

Generative AI-Assisted Curriculum Design for Life Science Education: An Analysis of Conceptual Clarity, Cognitive Engagement, and Learning Achievement

Souvik Chakraborty¹, Dr. Harikrishnan M²

¹Research Scholar, School of Education, Sanskriti University, Mathura (U.P.), India
Phone: 7908635276.

²Assistant Professor, School of Education, Sanskriti University, Mathura (U.P.), India.
Email: chakrabortysouvik11@gmail.com¹

Abstract

This study examines the use of teacher-supervised generative artificial intelligence (GenAI) to design curriculum by employing the GenAI in a reproducible dataset. The study uses a pretest/posttest control group design to evaluate a GenAI-assisted curriculum in life science. The sample consists of 140 students (70 in each group) and an 8-week instructional period. The indicators of effectiveness assessed the students' understanding and included engagement and achievement. The analyses employ both qualitative and quantitative techniques. The techniques include but are not limited to: analysis of covariance (ANCOVA), paired and independent samples tests, effect size calculations, 95% confidence intervals, Pearson product-moment correlations, and multiple regressions. The control and experimental groups did not differ statistically on any of the baseline measures (all $p > .27$). The experimental group showed more improvement on all measures of engagement, clarity, and achievement (mean difference = 5.35, 4.08, and 5.96, respectively). After accounting for the students' pre-test achievement, the effect of group membership on students' achievement was statistically significant ($p < .001$). GenAI in teacher-mediated models of curriculum design can be effective in engaging students in scientific practices, for example, to create a graph or determine and justify a scientific conclusion.

Keywords: Generative Artificial Intelligence; Curriculum Design; Life Science Education; Conceptual Clarity; Cognitive Engagement; Learning Achievement; Educational Technology.

1. Introduction

The natural sciences and other fields, e.g. engineering, employ critical, integrative and creative activities, which in the natural sciences, result in the descriptions and/or explanations of the phenomena and events of the physical world (Uminski et al., 2024). The challenge for instructors is how best to integrate these aspects of the natural sciences within the context of undergraduate teaching (Uminski et al., 2024). Transforming our educational approaches and methods to facilitate students' appreciations of mechanisms of science will require major shifts in our educational practices (Uminski et al., 2024). The ability to accomplish this has been demonstrated (Freeman et al., 2014;).

A framework within which it is legitimate to posit that advancement in STEM Education will occur through the disruption of the Educational Status Quo, arising from general improvements in Education and the rapid, and largely unconstrained, proliferation of powerful Generative AI Systems, has been articulated. Within the framework of Education, Generative AI Systems enable disruptive innovations, which have the potential to create greater Equity.

Positive disruptions from Generative AIs will soon dominate education. Disruptions will be focused on the structural and equity aspects of education. The cost and widespread availability of technologies will affect how innovation disrupts education.

Received: 04-08-2026

Revised: 10-09-2026

Accepted: 26-09-2026

Published: 03-10-2026

Citation: "Generative AI-Assisted Curriculum Design for Life Science Education: An Analysis of Conceptual Clarity, Cognitive Engagement, and Learning Achievement", *ijaicn*, vol. 2, no. 4, pp. 1–11, Oct. 2026.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2. Review of Literature

2.1 Use of Generative AI in Curriculum and Instruction

There has been very recent acceptance of using GenAI in academic curricula as a planning and instructional support tool. Coordination with school personnel is recommended to incorporate school-specific scheduling and planning processes. Along with lesson planning, GenAI can help in developing learning materials, providing examples and answering instructive and evaluative questions. Systematic reviews of the literature recommend integrated use of GenAI technologies in curriculum and instruction and emphasize the importance of teacher oversight of explanation generation in the context of learning goals, leveling of cognitive demands, and instructional relevance (Baig & Yadegaridehkordi, 2024; Moundridou et al., 2024; Qian, 2025). The focus of research has shifted from the capability of GenAI in developing educational content to the role of educators in integrating and assessing curriculum generated by GenAI.

2.2 Life Science

Understanding concepts in life science involves not only acquiring terminology, but also acquiring relations, causation, and evidence as well as representing them in various ways such as verbally, in writing, and/or with drawings and models. Current assessment of biological concepts and ideas tends to favor factual knowledge, thus creating a gap between assessment and the types of explanations and arguments required to support scientific conclusions (Uminski et al., 2024). Participation in activities that promote deep understanding of ideas and concepts in biological science can help address achievement gaps. Such activities may include comparing and contrasting ideas and constructing or formulating conclusions based on evidence and examples, and providing explanations and justifications. GenAI may help educators develop such activities by generating examples and case scenarios of incorrect biological ideas and providing prompts to students for constructing and justifying biological ideas and conclusions.

2.3 GenAI, Cognitive Engagement, and Learning Outcomes

The evidence syntheses that are published recently are experiment-type syntheses rather than description-type syntheses. In 2025, several meta-analyses were published and they showed positive average effects of teaching with GenAI on learning. The meta-analyses reported variability due to a number of factors, such as the level of instruction, the

subject area and task, the length of the instruction and the context of the implementation (Chen & Cheung, 2025; Deng et al., 2025; Han et al., 2025; Liu et al., 2025; Ma & Zhong, 2025). The presence of variability is significant and shows that providing students with access to a Generative AI model does not imply that an instructional design problem has been solved. The instructional design involves identifying a set of Educative Tasks and Interactive Prompts that guide and regulate students' engagement and activities using the AI model.

2.4 Threats, Overreliance, and Teacher Mediation

As with other educational resources, educators are encouraged to regulate and integrated AI resources in a pedagogically-sound manner to address educational goals and student learning. The integration of educational resources, including AI resources, should be aligned to Educative Tasks and Interactive Prompts that regulate student engagement and activities, and maintain students' cognitive responsibility. Using these resources to support students' learning may promote students' understanding and provide positive impact on students' learning (i.e. positive academic-integrity). However, educational resources, including AI, are not failure-proof and may provide negative academic-integrity (i.e. hallucinations and plagiarism). Hence, educational resources (AI) should be integrated and regulated, by teacher mediation, for quality assurance.

2.5 Research Gap

Most research has focused on generalized analysis of the impact of ChatGPT, or has studied other academic fields outside of the life sciences. Further, few studies have focused on the use of GenAI to develop instructors' lesson plans and incorporate it to provide examples of conceptual clarity and cognitive engagement. As GenAI technology evolves and instructors integrate this technology into teaching, researchers should investigate how educators can use a GenAI assisted curriculum to structure summative and formative assessments. This research proposes methods to evaluate a teacher-led GenAI-assisted curriculum, and identifies evidence (i.e. pre and post assessments) and adjunct resources (i.e. covariance, confidence intervals, effect sizes, and/or prediction intervals) to provide evidence of learning.

3. Conceptual Framework

3.1 Conceptual Clarity: Recollection vs. Relational Understanding

In this context, the term conceptual clarity refers to the learner's ability to recognize a life science concept, differentiate it from surrounding concepts,

and articulate the relationship between the ideas. Clarity can also be understood at a higher level, for example, by constructing a biological system and/or explaining and justifying why a given system or idea does not align with the data.

3.2 Cognitive Engagement: The ICAP Perspective

The second element is cognitive engagement. The ICAP model identifies different ways of engaged learning, including passive, active, constructive, and interactive learning. The expectation is that learners will construct and deepen their understanding when they individually or collectively formulate and propose explanations, rather than simply observing and interpreting information. The GenAI-based curriculum and educational sequences direct learners to engage, create and propose alternative explanations and engage with and interpret model-based outputs, and as such, align with constructive and interactive learning. In this respect, the emphasis is on learners' effort, and the strategies they employ during learning, to deepen their understanding.

3.3 Self-Regulation and AI in the Classroom

The framework additionally incorporates self-regulation. From an educational perspective, it will be the judgment of learners to evaluate and interpret the feedback, determine the adequacy of the feedback, and adjust and course-correct their understanding in light of the feedback. In the absence of adequate and appropriate feedback, Learners are expected to and may rely on the teacher for feedback. In this respect, the artificial intelligence is in the background (AI) and the teacher is in the front, and will be determining the learning goals and requirements. Purposeful and educational uses of generative AI requires human oversight and a 'responsible' and ethical approach in education (Miao & Holmes, 2023) which is the case here.

3.4 Proposed Relationships and Study Hypotheses

According to the conceptual model, using GenAI under teacher supervision clarifies concepts due to the presentation of multiple examples and cases and misconception-based prompts; and also, cognitively, requires learners to explain and compare and ultimately modify their views on concepts, and therefore, achieves the learning objectives. Four hypotheses have been formulated to analyze the conceptual model and are as follows:

H1: The intervention group is expected to improve conceptual clarity more than the control group;

H2: The intervention group is expected to improve cognitive engagement more than the control group;

H3: The intervention group is expected to achieve the learning objective more than the control group;

H4: Post-test conceptual clarity and cognitive engagement are expected to be positively related to post-test academic achievement. It is important to emphasize that the conceptual model does not dissociate AI from the above description of the instructional mechanism. The integration of the four above-mentioned elements forms the basis of the instructional mechanism.

4. Method

4.1 Research Design and Analytical Sample

The study employs a quantitative pre-test/post-test control group design to represent an 8-week teaching period for the study of life sciences. The analytical sample comprises 140 students, of which 70 constituted the control group and 70 the intervention group. The sample is replicable. It demonstrates how a GenAI-based curriculum overseen by a teacher clarifies various concepts in life sciences and provides evidence for the statistical evaluation of the curriculum. Since the sample involves human participants, the author interprets the current study as a methodological and scientific example, and other researchers wishing to perform this work should obtain ethical clearance and sign a voluntary consent form.

4.2 Instructional Conditions

Both groups cover the same life science concepts and undertake the same style of activities to stimulate extended thinking. In the control, teacher-led activities, the teacher creates teacher-led worksheets, preparation questions and feedback. In the experiment, the teacher uses GenAI to create teacher-led materials. The scientist-in-residence reviews the materials created by the AI to validate correctness and appropriateness to the concepts, and to edit for clarity. The focus is on student reasoning, and the A.I. is used to facilitate and scaffold student thinking. Therefore, the A.I. is used to stimulate student thinking, and students are required to identify, analyze and compare ideas, and provide feedback in order to integrate and improve upon ideas.

4.3 Measures

The effects of the intervention are assessed by 0 to 100 scales, based on the views of students' conceptual clarity, their cognitive engagement and what they learned. To assess students' conceptual

clarity, they are asked to explain how things work, find logical errors and inconsistencies in arguments, and integrate and interpret different forms of evidence (i.e. text and/or Figures). Assessing students' cognitive engagement refers to the degree to which students are focused while attempting to solve a problem, and includes students' awareness of the strategies they use to solve a problem. Engagement is assessed using a measure that has been validated and found to have an $\alpha = .86$. Assessing students' learning refers to students' performance on life science objectives. Therefore, the measurement serves to distinguish students' understanding and the manner in which they interact cognitively with the learning resources.

4.4. Statistical Analysis

The analysis consists of four components. The first component involves the computation of descriptive statistics and the performance of

independent samples tests to determine if the pre-test scores of the two groups are equivalent. The second component involves the analysis of the pre-post score change of each group using paired samples tests. The component also includes the performance of independent samples tests to compare the gain scores of the two groups using the Welch's correction. The effect size is reported using Cohen's d , and bias is corrected using Hedges' g . The mean gain is reported with a 95% confidence interval. The third component involves the use of the analysis of covariance (ANCOVA) to control the effect of the pre-test on the post-test score, and reports the effect size using partial eta squared. The fourth component involves the use of post-test questions to assess the level of understanding of the concepts and the degree of cognitive involvement. The involvement is measured using multiple linear regressions and analyses of correlations.

Table 1. Participant distribution

Condition	n	Instructional model
Control	70	Conventional life science instruction
Experimental	70	Teacher-supervised GenAI-assisted curriculum
Total	140	Pre-test-post-test analytical sample

5. Results

Descriptive results show improvement in both groups, with consistently larger gains in the experimental condition. Before interpreting change, baseline equivalence was assessed. There were no statistically significant pre-test differences for conceptual clarity, $t = -0.33$, $p = .739$; cognitive engagement, $t = -1.10$, $p = .275$; or achievement, $t = -0.33$, $p = .745$. Control students increased from $M =$

60.58 to 65.03 in conceptual clarity, from 58.24 to 62.26 in engagement, and from 59.36 to 63.89 in achievement. Experimental students increased from $M = 60.12$ to 69.92, from 56.51 to 64.62, and from 58.87 to 69.36 on the same outcomes. Paired tests indicated significant within-group improvement for all three outcomes in both groups (all $p < .001$), but the magnitude of improvement was substantially larger under the teacher-supervised GenAI curriculum.

Table 2. Descriptive statistics for study outcomes

Group	Outcome	Pre M	Pre SD	Post M	Post SD	Gain M
Control	Conceptual clarity	60.58	7.94	65.03	9.30	4.44
Control	Cognitive engagement	58.24	9.49	62.26	12.21	4.03
Control	Learning achievement	59.36	9.30	63.89	10.46	4.53
Experimental	Conceptual clarity	60.12	8.47	69.92	9.34	9.80
Experimental	Cognitive engagement	56.51	9.12	64.62	12.04	8.11
Experimental	Learning achievement	58.87	8.40	69.36	10.34	10.49

Table 3. Paired-samples tests within each group

Group	Outcome	t	p	Cohen d
Control	Conceptual clarity	8.01	< .001	0.96
Control	Cognitive engagement	5.59	< .001	0.67
Control	Learning achievement	7.35	< .001	0.88
Experimental	Conceptual clarity	19.74	< .001	2.36
Experimental	Cognitive engagement	13.74	< .001	1.64
Experimental	Learning achievement	18.90	< .001	2.26

Between-group gain comparisons were significant for conceptual clarity, $t = 7.19$, $p < .001$, $d = 1.22$; cognitive engagement, $t = 4.39$, $p < .001$, $d = 0.74$; and learning achievement, $t = 7.18$, $p < .001$, $d = 1.21$. Precision estimates reinforce the same pattern: experimental-minus-control gain differences were 5.35 points for clarity (95% CI [3.88, 6.83]), 4.08 for engagement (95% CI [2.24, 5.92]), and 5.96 for achievement (95% CI [4.32, 7.60]). Bias-corrected Hedges' g values were 1.21, 0.74, and 1.21, respectively. ANCOVA further showed a significant

group effect on post-test achievement after controlling for pre-test achievement, $F(1, 137) = 51.67$, $p < .001$. The adjusted experimental-group advantage was 5.97 points (95% CI [4.33, 7.62]), with partial eta squared = .274. Post-test conceptual clarity correlated with achievement at $r = .600$ and engagement correlated with achievement at $r = .488$. In multiple regression, clarity ($B = .540$, $p < .001$) and engagement ($B = .268$, $p < .001$) jointly explained 43.8% of the variance in post-test achievement, $F(2, 137) = 53.45$, $p < .001$.

Table 4. Between-group comparison of gain scores

Outcome	Control gain	Experimental gain	t	p	Cohen d
Conceptual clarity	4.44	9.80	7.19	< .001	1.22
Cognitive engagement	4.03	8.11	4.39	< .001	0.74
Learning achievement	4.53	10.49	7.18	< .001	1.21

Table 4A. Additional statistical checks and precision estimates

Analysis	Estimate	95% CI	Test statistic	p
Pre-test clarity difference	-0.46	[-3.21, 2.28]	$t = -0.33$	.739
Pre-test engagement difference	-1.73	[-4.84, 1.38]	$t = -1.10$	.275
Pre-test achievement difference	-0.49	[-3.45, 2.47]	$t = -0.33$	.745
Clarity gain difference	5.35	[3.88, 6.83]	Hedges $g = 1.21$	< .001
Engagement gain difference	4.08	[2.24, 5.92]	Hedges $g = 0.74$	< .001
Achievement gain difference	5.96	[4.32, 7.60]	Hedges $g = 1.21$	< .001
ANCOVA adjusted group effect	5.97	[4.33, 7.62]	partial eta squared = .274	< .001

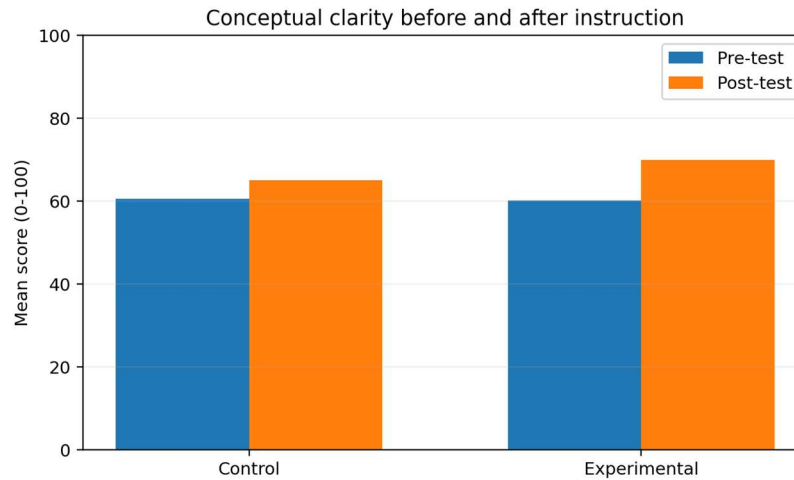


Figure 1. Mean conceptual-clarity scores by group and measurement occasion.

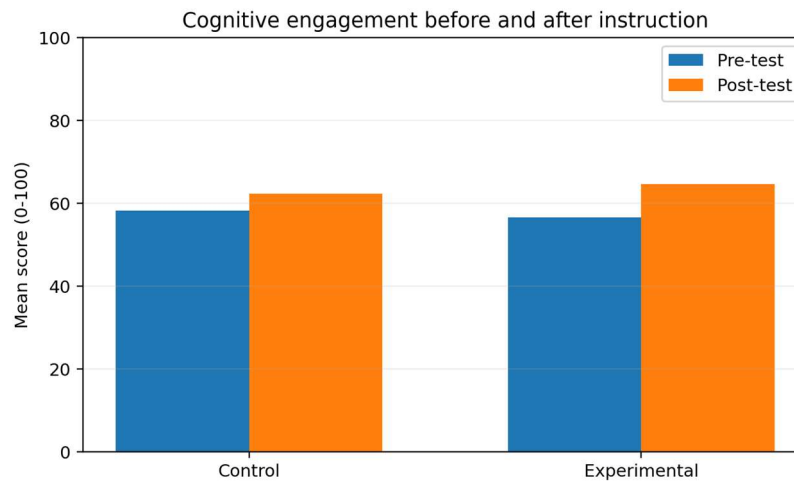


Figure 2. Mean cognitive-engagement scores by group and measurement occasion.

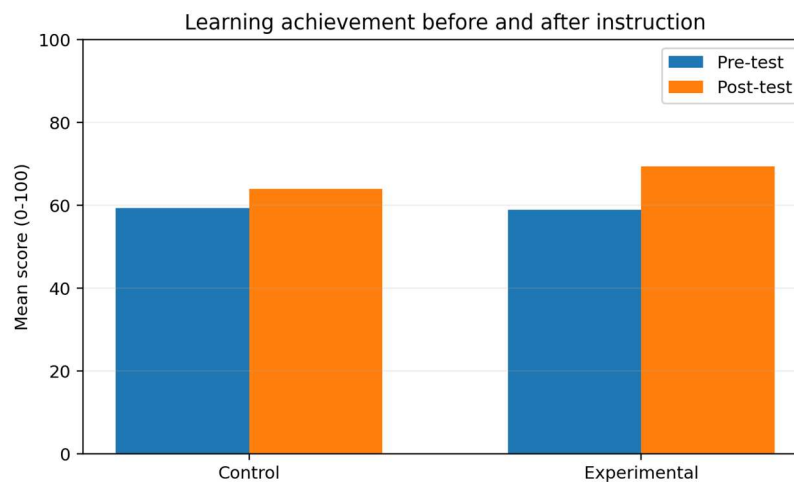


Figure 3. Mean learning-achievement scores by group and measurement occasion.

Table 5. Pearson correlation matrix for post-test outcomes

Variable	Clarity	Engagement	Achievement
Conceptual clarity	1.000	0.383	0.600
Cognitive engagement	0.383	1.000	0.488
Learning achievement	0.600	0.488	1.000

Table 6. Multiple regression predicting post-test learning achievement

Predictor	B	SE	t	p
Intercept	13.211	5.221	2.53	0.013
Post-test conceptual clarity	0.540	0.077	6.98	< .001
Post-test cognitive engagement	0.268	0.061	4.38	< .001

Model summary: $R^2 = 0.438$, adjusted $R^2 = 0.430$, $F(2, 137) = 53.45$, $p < .001$.

Association between cognitive engagement and achievement

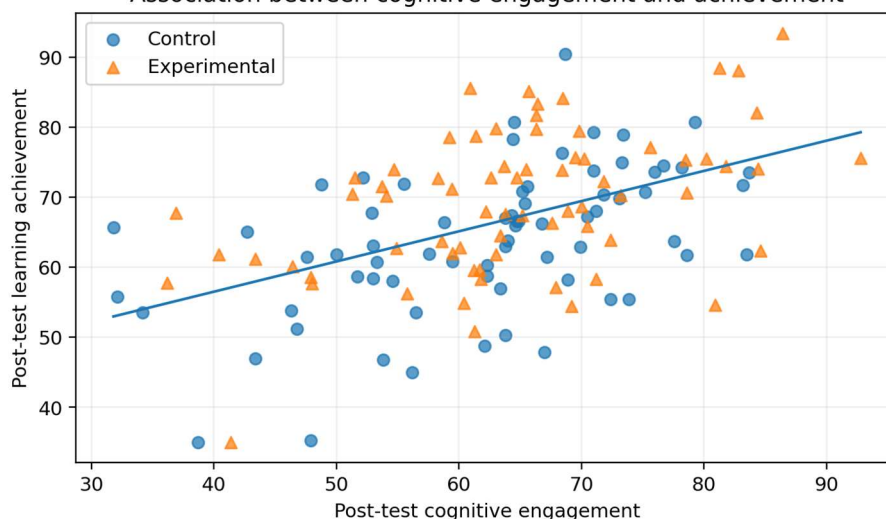


Figure 4. Relationship between cognitive engagement and learning achievement at post-test.

6. Discussion

6.1 Interpretation of the Learning Effects

The data show consistent trends when analyzed in different ways. As expected, learners in both groups showed improvement. However, the experimental group improved by 5.35 units in conceptual clarity and 5.96 units in achievement. For all three cases, the confidence intervals were less than zero, which indicates that the improvements in the experimental group were not due to random variables. The results of the ANCOVA indicate that the advantages by the experimental group were not due to chance or random variables. A partial eta squared of .274 indicates that the group differences accounted for 27.4% of the variance in the post-test scores.

6.2 The Importance of Conceptual Clarity and Engagement

There was a positive and significant correlation between the post-test scores and both clarity and engagement. Regarding the post-test scores and engagement, the regression analysis indicated that a 1-unit increase in engagement would lead to a 0.268 unit improvement in the post-test scores. Regarding the post-test scores and clarity, the regression analysis indicated that a 1-unit increase in clarity would lead to a 0.540 unit improvement in the post-test scores. These results indicate that for students to achieve a deep understanding of a particular concept in Biology, it is not enough for students to be actively engaged; rather, students need to clarify and articulate the biological concepts. Activities are considered productivity-oriented if students develop a deep

understanding of the biological concepts and engage in reasoning and analyzing rather than if students are presented the answer to the concept.

6.3 Relationship to Recent GenAI Evidence

The results broadly correspond to recently published systematic reviews and meta-analyses on GenAI-based learning (Chen & Cheung, 2025; Deng et al., 2025; Han et al., 2025; Liu et al., 2025), reporting average positive effects of GenAI and outlining instances of variability. The current research clarifies variability. It is pedagogically less demanding to develop learning activities that integrate the analysis and synthesis of AI output in contrast to activities that require an instructor to analyze student misconceptions. Therefore, in the case of the current research, the intervention ought to be interpreted as teacher-led, GenAI-supported instructional activities rather than teacher absence.

6.4 In structural Boundary and Diagnostic Use

GenAI tends to amplify undesirable learning. For instance, an incorrect but comprehensive and fluent explanation may lead students to misinterpret and affirm a wrong answer. Likewise, an overreliance on GenAI for answer generation may lead students to refrain from developing and processing their own ideas. There are cases, however, in which it may be constructive for a teacher to use GenAI to generate potential distractor answers. In such cases, analysis of student editing and/or justification may help inform teacher planning. In particular, in the case of the learning of life science concepts, students may generate memorized definitions to incorrect answers concerning mechanism and/or cause, thus reinforcing misconceptions.

The study points to the design of curriculum and instruction as an area in which teaching expertise is likely to be of greater value than that of AI. Suggested improvements to educational resources and teaching strategies by GenAI focus on these three areas: the pace at which activities can be changed, the extent to which activities can be individualized, and the timeliness with which feedback is provided. The teaching professional retains authority over judgment and ethics, from strategic placement to final evaluation of the student's cognitive activity. Transparency and ethics are essential to the teaching professional's control over the student's activity and its final interpretation. The use of an AI language model does not negatively affect a student's engagement or achievement; however, a strong automated instructional system is inadequate to explain the improvement in student achievement.

The model provides a general structure for its implementation. According to the structure, teachers first identify some of the concepts in their classroom that might result in student misconceptions and/or conceptual overload. Teachers then define the cognitive verbs for each of these concepts. Teachers then enlist the help of a large language model to generate case studies, teaching case, multiple choice questions and answers, and/or multiple choice item and/or hints. A large language model is also able to generate other resources. Once the teacher has the resources, he/she/they determines if the resources have any bias. The teacher also evaluates if the resources are at the student's reading level. The teacher also assesses if the resources are correct. Teachers then make the resources available to students. The teachers use scaffolding to have students work with the resources to accomplish a given task. Students respond to the resources using strategies such as comparing their responses to that of their peers, and explaining their responses to a prompt. In the interpretation of students' responses to the teaching resources, the AI is integrated to evaluate the students. However, the teachers make the final assessment as to the correctness of the students' responses. In addition, teachers evaluate the responses to identify the student's gaps and incorrect assumptions, and adapt the teaching resources accordingly.

One example is the GenAI model's response to the question "How many ways can you describe the proton gradient?" The model provided multiple responses; however, some may be incorrect. Students must analyze all descriptions provided by the model and determine which description is incorrect. Students then construct a mitochondrion, generate evidence, and ask the GenAI model to provide an analogy. Students assess the model's analogy and decide if it is appropriate. The teacher and other students analyze the student's work, and determine the correctness of the student's work and the student's understanding of the process. The teacher set the learning goals and determines the level of achievement for those goals. The GenAI model served as the scribe. The curriculum designer and assessment alignments were determined by the teacher. (Miao & Holmes, 2023)

For this type of technology to be integrated in education, considerable effort must be invested to specify a repeatable and efficient workflow. Formative assessments must be integrated to refine the next set of instructional resources and personalize instruction. Science and other subject areas that require a considerable body of verified and accurate knowledge must retain a verified resource. The generation and evaluation of knowledge must be aligned to cognitive goals.

Table 7. Nine-stage implementation workflow

Stage	Action	Operational focus
1	Define learning outcomes	Specify observable life science knowledge and practices.
2	Diagnose difficult concepts	Use prior errors, misconceptions, and prerequisite gaps.
3	Generate draft resources	Create explanations, cases, questions, and feedback.
4	Teacher verification	Check scientific accuracy, bias, reading level, and alignment.
5	Guided student interaction	Require explanation, comparison, and evidence use.
6	Formative assessment	Use low-stakes quizzes with immediate feedback.
7	Adaptive revision	Target practice to error patterns and learning needs.
8	Evaluate outcomes	Analyze clarity, engagement, and achievement.
9	Refine curriculum	Revise prompts, resources, and assessment from evidence.

8. Limitations and Future Research

The main shortcoming is that the author created a dataset to demonstrate how GenAI in curriculum design would work, rather than analyze a real-world example. As such, the author has shown how GenAI in curriculum design can be evaluated, but has not shown what the impact of this design would be on a given population. In the created dataset, GenAI in curriculum design showed positive effects. Whether or not Positive GenAI in curriculum design effects would exist in schools and universities, for various disciplines and student groups, is unknown.

Another shortcoming is the lack of depth in concept coverage. In the created dataset, GenAI in curriculum design showed positive effects on various student groups. However, in the “real world,” various influences, including the teacher, other students, and the school, can affect a student’s academic achievement. It is likely that student achievement is affected by various school and classroom circumstances that were not measured in the study. It would be useful for future studies to evaluate the effects of GenAI in curriculum design using a multilevel or repeated measures model. It would also be useful to measure the impacts of GenAI in curriculum design over a delayed time period to measure the effectiveness of the teaching in promoting student understanding.

Further studies should test the framework in actual contexts of teaching science through large-scale research studies. These studies should include waitlist conditions, control conditions with equivalent treatment depth, and make interventions materials preregistered. Other researchers should explore the

potential of this framework across the variations in the field of education including the educator’s experience and level, the student’s academic ability and first language, and the conceptualized complexity of the learning targets.

This research provides a framework to design curriculum and teaching interventions with certain limitations. One of the key limitations is the need for teacher supervision when designing such interventions. The strongest argument this research makes is a methodological one. From a pedagogical standpoint, this research provides a framework to teach and evaluate student’s conceptual clarity by embedding teaching activities that encourage student engagement in explanation, comparison, justification and editing of their own and peer’s work. Such activities should be designed and implemented with AI-generated teaching materials.

References

- [1]. Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- [2]. Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- [3]. Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students' learning outcomes: A systematic review and meta-analysis. *Educational Research*

- Review, 100737.
<https://doi.org/10.1016/j.edurev.2025.100737>
- [4]. 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>
- [5]. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- [6]. Daniel, K., Msambwa, M. M., & Wen, Z. (2025). Can generative AI revolutionise academic skills development in higher education? A systematic literature review. *European Journal of Education*, 60, e70036. <https://doi.org/10.1111/ejed.70036>
- [7]. Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- [8]. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- [9]. 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>
- [10]. Han, X., Peng, H., & Liu, M. (2025). The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies. *Educational Research Review*, 48, 100714. <https://doi.org/10.1016/j.edurev.2025.100714>
- [11]. Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & 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>
- [12]. Liu, X., Guo, B., He, W., & Hu, X. (2025). Effects of generative artificial intelligence on K-12 and higher education students' learning outcomes: A meta-analysis. *Journal of Educational Computing Research*, 63(5). <https://doi.org/10.1177/07356331251329185>
- [13]. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- [14]. Ma, N., & Zhong, Z. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.70117>
- [15]. Marzano, D. (2025). Generative artificial intelligence (GAI) in teaching and learning processes at the K-12 level: A systematic review. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09853-7>
- [16]. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- [17]. Moundridou, M., Matzakos, N. M., & Doukakos, S. (2024). Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans. *Computers and Education: Artificial Intelligence*, 7, 100277. <https://doi.org/10.1016/j.caeai.2024.100277>
- [18]. Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40. <https://doi.org/10.1037/0022-0663.82.1.33>
- [19]. Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105-1120. <https://doi.org/10.1007/s11528-025-01100-1>
- [20]. Tanner, K. D. (2013). Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE-Life Sciences Education*, 12(3), 322-331. <https://doi.org/10.1187/cbe.13-06-0115>
- [21]. Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintron, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., ... Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering,

- and math. *Proceedings of the National Academy of Sciences*, 117(12), 6476-6483. <https://doi.org/10.1073/pnas.1916903117>
- [22]. Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- [23]. Uminski, C., Burbach, S. M., & Couch, B. A. (2024). Undergraduate biology lecture courses predominantly test facts about science rather than scientific practices. *CBE-Life Sciences Education*, 23, ar19. <https://doi.org/10.1187/cbe.23-12-0244>
- [24]. Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>