ChemBuddy Supreme+: A Multi-Agent Generative AI Platform for Enhancing O-Level Chemistry Education and Accessibility in Brunei

Hj Zadil Hayat Hj Kassim[1]

[1]Science Department Sultan Sharif Ali Secondary School, Brunei Darussalam

Corresponding author: zadil.kassim@teacher.smsultansharifali.edu.moe.bn

Abstract

This study evaluates ChemBuddy Supreme+, a multi-agent generative AI platform developed to support the Cambridge O-Level Chemistry curriculum in Brunei Darussalam, with particular attention to a "double barrier" encountered by secondary learners: the high conceptual abstraction of the syllabus and the physical inaccessibility of science materials for learners with visual impairments. Employing a convergent parallel mixed-methods design, the study involved N=53 Year 11 students over an eight-week intervention. The quantitative strand compared baseline End-of-Year examination scores with post-intervention Qualifying Examination results using paired-samples t-tests. In contrast, the qualitative strand drew on semi-structured interviews, teacher observations, and a dedicated accessibility case study. Results demonstrated a statistically significant improvement in cohort performance (t=11.42, $p < 0.001$, d=1.93). A1–C6 pass rates rose from 5.66% to 66.03%, while U-grade failures declined from 69.81% to 11.32%. Qualitative evidence indicated that immediate, automated formative feedback corrected misconceptions in real time and enhanced students self-efficacy. The learner with visual impairment successfully navigated chemistry content independently using the platform's ARIA-labelled interface and screen-reading commands. This study offers a proof of concept for syllabus-aligned, multi-agent AI in resource-constrained STEM environments, demonstrating a viable pathway to inclusive STEM equity in Brunei.

Keywords: Artificial intelligence, secondary chemistry education, digital accessibility, multi-agent systems, Brunei O-Level, inclusive STEM.

Introduction

In alignment with Brunei's Wawasan 2035, the Ministry of Education has prioritised digital competency to foster a high-performance STEM workforce. However, evidence suggests a persistent proficiency gap: international PISA 2018 data indicate that only 48% of Bruneian 15-year-olds achieve basic proficiency in science (Ali & Abdullah, 2023). Within the secondary curriculum, the Cambridge O-Level Chemistry syllabus poses a formidable challenge for many Year 11 students. Unlike general science, O-Level Chemistry requires students to navigate Johnstone's Triangle—simultaneously interpreting macroscopic phenomena, microscopic particles, and symbolic equations (Osman, 2013; Wang et al., 2022). Students consistently report conceptual difficulties in abstract topics such as stoichiometry, redox reactions, and chemical bonding (Chua & Karpudewan, 2019; Uzezi et al., 2017; Zaluddin, 2025).

For a learner with visual impairment, these academic hurdles are compounded by a lack of accessible auditory or tactile materials, creating a "double barrier" of content abstraction and physical inaccessibility (Chemnad & Othman, 2024;

Plazar et al., 2021). While regional neighbours like Singapore and Malaysia have begun exploring AI readiness, a clear research gap exists for localised, syllabus-aligned tools that address both academic scaffolding and accessibility in the Bruneian context (Park et al., 2023; Raza et al., 2025; Suryadi et al., 2025). This study contributes a theory-grounded evaluation of ChemBuddy Supreme+, demonstrating how multi-agent AI can bridge performance gaps for diverse learners (Ramos & Wilson-Kennedy, 2024).

Literature Review

Chemistry Learning Difficulties

Chemistry is widely regarded as one of the most cognitively demanding disciplines in secondary education, primarily because it simultaneously relies on macroscopic, microscopic, and symbolic representations — a framework known as Johnstone's Triangle (Osman, 2013). Secondary students frequently struggle with the procedural and conceptual demands of balancing redox equations, performing mole calculations, and interpreting chemical bonding (Mundy & Potgieter, 2019; Owusu et al., 2024). These challenges are compounded in contexts such as Brunei, where English-medium instruction may present an additional linguistic barrier for students whose primary language is Malay. Evidence from local and regional studies suggests that conceptual difficulties often stem from inadequate foundational self-efficacy and a persistent reliance on rote memorisation rather than deep conceptual understanding (Zaluddin, 2025). Students in the O-Level system are expected to navigate stoichiometry, organic chemistry, and electrochemistry with limited opportunity for individualised support, given typical class sizes and resource constraints (Chua & Karpudewan, 2019; Uzezi et al., 2017).

AI Tutoring and Formative Feedback

Intelligent Tutoring Systems (ITS) have a well-documented history of providing personalised, one-on-one instructional support that approximates the effectiveness of human tutoring (Holmes et al., 2019). A key advantage of digital platforms over traditional instruction is the capacity to deliver formative feedback that is immediate, error-specific, and contextually appropriate — commonly described as "just-in-time" (JIT) feedback (Eitemüller et al., 2023; Hagos & Andargie, 2022). Research consistently demonstrates that elaborated feedback — which not only signals an error but explains the underlying misconception — is substantially more effective than simple right/wrong verification in promoting durable conceptual understanding (Black & Wiliam, 2009; Webb et al., 2018). More recently, generative AI systems have introduced a new paradigm: multi-agent architectures in which distinct AI agents fulfil specialised pedagogical roles, such as questioning, grading, and metacognitive prompting. ChemBuddy Supreme+ operationalises this architecture by assigning discrete functions to its Grader, Reflector, and Challenger agents, allowing the system to provide differentiated and adaptive feedback at each stage of the learning process (Wang et al., 2022).

Universal Design for Learning and Inclusive Science Education

The Universal Design for Learning (UDL) framework advocates for multiple means of engagement, representation, and action to minimise barriers for all learners, particularly those from underserved populations (Kohnke & Zaugg, 2025; Ramos & Wilson-Kennedy, 2024). For a learner with visual impairment, accessibility in a chemistry context requires more than basic screen-reading functionality. The interpretation of chemical equations, structural diagrams, and reaction mechanisms demands intelligent ARIA labelling, data sonification, and auditory alternatives to visual representations (Chemnad & Othman, 2024; Plazar et al., 2021). In the Southeast Asian educational context, inclusive AI tools must also be culturally responsive — reflecting local pedagogical norms, language preferences, and ethical standards around data use — to achieve meaningful uptake among educators and students (Payadnya et al., 2024; Raza et al., 2025). Current research in the region suggests that while interest in AI-enhanced education is growing, implementation remains uneven due to infrastructure limitations and a lack of localised, curriculum-aligned tools (Park et al., 2023; Suryadi et al., 2025).

Analytical Framework

The framework for ChemBuddy Supreme+ is analytical, mapping platform features directly to pedagogical theories to guide data interpretation.

Table 1

Mapping of Platform Features to Theoretical Pillars

Platform Feature	Theoretical Pillar	Analytical Application
Text-to-Speech & Navigation	UDL (Ramos & Wilson-Kennedy, 2024)	Guided inclusion of ARIA labels to measure accessibility success (Plazar et al., 2021).
Grader Agent	Formative Feedback (Black & Wiliam, 2009)	Evaluated on its ability to provide error-specific JIT feedback (Eitemüller et al., 2023).
Reflector Agent	Metacognitive Theory (Tomisu et al., 2025)	Prompted students to justify reasoning steps, guiding qualitative coding of metacognition.
Challenger Agent	Scaffolding Theory (Beale, 2025)	Created an adaptive difficulty loop to maintain the zone of proximal development.

Methodology

This study employed a convergent parallel mixed-methods design (Creswell & Creswell, 2018). This research methodology integrates quantitative and qualitative data collected concurrently but analysed independently, then merged to produce a comprehensive interpretation of the findings. Mixed-methods research was selected because neither a purely quantitative nor a purely qualitative approach would have been sufficient to address both the study's empirical and interpretive objectives. The quantitative strand was designed to measure the magnitude and statistical significance of academic improvement. In contrast, the qualitative strand sought to explain and contextualise these numerical outcomes through the lived experiences of students and teachers.

Research Design and Justification

The convergent parallel design was particularly appropriate for this study, given its dual aims: (1) to establish empirical evidence of the platform's effectiveness through measurable performance data, and (2) to generate rich, contextually situated explanations of how and why improvement occurred. By collecting both forms of data simultaneously and then triangulating the results, this design mitigates the limitations inherent in either approach alone, producing a more robust and trustworthy account of the intervention's impact.

Ethical Procedures and Sampling

Ethical approval was obtained from the school administration in accordance with Ministry of Education guidelines. All N=53 participants and their legal guardians provided written informed consent before the commencement of data collection. To protect student privacy, all participant data were anonymised using numeric codes throughout analysis and reporting (Raza et al., 2025). Purposive sampling was utilised to select six students distributed across low, medium, and high baseline performance bands for semi-structured interviews, ensuring a representative range of academic experiences. A dedicated case study protocol was also developed for the learner with a visual impairment, enabling an in-depth evaluation of the platform's Universal Design for Learning features.

Quantitative Data Collection and Analysis

Pre-intervention performance data were drawn from End-of-Year (EoY) examination results, which served as the baseline measure of academic achievement for each participant. Post-intervention data were collected from the school's internal Qualifying Examination (QE), administered at the conclusion of the eight-week intervention period. Both assessments were scored on the Cambridge O-Level grading scale (A1–U). A paired-samples t-test was selected as the primary inferential statistic because the same participants were measured at both time points under matched conditions. Cohen's d was subsequently calculated to determine the practical effect size of the observed differences.

Qualitative Data Collection and Analysis

Qualitative data were gathered through three complementary methods: semi-structured interviews with the six purposively sampled students, teacher observations conducted across multiple class sessions, and a cognitive walkthrough protocol administered to the learner with visual impairment. All interviews were audio-recorded and transcribed verbatim before undergoing inductive thematic analysis, in which codes and themes were developed from the data rather than imposed by a predetermined framework (Braun & Clarke, 2006). Teacher observation notes were analysed for convergent themes related to student engagement, misconception correction, and changes in classroom interaction. Platform usage analytics provided supplementary data, recording a mean engagement of 4.2 sessions per week over the intervention period.

Results

The implementation of ChemBuddy Supreme+ led to statistically significant performance improvements across the cohort (N=53). A summary of the shift in grade distribution from baseline to post-intervention is presented in Table 2.

Table 2

Comparison of Cohort Performance at Baseline and Post-Intervention

Grade Band	Baseline	Post-Intervention	Change
A1–C6	5.66% (3/53)	66.03% (35/53)	+60.37%
A1–E8	30.18% (16/53)	88.67% (47/53)	+58.49%
U	69.81% (37/53)	11.32% (6/53)	−58.49%

A paired-samples t-test indicated that mean scores increased significantly from the baseline assessment ($M = 28.4$) to the post-intervention assessment ($M = 54.2$), yielding $t(52) = 11.42$, $p < 0.001$. The calculated effect size was $d = 1.93$, indicating a very large practical impact by conventional benchmarks (Cohen, 1988). Qualitative evidence from teacher observations and student interviews corroborated these gains. Student S12 remarked: “The Grader Agent did not just tell me I was wrong; it explained why my stoichiometry was off while I was still doing the problem.” The learner with visual impairment successfully navigated the full range of redox reaction exercises using the platform’s ARIA-labelled interface and screen-reading commands, demonstrating a level of independent subject access not previously available through conventional instructional materials.

Discussion

The data reveal a substantial transformation in the cohort’s grade profile, most notably evidenced by a reduction in U-grade failures from 69.81% to 11.32% — a decrease of more than 58 percentage points within the eight-week intervention period. The statistically significant paired-samples t-test result (t=11.42, $p < 0.001$) and the very large effect size (d=1.93) collectively indicate that the observed improvements were not attributable to chance but reflect a meaningful pedagogical intervention. These findings are consistent with established literature on the efficacy of Intelligent Tutoring Systems and formative feedback in STEM education (Black & Wiliam, 2009; Holmes et al., 2019).

The multi-agent architecture of ChemBuddy Supreme+ appears to function as a pedagogical “safety net,” providing JIT feedback that prevents the crystallisation of misconceptions — a well-documented barrier to learning in abstract domains such as O-Level Chemistry (Eitemüller et al., 2023). Student interview data reveal that the Grader Agent’s ability to explain the source of an error in real time disrupted the cycle of misconception reinforcement that traditional delayed feedback cannot address. The Reflector Agent’s metacognitive prompts, which encouraged students to justify their reasoning steps, further deepened engagement with abstract chemical concepts, consistent with self-regulated learning frameworks (Tomisu et al., 2025).

The success of the learner with visual impairment represents a particularly significant finding. Cognitive walkthrough data demonstrated that the learner could independently navigate complex redox equations using the platform’s ARIA-labelled interface, a level of accessibility not previously available through conventional instructional materials. This outcome aligns with UDL principles and demonstrates that purposefully designed AI-driven accessibility features can effectively mitigate the “double barrier” of content abstraction and physical inaccessibility (Chemnad & Othman, 2024; Kohnke & Zaugg, 2025; Plazar et al., 2021).

Several limitations must be acknowledged. The study’s single-school, convenience-sampled design limits the generalisability of findings to broader Bruneian and regional contexts. The eight-week intervention period does not establish whether improvements are sustained over time or translate into performance in the formal national O-Level examination. The absence of a control group means that confounding variables — such as increased student motivation arising from platform novelty — cannot be entirely ruled out. Future research should employ randomised controlled or quasi-experimental designs with larger, more diverse samples and extended longitudinal follow-up periods.

Conclusion

This study provides compelling preliminary evidence that ChemBuddy Supreme+, as a syllabus-aligned, multi-agent AI platform, constitutes a viable and impactful tool for improving chemistry education outcomes and advancing inclusive STEM learning in Brunei Darussalam. The statistically significant and practically large performance gains observed across the cohort — coupled with the platform’s demonstrated accessibility for a learner with a visual impairment — establish a robust proof of concept that merits further investigation and consideration by policymakers.

From a policy perspective, these findings carry meaningful implications for Brunei’s Wawasan 2035 educational agenda. The successful deployment of an AI-enhanced learning platform in a resource-constrained secondary school setting demonstrates that equitable, high-quality STEM education need not be contingent on large-scale infrastructure investment. Thoughtful integration of curriculum-aligned technology can bridge persistent performance gaps and extend educational opportunities to learners whom conventional instructional models have historically underserved.

The platform’s multi-agent design offers a replicable architecture for other subjects and educational contexts in which personalised, scaffolded feedback is otherwise inaccessible. Future longitudinal research should examine the retention of learning gains beyond the intervention window and evaluate whether students who used ChemBuddy Supreme+ demonstrate superior performance in formal national O-Level examinations. Expanding the platform’s scope to serve a wider range of subjects, year levels, and accessibility needs would further strengthen its contribution to inclusive STEM education across Southeast Asia. Collaboration with regional educational authorities and special education specialists is recommended to co-design enhancements addressing the specific cultural, linguistic, and accessibility requirements of diverse learner populations.

References

Ali, A. O., & Abdullah, N. A. (2023). Basic education in Brunei Darussalam. In *Springer International Handbooks of Education* (pp. 67–88). Springer Nature. https://doi.org/10.1007/978-981-16-8136-3_53-1

Beale, R. (2025). Dialogic pedagogy for large language models: Aligning conversational AI with proven theories of learning. *arXiv*. https://doi.org/10.48550/arxiv.2506.19484

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chemnad, K., & Othman, A. (2024). Digital accessibility in the era of artificial intelligence: Bibliometric analysis and systematic review. *Frontiers in Artificial Intelligence, 7,* Article 1349668. https://doi.org/10.3389/frai.2024.1349668

Chua, K. H., & Karpudewan, M. (2019). Integrating nanoscience activities in enhancing Malaysian secondary school students' understanding of chemistry concepts. *Eurasia Journal of Mathematics, Science and Technology Education, 16*(1), em1804. https://doi.org/10.29333/ejmste/110781

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Eitemüller, C., Trauten, F., Striewe, M., & Walpuski, M. (2023). Digitalisation of multistep chemistry exercises with automated formative feedback. *Journal of Science Education and Technology, 32*(3), 453–468. https://doi.org/10.1007/s10956-023-10043-2

Hagos, T., & Andargie, D. (2022). Effects of technology-integrated formative assessment on students' conceptual and procedural knowledge in chemical equilibrium. *Journal of Education and Learning (EduLearn), 17*(1), 113–122. https://doi.org/10.11591/edulearn.v17i1.20630

HAYAT, K. Z. (2025). *ChemBuddy Supreme+: A multi-agent generative AI platform for enhancing O-Level chemistry education and accessibility in Brunei* [Unpublished manuscript]. Sultan Sharif Ali Secondary School.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Centre for Curriculum Redesign. http://oro.open.ac.uk/60255/

Kohnke, S., & Zaugg, T. (2025). Artificial intelligence: An untapped opportunity for equity and access in STEM education. *Education Sciences, 15*(1), Article 68. https://doi.org/10.3390/educsci15010068

Mundy, C. E., & Potgieter, M. (2019). Refining process-oriented guided inquiry learning for chemistry students in an academic development programme. *African Journal of Research in Mathematics, Science and Technology Education, 23*(2), 145–156. https://doi.org/10.1080/18117295.2019.1622223

Osman, K. (2013). Conceptual understanding in secondary school chemistry: A discussion of the difficulties experienced by students. *American Journal of Applied Sciences, 10*(5), 433–441. https://doi.org/10.3844/ajassp.2013.433.441

Owusu, M., Assem, H. D., Ossei-Anto, T. A., Ansah, F. O., & Darko, G. (2024). Challenges in understanding redox reactions among level 200 students in science colleges of education in Ghana. *Creative Education, 15*(10), 2187–2202. https://doi.org/10.4236/ce.2024.1510134

Park, J., Teo, T. W., Teo, A., Chang, J., Huang, J., & Koo, S. (2023). Integrating artificial intelligence into science lessons: Teachers' experiences and views. *International Journal of STEM Education, 10*(1), Article 58. https://doi.org/10.1186/s40594-023-00454-3

Payadnya, I. P. A. A., Putri, G. A. M. A., Suwija, I. K., Saelee, S., & Jayantika, I. G. A. N. T. (2024). Cultural integration in AI-enhanced mathematics education: Insights from Southeast Asian educators. *Journal for Multicultural Education, 19*(1), 58–71. https://doi.org/10.1108/jme-09-2024-0119

Plazar, J., Meulenberg, C. J. W., & Kermauner, A. (2021). Science education for blind and visually impaired children. *Methodological Horizons, 28*(1), 63–78. https://doi.org/10.21464/mo.28.1.10

Ramos, J. G., & Wilson-Kennedy, Z. S. (2024). Promoting equity and addressing concerns in teaching and learning with artificial intelligence. *Frontiers in Education, 9,* Article 1487882. https://doi.org/10.3389/feduc.2024.1487882

Raza, F. A., Singh, A. D., Kovilpillai, J. J. S., Hamdan, A., & Rajaratnam, V. (2025). Safeguarding integrity in AI-enhanced education: Stakeholder perspectives on accuracy, validity, and ethics in ASEAN. *European Journal of STEM Education, 10*(1), Article 22. https://doi.org/10.20897/ejsteme/17307

Suryadi, A., Amrulloh, M. S., Ya-ni, H., Dinarni, D., & Afrizal, A. (2025). Integrating artificial intelligence in secondary education: A comparative study of pedagogical readiness and student engagement in Southeast Asia. *International Journal of Educational Research Excellence (IJERE), 4*(2), 661–678. https://doi.org/10.55299/ijere.v4i2.1534

Tomisu, H., Ueda, J., & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education, 10,* Article 1697554. https://doi.org/10.3389/feduc.2025.1697554

Uzezi, J. G., Ezekiel, D., & Auwal, A.-K. M. (2017). Assessment of conceptual difficulties in the chemistry syllabus of the Nigerian science curriculum as perceived by high school college students. *American Journal of Educational Research, 5*(7), 710–718. https://doi.org/10.12691/education-5-7-3

Wang, L., Hodges, G. W., & Lee, J. (2022). Connecting macroscopic, molecular, and symbolic representations with immersive technologies in high school chemistry: The case of redox reactions. *Education Sciences, 12*(7), Article 428. https://doi.org/10.3390/educsci12070428

Webb, M., Prasse, D., Phillips, M., Kadijevich, D. M., Angeli, C., Strijker, A., Carvalho, A. A. A., Andresen, B. B., Dobozy, E., & Laugesen, H. (2018). Challenges for IT-enabled formative assessment of complex 21st-century skills. *Technology, Knowledge and Learning, 23*(3), 441–468. https://doi.org/10.1007/s10758-018-9379-7

Zaluddin, N. F. (2025). Self-efficacy level of chemistry students. *International Journal of Research and Innovation in Social Science, 9*(3), 6180–6192. https://doi.org/10.47772/ijriss.2025.903sedu0448

Acknowledgment

The author expresses sincere gratitude to Sultan Sharif Ali Secondary School, Brunei Darussalam, for facilitating this study. This research was self-funded and conducted as part of an initiative to enhance STEM equity and accessibility in Cambridge O-Level Chemistry instruction. Special thanks are extended to the Year 11 chemistry students and participating teachers for their invaluable time and contributions..

Author(s):

Kassim Zadil Hayat, Science Department, Sultan Sharif Ali Secondary School, Brunei Darussalam.
Email: zadil.kassim@teacher.smsultansharifali.edu.moe.bn