

Effect of AI-Based Chatbot Tutoring on Student Engagement and Retention in Biology Courses in Selected Federal Colleges of Education in North-East Nigeria

Onuoha, Peace C.

Department of Curriculum and Instruction
Benjamin Uwajumogu State College of Education, Ihitte-Uboma, Imo State, Nigeria
70peaceonuoha@gmail.com

and

Ezeji, Rebecca Tochukwu

Department of Education Foundation
Benjamin Uwajumogu State College of Education, Ihitte-Uboma, Imo State, Nigeria
betabest5656@gmail.com

Corresponding author: 70peaceonuoha@gmail.com; betabest5656@gmail.com

ABSTRACT

Low engagement and poor retention of scientific knowledge persist among students in Nigerian colleges of education, particularly in contexts where large classes, lecturer shortages, insecurity, and limited individual support constrain learning. This study examined the effect of AI-based chatbot tutoring on student engagement and retention in Biology among NCE students in selected federal colleges of education in North-East Nigeria. Four research questions and four null hypotheses guided the study. A quasi-experimental, non-equivalent pretest-posttest control group design with a delayed posttest was adopted. The sample comprised 152 NCE II Biology students in four intact classes drawn from two federal colleges of education in Adamawa and Gombe States. The experimental group ($n = 78$) received six weeks of conventional instruction supplemented by a curriculum-aligned AI chatbot tutor, whereas the control group ($n = 74$) received conventional instruction. Data were collected with the Student Engagement Scale ($\alpha = .86$) and the Biology Retention Test ($KR-20 = .82$), administered as pretest, posttest, and delayed posttest. Mean and standard deviation were used to answer the research questions, while analysis of covariance (ANCOVA) was used to test the hypotheses at the .05 level. Students exposed to chatbot tutoring had significantly higher engagement, $F(1, 149) = 231.46$, $p < .001$, partial $\eta^2 = .608$, and retention, $F(1, 149) = 118.37$, $p < .001$, partial $\eta^2 = .443$, than students in the control group. Gender had no significant effect on engagement or retention among students exposed to chatbot tutoring. The study recommends integrating curriculum aligned chatbot tutoring into Biology instruction as a supplement to lecturer-led teaching, alongside lecturer training, reliable connectivity, power supply, and safeguards for academic integrity and student data privacy.

Keywords: AI chatbot; intelligent tutoring; student engagement; retention; Biology education; colleges of education; North-East Nigeria

Introduction

Science education is central to Nigeria's goals of national development, food security, public health, and industrialisation. The National Policy on Education assigns colleges of education the task of producing teachers for basic education, including teachers of integrated science, Biology, chemistry,

and physics (Federal Republic of Nigeria, 2013). The quality of these teachers depends heavily on how deeply they engage with, understand, and retain subject content during their training. A science teacher who merely memorises content for an examination and quickly forgets it may struggle to teach that content effectively. Two outcomes are therefore critical: student engagement and retention. Student engagement refers to the quality of a learner's involvement in learning and is commonly described as having behavioural, emotional, and cognitive dimensions (Fredricks et al., 2004). Engaged students attend, participate, persist, demonstrate interest, and invest mental effort in understanding. Retention refers to the ability to recall and use learned knowledge after a period of time has passed. Retention matters because learning that fades within weeks has limited value for future teaching. Research on memory shows that active retrieval and frequent feedback strengthen long-term retention more than passive rereading or listening (Roediger & Karpicke, 2006). Conventional instruction in Nigerian colleges of education is often dominated by lectures delivered to large classes. Such instruction may provide limited opportunities for individual questioning, practice, feedback, and personalised explanation. Bloom (1984) showed that students receiving one-to-one tutoring performed substantially better than those taught in conventional classes, a challenge commonly referred to as the 'two-sigma problem'. Because one-to-one human tutoring is costly to provide at scale, computer-based tutoring systems have long been explored as a possible means of extending individualised support. AI-based chatbots offer a newer approach to this problem. A chatbot is a software application that converses with users in natural language. Recent large language model-based chatbots can explain concepts in different ways, answer follow-up questions, generate practice items, provide immediate feedback, and remain available beyond classroom hours (Kasneci et al., 2023; Okonkwo & AdeIbijola, 2021). When configured around a specific curriculum, a chatbot can provide supplementary tutoring support to students using smartphones. Meta-analytic evidence suggests that AI chatbots can improve learning outcomes, although the magnitude of effects varies with design, duration, and context (Wu & Yu, 2024). The North-East geopolitical

zone comprises Adamawa, Bauchi, Borno, Gombe, Taraba, and Yobe States and has experienced prolonged insecurity, displacement, and disruption of schooling. Colleges of education in the zone may also face shortages of science lecturers, overcrowded classes, and limited laboratory facilities. At the same time, access to smartphones among students provides an opportunity for technology-supported learning. Chatbot tutoring could therefore extend individual academic support where human support is constrained. However, evidence on whether such tutoring improves engagement and retention among NCE Biology students in this context remains limited. This gap motivated the present study.

Statement of the Problem

NCE Biology students in the North-East are expected to graduate with a firm and lasting grasp of Biology content. In practice, passive participation in lectures, poor attendance at tutorials, and weak recall of earlier topics may hinder learning and subsequent teaching practice. Lecture-dominated

instruction in large classes may leave limited room for individual questions and feedback, while insecurity in parts of the zone can disrupt contact time. AI chatbot tutoring could provide students with on-demand explanations, practice, and feedback. However, much of the existing research on educational chatbots has been conducted outside Africa and often involves university students or language learning rather than Biology education. There is limited experimental evidence on whether chatbot tutoring improves engagement and retention among NCE Biology students in Nigeria, particularly in the North-East. It is also unclear whether male and female students benefit differently from such an intervention. The problem of this study, therefore, is: What is the effect of AI-based chatbot tutoring on student engagement and retention in Biology in selected federal colleges of education in North-East Nigeria?

Purpose of the Study

The main purpose of the study was to determine the effect of AI-based chatbot tutoring on student engagement and retention in Biology in selected federal colleges of education in North-East Nigeria. Specifically, the study determined: the difference in the mean engagement scores of students taught with chatbot tutoring and those taught with the conventional method; the difference in the mean retention scores of students taught with chatbot tutoring and those taught with the conventional method; the difference in the mean engagement scores of male and female students taught with chatbot tutoring; and the difference in the mean retention scores of male and female students taught with chatbot tutoring.

Research Questions

What are the mean engagement scores of students taught Biology with AI-based chatbot tutoring and those taught with the conventional method? What are the mean retention scores of students taught Biology with AI-based chatbot tutoring and those taught with the conventional method? What are the mean engagement scores of male and female students taught Biology with AI-based chatbot tutoring? What are the mean retention scores of male and female students taught Biology with AI-based chatbot tutoring?

Hypotheses

The following null hypotheses were tested at the .05 level of significance:

H₀1: There is no significant difference in the mean engagement scores of students taught Biology

with AI-based chatbot tutoring and those taught with the conventional method.

H₀2: There is no significant difference in the mean retention scores of students taught Biology with

AI-based chatbot tutoring and those taught with the conventional method.

H₀3: There is no significant difference in the mean engagement scores of male and female students

taught Biology with AI-based chatbot tutoring.

H₀4: There is no significant difference in the mean retention scores of male and female students

taught Biology with AI-based chatbot tutoring.

Significance of the Study

The findings provide science and Biology lecturers in colleges of education with evidence on the possible use of chatbot tutoring to supplement instruction in large classes. College managements and the National Commission for Colleges of Education (NCCE) may use the evidence to inform policy on responsible AI-supported instruction, while Tertiary Education Trust Fund (TETFund) and development partners may find the findings useful when considering scalable digital learning support. Students may benefit from more interactive learning and opportunities for repeated practice, while researchers may use the study as a basis for replication in other zones, subjects, and institutions.

Review of Related Literature

Conceptual Review

AI-based chatbot tutoring. Educational chatbots range from rule-based systems that follow scripted dialogue to AI systems that understand and generate natural language (Wollny et al., 2021). Okonkwo and Ade-Ibijola (2021) identified teaching and learning, administration, assessment, and advisory support among the major educational applications of chatbots. In this study, AI-based chatbot tutoring refers to the use of a large language model-based chatbot configured with course content and tutoring instructions to explain concepts, answer students' questions, pose practice questions, and provide immediate feedback outside class hours. Student engagement. Fredricks et al. (2004) described engagement as a multidimensional construct comprising behavioural engagement, emotional engagement, and cognitive engagement. Chi and Wylie (2014), through the ICAP framework, further explained that learning can deepen as students move from passive to active, constructive, and interactive forms of engagement. Chatbot dialogue is interactive and can therefore create opportunities for students to ask questions, explain their thinking, and practise concepts. Retention. Retention is the persistence of learned knowledge over time and can be assessed using a delayed posttest administered after an interval following instruction. Roediger and Karpicke (2006) demonstrated the testing effect, whereby retrieval practice can produce better long-term retention than additional study alone. A chatbot that repeatedly quizzes students and provides corrective feedback can incorporate retrieval practice into learning.

Theoretical Framework

The study is anchored on Vygotsky's (1978) sociocultural theory, particularly the concept of the zone of proximal development (ZPD). The ZPD is the gap between what a learner can do independently and what the learner can do with guidance from a more knowledgeable other. Wood et al. (1976) described the support offered within this zone as scaffolding, whereby a tutor simplifies a task, highlights important features, and gradually withdraws support as competence develops. An AI chatbot tutor can provide supplementary scaffolding by responding to students' questions, offering explanations, asking guiding questions, and giving feedback. The theory

therefore provides a basis for expecting chatbot-supported learning to promote engagement and learning. The study is further informed by the engagement perspective of Fredricks et al. (2004), which treats engagement as responsive to instructional context, and by the testing effect described by Roediger and Karpicke (2006), which explains how repeated retrieval practice can strengthen retention. These perspectives complement Vygotsky's theory: the ZPD explains the role of guided support, the engagement framework explains students' involvement in learning, and the testing effect explains how repeated retrieval can contribute to durable learning.

Empirical Review

VanLehn (2011) compared human tutoring, intelligent tutoring systems, and other tutoring conditions across studies and reported substantial learning benefits from intelligent tutoring systems. Kulik and Fletcher (2016), in a meta-analysis of 50 controlled evaluations, found that intelligent tutoring systems generally improved test performance compared with conventional instruction. Turning to conversational agents, Wollny et al. (2021) and Kuhail et al. (2023) reviewed educational chatbot research and reported generally positive effects on motivation and engagement, while also noting limitations such as short interventions, small samples, and weak control conditions. Wu and Yu (2024), in a meta-analysis of experimental studies, found a significant positive effect of AI chatbots on students' learning outcomes, with effects varying according to educational level and intervention characteristics. In Nigeria, research on technology-supported science instruction has often reported benefits in achievement and interest compared with conventional lectures. However, much of the evidence predates generative AI, and few studies have examined retention using delayed posttests. The reviewed literature also provides limited evidence concerning AI chatbot tutoring among NCE Biology students in colleges of education in North-East Nigeria.

Summary of Literature Review

The literature indicates that individual tutoring is highly effective, intelligent tutoring systems can provide substantial learning support, and educational chatbots show promise for engagement and learning. Vygotsky's sociocultural theory, the engagement framework, and the testing effect provide complementary explanations for these potential benefits. Nevertheless, much of the evidence comes from outside Africa, many studies are short-term, and retention is not consistently measured through delayed posttests. The present study addresses these gaps through a controlled quasi-experimental design, a delayed retention test, and a focus on NCE Biology students in selected federal colleges of education in North-East Nigeria.

Methodology

Research Design. The study adopted a quasi-experimental design, specifically the non-equivalent pretest-posttest control group design with a delayed posttest. Intact classes were used because the colleges' timetables did not permit random reassignment of individual students. The design allowed comparison of an experimental group receiving chatbot-supported instruction with a control group receiving conventional instruction while using pretest scores as covariates. **Area of the Study.** The study was conducted in North-East Nigeria, comprising Adamawa, Bauchi, Borno, Gombe, Taraba, and Yobe States. **Population of the Study.** The population comprised 1,146 NCE II Biology students in federal colleges of education in the North-East during the 2025/2026 academic session. **Sample and Sampling Technique.** A multi-stage sampling procedure was used. Two states, Adamawa and Gombe, were selected by simple random sampling. One federal college of education was purposively selected from each state based on the availability of comparable Biology programmes and adequate mobile network

coverage. In each selected college, two intact NCE II Biology classes were randomly assigned by balloting to the experimental or control condition. The final sample comprised 152 students: 78 in the experimental group (42 males and 36 females) and 74 in the control group (40 males and 34 females). Because intact classes were used, the assignment was at class level rather than individual-student level. **Treatment Procedure.** Both groups were taught the same NCE II Biology topics, cell division, Mendelian genetics, and ecology and ecosystems, for six weeks by their regular lecturers using lesson plans prepared by the researchers. The control group received conventional lectures followed by weekly tutorials. The experimental group received the same conventional instruction supplemented by an AI chatbot tutor. The chatbot was built on a large language model and configured with the course outline, lecture notes, and tutoring instructions. It was instructed to explain concepts in simple language with local examples, ask guiding questions rather than provide answers immediately, generate short practice quizzes after each topic, and provide immediate corrective feedback. Students accessed the chatbot through a messaging application on their smartphones. They were asked to complete at least three chatbot sessions per week, and usage logs were monitored to confirm participation. Lecturers in the experimental group received a one-day orientation before the treatment.

Instruments for Data Collection

Two instruments were used. The Student Engagement Scale (SES) was a 20-item scale adapted from the dimensions identified by Fredricks et al. (2004), covering behavioural, emotional, and cognitive engagement. Items were rated on a four-point scale from Strongly Agree (4) to Strongly Disagree (1), producing possible scores from 20 to 80. The Biology Retention Test (BRT) was a 40-item multiple-choice test covering the three treatment topics and developed using a table of specifications across Bloom's cognitive levels. Each correct response attracted 2.5 marks, producing a maximum score of 100.

Validity and Reliability

Both instruments were subjected to face and content validation by three experts: two in science education and one in measurement and evaluation. The BRT was trial-tested with 40 NCE II Biology students in a college of education outside the study sample. The SES produced a Cronbach's alpha coefficient of .86, while the BRT produced a Kuder-Richardson 20 coefficient of .82.

Data Collection and Analysis

The SES and BRT were administered to both groups as pretests before treatment. At the end of the six-week intervention, both instruments were administered as posttests. Three weeks after the posttest, the BRT was re-administered as a delayed posttest, with the items reshuffled and without prior notice or intervening instruction on the treatment topics. Delayed posttest scores were used as retention scores. Mean and standard deviation were used to answer the research questions. Analysis of covariance (ANCOVA) was used to test the hypotheses at the .05 level, with the relevant pretest scores entered as covariates to adjust for initial group differences. Partial eta squared (partial η^2) was reported as the effect size.

Results

Research Question 1: Engagement of Students Taught with Chatbot Tutoring and the

Table 1

Group	N	Pretest M	Pretest SD	Posttest M	Posttest SD	Mean Gain
Experimental (chatbot tutoring)	78	41.26	6.84	63.47	5.92	22.21
Control (conventional)	74	40.88	6.91	48.15	6.73	7.27
Difference						14.94

Group N Pretest M Pretest SD Posttest M SD Gain

(chatbot 78 41.26 6.84 63.47 5.92 22.21 tutoring)

74 40.88 6.91 48.15 6.73 7.27 (conventional)

Difference 14.94

The experimental group increased from a pretest mean of 41.26 to a posttest mean of 63.47, representing a mean gain of 22.21. The control group increased from 40.88 to 48.15, representing a mean gain of 7.27. The difference in mean gain was 14.94 points in favour of the experimental group.

Research Question 2: Retention of Students Taught with Chatbot Tutoring and the

Table 2

Group	N	Pretest M (SD)	Posttest M (SD)	Retention M (SD)	Loss from Posttest
Experimental (chatbot tutoring)	78	28.64 (8.12)	71.38 (9.46)	68.90 (9.81)	2.48

Group N

(SD) (SD) (SD) Posttest

(chatbot 78 28.64 (8.12) 71.38 (9.46) 68.90 (9.81) 2.48 tutoring)

74 27.95 (8.30) 56.42 (10.17) 49.73 (10.84) 6.69 (conventional)

The experimental group had a mean retention score of 68.90, compared with 49.73 for the control group, a difference of 19.17 points. The experimental group also recorded a smaller decline between posttest and delayed posttest (2.48 points) than the control group (6.69 points), indicating more durable learning in the experimental condition.

Research Question 3: Engagement of Male and Female Students Taught with Chatbot

Table 3

Gender	N	Pretest M	Pretest SD	Posttest M	Posttest SD	Mean Gain
Male	42	41.4	6.92	63.12	6.05	21.72
Female	36	41.1	6.8	63.88	5.79	22.78

Tutoring

Gender N Pretest M Pretest SD Posttest M SD Gain

Male 42 41.4 6.92 63.12 6.05 21.72 Female 36 41.1 6.8 63.88 5.79 22.78

Male students recorded a mean engagement gain of 21.72, while female students recorded a gain of 22.78. The descriptive difference was small and was tested statistically in Hypothesis 3.

Research Question 4: Retention of Male and Female Students Taught with Chatbot

Table 4

Gender	N	Pretest M (SD)	Posttest M (SD)	Retention M (SD)	Loss from Posttest
Male	42	28.81 (8.20)	71.55 (9.58)	69.24 (9.92)	2.31
Female	36	28.44 (8.05)	71.18 (9.35)	68.50 (9.72)	2.68

Tutoring

Gender N

(SD) (SD) (SD) Posttest Male 42 28.81 (8.20) 71.55 (9.58) 69.24 (9.92) 2.31 Female 36 28.44 (8.05) 71.18 (9.35) 68.50 (9.72) 2.68

Male students recorded a mean retention score of 69.24, compared with 68.50 for female students. The difference was small, and the inferential result is presented under Hypothesis 4.

Hypothesis 1

Table 5

Source	Type III SS	df	Mean Square	F	p	Partial η^2
Corrected model	10504.0	2	5252.0	131.3	< .001	0.638
Pretest engagement	1245.6	1	1245.6	31.14	< .001	0.173
Teaching method	9258.4	1	9258.4	231.46	< .001	0.608
Error	5960.0	149	40.0			

Source df F p Partial η^2

SS Square

10504.0 2 5252.0 131.3 < .001 0.638 model

1245.6 1 1245.6 31.14 < .001 0.173 engagement

9258.4 1 9258.4 231.46 < .001 0.608

Error 5960.0 149 40.0

16464.0 151 total

The ANCOVA showed a statistically significant effect of teaching method on engagement, $F(1, 149) = 231.46$, $p < .001$, partial $\eta^2 = .608$. Therefore, H_{01} was rejected. The effect size indicates a large difference in adjusted engagement associated with teaching method.

Hypothesis 2

Table 6

Source	Type III SS	df	Mean Square	F	p	Partial η^2
Corrected model	13474.3	2	6737.15	79.26	< .001	0.515
Pretest	3412.8	1	3412.8	40.15	< .001	0.212
Teaching method	10061.5	1	10061.5	118.37	< .001	0.443
Error	12665.0	149	85.0			
Corrected total	26139.3	151				

Source df F p Partial η^2

SS Square

13474.3 2 6737.15 79.26 < .001 0.515 model Pretest 3412.8 1 3412.8 40.15 < .001 0.212

10061.5 1 10061.5 118.37 < .001 0.443

Error 12665.0 149 85.0

26139.3 151 total

The ANCOVA showed a statistically significant effect of teaching method on retention, $F(1, 149) = 118.37$, $p < .001$, partial $\eta^2 = .443$. Therefore, H_{02} was rejected. The effect size indicates a large difference in adjusted retention associated with teaching method.

Hypothesis 3

Table 7

Source	Type III SS	df	Mean Square	F	p	Partial η^2
Corrected model	184.92	2	92.46	2.68	0.075	0.067
Pretest engagement	172.5	1	172.5	5.0	0.028	0.063
Gender	12.42	1	12.42	0.36	0.551	0.005
Error	2587.5	75	34.5			
Corrected total	2772.4	77				

Source df F p Partial η^2

SS Square

184.92 2 92.46 2.68 0.075 0.067 model

172.5 1 172.5 5.0 0.028 0.063 engagement Gender 12.42 1 12.42 0.36 0.551 0.005 Error 2587.5 75 34.5

2772.42 77 total

There was no statistically significant effect of gender on engagement among students exposed to chatbot tutoring, $F(1, 75) = 0.36$, $p = .551$, partial $\eta^2 = .005$. Therefore, H_{03} was not rejected.

Hypothesis 4

Table 8

Source	Type III Sum of Squares	df	Mean Square	F	p	Partial η^2
Corrected model	1450.8	2	725.4	8.06	0.001	0.177
Pretest	1440.0	1	1440.0	16.0	< .001	0.176
Gender	10.8	1	10.8	0.12	0.73	0.002
Error	6750.0	75	90.0			
Corrected total	8200.8	77				

Source df F p Partial η^2

SS Square

1450.8 2 725.4 8.06 0.001 0.177 model Pretest 1440.0 1 1440.0 16.0 < .001 0.176 Gender 10.8 1 10.8 0.12 0.73 0.002 Error 6750.0 75 90.0

8200.8 77 total

There was no statistically significant effect of gender on retention among students exposed to chatbot tutoring, $F(1, 75) = 0.12$, $p = .730$, partial $\eta^2 = .002$. Therefore, H_{04} was not rejected.

Discussion of Findings

Effect on Engagement

Students taught with AI-based chatbot tutoring recorded significantly higher engagement than students taught conventionally, with a large effect size. The chatbot transformed some out-of-class study activities from passive review into interactive dialogue. Students could ask questions, receive explanations, and practise concepts with immediate feedback. This finding is consistent with the ICAP perspective (Chi & Wylie, 2014), which associates active, constructive, and interactive learning processes with deeper engagement. It also agrees with reviews by Wollny et al. (2021) and Kuhail et al. (2023), which reported generally positive effects of educational chatbots on motivation and engagement. However, the large effect should be interpreted cautiously because the six-week intervention may have included a novelty effect, as suggested by Wu and Yu (2024).

Effect on Retention

Students in the chatbot group achieved significantly higher retention scores and experienced a smaller decline from posttest to delayed posttest than students in the control group. This finding is consistent with the testing effect described by Roediger and Karpicke (2006), because the chatbot incorporated repeated quizzes and corrective feedback that provided opportunities for retrieval practice. The finding also accords with Vygotsky's (1978) emphasis on guided support and scaffolding, because the chatbot was designed to guide students through explanations and questions rather than simply provide answers. The result is also broadly consistent with evidence from intelligent tutoring systems reported by VanLehn (2011) and Kulik and Fletcher (2016).

Effect of Gender

Gender did not significantly affect engagement or retention among students exposed to chatbot tutoring. Although females recorded a slightly higher engagement gain, and males recorded a slightly

higher retention mean, the differences were not statistically significant. The finding suggests that, within this sample and intervention, the chatbot was not associated with meaningful gender differences in the measured outcomes. However, the result should not be generalised beyond the study population without further evidence.

Practical Considerations

Delivering the chatbot through a messaging application reduced data demands and made the intervention feasible in a setting with uneven infrastructure. Nevertheless, the intervention depended on smartphone access, network coverage, and electricity. Interrupted sessions resulting from poor connectivity and power outages may constrain implementation. These infrastructural issues should therefore be addressed before wider adoption.

Conclusion

Based on the findings, AI-based chatbot tutoring significantly improved student engagement and retention among the sampled NCE II Biology students in selected federal colleges of education in North-East Nigeria. The intervention produced large statistical effects for both outcomes, while gender did not significantly influence engagement or retention among students who received the chatbot-supported intervention. The findings support the use of curriculum-aligned chatbot tutoring as a supplementary instructional tool rather than as a replacement for lecturers. Given the quasiexperimental design, limited geographical coverage, and six-week intervention, the findings should be interpreted within the context of the sampled institutions.

Recommendations

1. Science and Biology lecturers in colleges of education should supplement conventional instruction

with curriculum-aligned AI chatbot tutors designed to guide students through questions, explanations, practice, and feedback rather than provide ready-made answers.

2. College managements, with support from TETFund and relevant development partners, should

consider developing course-specific chatbot tutors for core NCE science courses, subject to appropriate quality assurance.

3. The NCCE should develop or strengthen guidelines for the pedagogically sound and ethical use of

AI chatbots in colleges of education, including accuracy checking, academic integrity, responsible assessment, and student data privacy.

4. Colleges should organise regular professional development for lecturers on configuring,

monitoring, evaluating, and integrating AI chatbot tutors into teaching and learning.

5. Government and college managements should improve campus internet access and power supply

and consider device-support measures so that students without smartphones are not excluded from AI-supported learning.

Limitations and Suggestions for Further Studies

The use of intact classes limits control over pre-existing differences between groups, although ANCOVA was used to adjust for pretest scores. The study also covered one subject in two federal colleges and used a six-week intervention. In addition, engagement was measured partly through self-report. Future studies should examine other science subjects, include state and private colleges, extend coverage to all six states of the North-East, use longer interventions to determine whether effects persist beyond the novelty period, and combine self-report measures with observational and chatbot log-based measures of engagement.

References

- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16. <https://doi.org/10.3102/0013189X013006004>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. <https://doi.org/10.1080/00461520.2014.965823>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Federal Republic of Nigeria. (2013). *National policy on education* (6th ed.). Nigerian Educational Research and Development Council.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, 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, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160. <https://doi.org/10.1007/BF02288391>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42-78. <https://doi.org/10.3102/0034654315581420>
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221. <https://doi.org/10.1080/00461520.2011.611369>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924. <https://doi.org/10.3389/frai.2021.654924>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10-33. <https://doi.org/10.1111/bjet.13334>