted group) were provided with specific AI tools (with their QR codes presented on the screen).

1. DeepSeek: A free AI tool with unlimited prompts.
2. ChatGPT: free version with a 10-prompt limit unless the student had a paid subscription.

We allowed this slight variation in AI tools to reflect real-world dynamics because AI-assisted practices are heterogeneous in actual settings. By permitting students to use their preferred tools, we aimed to model realistic routine practices.

Three pediatric interceptive orthodontic clinical cases representing common clinical scenarios were selected by faculty members certified by the American Board of Orthodontics and Pediatric Dentistry. Standardized multiple-choice questions (MCQs) based on evidence-based guidelines were developed to assess task performance and clinical decision-making. Content validity was examined by five expert faculties using the Content Validity Index (CVI) (Polit and Beck, 2006) across six dimensions: ambiguity, importance, simplicity, relevance, clarity, and inclusion in the final version. The CVI values for the individual items ranged from 0.80 to 1.00. The average CVI values across each dimension were as follows: ambiguity (1.00), importance (1.00), simplicity (1.00), relevance (0.99), clarity (0.99), and inclusion (0.99). The overall mean CVIs across all dimensions was 0.99, indicating excellent content validity. In addition, the five experts independently rated the complexity of each clinical case as simple, moderate, or complex. For Case 1, three experts rated the case as moderate and two as simple, resulting in a simplicity rating of 40%. Case 2 received a simplicity rating of 60%. Case 3 was unanimously rated as simple (100%). This approach ensured a gradient of case complexity, allowing the test to capture multiple levels of realistic clinical scenarios relevant to general dental practice (details provided in Supplementary materials).

2.4 Measures and outcomes

The dental interns completed the following:

- A task that tested aspects of diagnostic reasoning and treatment planning for three clinical cases; administered either digitally using Google Forms for the AI-assisted group or on paper for baseline knowledge group.
- Immediately after task completion, both groups completed the NASA-TLX survey to assess their perceived task load, which was completed digitally using Google Forms.
- Subsequently, the AI-assisted group completed the TAM survey to digitally evaluate AI acceptance and usability using Google Forms.

NASA-TLX is a widely validated tool for assessing perceived task load in healthcare and education across six dimensions (Hart and Staveland, 1988).

1. Mental Demand - How challenging was the task mentally?
2. Physical Demand - How much physical effort was required to perform the task?
3. Temporal Demand - How pressured did you feel by time constraints?
4. Performance - How effectively did you meet the task objectives?
5. Effort - How much exertion was required to finish the task?
6. Frustration Level - How much discouragement, irritation, or stress did you experience?

Raw NASA-TLX Score:

- The sum of all six ratings was divided by six to obtain the average perceived task load score.

Formula:

$$\text{Raw TLX Score} = \sum\left(\text{all six ratings}\right)/6$$

The TAM was used to assess the PU and PEOU of AI in clinical decision-making (Ibrahim and Shiring, 2022). PU was defined as “the degree to which a person believes that using a particular system would enhance job performance” (Davis, 1989) whereas PEOU was defined as “the degree to which a person believes that using a particular system would be free of effort” (Davis, 1989).

The study outcomes are:

1. Primary outcome: Task performance score, reported as the total number of correct answers out of 15 multiple-choice questions (score range: 0–15), reflecting accuracy in three structured case-based scenarios (each question was weighted equally).
2. Secondary outcome: Perceived task load score, measured using NASA-TLX scores across six task load dimensions (score range: 0–100).

2.5 Statistical analysis

Descriptive analyses were conducted to summarize the data using frequencies, percentages, means, standard deviations (SD), medians, and their associated 25th and 75th interquartile ranges (IQR). Cronbach’s α was calculated to assess the internal consistency of the TAM constructs among the students in the AI-assisted group. The Shapiro–Wilk test was used to assess the normality of continuous variables, including task performance and task load scores. Both variables demonstrated statistically significant deviations from normality ($p < 0.05$). Consequently, Wilcoxon rank-sum tests were applied to compare task performance and NASA-TLX task load scores between the AI-assisted and baseline knowledge groups. Bonferroni correction was applied to reduce the risk of type I error due to multiple testing on families of secondary outcomes. The three case-level task performance scores in Table 1 were evaluated using a Bonferroni-corrected significance threshold of $p < 0.017$ (=0.05/3), and the six NASA-TLX subscales in Table 2 were evaluated using a threshold of $p < 0.008$ (=0.05/6). The total test score in Table 1 and the overall NASA-TLX workload score in Table 2 were the primary outcomes and were reported without adjustment for multiple comparisons. All statistical tests were performed using two-tailed tests. For multivariable analysis, the unit of analysis was the participant. Each participant had a single outcome score for test performance, which was measured as the total number of correct answers out of 15 multiple-choice questions (bounded count ranged between 0 and 15). Generalized linear models (GLM) with a binomial family and logit link (Stata: glm, family(binomial 15) link(logit)) were selected to model the non-normally distributed outcome (test performance). Since each participant contributed one observation to the regression, the assumption of independence applies between participants. Two models were conducted using the binomial

TABLE 1 Comparison of task performance (number of correct answers out of 15 multiple-choice questions) of dental interns by group (AI-assisted vs. baseline knowledge control) using Wilcoxon rank-sum test.

Test Performance	Median (25th-75th interquartile range)			*P*-value (*)
	Total $n = 132$	Control group $n = 65$ (49.2%)	AI-assisted group $n = 67$ (50.8%)	
Case 1 score (out of 5)	4 [3–5]	4 [3–4]	5 [4–5]	**<0.0001**
Case 2 score (out of 5)	3 [2–4]	3 [2–3]	4 [3–4]	**0.0003**
Case 3 score (out of 5)	5 [4–5]	5 [4–5]	5 [5–5]	**0.0140**
Total score (out of 15)	12 [10–14]	11 [9–13]	13 [11–14]	**<0.0001 (**)**

AI, Artificial Intelligence. The bold font indicates statistical significance. *Raw *p*-values are reported. Significance was evaluated using a Bonferroni-adjusted threshold of $\alpha = 0.05/3 = 0.017$, applied to the case-level scores, which are considered as a family of three secondary outcomes. **The total test score was treated as a separate primary outcome; therefore, Bonferroni correction was not applied, and the statistical significance threshold was set at $\alpha = 0.05$.

TABLE 2 Comparison of NASA task load index (TLX) scores for dental interns by group (AI-assisted vs. baseline knowledge control) using Wilcoxon rank-sum test.

NASA-TLX dimension	Median score (25th-75th interquartile range) Scores are out of 100			*P*-value (*)
	Total $n = 132$	Control group $n = 65$	AI-assisted group $n = 67$	
Mental Demand	50 [20–80]	65 [50–90]	35 [10–70]	**0.0004**
Physical Demand	20 [20–62.5]	40 [10–80]	10 [0–40]	**0.0070**
Temporal Demand	20 [5–55]	30 [5–70]	10 [5–45]	0.0202
Performance	55 [10–90]	75 [25–85]	40 [0–90]	0.1162
Effort	25 [5–60]	50 [20–75]	10 [0–35]	**<0.0001**
Frustration	10 [0–40]	10 [0–60]	0 [0–25]	0.0266
Overall Load Score	35 [17.1–54.6]	41.7 [30.0–62.5]	21.7 [8.3–46.7]	**<0.0001 (**)**

AI, Artificial Intelligence NASA-TLX score: High number equals higher perceived task load (0 = Zero Load). The bold font indicates statistical significance. *Raw *p*-values are reported. Significance was evaluated using a Bonferroni-adjusted threshold of $\alpha = 0.05/6 = 0.008$, applied across the six NASA-TLX subscales, which are considered as a family of six secondary outcomes. **The total test score was treated as a separate primary outcome; therefore, Bonferroni correction was not applied, and the statistical significance threshold was set at $\alpha = 0.05$.

GLM. Model 1 (the unadjusted model) estimated the total effect of AI assistance on task performance. In model 2, the GLM evaluated the association between the outcome (task performance) and the main predictor (group allocation: AI assistance vs. baseline knowledge), while adjusting for the NASA-TLX task load scores and sex. Model 2 was conducted to estimate the direct effect of AI assistance while holding the perceived task load and sex constant. Perceived task load was scaled and reported per 10-point increase in the NASA-TLX score to facilitate interpretation. The measure of estimation was the exponentiated coefficients presented as crude and adjusted odds ratios (ORs/AORs) with their corresponding 95% confidence intervals (CIs). The statistical significance for the regression analysis was set at $p < 0.05$. All statistical analyses were performed using Stata/IC version 15.1 (StataCorp, College Station, TX, USA).

3 Results

The final analyzed sample included 132 dental interns, of whom 53.8% were female ($n = 71$) and 46.2% were male ($n = 61$). Participants were randomized into AI-assisted ($n = 67$, 50.8%) and baseline knowledge ($n = 65$, 49.2%) groups. As enrolled dental interns of the 2024–2025 academic year, the participants were between 22 and 24 years of age. Table 1 presented the task performance of the AI-assisted and baseline knowledge groups. The AI-assisted group achieved significantly higher overall task scores than the baseline knowledge group (median score 13 [IQR 11–14] vs. 11 [IQR 9–13]; $p < 0.0001$). Case-level analysis showed consistently superior performance in the AI-assisted group for the first two cases (Bonferroni-corrected significance threshold = 0.017; $p < 0.0001$ and $p = 0.0003$; respectively), whereas no significant difference was observed for the third case ($p = 0.0140$).

Table 2 compared the perceived task load scores measured using the NASA-TLX. The AI-assisted groups reported significantly lower overall perceived task load scores than the baseline knowledge group (median score 21.7 [IQR 8.3–46.7] vs. 41.7 [IQR 30.0–62.5]; $p < 0.0001$). The scores for mental demand, physical demand and effort were substantially lower in the AI-assisted group (Bonferroni-corrected significance threshold = 0.008; *p*-values were<0.008). Although temporal demand and frustration scores were also lower in the AI-assisted group, these differences were not statistically significant (*p*-values were>0.008).

Table 3 summarized the descriptive statistics for the TAM constructs among dental interns in the AI-assisted group ($n = 63$; four of the 67 interns did not complete the TAM survey). Possible scores ranged from 1 (Highly Unlikely) to 7 (Highly Likely). The PU of AI in clinical decision-making was rated highly, with item means ranging from 5.73 to 5.83 (SD ~ 1.3–1.5), and excellent internal consistency (Cronbach's $\alpha = 0.95$). PEOU was also rated highly, with item means ranging from 5.95 to 6.40 (SD ~ 1.0–1.3), and excellent reliability

TABLE 3 Descriptive statistics for technology acceptance model constructs (TAM) (artificial intelligence assisted group, *n* = 63; four of the 67 interns did not complete the TAM survey).

Construct	Mean	Standard deviation	Cronbach's α
Perceived Usefulness (PU)			
Using AI enabled me to make clinical decisions more quickly.	5.76	1.48	0.946
Using AI improved my clinical decision performance.	5.76	1.28	
Using AI in making clinical decisions increased my productivity.	5.79	1.37	
Using AI enhanced my effectiveness in making clinical decisions.	5.83	1.26	
Using AI made it easier to make clinical decisions.	5.78	1.37	
I find AI useful in making clinical decisions.	5.73	1.32	
Perceived Ease of Use (PEOU)			
Learning to operate AI is easy for me.	6.03	1.17	0.915
I find it easy to get AI to do what I want it to do.	5.95	1.26	
My interaction with AI was clear and understandable.	6.05	1.03	
I found AI to be flexible to interact with.	6.08	1.17	
It would be easy for me to become skillful at using AI.	6.14	1.12	
I find AI easy to use.	6.40	0.99	
Behavioral Intention to Use (BI)			
I intend to use AI regularly in the future.	6.29	1.01	

TABLE 4 Generalized linear regression with a binomial family and logit link predicting task performance (total number of correct responses out of 15 multiple-choice questions), among dental interns (*n* = 132; academic year 2024–2025).

Predictors	Outcome (task performance) as the total number of correct answers Odds ratios (ORs) and 95% confidence intervals [95%CI]			
	Model 1		Model 2	
	Crude OR of answering a question correctly [95%CI]	*P*-value	Adjusted OR of answering a question correctly [95%CI]	*P*-value
Group allocation				
Control	Reference		Reference	
Artificial Intelligence Assisted	1.96 [1.58–2.44]	<0.0001	1.67 [1.33–2.10]	<0.0001
Sex				
Male			Reference	
Female			0.91 [0.73–1.13]	0.394
Perceived Task Load (NASA-TLX; 0–100 scale, per 10-point increase)			0.899 [0.86–0.94]	<0.0001
Constant	2.65 [2.22–2.94]	<0.0001	4.43 [3.35–5.86]	<0.0001

(Cronbach's α = 0.92). The behavioral Intention to Use AI showed strong endorsement (mean 6.29, SD 1.01), indicating a substantial willingness among dental interns to adopt AI in future clinical decision-making.

Table 4 presented the GLM with a binomial family examining the association between the test performance scores and AI assistance. Model 1 estimated the overall impact of AI assistance on the test performance. The students in the AI-assisted group had 1.98 times higher odds of answering a question correctly compared to those in the baseline knowledge group (OR = 1.96; 95% CI: 1.58–2.44; $p < 0.0001$). Meanwhile, Model 2 estimated the direct effect of AI assistance on task performance, adjusting for the perceived task load scores and sex. In the adjusted model, students in the AI-assisted group had 1.67 times higher odds of answering a question correctly compared to those in the baseline knowledge group (AOR = 1.67; 95% CI: 1.33–2.10; $p < 0.0001$). A higher perceived task load was significantly associated with lower test performance, with every 10-unit increase in task load score, there was a 10.1% decrease in the odds of answering a question correctly (AOR = 0.899; 95% CI: 0.86–0.94; $p < 0.0001$). Sex was not significantly associated with task performance in either model ($p > 0.05$).

For a more intuitive and practical interpretation, we calculated the model-estimated means (Stata: margins). The unadjusted model-estimated means for task performance score was 12.51 correct answers

10.3389/feduc.2026.1754136

(95% CI: 12.16–12.85) in the AI-assisted group compared to 10.78 (95% CI: 10.36–11.21) in the baseline knowledge group. Meanwhile, the adjusted model-estimated means for the task performance score was 12.33 correct answers (95% CI: 11.95–12.70) in the AI-assisted group compared to 11.03 (95% CI: 10.61–11.45) in the baseline knowledge group. This corresponds to an average of 1–2 more correct answers out of 15 (all $p < 0.0001$) for the AI-assisted group compared to the control group.

4 Discussion

This pragmatic experiment assessed the impact of AI assistance on immediate task performance, perceived task load, and acceptance of AI among dental interns. The AI-assisted group demonstrated statistically significantly higher task performance scores than the baseline knowledge group, even after adjusting for perceived task load. Perceived task load, measured using the NASA-TLX, was significantly lower among students in the AI-assisted group than in the baseline-knowledge group. Acceptance of AI, assessed using the TAM survey, was rated highly for PU, PEOU, and behavioral intention to use (BIU) AI. Therefore, we rejected the null hypothesis and found that among a sample of dental interns in an educational setting, the AI-assisted group demonstrated improved immediate task performance and reduced perceived task load compared to the baseline knowledge group.

In this study, the overall task scores were significantly higher in the AI-assisted group than in the baseline-knowledge group. Moreover, the adjusted regression model in this study suggested that dental interns in the AI-assisted group had 1.67 times higher odds of answering a question correctly compared to those in the baseline knowledge group. These findings align with those of a previous randomized clinical trial (RCT) conducted at Georgetown University School of Medicine (Kalam et al., 2025), where students in the AI-assisted group achieved significantly improved quiz performance across all cases compared with the unaided clinical decision-making approach. That study assigned first-year medical students to one of three groups: ChatGPT-4, external non-AI online resources, or institutional resources (textbooks and lectures). Participants in the ChatGPT group outperformed those in the other groups in case-based assessments, supporting the conclusion that AI assistance enhances clinical decision-making and academic performance (Kalam et al., 2025). However, some methodological differences distinguish the two studies. The present study included a larger sample of dental interns and focused on specialty cases involving pediatric and interceptive orthodontics. Additionally, a recent mixed-methods study at the European University Cyprus (Kavadella et al., 2024), where 77 s-year dental students completing a “Radiation Biology and Radiation Protection” assignment achieved significantly higher scores when using ChatGPT compared to traditional literature searches ($p = 0.045$). Although the Cyprus study focused on collaborative learning and knowledge retention, the present study addressed AI integration into individual clinical decision-making in pediatric orthodontics. Despite these variations, these studies emphasize the potential of AI to improve clinical decision-making and learning outcomes in health professions education.

In the current study, the overall perceived task load scores were significantly lower in the AI-assisted group than in the baseline knowledge group. Additionally, the regression model demonstrated that a higher perceived task load was significantly associated with lower test performance. For every 10-unit increase in the perceived task load score, there was a 10.1% decrease in the odds of answering a question correctly. Our findings correspond with the results of a recent RCT in which 187 undergraduate dental students in China were randomly assigned to a ChatGPT-assisted group or a control group using traditional video-based learning (Huang et al., 2025). After 1 week of intervention, participants in the ChatGPT group demonstrated reduced cognitive load, measured objectively via pupil diameter changes using eye-tracking and subjectively through questionnaires, compared to those in the control group (Huang et al., 2025). The consistent benefits across different contexts highlight the broader educational value of AI technology. Overall, these findings support the current results and suggest that AI-assisted decision-making may streamline the diagnostic process by providing structured suggestions. Such assistance could reduce the time spent on manual data retrieval, potentially allowing students to focus more on interpretative and reasoning tasks.

Regarding AI acceptability and usability, the overall PU, PEOU, and BIU of AI for clinical decision-making were rated highly, reflecting the intention of students to adopt AI in the future. These findings should be interpreted as exploratory and descriptive, and not as evidence of strong predictive ability or sustained acceptability for long-term adoption. TAM was administered only to the AI-assisted group and immediately after task completion, introducing potential confirmation bias. While the internal consistency of TAM constructs was high, which is consistent with prior TAM literature (Ashmawy et al., 2025; Ibrahim and Shiring, 2022), they do not mitigate the design limitations noted above. Future research may benefit from administering TAM to both groups and longitudinally after delayed exposure. These findings align with those of a previous study, in which 59.7% of dental students and dentists agreed that the dental curriculum should be updated with AI and reported positive attitudes toward its integration into dental practice (Kalaimani et al., 2023). A recent study further reinforced these observations, as 428 dental educators participated in a global survey assessing perceptions of AI chatbots and large language models such as ChatGPT (Uribe et al., 2024). Most participants (64%) acknowledged the potential of these tools in dental education, recognizing their usefulness for knowledge acquisition, research, and clinical decision making. Concerns persisted, particularly regarding the risk of reduced human interaction and the absence of explicit guidelines or structured training for curriculum integration. A scoping review of the barriers and facilitators of AI adoption in healthcare supported these concerns, emphasizing that user acceptance improves when AI is perceived as useful and easy to use, while successful integration requires addressing challenges such as insufficient training, limited institutional support, data privacy concerns, and trust in technology (Hassan et al., 2024). Collectively, these studies indicate that although learners and educators increasingly hold positive attitudes toward AI, sustained adoption in dental education will depend on curriculum reforms, targeted training, and robust technological support to ensure the safe, effective, and ethical implementation of AI. Evidence from Turkey adds further support (Eroğlu Çakmakoğlu and Günay, 2025), as 72.3% of dental students believed that AI would bring significant changes to the profession, with 64.5% supporting AI-based diagnostic systems despite nearly half of them expressing ethical concerns. Similarly, a study of healthcare students in Korea highlighted that AI literacy, although slightly below average, was positively associated with favorable attitudes toward AI and the intention to apply it in clinical contexts (Si, 2025). Notably, students' interest in AI correlated more strongly with literacy than with prior training, emphasizing the importance of fostering engagement through structured education.

In this study, ChatGPT was selected because it became the fastest-growing application in history, reaching over 100 million active users within 2 months as of January 2023 (Buriak et al., 2023). However, some students might have subscriptions to OpenAI, which allows unlimited use of ChatGPT. This concern was addressed by providing the option of using DeepSeek as a free AI tool, which was included because of its increasing popularity, particularly because of its free accessibility at the time of experiment conduction. Additionally, we allowed this slight variation in AI tools to reflect real-world dynamics, aligning with the pragmatic experiment design of choice. By permitting students to use their preferred tools, we aimed to model realistic behavior because AI-assisted practices are heterogeneous in real settings. The intervention also aimed to evaluate the general concept of AI assistance rather than a single isolated system to support the broader applicability of the study's findings. Despite this, the heterogeneity in AI interventions used may have limited the reproducibility and interpretability of the study findings. Future randomized trials may benefit from evaluating isolated tool-specific effects using limited and standardized AI tools.

This study had some limitations. First, we did not measure prior AI familiarity, which may have contributed to variations in task performance among the AI-assisted group. However, at the time of data collection, none of the enrolled students had received formal training in AI tools as part of their curriculum, and exposure to generative AI was expected to be comparable across a relatively homogenous group of dental interns from the same school and class. Future research should include baseline measures of AI literacy and self-efficacy. Second, we did not quantify susceptibility to AI errors; future studies should assess error propagation as a safety outcome and examine skill transfer to high-fidelity and more complex clinical evaluations. Third, although the experiment was designed to reflect real classroom practice, the difference in the intervention group completing an online quiz while the control group completed a paper-based quiz might have influenced the participants performance and perceived task load. These differences could be unrelated to AI assistance such as (ergonomics, navigation speed, or user interface) and are potential confounders. Additionally, despite the incorporation of pragmatic elements, the study was implemented within a controlled educational setting using surrogate outcomes without real clinical consequences, limiting its applicability in real-world settings. Finally, this study was conducted in a single dental institution; to improve generalizability and external validity, subsequent research should be conducted across multiple settings or institutions.

Given the limited evidence currently available on the impact of AI on clinical cognition in dentistry, this study addresses an important gap in the literature. The findings add to the broader discourse on digital transformation in healthcare education and provide preliminary findings that may support future research on the long-term effects of AI integration on clinical competency and decision making. Future investigations should include longitudinal studies that assess the sustained influence of AI on perceived task load and clinical performance, its impact on critical thinking, and strategies to balance AI support with independent problem-solving skills over time. In conclusion, among a sample of dental interns in an educational setting, the AI-assisted group demonstrated improved task performance and reduced perceived task load using a structured case-based clinical scenario compared to the baseline knowledge group. Additionally, AI assistance received strong acceptance from the students. This study serves as a preliminary step toward understanding how AI tools can support clinical learning and decision-making processes in educational settings.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethical Research Committee of the Faculty of Dentistry at King Abdulaziz University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

GB: Validation, Funding acquisition, Resources, Project administration, Writing – review & editing, Conceptualization, Methodology, Writing – original draft, Investigation, Visualization. DB: Validation, Writing – review & editing, Conceptualization, Writing – original draft, Data curation, Methodology, Formal analysis. NS: Writing – review & editing, Writing – original draft, Methodology, Investigation. DA: Validation, Conceptualization, Investigation, Writing – review & editing, Writing – original draft, Methodology.

Funding

The author(s) declared that financial support was received for this work and/or its publication.This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. IPP: 1471–165-2025. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2026.1754136/full#supplementary-material

References

Al-Saud, L. M. (2023). Simulated skill complexity and perceived cognitive load during preclinical dental training. *Eur. J. Dent. Educ.* 27, 992–1003. doi: 10.1111/eje.12891

Ashmawy, R., Zeina, S., Kamal, E., Shelbaya, K., Gawish, N., Sharaf, S., et al. (2025). A reliable tool for assessment of acceptance of e-consultation service in hospitals: the modified e-consultation technology acceptance model (TAM) questionnaire. *J. Egypt. Public Health Assoc.* 100:6. doi: 10.1186/s42506-025-00187-x

Bennett, J. A. (2005). The consolidated standards of reporting trials (CONSORT): guidelines for reporting randomized trials. *Nurs. Res.* 54, 128–132. doi: 10.1097/00006199-200503000-00007

Bhadila, G. Y., Alhomied, M., Mahmoud, A., and Farsi, N. J. (2025). Accuracy of artificial intelligence in making diagnoses and treatment decisions in pediatric dentistry. *Pediatr. Dent.* 47, 73–78.

Buriak, J. M., Akinwande, D., Artzi, N., Brinker, C. J., Burrows, C., Chan, W. C. W., et al. (2023). Best practices for using AI when writing scientific manuscripts. *ACS Nano* 17, 4091–4093. doi: 10.1021/acsnano.3c01544

Davis, F. D. (1989). "Technology acceptance model: TAM" in Information seeking behavior and technology adoption. eds. M. N. Al-Suqri and A. S. Al-Aufi (Cham: Springer), 5.

Eroğlu Çakmakoğlu, E., and Günay, A. (2025). Dental students' opinions on use of artificial intelligence: a survey study. *Med. Sci. Monit.* 31:e947658. doi: 10.12659/msm.947658

Gandhi, T. K., Classen, D., Sinsky, C. A., Rhew, D. C., Vande Garde, N., Roberts, A., et al. (2023). How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open* 6:oad079. doi: 10.1093/jamiaopen/ooad079

Gracea, R. S., Winderickx, N., Vanheers, M., Hendrickx, J., Preda, F., Shujaat, S., et al. (2025). Artificial intelligence for orthodontic diagnosis and treatment planning: a scoping review. *J. Dent.* 152:105442. doi: 10.1016/j.jdent.2024.105442

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. New York, NY: Elsevier.

Hassan, M., Kushniruk, A., and Borycki, E. (2024). Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. *JMIR Hum. Factors* 11:e48633. doi: 10.2196/48633

Huang, S., Wen, C., Bai, X., Li, S., Wang, S., Wang, X., et al. (2025). Exploring the application capability of ChatGPT as an instructor in skills education for dental medical students: randomized controlled trial. *J. Med. Internet Res.* 27:e68538. doi: 10.2196/68538

Ibrahim, A., and Shiring, E. (2022). The relationship between educators' attitudes, perceived usefulness, and perceived ease of use of instructional and web-based technologies: implications from technology acceptance model (TAM). *Int. J. Technol. Educ.* 5, 535–551. doi: 10.46328/ijte.285

Kalaimani, G., Chockalingam, R. M., and Karthick, P. (2023). Evaluation of knowledge, attitude, and practice (KAP) of artificial intelligence among dentists and dental students: a cross-sectional online survey. *Cureus* 15:e44656. doi: 10.7759/cureus.44656

Kalam, K. A., Masoud, F. D., Muntaser, A., Ranga, R., Geng, X., and Goyal, M. (2025). ChatGPT as a learning tool for medical students: results from a randomized controlled trial. *Cureus* 17:e85767. doi: 10.7759/cureus.85767

Kavadella, A., Dias Da Silva, M. A., Kaklamanos, E. G., Stamatopoulos, V., and Giannakopoulos, K. (2024). Evaluation of ChatGPT's real-life implementation in undergraduate dental education: mixed methods study. *JMIR Med. Educ.* 10:e51344. doi: 10.2196/51344

Lee, A. T., Ramasamy, R. K., and Subbarao, A. (2025). Understanding psychosocial barriers to healthcare technology adoption: a review of TAM technology acceptance model and unified theory of acceptance and use of technology and UTAUT frameworks. *Healthcare (Basel)* 13:250. doi: 10.3390/healthcare13030250

Polit, D. F., and Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health*, 29, 489–97.

Samaranayake, L., Tuygunov, N., Schwendicke, F., Osathanon, T., Khurshid, Z., Boymuradov, S. A., et al. (2025). The transformative role of artificial intelligence in dentistry: a comprehensive overview. Part 1: fundamentals of AI, and its contemporary applications in dentistry. *Int. Dent. J.* 75, 383–396. doi: 10.1016/j.identj.2025.02.005

Schwartz, D., and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *J. Chronic Dis.* 20, 637–648. doi: 10.1016/0021-9681(67)90041-0

Semerci, Z. M., and Yardımcı, S. (2024). Empowering modern dentistry: the impact of artificial intelligence on patient care and clinical decision making. *Diagnostics* 14:260. doi: 10.3390/diagnostics14121260

Si, J. (2025). Exploring AI literacy, attitudes toward AI, and intentions to use AI in clinical contexts among healthcare students in Korea: a cross-sectional study. *BMC Med. Educ.* 25:1233. doi: 10.1186/s12909-025-07766-8

Torgerson, C. J., and Torgerson, D. J. (2007). The need for pragmatic experimentation in educational research. *Econ. Innov. New Technol.* 16, 323–330. doi: 10.1080/10438590600982327

Tuygunov, N., Samaranayake, L., Khurshid, Z., Rewthamrongsris, P., Schwendicke, F., Osathanon, T., et al. (2025). The transformative role of artificial intelligence in dentistry: a comprehensive overview part 2: the promise and perils, and the international dental federation communique. *Int. Dent. J.* 75, 397–404. doi: 10.1016/j.identj.2025.02.006

Uribe, S. E., Maldupa, I., Kavadella, A., El Tantawi, M., Chaurasia, A., Fontana, M., et al 2024. Artificial intelligence chatbots and large language models in dental education: worldwide survey of educators. Eur. J. Dent. Educ., 28, 865–876, doi: DOI: 10.1111/eje.13009, PMID: 38586899.

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., et al. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 337:a2390. doi: 10.1136/bmj.a2390