



Development and Validation of an AI-Based Model for Identifying Senior Secondary School Students' Errors in Chemistry Calculations in Lokoja Metropolis, Kogi State, Nigeria

*Iliyasu Bakari, Sor Endurance Nubagbi, Imenka Wisdom Chukwuemeka, Nasiru Zuleihat Asipita and Abel Ojochenemi Doris

Department of Science Education, Federal University Lokoja, Lokoja, Kogi State, Nigeria.

*Corresponding authors' email: iliyasu.bakari@fulokoja.edu.ng

ABSTRACT

This study evaluated the development and validation of an AI-based model for identifying senior secondary school students' errors in chemistry calculations in Lokoja Metropolis, Kogi State, Nigeria. The study adopted a Research and Development (R&D) design. The study population comprised all 24,152 SS1 and SS2 students, from which a sample of 269 students was selected. The instruments used for data collection were the researcher-developed Chemistry AI Test (CALT) and Identification of Students' Errors in Chemistry Calculations (ISECC) questionnaire. The instruments were validated by two chemistry teachers and one Measurement and Evaluation expert. Reliability was established through pilot testing, and Kuder–Richardson Formula 20 and Cronbach's alpha yielded coefficients of 0.79 and 0.81, respectively. Data were analysed using descriptive statistics, while students' working steps were analysed using the ChemError-AI model to classify errors into mathematical/arithmetic, conceptual, formula, unit, interpretation, and substitution categories. Findings showed that the ChemError-AI model correctly classified 58 out of 64 unseen test cases, representing 90.6% accuracy. The findings indicate that the model can provide preliminary information about the types of errors students make in chemistry calculations. It was concluded that students experience difficulties at different stages of chemistry problem solving and that these difficulties extend beyond mathematical operations. It was recommended, among other measures, that teachers should receive practical training on how to interpret AI-generated classifications and combine AI-supported information with their professional judgement before making instructional decisions.

Received: 24 Sep. 2026

Accepted: 08 Oct. 2026

Published: 10 Oct. 2026

Keywords:

Chemistry Calculations, Chemerror-AI, Error Classification, Conceptual Errors, Procedural Errors, Computational Errors

INTRODUCTION

Chemistry calculations involve integrating chemical concepts with mathematical techniques to solve quantitative problems. Topics such as the mole concept, stoichiometry, concentration, gas laws, thermochemistry, and acid–base reactions require students not only to perform numerical operations but also to understand the underlying chemical principles, identify relevant information, select appropriate relationships or formulas, substitute values correctly, perform calculations accurately, and interpret the resulting answers within the context of the problem.

Studies have shown that many students struggle to integrate these conceptual and procedural elements when solving chemistry calculations. For instance, Rusek et al. (2022) found that students often experience difficulties in identifying problems, understanding the relevant concepts, and applying appropriate mathematical procedures. These difficulties suggest a gap in the integration of conceptual understanding, procedural skills, and mathematical foundations required for successful chemistry problem solving.

The challenges associated with chemistry calculations are also evident in the Nigerian educational context. Upahi and Olorundare (2012) reported that senior secondary school students experienced difficulties in solving stoichiometric problems, particularly in connecting known and unknown quantities and identifying the relevant information needed to solve problems. More recent Nigerian studies have continued to highlight students' difficulties with applying chemical

formulas and mathematical procedures. For instance, Afolabi and Anaekwe (2025) reported difficulties in writing and balancing chemical equations, including challenges in recalling appropriate chemical formulas and applying relevant procedures. Oladejo et al. (2022) also reported that fear of calculations was among the challenges perceived by secondary school students in chemistry.

An important concern is not only whether students obtain the correct or incorrect answer but also the nature and source of the errors made during calculations. A student may make a conceptual error by misunderstanding the chemical principle required to solve a problem. A formula error may occur when an inappropriate equation or mathematical relationship is selected. A substitution error may occur when an incorrect value is inserted into an appropriate formula. Mathematical or arithmetic errors involve incorrect numerical operations. Interpretation errors may occur when a student misunderstands the question or the meaning of a calculated result. Unit errors may involve omitting units, using inappropriate units, or failing to carry out necessary unit conversions.

Research in chemistry has identified several types of errors made by students when solving quantitative problems. Reports from WAEC Chief Examiners between 2020 and 2023 have highlighted difficulties in areas such as mole concepts, stoichiometry, titration, gas laws, and thermochemical calculations. Common difficulties include misinterpreting questions, applying formulas incorrectly,

performing algebraic manipulations, omitting units, and misreading chemical data. Basuki (2020) identified several error categories in students' solutions to stoichiometry problems, including conceptual misunderstandings, problems with data use, technical calculation errors, and errors in drawing conclusions. Abdullah et al. (2021) also identified patterns involving comprehension, transformation, process skills, and encoding. Understanding distinctions among these errors is important for effective diagnosis and remediation. An incorrect final answer alone may not reveal the specific difficulty experienced by a student. For example, two students may obtain the same incorrect answer even though one selected an inappropriate formula while the other selected the correct formula but made an arithmetic error. Similarly, a student may perform the calculation correctly but fail to interpret the result or report the appropriate unit. Error-pattern analysis can therefore provide more detailed information about students' learning difficulties.

Identifying different types of errors can be challenging when teachers have to examine a large number of student responses. Manual analysis requires careful attention to the chemical and mathematical processes involved and can be time-consuming, particularly when students use different solution procedures. A diagnostic method that can identify patterns in students' responses could therefore provide useful support to teachers. Such a method would not replace teachers' professional judgement but could help them identify common areas of difficulty and target instructional support.

Artificial intelligence (AI) presents a computational approach to this diagnostic challenge. In this study, AI is used to learn from previously classified student responses and apply the learned patterns to new responses for categorization into specific error types. Instead of treating an incorrect answer as a single outcome, an AI-based diagnostic model can distinguish among conceptual, formula, mathematical/arithmetic, substitution, interpretation, and unit errors. However, the usefulness of such a model depends on its ability to classify unseen cases accurately and consistently across the different error categories.

Evaluation of an AI-based diagnostic model requires more than counting correct classifications. Metrics such as accuracy, precision, recall, F1-score, and the confusion matrix provide complementary information about classification performance. A confusion matrix can show whether cases from one error category are being systematically misclassified as another. Kingsley et al. (2025) reported the use of confusion matrices, precision, recall, and F1-score in evaluating classification models. Tadese et al. (2025) also demonstrated the use of a confusion matrix to identify classification outcomes and the use of precision, recall, and F1-score to evaluate machine-learning performance. This is important in chemistry because some calculation errors may be closely related. For example, selecting an inappropriate formula can lead to incorrect substitution and ultimately to an incorrect final answer. A diagnostic model should therefore distinguish among error types rather than merely classify responses as correct or incorrect.

Although previous studies have examined students' difficulties and error patterns in chemistry calculations, there is limited evidence in the present study context on the use of an AI-based model to classify different categories of calculation errors among senior secondary school students. Most existing studies have focused on identifying or describing students' difficulties and error patterns. Systematically classifying conceptual, formula, mathematical/arithmetic, substitution, interpretation, and unit

errors using an AI-based model provides a further diagnostic perspective. Therefore, this study aims to develop and evaluate an AI-based diagnostic model for classifying errors in chemistry calculations among senior secondary school students in Lokoja Metropolis, Kogi State, Nigeria.

MATERIALS AND METHODS

The study adopted a Research and Development (R&D) design. The design was considered appropriate because the study involved the identification and classification of students' errors in chemistry calculations and the development, testing, and evaluation of an AI-based diagnostic model. Jacobsen and McKenney (2024) describe R&D as a systematic approach involving the investigation of an identified problem and the design, development, testing, evaluation, and refinement of an educational innovation. Descriptive procedures were also used to summarise the identified errors and teachers' questionnaire responses. The population comprised all SS1 and SS2 students in Lokoja Metropolis, Kogi State, Nigeria. Records obtained from the Kogi State Ministry of Education, Lokoja Zonal Office, in 2026 indicated a population of 24,152 SS1 and SS2 students (8,550 males and 15,602 females) and 31 chemistry teachers. A total of 269 students participated in the study and were selected using stratified and simple random sampling techniques. Their responses to the chemistry calculation tasks were examined using the Chemistry Error Diagnostic Artificial Intelligence model (ChemError-AI), developed for this research to identify and classify errors in students' solution procedures. The analysis produced 340 identifiable error instances across six predefined categories: conceptual, formula, mathematical/arithmetic, substitution, interpretation, and unit errors. Of the 340 error instances, 276 (81.2%) were used as training/seen cases, while 64 (18.8%) were reserved as unseen testing cases. The 64 evaluation cases did not represent an additional group of 64 students; they constituted a held-out subset of the error instances generated from the responses of the 269 students. Because the teacher population was relatively small, all 31 chemistry teachers were included in the study.

Two instruments were used for data collection: the researcher-developed Chemistry AI Test (CALT) and the Identification of Students' Errors in Chemistry Calculations (ISECC) questionnaire. The CALT was used to assess students' performance in chemistry calculations and provide working steps for identifying errors. It contained 40 WAEC-style multiple-choice calculation questions with options A–D. A table of specifications was used to distribute the questions across mole concept, stoichiometry, titration analysis, gas laws, and thermochemistry. Space was provided for students to show their working steps to facilitate the identification and classification of errors. Each solution was examined systematically to identify the stage at which an error occurred. The identified errors were classified into six categories: conceptual, formula, mathematical/arithmetic, substitution, interpretation, and unit errors. Conceptual errors referred to incorrect understanding or application of a chemistry concept; formula errors involved the selection or application of an inappropriate formula; mathematical/arithmetic errors involved incorrect mathematical operations or calculations; substitution errors involved incorrect insertion of values into an appropriate formula; interpretation errors involved incorrect interpretation of the question or calculated result; and unit errors involved the use, omission, or inappropriate conversion of units.

The second instrument was a structured questionnaire titled Identification of Students' Errors in Chemistry Calculations (ISECC). It was designed by the researcher to obtain empirical data from chemistry teachers about students' difficulties in chemistry calculations and how AI-based diagnostic models can assist in identifying and analysing students' learning difficulties. It consisted of two sections. Section A contained demographic information such as gender, qualification, and years of teaching experience. Section B contained 30 items on how AI-based diagnostic models can assist teachers in understanding students' learning difficulties through error identification, performance analysis, personalised feedback, prediction of learning challenges, and continuous assessment. Responses were measured using a four-point Likert scale ranging from Strongly Agree (SA), Agree (A), Disagree (D), to Strongly Disagree (SD). A mean score of 2.50 or above was accepted, whereas a mean below 2.50 was rejected. The instruments were validated by two chemistry teachers and one expert in Measurement and Evaluation from the Faculty of Education, Federal University Lokoja. The experts examined the instruments for clarity, relevance, and appropriateness in relation to the objectives of the study, and their suggestions were incorporated into the final versions before administration.

The reliability of the instruments was established through a pilot study conducted with respondents outside the main sample. The CALT was pilot-tested with a small group of SS2 students, while the ISECC questionnaire should be pilot-tested with chemistry teachers who are comparable to the study respondents. The reliability coefficients reported were 0.79 for the CALT using Kuder–Richardson Formula 20 and 0.81 for the ISECC using Cronbach's alpha. These

coefficients indicated satisfactory internal consistency for the instruments within the context of the study. The researchers visited the selected schools and administered the instruments to the selected respondents. The collected data were analysed using descriptive statistics, including frequencies, percentages, means, and standard deviations. Responses to the teachers' questionnaire were coded and scored using the four-point Likert scale.

RESULTS AND DISCUSSION

Development of the AI-Based Diagnostic Model

The ChemError-AI model was developed using a decision-tree classification algorithm implemented in Python as a multiclass classification system. The model was designed to classify identified chemistry calculation errors into six categories: conceptual, formula, mathematical/arithmetic, substitution, interpretation, and unit errors. The labelled error instances obtained from students' responses constituted the dataset for model development. Of the 340 identified error instances, 276 (81.2%) were used as training/seen cases, while 64 (18.8%) were withheld as unseen testing cases. The 64 testing cases were not used in training or model development. After the model had been developed using the training cases, the 64 withheld cases were submitted to the model for classification. The predicted classifications were then compared with the predetermined reference classifications.

Research Question One

What is the performance of the ChemError-AI model in identifying the different categories of Chemistry calculation errors?

Table 1: Distribution of Identified Chemistry Calculation Errors

Error Type	Frequency	Percentage
Conceptual	71	20.9
Formula	53	15.6
Mathematical/Arithmetic	89	26.2
Substitution	36	10.6
Interpretation	41	12.1
Unit	50	14.7
Total	340	100%

Table 1 shows that mathematical/arithmetic errors were the most frequently identified errors, accounting for 89 (26.2%) of the 340 error instances. Conceptual errors accounted for 71 (20.9%), followed by formula errors with 53 (15.6%), unit errors with 50 (14.7%), interpretation errors with 41 (12.1%),

and substitution errors with 36 (10.6%). The distribution indicates that students' difficulties in chemistry calculations occurred at different stages of the problem-solving process and were not limited to mathematical computation.

Table 2: Distribution of Error Instances used for Model Training and Testing

Dataset	Frequency	Percentage
Training/seen cases	276	81.2
Testing/unseen cases	64	18.8
Total	340	100%

Table 2 indicates that, of the 340 identified error instances, 276 (81.2%) were used as training/seen cases, while 64 (18.8%) were reserved as unseen testing cases. The 64 testing cases were withheld from model training and were used only to evaluate the classification performance of the developed

AI-based diagnostic model. The 64 cases represent error instances rather than 64 additional students; they were derived from the responses of the 269 students who participated in the study.

Table 3: Distribution of the AI Model in Precision, Recall, and F1-Score Categories

Error Type	Precision	Recall	F1-Score
Conceptual	87.5%	87.5%	87.5%
Formula	87.5%	87.5%	87.5%
Mathematical/Arithmetic	92.3%	92.3%	92.3%
Substitution	100.0%	90.0%	94.7%
Interpretation	80.0%	92.3%	85.7%
Macro average	100.0%	91.7%	95.7%
Total	91.2%	90.2%	90.6%

Table 3 indicates that the ChemError-AI model achieved an overall accuracy of 90.6%, correctly classifying 58 of the 64 testing cases. The model recorded the highest precision for substitution and unit errors (100.0% each), indicating that all cases predicted in these categories were correctly classified. Mathematical/arithmetic errors also showed relatively strong performance, with precision and recall of 92.3%. The relatively lower precision for interpretation errors (80.0%) suggests that some cases predicted as interpretation errors belonged to other categories. Overall, the results show

variation in classification performance across error categories and indicate that the model should be interpreted as a preliminary diagnostic support tool rather than a replacement for teachers' professional judgement.

Research Question Two

What are teachers perception of Usefulness of the AI model in understanding the nature of students difficulties in chemistry calculations?

Table 4: Mean and Standard Deviation Statistics of Teachers' Perceptions of AI Model Support for Understanding Students' Difficulties in Chemistry Calculations (N = 31)

S/N	STATEMENT	Mean ($\bar{x}$)	Std. Dev.	Remarks
1	An AI-based diagnostic model can help identify common student errors in chemistry	3.14	0.64	Accepted
2	AI can assist teachers in categorizing students' errors by topic (e.g., stoichiometry, mole concepts, gas laws)	2.75	0.56	Accepted
3	AI can assist teachers in categorizing students' errors by topic (e.g., stoichiometry, molar concepts, gas laws)	3.30	0.70	Accepted
4	AI feedback can help teachers design targeted interventions for struggling students	3.23	0.68	Accepted
5	AI models can save teachers time in diagnosing students' learning difficulties	3.50	0.69	Accepted
6	Integrating AI into teaching practice would improve lesson planning and student support	3.11	0.72	Accepted
7	AI-based analysis can highlight recurring misconceptions that teachers might otherwise overlook	2.73	0.45	Accepted
8	AI tools can provide individualized feedback that supports differentiated instruction	3.46	0.59	Accepted
9	AI models can enhance collaboration among teachers by sharing diagnostic insights	2.80	0.61	Accepted
10	The use of AI in chemistry teaching would improve students' overall academic performance	3.40	0.68	Accepted
Cluster mean and standard deviation		3.14	0.54	

Table 4 shows teachers' perceptions of how the AI model can help them better understand students' difficulties in chemistry calculations in Lokoja Metropolis. The highest-rated item was "AI models can save teachers time in diagnosing students' learning difficulties," with a mean score of 3.50 and a standard deviation of 0.69. Other highly rated items were "AI tools can provide individualized feedback that supports differentiated instruction," with a mean of 3.46 and a standard deviation of 0.59; "The use of AI in chemistry teaching would improve students' overall academic performance," with a mean of 3.40 and a standard deviation of 0.68; and "AI can assist teachers in categorizing students' errors by topic (e.g., stoichiometry, mole concepts, gas laws)," with a mean of 3.30 and a standard deviation of 0.70. The lower-rated items included the statements on detailed error-frequency reports ($M = 2.75$, $SD = 0.56$), recurring misconceptions ($M = 2.73$, $SD = 0.45$), collaboration among teachers ($M = 2.80$, $SD = 0.61$), and lesson planning and student support ($M = 3.11$, $SD = 0.72$). All items were accepted because their mean scores exceeded the decision threshold of 2.50.

Discussion

The finding that the ChemError-AI model correctly classified 58 of 64 unseen testing cases, representing an overall accuracy of 90.6%, suggests that the model demonstrated preliminary classification performance under the reported test conditions. Its potential usefulness lies in converting students' responses into identifiable error categories, thereby helping teachers move beyond merely recording whether an answer is right or wrong. A diagnostic system of this nature can provide preliminary information about areas in which students may need support and can facilitate more targeted remediation. This finding is consistent with Almasri (2024), whose systematic review reported that AI is being used to support assessment, feedback, and instructional decision-making in science education. It is also consistent with the OECD (2024), which noted the potential of AI and digital technologies to support teachers with assessment and other non-teaching tasks. Cooper et al. (2023) reported variation in AI performance across chemistry calculation tasks, while Kieser et al. (2024) found that ChatGPT performed strongly on many chemistry assessment tasks but also produced errors and did not always correspond with human grading. These

findings underscore the importance of human verification, independent validation, and monitoring of model errors. The present result concerns a specifically trained diagnostic classification model rather than a general-purpose generative AI system. The variation in precision, recall, and F1-score across categories further indicates that model performance may differ according to the type of error being classified.

The finding also showed that the majority of senior secondary school teachers in Lokoja Metropolis, Kogi State, agreed that an AI model can help them better understand students' difficulties in chemistry calculations. The highest-rated item was the statement that AI models can save teachers time in diagnosing students' learning difficulties, with a mean score of 3.50. This suggests that teachers particularly recognised the potential time-saving value of AI-supported diagnosis. The result is consistent with Holmes et al. (2023), who discussed the potential of AI to support teaching, assessment, and feedback. It is also consistent with the argument of Sanusi (2025) that AI integration in STEM education may support personalised instruction and assessment. Roy et al. (2024) reported that teachers using ChatGPT for lesson preparation experienced reduced preparation time during the later stages of their trial, while the OECD (2024) noted that AI may reduce teacher workload through support for marking, administration, communication, and related tasks. Samaila et al. (2024) similarly examined teachers' awareness and readiness to use AI assessment methods in Nigeria. However, UNESCO (2023) emphasised that AI in education should be human-centred and accompanied by attention to privacy, ethical use, safety, equity, and institutional validation. Therefore, the teachers' positive perceptions should not be interpreted as evidence that AI can independently diagnose students without supervision. Rather, the findings indicate that teachers perceive AI as a potentially useful support mechanism that should complement, rather than replace, professional judgement.

CONCLUSION

Based on the findings of the study, it was concluded that the ChemError-AI model demonstrated preliminary classification performance under the reported test conditions, correctly classifying 58 of 64 unseen testing cases, representing 90.6% accuracy. The findings also showed that teachers generally perceived AI-supported diagnosis positively, particularly in relation to its potential to save time when identifying students' learning difficulties. However, because the model did not correctly classify every testing case and was evaluated on a relatively small held-out set of 64 error instances, it should be regarded as a supportive diagnostic tool rather than a substitute for chemistry teachers. Human judgement, independent validation, continued performance monitoring, and appropriate ethical safeguards are necessary before broader classroom deployment.

Based on the findings and conclusions of this study, the following recommendations were made:

- i. The ChemError-AI model should be considered for classroom deployment only after independent validation on new and sufficiently diverse datasets and with appropriate teacher oversight.
- ii. Teachers should receive practical training on how to interpret AI-generated classifications and combine AI-supported information with their professional judgement before making instructional decisions.
- iii. Chemistry teachers should diagnose students' calculation errors beyond the final numerical answer by examining conceptual understanding, formula selection, substitution, unit handling, interpretation, and mathematical operations.
- iv. Schools should consider simple AI-supported diagnostic tools as supplementary assessment resources for providing timely and individualised feedback to students, subject to appropriate validation and oversight.
- v. School authorities and relevant education agencies should improve access to ICT infrastructure, reliable internet connectivity, and digital learning resources required for effective AI-supported instruction.

REFERENCES

- Abdullah, A., Azmi, J., & Ardiansyah, A. (2021). Analisis pola kesalahan jawaban siswa pada materi hitungan kimia. *Orbital: Journal Pendidikan Kimia*, 5(1), 13–27. <https://doi.org/10.19109/ojpk.v5i1.7237>
- Afolabi, S. S., & Anaekwe, M. C. (2025). Identification and remediation of difficulties in writing and balancing chemical equations as perceived by chemistry students in Kaduna State. *UniJos Journal of Contemporary Studies in Education*, 1(2), 102–106
- Agogo, P. O., & Maduawesi, C. C. (2021). Artificial intelligence-driven diagnostic approaches for identifying learners' error patterns in mathematics education.
- Agbarakwe, H. A., & Chibueze, O. O. (2024). Leveraging artificial intelligence for enhanced assessment and feedback mechanisms in Nigeria higher education system. *International Journal of Research and Innovation in Social Science*, 8(9), 142–151. <https://doi.org/10.47772/IJRISS.2024.809012>
- Ahiakwo, M. J., & Isiguzo, C. Q. (2015). Students' conceptions and misconceptions in chemical kinetics in Port Harcourt Metropolis of Nigeria. *African Journal of Chemical Education*, 5(2), 112–130.
- Al-Balushi, S. M., & Al-Abdali, N. S. (2019). Using technology-supported learning to enhance inquiry and address students' learning difficulties in science.
- Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, 54, 977–997. <https://doi.org/10.1007/s11165-024-10176-3>
- Ayoade, A. A. (2019). Assessment practices and the identification of students' misconceptions and learning difficulties.
- Basuki, K. H. (2020). Analisis Kesalahan Mahasiswa dalam menyelesaikan soal stoikiometri. *Journal Ilmiah Wahana Pendidikan*, 6(4), 641–651. <https://doi.org/10.5281/zenodo.4298027>
- Cooper, M. M., Stieff, M., & others. (2023). Comparing the performance of college chemistry students with ChatGPT for calculations involving acids and bases. *Journal of Chemical Education*, 100(10), 3934–3944. <https://doi.org/10.1021/acs.jchemed.3c00500>
- Federal Republic of Nigeria. (2019). National policy on education (6th Ed.). NERDC Press.

- Holmes, W., Wieman, C., & Bonn, C. (2023). Artificial intelligence in education: Opportunities and challenges for teaching, learning and assessment.
- Jacobsen, M., & McKenney, S. (2024). Research and development in education: Designing, developing, testing and refining educational innovations.
- Kieser, J., et al. (2024). ChatGPT as an instructor's assistant for generating and scoring exams. *Journal of Chemical Education*, 101(9), 3780–3788. <https://doi.org/10.1021/acs.jchemed.4c00231>
- Kingsley, C. J. O., Uthman, M., Bebeji, I. A., & Udeh, P. C. (2025). Fault diagnosis system using CNN-LSTM networks. *UNIABUJA Journal of Engineering and Technology*, 2(2), 224–232.
- Krenn, B., et al. (2020). Artificial intelligence, adaptive learning and intelligent assessment in STEM education.
- Nwani, O. E. (2019). Chemistry education and its relevance to national development.
- Olubunmi, O. A., & Aarinola, A. A. (2022). Challenges affecting effective science and Chemistry teaching in Nigerian secondary schools.
- Oladejo, M. A., et al. (2022). Concept difficulty in secondary school chemistry: An intra-play of gender, school location and school type. *Journal of Technology and Science Education*. 12 (1), 20-24
- Reinhold, P., et al. (2005). Evaluating students' conceptual understanding of balanced equations and stoichiometric ratios using a particulate drawing. *Journal of Chemical Education*.
- Rusek, M., et al. (2026). Challenges in chemistry calculations at university entry: Evidence from a multi-institutional study. *Journal of Chemical Education*, 103(4), 1826–1834. <https://doi.org/10.1021/acs.jchemed.6c00022>
- Rusek, M., Vojř, K., Bártová, I., Klečková, M., Šírotek, V., & Štrofová, J. (2022). To what extent do freshmen university chemistry students master chemistry calculations? *Acta Chimica Slovenica*, 69(2), 371–377. <https://doi.org/10.17344/ACSI.2021.7250>
- Roy, P., Poet, H., Staunton, R., Aston, K., & Thomas, D. (2024). ChatGPT in lesson preparation: A teacher choices trial. Education Endowment Foundation.
- OECD. (2024). Education policy outlook 2024: Reshaping teaching and learning from ABCs to AI. OECD Publishing. <https://doi.org/10.1787/dd5140e4-en>
- Organisation for Economic Co-operation and Development. (2021). OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots. OECD Publishing.
- Sanusi, T. (2025). Artificial intelligence integration in STEM education: Personalized learning and assessment.
- Samaila, K., Abdulfattah, K., Babatunde, O. M., & Akindele, N. A. (2024). Teachers' awareness and readiness to use AI assessment methods in Kwara State, Nigeria. *International Journal of Innovative Technology Integration in Education*, 7(2).
- Scott, F. J. (2012). Is mathematics to blame? An investigation into high school students' difficulty in performing calculations in chemistry. *Chemistry Education Research and Practice*, 13, 330–336. <https://doi.org/10.1039/C2RP00001F>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
- Upahi, J. E., & Olorundare, A. S. (2012). Difficulties faced by Nigerian senior school chemistry students in solving stoichiometric problems. *Journal of Education and Practice*, 3(12), 31-35
- Tadese, Z. B., Nimani, T. D., Mare, K. U., Gubena, F., Wali, I. G., & Sani, J. (2025). Exploring machine learning algorithms for predicting fertility preferences among reproductive age women in Nigeria. *Frontiers in Digital Health*, 6, Article 1495382. <https://doi.org/10.3389/fdgth.2024.1495382>
- Taskin, V., & Bernholt, S. (2014). Students' understanding of chemical formulae: A review of empirical research. *International Journal of Science Education*, 36(1), 157–185. <https://doi.org/10.1080/09500693.2012.744492>
- Weegar, R., & Idestam-Almquist, P. (2024). Reducing workload in short answer grading using machine learning. *International Journal of Artificial Intelligence in Education*, 34, 247–273. <https://doi.org/10.1007/s40593-022-00322-1>
- Yang, Y., Huang, F., & Liang, Z. (2024). Effects of reduction on teacher's teaching load in the lesson development of AI-led instruction: A pilot study. *Journal of East China Normal University (Educational Sciences)*, 42(2), 46–62. <https://doi.org/10.16382/j.cnki.1000-5560.2024.02.004>

