

EFFECT OF TEACHER-GUIDED USE OF AI-CHATBOTS ON SECONDARY SCHOOL STUDENTS' ACADEMIC ACHIEVEMENT IN CHEMISTRY IN ONITSHA EDUCATION ZONE

Pius, Promise Obinna

Alvan Ikoku Federal University of Education

promisepius@gmail.com

Abstract

The persistent underachievement of secondary school students in chemistry, particularly in abstract concepts requiring visualization and active inquiry, necessitates investigation into innovative, technology-embedded pedagogies. This study examined the effect of teacher-guided use of AI-chatbots (TGUAC) on secondary school students' academic achievement in chemistry, with gender as a moderating variable. A quasi-experimental, non-randomized pre-test post-test control group design was adopted. The sample comprised 51 SS2 chemistry students drawn from two intact classes in two purposively selected co-educational public secondary schools from Onitsha Education Zone, Anambra State, Nigeria. The experimental group received instruction through TGUAC using ChatGPT-3.5, while the control group received conventional lecture-demonstration methods. The Chemistry Achievement Test, a 40-item multiple-choice instrument validated by three experts and reliability- coefficient of 0.81 using Kuder-Richardson formula 20, served as the data collection instrument. Data were analysed using mean, standard deviation, and Analysis of Covariance at 0.05 alpha level. Findings revealed that students taught using TGUAC achieved significantly higher mean achievement scores than those taught using lecture methods. Gender had no significant effect on achievement, and there was no significant interaction between instructional method and gender. The study concluded that teacher-guided use of AI-chatbots is an effective, gender-equitable instructional strategy for enhancing chemistry achievement. Recommendations include integrating TGUAC into secondary school chemistry curricula and ensuring equitable access to AI-chatbot platforms across student populations.

Keywords: Teacher-guided AI-chatbots, academic achievement, chemistry education, gender

Introduction

Science has transformed societies through continuous discoveries and innovations. This progress depends on effective science education that develops future scientists and critical thinkers. This exponential progress relies fundamentally on the deliberate cultivation of future scientists, engineers, and critically minded citizens, making science education essential to sustainable development. As a result, science education now extends beyond the transmission of facts to fostering scientific literacy, problem-solving skills, and the ability to understand complex abstract systems. It functions as the primary means of transmitting the norms and methods of the scientific community to society, thereby preparing a workforce capable of thriving in a knowledge-based global economy and reinforcing its importance in national development agendas. Within this framework, chemistry occupies a central yet challenging role, serving as the bridge between the physical and life sciences by explaining the composition, structure, properties, and transformations of matter. A meaningful understanding of biology depends heavily on chemical principles, yet students often encounter substantial cognitive difficulties when transitioning to chemistry due to its abstract concepts and the need to navigate macroscopic, submicroscopic, and symbolic representations. These cognitive demands frequently lead to misconceptions, low self-efficacy, and persistent underachievement, highlighting the limitations of traditional lecture-based methods that emphasize rote memorization over deep conceptual understanding.

Consequently, the call for educational reform has intensified, advocating strategic embedding of technology within pedagogical frameworks. The digital revolution has endowed educators with interactive simulations, virtual laboratories, learning management systems, and mobile applications designed to make learning more interactive, visualised, and personalised. Within this technological renaissance, artificial intelligence has emerged as a potentially transformative force in educational practice. Specifically, conversational AI and chatbots have garnered significant scholarly attention. Joyner (2024) defined conversational artificial intelligence as language-based AI tools—including chatbots, intelligent tutoring systems, and large language models—that facilitate interactive dialogue between machine and user for educational purposes. Joyner (2024) further conceptualised chatbots' instructional role through three frameworks: AI as Teacher, where the chatbot directly imparts knowledge or assesses understanding; AI as Partner, where the chatbot collaborates with the student to solve problems or explore concepts; and AI as Goal, where manipulation and understanding of the AI itself becomes a

learning objective. This conceptualisation positions chatbots as instructional agents capable of transforming classroom interaction. The present study's intervention—teacher-guided use of AI-chatbots—is understood as a hybrid of Joyner's AI as Teacher and AI as Partner frameworks, wherein the teacher intentionally scaffolds student interaction with the chatbot to optimise learning outcomes, ensuring technology complements rather than replaces pedagogical expertise.

The potential efficacy of AI chatbots in enhancing academic achievement in chemistry is increasingly substantiated by empirical research. Academic achievement, broadly defined as the extent to which students attain specified educational goals and master subject content, is frequently operationalised through standardised test scores and validated achievement instruments. Recent quasi-experimental investigations provide preliminary evidence supporting AI integration in chemistry pedagogy. Sajadi (2025), examining AI tools' effect on chemistry achievement, utilised ChatGPT chatbots augmented with Wolfram Alpha and ChemDraw plugins. Employing a quasi-experimental pre-test post-test control group design with twelfth-grade science students, Sajadi (2025) reported significant improvement in the experimental group's academic performance, concluding such chatbots are effective tools for promoting learning and engagement with chemistry concepts. Similarly, in chemical engineering education, Keith, Keiller, Windows-Yule, Kings, and Robbins (2025) implemented the IDEE framework to integrate large language models into laboratory sessions. Keith et al. (2025) found students developed ability to critique AI output and gained valuable insight into the technology's limitations and ethical use. Further supporting this trend, researchers at the Technical University of Denmark (2025) developed and evaluated ChatGMP, demonstrating that AI chatbots could effectively automate repetitive instructional tasks, providing information equivalent in quality to human teachers and enabling comparable learning outcomes. These findings collectively suggest AI-chatbots are viable instructional partners capable of supporting meaningful learning in chemical sciences. However, the extent to which such findings generalise to diverse educational contexts—particularly developing nations with unique infrastructural and sociocultural realities—remains an open empirical question.

Despite these technological promises, the effectiveness of any instructional strategy cannot be fully understood without considering learners' demographic characteristics. Among these, gender remains a persistent and critical variable in science education research. Historically, science, technology, engineering, and mathematics fields have been characterised by significant gender disparities, with

female students often exhibiting lower participation, confidence, and achievement, particularly in physical sciences. This disparity is not attributable to innate cognitive differences but largely to sociocultural factors including stereotype threat, differential teacher expectations, and lack of relatable role models. In chemistry education specifically, research has yielded inconsistent findings regarding gender disparities. While some studies continue reporting significant differences favouring male students in specific content areas or problem-solving contexts, others have found no significant gender gap, particularly when innovative and culturally responsive pedagogies are employed. Notably, Oladejo, Olateju, Okebukola, and Braimoh (2025), in a rigorous quasi-experimental study, explored the culturo-techno-contextual approach's potency in levelling the playing field in chemistry achievement. Oladejo et al. (2025) found no significant difference in post-test achievement scores of male and female students, concluding the approach was a viable tool for bridging gender differences in chemistry learning. This finding aligns with earlier research within the present study's geographical context: George and Abumchukwu (2021), investigating the generative learning model's impact on academic self-concept in chemistry in Onitsha Education Zone, similarly reported no significant gender difference in students' mean self-concept scores. These results do not negate gender disparities but suggest such disparities are malleable and mitigable through appropriately designed instructional strategies.

This body of evidence provides a compelling case for continued and explicit investigation of gender as a moderating variable in intervention studies. The fact that some pedagogies successfully eliminate gender gaps indicates that baseline instruction may inadvertently perpetuate or fail to address these inequities. Therefore, it is methodologically and ethically insufficient to measure merely the main effect of an intervention—such as teacher-guided use of AI-chatbots—on academic achievement without simultaneously testing whether the intervention interacts with student gender. Such interaction would indicate that the AI-chatbot strategy is differentially effective for male and female students. If the strategy is significantly more effective for one gender, it may inadvertently widen the very disparities the education system seeks to eliminate. Conversely, if equally effective for both genders, it provides empirical validation for its adoption as an equitable instructional practice. It is imperative to move beyond the null hypothesis of gender neutrality and actively test these effects to inform evidence-based policy and classroom practice. The persistence of gender-based stereotypes in science, particularly the pervasive narrative queried by Oladejo et al. (2025) in their titular question “I am Black and I am a girl-child, can I still be a scientist?”, necessitates that every pedagogical innovation be scrutinised for its potential to either reinforce

or dismantle these traditional narratives. Thus, the present study incorporates gender as a moderating variable to determine whether the effect of teacher-guided AI-chatbot use on chemistry achievement differs between male and female students in the Onitsha Education Zone.

It is against this complex backdrop of scientific imperative, pedagogical challenge, technological promise, and social equity that the present study is situated. The Onitsha Education Zone, a significant educational hub within Anambra State, Nigeria, presents a relevant and urgent context for this investigation. Recent local research has highlighted persistent issues within this zone. Nwankwo and Enwemeka (2025) examined the relationship between student self-concept and academic performance in chemistry, confirming that while innovative teaching methods are being explored regionally, student achievement in chemistry remains a primary concern for educators and policymakers. Similarly, George and Abumchukwu (2021) demonstrated the generative learning model's potential to enhance self-concept, yet the fundamental challenge of low academic achievement endures. The specific intervention of teacher-guided use of AI-chatbots, while demonstrated effective in other national contexts by Sajadi (2025), Keith et al. (2025), and the Technical University of Denmark (2025), has not been rigorously tested within the unique sociocultural and infrastructural realities of Nigerian secondary schools, nor specifically within the Onitsha Education Zone. Furthermore, the critical question of whether such intervention interacts with gender to produce differential academic outcomes remains entirely unexamined in this context. Given that previous local research by George and Abumchukwu (2021) found gender to be a non-significant factor under certain innovative conditions, it is plausible that AI-chatbots could similarly serve as an equalising pedagogical tool. However, this assumption must be empirically tested rather than assumed. Therefore, this study seeks to fill a significant gap in the literature by determining the effect of teacher-guided use of AI-chatbots on secondary school students' academic achievement in chemistry, with specific focus on the moderating influence of gender. The findings are expected to contribute to the global discourse on AI in education, provide localised empirical evidence for curriculum planners and science teachers in Anambra State, and offer insights into the design of equitable, technology-enhanced learning environments that serve all students irrespective of gender.

Purpose of the Study

The aim of this research was to examine effect of teacher-guided use of AI-chatbots on secondary school students' academic achievement in chemistry in

Onitsha education zone. Specifically, the study sought to determine the:

1. Mean achievement scores of students taught Chemistry using teacher-guided use of AI-Chatbots (TGUAC) and those taught using lecture method.
2. Mean achievement scores of male and female students taught Chemistry using TGUAC and those taught using the lecture method.
3. Interaction effects of teaching methods and gender on students' academic achievement in Chemistry.

Research Questions

1. What are the mean achievement scores of students taught Chemistry using teacher-guided use of AI-Chatbots (TGUAC) and with those taught using lecture method?
2. What are the mean achievement scores of male and female students taught Chemistry using TGUAC and with those taught using the lecture method?

Hypotheses

1. There is no significant difference in the mean achievement scores of students taught Chemistry using teacher-guided use of AI-Chatbots (TGUAC) and those taught using lecture method.
2. There is no significant difference in the mean achievement scores of male and female students taught Chemistry using TGUAC and with those taught using the lecture method.
3. There is no significant interaction effects of teaching methods and gender on students' academic achievement in Chemistry.

Method

This study adopted a quasi-experimental, non-randomized pre-test post-test control group design (Creswell, 2014). The study was conducted in Onitsha Education Zone, Anambra State, Nigeria. The population comprised all 4,286 senior secondary two (SS2) chemistry students secondary schools within Onitsha Education Zone. The sample consisted of 51 SS2 chemistry students drawn from two intact classes across two purposively selected co-educational public secondary schools that satisfied specific inclusion criteria: co-educational status, functional computer laboratories or reliable internet-enabled device access, qualified chemistry teachers willing to participate, and consistent SSCE presentation history. From each school, two intact classes were randomly assigned to experimental or control groups via simple ballot. The primary instrument was the Chemistry Achievement Test (CAT), a 40-item multiple-choice objective test with four response options measuring achievement in chemical bonding, mole concept,

stoichiometry, and periodic trends. Each correct response scored 2.5 mark, maximum score 100. Items were rearranged for post-test to minimize practice effects. The CAT was developed based on the SS2 chemistry scheme of work, standardized textbooks, and past West African Senior School Certificate Examination questions. The CAT was subjected to face and content validation by three experts: two chemistry education specialists and one measurement and evaluation specialist from a federal university. Reliability was established using Kuder-Richardson formula 20 which yielded a reliability coefficient of 0.81.

A Treatment Implementation Guide was developed to standardize the experimental group intervention, outlining teacher roles, chatbot interaction protocols, prompt types, and scaffolding strategies. A Conventional Lesson Plan Guide was developed for the control group to ensure identical content coverage without AI-chatbot integration. ChatGPT-3.5, configured for chemistry-related queries with explanations, worked examples, and guided inquiry prompts, served as the AI-chatbot. The six-week study commenced after formal permission was obtained from the school authorities and informed consent from parents and students. In week one, two chemistry teachers from experimental schools underwent a two-hours researcher-facilitated training workshop on AI-chatbot theoretical underpinnings, Joyner's (2024) teacher-guided framework, ChatGPT operation, scaffolding strategies, and troubleshooting.

In week one, the CAT pre-test was administered under standard examination conditions to both groups. From weeks two to five, a four-weeks intervention comprising four 40-minute periods weekly, totaling 16 sessions, was implemented. The experimental group received teacher-guided AI-chatbot instruction. The teacher introduced topics, presented problems, and guided students in small mixed-gender groups of three to four in ChatGPT interactions to explore explanations, generate examples, solve problems, and verify answers. Each group accessed at least one internet-enabled device. The teacher circulated, assisted with query formulation, prompted justification of chatbot responses, clarified misconceptions, and synthesized information with whole-class discussion. The control group received identical content coverage and instructional time using conventional lecture-demonstration with whiteboard, textbooks, and standard instructional materials without AI-chatbot access. The researcher conducted regular unannounced fidelity monitoring using checklists. In week six, the CAT post-test with rearranged items was administered under identical conditions to both groups.

Teacher variability was controlled by using the regular chemistry teachers

within each school. Initial group differences were controlled by using pre-test scores as covariates in ANCOVA. The Hawthorne effect was controlled by informing both groups of research participation without disclosing the hypothesis. Diffusion of treatment was controlled by selecting groups from different schools or from the same school with strict non-communication instructions. Maturation and history effects were controlled by the six-week duration and identical environmental conditions. Instrumentation effect was controlled by using the same test with rearranged items and blinded scoring. Mortality was controlled by including only students with minimum 80% attendance and complete pre-test and post-test participation. Descriptive statistics (means and standard deviations) were used to answer the research questions. Null hypotheses were tested at 0.05 level of significance using Analysis of Covariance (ANCOVA) with pre-test scores as covariate.

Results

Table 1: Mean Achievement Scores of Students taught Chemistry using Teacher-guided use of AI-Chatbots (TGUAC) and Lecture Method

Group	N	Pretest	Prete	Posttest	Posttest	Gained
		Mean	st SD	Mean	SD	Mean
TGUAC	27	27.67	2.86	66.89	9.60	39.22
LM	24	28.00	4.22	50.54	3.00	22.54

According to Table 1, students who were taught chemistry using teacher-guided use of AI-Chatbots (TGUAC) had higher mean gain achievement score (39.22) than those taught using lecture method (22.54).

Table 2: Mean Achievement scores of Male and Female Students taught Chemistry using TGUAC and Lecture Method

Method	Gender	N	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Gained Mean
TGUAC	Male	13	26.69	3.53	63.54	4.59	36.85
	Female	14	28.57	1.87	70.00	3.77	41.43
LM	Male	14	29.14	4.58	51.50	3.90	22.36
	Female	10	26.40	3.03	49.20	5.70	22.80

As shown in Table 2, male students taught chemistry using TGUAC had higher mean gain achievement score (36.85) than male students taught using lecture method (22.36). Female students taught chemistry using TGUAC had higher mean gain achievement score (41.43) than those taught using lecture method (22.80).

Table 3: ANCOVA Test of Significance of Difference in the Mean Achievement Score of Students taught Chemistry using TGUAC and Lecture Method

Source	SS	df	Mean Square	F	Sig.	Decision
Corrected Model	3719.860 ^a	4	929.965	6.599	.000	
Intercept	16125.929	1	16125.929	114.436	.000	
Pretest	12.152	1	12.152	.086	.770	
Method	3375.089	1	3375.089	23.951	.000	Sig.
Gender	55.340	1	55.340	.393	.534	Not Sig.
Method * Gender	223.115	1	223.115	1.583	.215	Not Sig.
Error	6482.179	46	140.917			
Total	188915.000	51				
Corrected Total	10202.039	50				

Table 3 shows that there is a significant main effect of the treatment on students' achievement in Chemistry, $F(1, 46) = 23.951$, $p < 0.01$. The null hypotheses was rejected meaning that there is a significant difference in the mean achievement scores of students taught Chemistry using teacher-guided use of AI-Chatbots (TGUAC) and with those taught using lecture method in favour of TGUAC.

Additionally, Table 3 indicates that gender did not have a statistically significant effect on students' achievement in Chemistry, $F(1, 46) = 0.393$, $P > 0.05$.

Furthermore, the table reveals no significant interaction between instructional method and gender on students' achievement in Chemistry, $F(1, 46) = 1.582$, $P > 0.05$. The interaction is as illustrated in Figure 1.

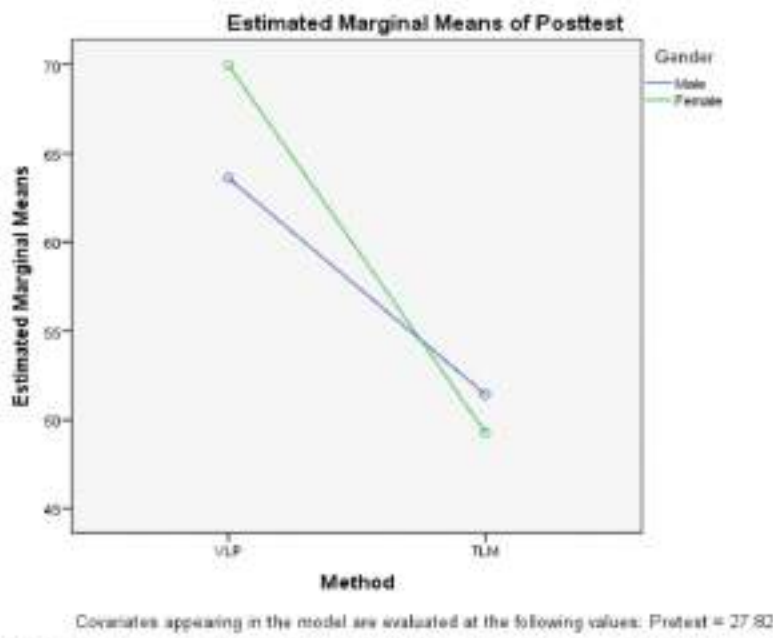


Figure 1: Plot of interaction effect of instructional methods and gender on students' achievement in Chemistry

The interaction effect of instructional methods and gender on students' Chemistry achievement, depicted in Figure 1, is both non-significant and disordinal. This indicates that the effect of instructional methods varied with gender but the variations were not significant.

Discussion

Students taught chemistry using teacher-guided AI-chatbots (TGUAC) attained significantly higher mean gain achievement scores than those taught with the conventional lecture method, indicating that the integration of conversational agents scaffolded by teacher facilitation produced a measurable advantage in chemistry content mastery. The superiority of the TGUAC strategy is attributed to the interactive, responsive, and individualized nature of chatbot-mediated learning, which positions students as active inquirers who engage in dialogue, formulate questions, evaluate responses, and construct knowledge through iterative interaction, unlike the passive reception characteristic of lecture methods. This finding aligns with Joyner's (2024) conceptual framework positioning chatbots as AI Teacher and AI Partner, fundamentally altering classroom dynamics through distributed instructional agency between human educator and artificial intelligence.

It is consistent with Sajadi's (2025) quasi-experimental study reporting significantly higher chemistry achievement among twelfth-grade students using ChatGPT chatbots augmented with Wolfram Alpha and ChemDraw plugins. Keith, Keiller, Windows-Yule, Kings, and Robbins (2025) similarly found that engagement with large language models in chemical engineering laboratory sessions enhanced critical thinking and content acquisition, supporting the present finding that AI-chatbots serve as viable instructional partners.

Both male and female students taught chemistry using TGUAC achieved higher mean gain scores than their same-gender counterparts in the lecture method group, demonstrating that the intervention was effective across both gender categories and confirming that pedagogical benefits of teacher-guided AI-chatbot use are accessible to all learners irrespective of sex. The consistency of the treatment effect across gender subgroups suggests that the TGUAC strategy provides equitable opportunities for academic achievement, a finding particularly significant given historical concerns regarding gender disparities in science achievement as it demonstrates that under optimized, technology-enhanced, inquiry-oriented instructional conditions, both male and female students are capable of comparable academic gains. This result further validates the position of Oladejo, Olateju, Okebukola, and Braimoh (2025) that gender gaps in chemistry achievement are not immutable but are responsive to appropriately designed pedagogical interventions, with TGUAC constituting one such intervention capable of raising achievement levels for both genders simultaneously. The finding that gender did not have a statistically significant effect on students' achievement in chemistry indicates that, within the context of this study, being male or female conferred no inherent advantage or disadvantage in learning chemistry when instructional exposure was controlled. This result challenges essentialist notions portraying male students as naturally predisposed to superior performance in the physical sciences and suggests that observed gender disparities in broader educational statistics are likely attributable to differential learning opportunities, socio-cultural expectations, and stereotype-based threats rather than immutable cognitive differences.

This finding is consistent with the results reported by Oladejo et al. (2025), who found no significant difference in post-test achievement scores of male and female students following exposure to the culturo-techno-contextual approach. It is also congruent with the earlier findings of George and Abumchukwu (2021), who reported no significant gender difference in students' academic self-concept in chemistry. The finding that there was no significant interaction between instructional method and gender on students' achievement in chemistry demonstrates that the effectiveness of the TGUAC strategy did not depend on

student gender, with the superiority of teacher-guided AI-chatbot method over the lecture method being consistent across both gender categories. This absence of interaction is methodologically significant as it confirms homogeneity of treatment effect across levels of the moderating variable, satisfying ANCOVA assumptions and enhancing interpretability of main effects, and practically significant as it indicates that TGUAC is not differentially beneficial to one gender at the expense of the other and its adoption would not inadvertently widen existing gender gaps. This finding aligns with the conclusion of Oladejo et al. (2025), who reported no significant interaction between their culturo-techno-contextual approach and gender, suggesting that innovative, student-centred pedagogies tend to be gender-neutral in their effectiveness. It also extends the findings of George and Abumchukwu (2021) by demonstrating that not only do certain instructional methods eliminate gender differences in affective outcomes such as self-concept but also in achievement.

Conclusion

The findings of this study provide compelling evidence that teacher-guided use of AI-chatbots is an effective, equitable, and pedagogically sound strategy for enhancing secondary school students' academic achievement in chemistry, and produces superior learning outcomes relative to conventional lecture-based instruction. It also established the benefits of TGUAC accrue equally to male and female students, confirming the gender-neutral effectiveness of the intervention.

Recommendations

1. The Ministry of Education and curriculum planners should integrate teacher-guided use of AI-chatbots into the secondary school chemistry curriculum and teacher education programmes as a complementary instructional strategy.
2. School administrators and educational authorities should ensure equitable access to internet-enabled devices and AI-chatbot platforms for all students, with particular attention to removing infrastructural and sociocultural barriers that may hinder female students' full participation.
3. Researchers and funding agencies should prioritise longitudinal, multi-subject, and cross-contextual investigations into the sustainability, scalability, and optimal configuration of teacher-guided AI-chatbot instruction across diverse African educational settings.

References

- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- Joyner, D. A. (2024). *A teacher's guide to conversational AI: Enhancing assessment, instruction, and curriculum with chatbots*. Routledge.
- Keith, M., Keiller, E., Windows-Yule, C., Kings, I., & Robbins, P. (2025). Harnessing generative AI in chemical engineering education: Implementation and evaluation of the large language model ChatGPT v3.5. *Education for Chemical Engineers*, 51, 20–33. <https://doi.org/10.1016/j.ece.2025.01.002>
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610. <https://doi.org/10.1177/001316447003000308>
- Nwankwo, M. C., & Enwemeka, E. O. (2025). Self-concept as a correlate of academic performance in chemistry among secondary school students in Onitsha Education Zone. *Unizik Journal of Educational Research, Science and Vocational Studies*, 3(1), 112–125.
- George, P.C., & Abumchukwu, A.A. (2021). Impact of generative learning model on academic self concept of secondary school students in chemistry in Onitsha Education Zone. *UNIZIK Journal of Educational Research and Policy Studies*, 7(1), 309–322.
- Oladejo, A. I., Olateju, T. T., Okebukola, P. A., & Braimoh, D. S. (2025). I am Black and I am a girl-child, can I still be a scientist? The potency of culturo-techno-contextual approach in bridging gender differences in chemistry achievement. *School Science and Mathematics*, 125(2), 188–206. <https://doi.org/10.1111/ssm.12678>
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). Open University Press.
- Sajadi, M. (2025). The effect of AI tools on academic achievement in chemistry education. *Quarterly Journal of Chemical Education*, 13(2), 45–58.
- Technical University of Denmark. (2025). ChatGMP: A case of AI chatbots in chemical engineering education towards the automation of repetitive tasks. *Computers & Education: Artificial Intelligence*, 8, Article 100354.