

Comparative Evaluation of Five Artificial Intelligence Platforms for Orthodontic Patient Education: A Pre–Post Comparative Study

Running title: AI Platforms in Orthodontic Education

Dr. Benith Alanraj J¹, Dr. Varadharaja M.M.², Dr. Mahalakshmi R.²,
Dr. Palaniappan C.², Dr. Nivashini², Dr. Sridevi², Dr. Madhumalar²

1

(Final Year Postgraduate, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Professor And Head, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Reader, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Senior Lecturer, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Senior Lecturer, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Senior Lecturer, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

2

(Senior Lecturer, Department of Orthodontics and Dentofacial Orthopaedics, CSI College Of Dental Sciences and Research, Madurai -625001, Tamil Nadu, India)

Abstract

Objective: To compare changes in orthodontic knowledge scores after patient-education sessions delivered through five artificial intelligence (AI) platforms: Gemini, Meta AI, ChatGPT, Perplexity, and Grok.

Materials and Methods: This comparative pre–post study included 150 patients aged 14–30 years who were allocated to five groups of 30 participants each. The groups received orthodontic education through Gemini, Meta AI, ChatGPT, Perplexity, or Grok. Knowledge was assessed before and after the educational intervention using the same questionnaire. The source study reports analysis in SPSS 21.0, assessment of normality, and use of descriptive statistics, one-way analysis of variance, and Tukey post hoc testing, with statistical significance set at $p \leq 0.05$.

Results: Mean knowledge scores increased in all five groups. Mean $\pm$ standard deviation scores before and after the intervention were 7.23 ± 1.59 and 13.03 ± 1.75 for Gemini; 9.00 ± 2.27 and 11.97 ± 1.90 for Meta AI; 7.70 ± 2.64 and 13.07 ± 1.64 for ChatGPT; 7.27 ± 1.55 and 12.47 ± 1.43 for Perplexity; and 7.73 ± 1.26 and 13.13 ± 1.38 for Grok. The reported mean changes were 5.80 ± 1.96 , 2.97 ± 1.87 , 5.37 ± 2.80 , 5.20 ± 2.19 , and 5.40 ± 1.61 , respectively. The source document reports within-group p values below 0.001; the underlying test statistics require verification before submission.

Conclusions: Knowledge scores improved after education with each of the five AI platforms. The largest descriptive mean improvement was observed in the Gemini group, followed by Grok and ChatGPT. Because baseline scores differed and complete between-group inferential statistics were not provided in the source material, superiority of one platform over another cannot be concluded from the available data. AI platforms may be considered adjuncts to clinician-led education, subject to accuracy checks and professional oversight.

Keywords: artificial intelligence; orthodontics; patient education; health literacy; large language models; knowledge assessment

Date of Submission: 14-10-2026

Date of Acceptance: 25-10-2026

I. Introduction

Patient education is an important component of orthodontic care. Understanding the nature of malocclusion, the objectives and limitations of treatment, appliance care, oral hygiene, expected discomfort, treatment duration, and retention may help patients participate more effectively in their care. Conventional education commonly involves chairside explanations, printed material, and audiovisual resources. Although these approaches remain essential, they may not always provide immediate answers to every question that arises between appointments.

Conversational artificial intelligence (AI) systems can generate natural-language responses to patient questions and may offer an accessible supplement to conventional education. Their potential advantages include rapid access, conversational interaction, and the ability to rephrase information. However, response quality may vary across platforms, and fluent language does not guarantee clinical accuracy, completeness, or suitability for an individual patient. Recent orthodontic research has examined both the influence of AI-supported education on patient outcomes and the adequacy of chatbot responses to common orthodontic questions. These studies support further evaluation while emphasizing the continued importance of clinician oversight. [1,2]

Direct comparisons of several widely used general-purpose AI platforms in a patient-level educational setting may help clarify whether they are associated with similar short-term knowledge gains. The present study compared pre-intervention and post-intervention orthodontic knowledge scores among patients receiving education through Gemini, Meta AI, ChatGPT, Perplexity, or Grok.

Aim

To compare the effectiveness of five AI platforms—Gemini, Meta AI, ChatGPT, Perplexity, and Grok—in improving orthodontic patient knowledge, as measured by questionnaire scores before and after an educational intervention.

II. Materials and Methods

Study design and participants

The source material describes a comparative study involving 150 participants aged 14–30 years who were undergoing orthodontic treatment. The reported inclusion criteria were male and female patients aged 14–30 years undergoing orthodontic treatment. Exclusion criteria were previous orthodontic treatment, cognitive impairment or learning disability, and syndromic conditions. The study setting, recruitment dates, sampling approach, sample-size calculation, and participant flow were not specified in the supplied presentation and must be added from the original study records.

Group allocation and educational intervention

Participants were reported to have been randomly allocated to five groups ($n = 30$ per group): Group 1 received education through Gemini; Group 2 through Meta AI; Group 3 through ChatGPT; Group 4 through Grok; and Group 5 through Perplexity. The educational content, exact prompts, language used, duration of each session, device and account types, platform/model versions, date of testing, and any restrictions on follow-up questions were not documented in the supplied presentation. These details should be reported to allow reproducibility and to account for changes in AI systems over time.

Outcome assessment

Orthodontic knowledge was assessed before and after the educational session using the same questionnaire. The questionnaire's item content, number of questions, score range, validation process, language, and reliability were not included in the supplied presentation and should be provided in the final manuscript or supplementary material. Reuse of the same questionnaire may introduce a testing effect and should be acknowledged.

Statistical analysis

According to the source presentation, data were entered into Microsoft Excel and analyzed using IBM SPSS Statistics for Windows, version 21.0. Normality was assessed using the Shapiro–Wilk and/or Kolmogorov–Smirnov test. Descriptive statistics were reported as means and standard deviations. One-way analysis of variance followed by Tukey post hoc comparisons was described for between-group analysis, with a significance threshold of $p \leq 0.05$. The presentation also reports $p < 0.001$ for pre–post comparisons within each group. The reported t statistic is identical across all five groups; this should be checked against the original statistical output. The final manuscript should report the appropriate paired test for within-group change, the ANOVA F statistic and degrees of freedom, exact p values, effect sizes, and confidence intervals.

Ethics

[Insert the name of the institutional ethics committee, approval number and date, informed-consent procedures, and details of parental consent/participant assent for participants younger than 18 years. Do not submit until these details have been confirmed. If the study was exempt, include the formal exemption determination.]

III. Results

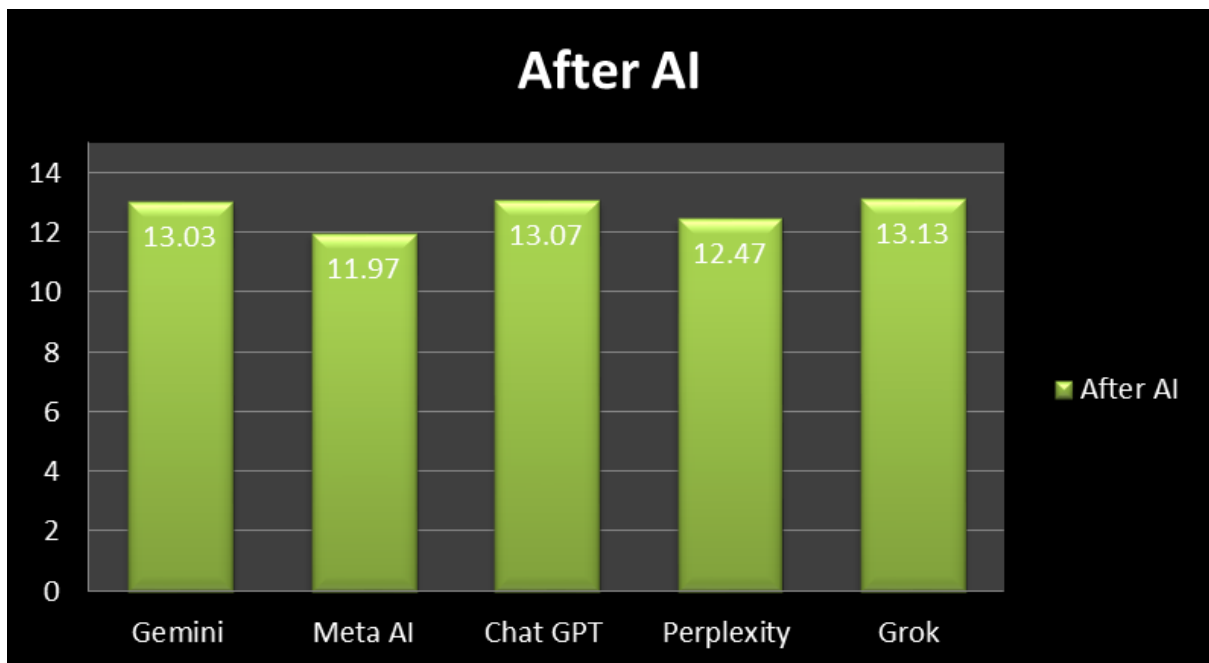
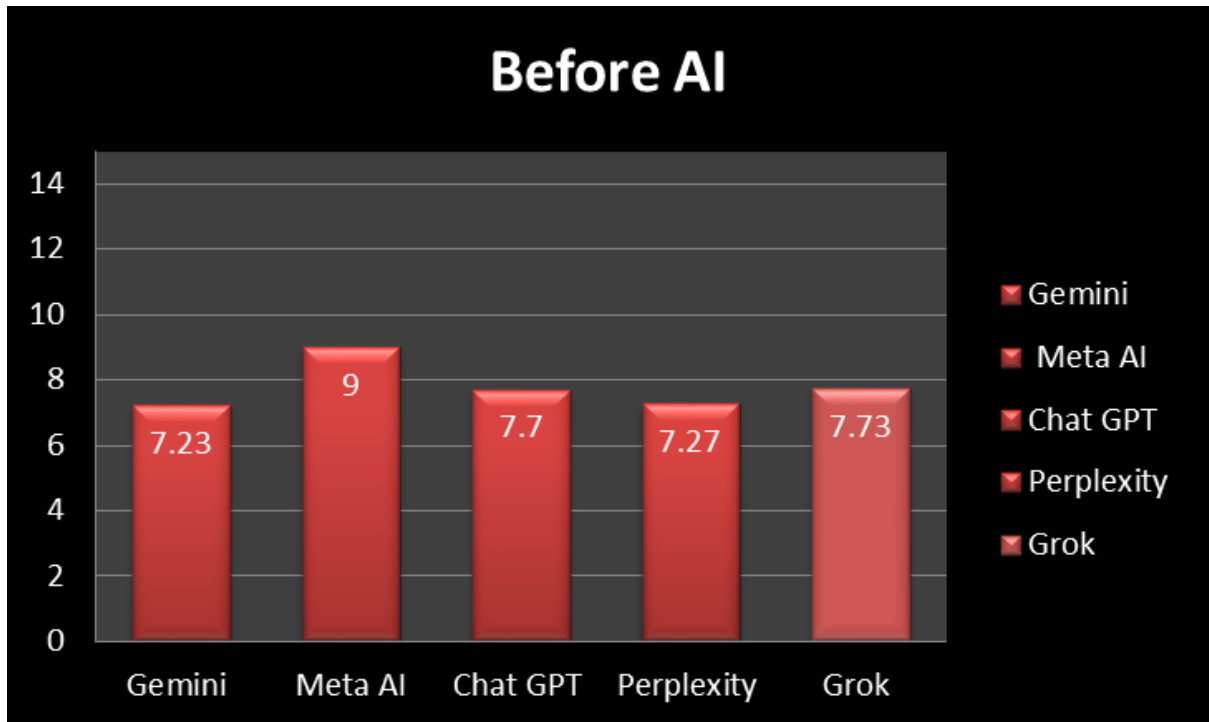
A total of 150 participants were reported across five groups, with 30 participants in each group. Participant demographic characteristics and a participant-flow diagram were not available in the supplied presentation.

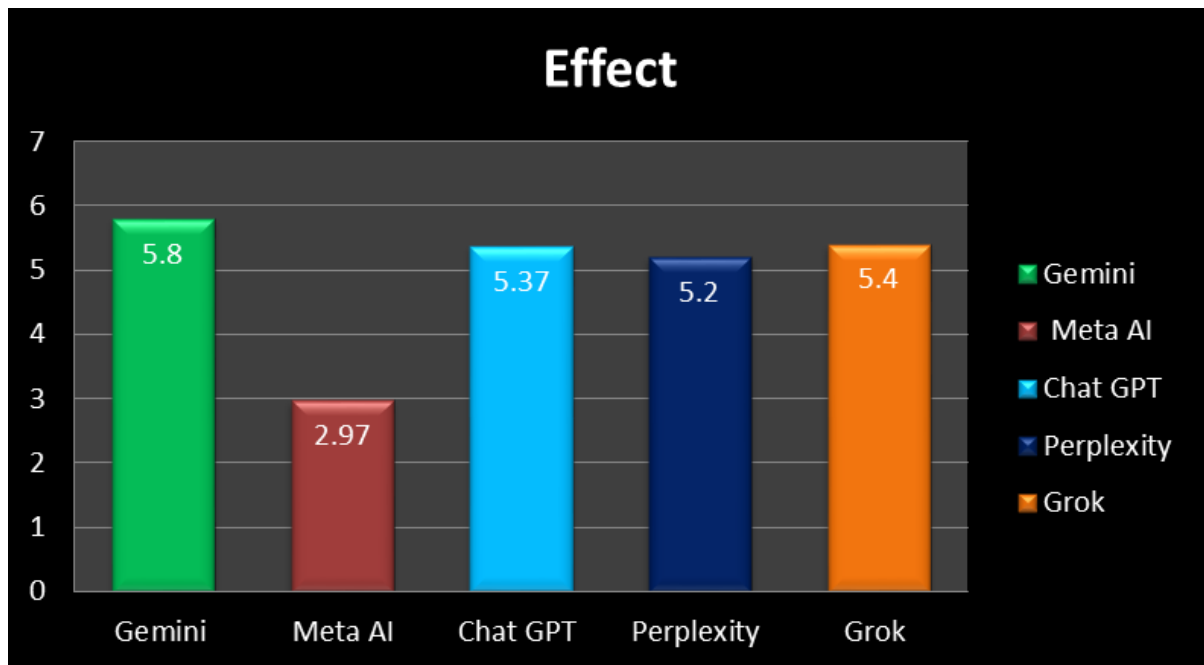
Baseline mean knowledge scores were highest in the Meta AI group (9.00 ± 2.27), followed by Grok (7.73 ± 1.26), ChatGPT (7.70 ± 2.64), Perplexity (7.27 ± 1.55), and Gemini (7.23 ± 1.59). These baseline differences are important when interpreting post-intervention scores and change scores.

AI platform (n = 30)	Pre-intervention mean $\pm$ SD	Post-intervention mean $\pm$ SD	Mean change $\pm$ SD
Gemini	7.23 ± 1.59	13.03 ± 1.75	5.80 ± 1.96
Meta AI	9.00 ± 2.27	11.97 ± 1.90	2.97 ± 1.87
ChatGPT	7.70 ± 2.64	13.07 ± 1.64	5.37 ± 2.80
Perplexity	7.27 ± 1.55	12.47 ± 1.43	5.20 ± 2.19
Grok	7.73 ± 1.26	13.13 ± 1.38	5.40 ± 1.61

Table 1. Knowledge scores before and after AI-assisted orthodontic education, as reported in the supplied presentation. SD = standard deviation. Confirm all values against the original dataset before submission.

Mean scores increased descriptively in every group. The largest reported mean change was in the Gemini group (5.80 ± 1.96), followed by Grok (5.40 ± 1.61), ChatGPT (5.37 ± 2.80), Perplexity (5.20 ± 2.19), and Meta AI (2.97 ± 1.87). Post-intervention mean scores were highest for Grok (13.13 ± 1.38), followed by ChatGPT (13.07 ± 1.64), Gemini (13.03 ± 1.75), Perplexity (12.47 ± 1.43), and Meta AI (11.97 ± 1.90). The source presentation reports $p < 0.001$ for each within-group pre–post comparison. However, it does not provide sufficient test output to independently verify these p values or to determine which between-group differences were statistically significant.





IV. Discussion

In this five-platform comparison, mean orthodontic knowledge scores increased after the educational intervention in all groups. The observed pattern suggests that AI-mediated explanations may help patients engage with basic orthodontic information. The finding is clinically relevant because orthodontic treatment requires sustained patient participation, including appliance care, oral hygiene, adherence to instructions, and understanding of retention.

The largest descriptive improvement was observed in the Gemini group, whereas the highest post-intervention mean score was observed in the Grok group. These are different outcomes and should not be treated as interchangeable. Moreover, the Meta AI group had the highest baseline score but the smallest reported mean improvement. Differences in baseline knowledge, ceiling effects, educational background, and random variation could contribute to this pattern. A valid comparison of effectiveness requires analysis of the between-group differences in change, with confidence intervals and appropriate adjustment for baseline scores.

The findings are consistent with the broader premise that conversational AI may have a supplementary role in orthodontic education, but they should be interpreted alongside studies evaluating response quality and patient outcomes. Santonocito and colleagues reported that a guideline-informed chatbot was associated with improved oral-hygiene-related outcomes compared with information leaflets in a preliminary randomized trial. [1] In a separate study, Salmanpour and colleagues found that expert responses were generally rated more highly than ChatGPT-4.0 and Microsoft Copilot responses by orthodontists and patients, underscoring that perceived usefulness and accessibility do not necessarily establish equivalence to expert advice. [2] These studies assess different interventions and outcomes, so their results should not be directly equated with the present knowledge-score comparison.

Several methodological issues should be addressed before firm conclusions are drawn. First, the same questionnaire was used before and after the intervention, which may increase scores through recall or familiarity. Second, the source material does not describe questionnaire validation, reliability, the educational prompts, platform versions, or the intervention duration. These details are particularly important because AI platforms are updated frequently and may generate different answers in response to the same question at different times. Third, the reported baseline scores differed across groups. Finally, the statistical summary contains an identical t statistic for all groups and does not include the full ANOVA or post hoc results. These issues preclude a reliable claim that one platform was statistically superior.

From a clinical perspective, AI tools may be useful for reinforcing routine explanations, providing plain-language summaries, and helping patients formulate questions for their orthodontist. They should not be used as autonomous sources of diagnosis, individualized treatment planning, medication advice, or management of orthodontic emergencies. Clinicians should review educational content for accuracy, explain case-specific decisions directly, and encourage patients to contact the practice when symptoms or appliance problems arise.

Limitations

The study evaluated short-term questionnaire scores and did not assess long-term knowledge retention, behavioral change, oral-hygiene outcomes, treatment adherence, or patient satisfaction. The supplied presentation does not report the questionnaire's validity or reliability, demographic comparability between groups, exact AI model versions, standardized prompts, duration of exposure, or detailed inferential statistics. The sample was limited to 150 participants within a specified age range and may not represent all orthodontic patients. The use of the same pre- and post-intervention questionnaire may have introduced a testing effect. These limitations should be revised to reflect the complete study records.

Clinical Implications

AI platforms may be considered supplementary channels for reinforcing general orthodontic information, provided that the information is checked and the patient is directed back to the treating clinician for individualized advice. Practices considering AI-supported education should establish a process for reviewing content, protecting patient privacy, documenting the platform and version used, and identifying questions that require direct clinician assessment.

V. Conclusion

Knowledge scores increased after educational sessions delivered through Gemini, Meta AI, ChatGPT, Perplexity, and Grok. Gemini showed the largest reported mean improvement, while Grok had the highest post-intervention mean score. Because baseline scores varied and the available report does not provide complete, verifiable between-group statistical results, the data do not establish the superiority of any single platform. AI tools may complement orthodontic patient education but should not replace individualized explanations and professional clinical judgment.

References

- [1]. Santonocito S, Cicciù M, Ronsivalle V. Evaluation of the impact of AI-based chatbot on orthodontic patient education: a preliminary randomised controlled trial. *Clin Oral Investig*. 2025;29:278. doi:10.1007/s00784-025-06356-8.
- [2]. Salmanpour F, Camcı H, Geniş Ö. Comparative analysis of AI chatbot (ChatGPT-4.0 and Microsoft Copilot) and expert responses to common orthodontic questions: patient and orthodontist evaluations. *BMC Oral Health*. 2025;25:896. doi:10.1186/s12903-025-06194-w.