Scaffolding Agency in STEM: The Effects of AI-Supported Self-Regulated Assessment on Performance and Self-Efficacy

Wai Kit, Mak
[1]Stewards Pooi Tun Secondary School
makwk@g.pooitun.edu.hk

Abstract: *This study investigates the impact of varying levels of selection autonomy on STEM performance and self-efficacy among 126 junior secondary students. Utilizing a three-phase repeated measures design, students transitioned from teacher-led assessment to AI-assisted choice via a role-play mentor, and finally to unguided self-selection. Results from MANOVA and paired t-tests reveal a significant stepwise decline in academic performance as autonomy increased, uncovering an "Autonomy Paradox" where high-achievers under structured conditions struggled most in autonomous settings. However, STEM self-efficacy remained stable across all phases, suggesting that the agency provided by AI and self-selection acts as a psychological buffer. These findings highlight the critical role of AI as a cognitive scaffold in the Zone of Proximal Development, facilitating a shared-regulation model that balances learner agency with academic achievement.*

Keywords: Student Agency, Artificial Intelligence in Education, STEM Self-Efficacy, Cognitive Load Theory, Learning Oriented Assessment

1. Introduction

In the rapidly evolving landscape of STEM education, the integration of Artificial Intelligence (AI) has shifted the focus from traditional teacher-led instruction to more personalized, learner-centric environments. Central to this transformation is the concept of Student Agency, which refers to the level of autonomy and choice students exercise in their learning process. While educational reformers advocate for increased agency to enhance motivation, the transition from structured environments to open-ended, autonomous tasks often presents a significant cognitive challenge for junior secondary students.

The core problem addressed in this study is the "Autonomy Paradox": providing students with the freedom to select their own assessment tasks may foster a sense of ownership, but without proper guidance, it can lead to a collapse in academic performance due to excessive cognitive load. This study draws upon Cognitive Load Theory (Sweller, 1988) to examine how unguided choice may overwhelm students' limited working memory, and Self-Determination Theory (Deci & Ryan, 2000) to explore how such agency impacts their STEM self-efficacy.

The emergence of AI-driven tools offers a potential solution to this paradox. By acting as a digital scaffold, AI can provide personalized recommendations that guide students toward tasks within their Zone of Proximal Development (ZPD). However, little empirical research has compared the actual performance and psychological shifts when students move through varying levels of autonomy—from teacher-selected tasks to AI-assisted choices, and finally to unguided self-selection.

This study involves 126 Form 1 and Form 2 students in a Hong Kong secondary school, engaging in STEM topics ranging from Micro:bit hardware to Generative AI multimedia. By analyzing their performance and self-efficacy across three distinct phases, this research aims to answer:

1. How does the level of selection autonomy (Teacher-led vs. AI-assisted vs. Learner-selected) affect objective STEM performance?
2. Does the shift in assessment mode impact students' STEM self-efficacy?
3. Is there a correlation between traditional academic excellence and the ability to self-regulate in autonomous assessment environments?

By investigating these questions, the research seeks to provide a roadmap for educators to implement Shared-Regulation models, where AI serves as a bridge between the precision of teacher instruction and the motivational benefits of student agency.

2. Literature Review

2.1. Student Agency and Assessment Choice

The integration of student agency into evaluative practices is primarily grounded in Self-Determination Theory (SDT), which posits that autonomy, competence, and relatedness are essential for fostering intrinsic motivation (Deci & Ryan, 2000). From the perspective of "Assessment as Learning," providing students with the authority to select their own evaluative tasks transforms them from passive subjects into active participants. However, a significant critique of unlimited autonomy is the Choice Overload Hypothesis; without structured guidance, students may experience increased cognitive burden (Sweller, 1988) or exhibit a "path of least resistance" bias, opting for cognitively undemanding tasks that offer minimal pedagogical growth (Kornell & Bjork, 2008).

2.2. Self-Efficacy in STEM Learning

According to Bandura's (1997) social cognitive theory, self-efficacy is constructed through four primary sources: mastery experiences, vicarious experiences, social persuasion, and physiological states. In STEM education, assessment design serves as a critical intervention point for these sources. While traditional high-stakes testing often triggers test anxiety—a physiological state that undermines confidence—agency-based assessment can foster a sense of mastery by aligning tasks with student interests. Furthermore, addressing self-efficacy is crucial for mitigating gender disparities in STEM, as female students frequently report lower confidence despite comparable technical performance, suggesting that the evaluative environment significantly shapes self-perception (Dasgupta & Stout, 2014; Schunk, 1991).

2.3. AI-Powered Task Cycles: From Teacher to Algorithm

The traditional Task Cycle in Learning Oriented Assessment (LOA) designates the instructor as the primary decision-maker responsible for task design and the interpretation of student performance (Jones & Saville, 2016). Recent advancements suggest that AI can effectively augment or even replace specific segments of this cycle. AI-powered adaptive systems can analyze student performance data in real-time to generate personalized content, effectively automating the clinical judgment traditionally exercised by educators. However, this transition necessitates a Shared-Regulation model; AI must provide the data-driven navigation while the student retains the executive decision-making power, thereby preserving learner agency without sacrificing precision (Luckin et al., 2016).

2.4. Question Quality, Difficulty Calibration, and the ZPD

The integration of student agency into evaluative practices is primarily grounded in Self-Determination Theory (SDT), which posits that autonomy, competence, and relatedness are essential for fostering intrinsic motivation (Deci & Ryan, 2000). From the perspective of "Assessment as Learning," providing students with the authority to select their own evaluative tasks transforms them from passive subjects into active participants. However, a significant critique of unlimited autonomy is the Choice Overload Hypothesis; without structured guidance, students may experience increased cognitive burden (Sweller, 1988) or exhibit a "path of least resistance" bias, opting for cognitively undemanding tasks that offer minimal pedagogical growth (Kornell & Bjork, 2008).

2.5. Developmental Maturity and Grade Level

As students progress from Form 1 to Form 2, their metacognitive abilities and self-regulatory skills undergo significant maturation (Zimmerman, 2002). Research indicates that younger learners require highly structured scaffolds, whereas older students demonstrate a heightened demand for autonomy and agency (Paris & Newman, 1990; Eccles et al., 1993). This study examines whether this developmental trajectory moderates the effectiveness of AI-supported choice, hypothesizing that older students are better equipped to leverage AI recommendations to bridge their own ZPD.

3. Methodology

3.1. Participants and Research Design

The study employed a one-group repeated measures design involving 126 junior secondary students (N=126) from a school in Hong Kong. The participants included students from Form 1 (n=73) and Form 2 (n=53). While the pedagogical framework remained consistent, the STEM content was grade-specific to align with the curriculum: Form 1 focused on Micro:bit hardware programming, whereas Form 2 engaged with Generative AI and Multimedia Elements. This dual-subject approach ensures that the findings regarding student agency and AI assistance are applicable across different domains of STEM education.

3.2 Task Design and Difficulty Calibration

To maintain internal validity, all assessment items were drawn from a centralized Item Bank categorized by three difficulty levels: Easy, Intermediate, and Challenging. Each quiz across the three phases was strictly configured to follow a fixed ratio of 2:2:1 (two easy, two intermediate, and one challenging item). This calibration ensures that any observed variance in student performance (Findings IV) is attributable to the selection mode rather than fluctuations in task difficulty.

3.3. The Three-Phase Experimental Procedure

The experiment was conducted in three sequential phases, representing the transition from teacher-led precision to unguided autonomy:

Phase 1: Teacher-led Assessment (TLA). The five items were pre-selected by the instructor, representing the traditional pedagogical baseline where learner agency is minimal.

Phase 2: AI-assisted Choice (AAC) via Role-Play. In this intervention phase, an AI agent assumed the persona of a "Learning Mentor." Students interacted with the AI, which provided a filtered list of recommended questions based on the student's prior performance. This "AI Role-Play" mechanism was designed to simulate the Shared-Regulation model (Luckin et al., 2016), where the AI suggests the most suitable path within the student's Zone of Proximal Development (ZPD) while the student retains the final decision-making power.

Phase 3: Unguided Learner Selection (ULS). Scaffolds removed; students independently selected items, testing their metacognitive ability to self-regulate (Paris & Newman, 1990).

3.4. Data Collection and Analysis

Academic performance was measured through the scores of the three quizzes. Simultaneously, the STEM Cognitive General Self-Efficacy Scale (STEM-CGSE, adapted from Leung & Leung, 2011) was administered after each phase to track longitudinal shifts in self-efficacy. The STEM-CGSE was contextually adapted for this study, focusing on students' confidence in their ability to tackle the specific STEM assessment tasks.

The scale included ten self-report items measuring various aspects of STEM self-efficacy, including task persistence ("I believe with enough effort, I can usually solve difficult problems in this STEM assessment"), problem-solving confidence ("I believe I can think of alternative solutions when my program/project encounters problems"), and Resilience Efficacy ("I am confident I can remain calm when facing difficulties and find errors to fix them").

Resilience Efficacy was calculated as a sub-score derived from four specific items of the scale that focus on students' confidence in: (1) handling unexpected technical situations, (2) maintaining composure when facing difficulties, (3) adapting to challenges with alternative approaches, and (4) self-regulating when encountering obstacles. Each item was rated on a 4-point Likert scale (1=strongly disagree, 4=strongly agree), and the Resilience Efficacy score was computed as the mean of these four items, with higher scores indicating greater perceived capacity to persist through technical challenges—a critical skill in autonomous learning environments.

Data were analyzed using Repeated Measures MANOVA and Pearson Correlation in SPSS to examine the interplay between objective performance (quiz scores) and subjective confidence (self-efficacy ratings) across the three assessment modes.

4. Findings & Discussion

4.1. Academic Performance across Assessment Modes

The first objective of this study was to compare STEM performance across three modes of question selection. As shown in Table 1, descriptive statistics indicate a clear downward trend in scores as student autonomy increased.

Table 1. *Descriptive Statistics for Performance and Self-Efficacy (N=126).*

Variable	Mean (M)	Std. Dev (SD)	Min	Max
Quiz 1 (Teacher-led)	3.80	0.76	2.0	5.0
Quiz 2 (AI-assisted)	3.48	0.80	2.0	5.0
Quiz 3 (Learner-selected)	3.06	0.52	2.0	5.0
STEM Self-Efficacy (Overall)	2.74	0.69	1.0	4.0

The To test for significance, a Repeated Measures MANOVA was conducted. The results (Table 2) reveal a significant main effect of the assessment mode on performance with a large effect size ($\eta p^2 = .383$). Crucially, the interaction between mode and grade level was not significant ($p = .160$), suggesting that both Form 1 and Form 2 students were similarly affected by the shift in autonomy.

Table 2. *Repeated Measures MANOVA for Performance and Grade Effects.*

Source	df	F	p	ηp^2
Assessment Mode	2	38.185	< .001***	.383
Grade Level	1	0.589	.444	.005
Mode X Grade	2	1.857	.160	.029

Post-hoc paired samples t-tests (Table 3) confirmed that the decline between each phase was statistically significant. The moderate-to-large effect size (Cohen's d = 0.627) for the drop from Q1 to Q2 suggests that even with AI support, students struggle to match teacher-calibrated performance.

Table 3. *Paired Samples T-Tests for Performance Comparison.*

Comparison	Mean Diff.	t	p	Cohen's d
Quiz 1 vs. Quiz 2	0.325	7.039	< .001***	0.627
Quiz 2 vs. Quiz 3	0.413	3.943	< .001***	0.351

4.2. The Stability of STEM Self-Efficacy

While performance declined, STEM self-efficacy remained stable. This suggests a "psychological buffering" effect; the agency provided in the later phases compensated for lower scores, maintaining the students' belief in their technical abilities (Deci & Ryan, 2000).

Table 4. *Repeated Measures MANOVA for Self-Efficacy.*

Source	df	F	p	ηp^2
Assessment Mode	2	0.365	.695	.006
Grade Level	1	0.050	.824	.000

This stability (M≈2.75) suggests a "psychological buffering" effect. According to Self-Determination Theory, the agency provided in Q2 and Q3 may have compensated for lower scores, maintaining students' belief in their technical coping abilities despite a more challenging.

4.3. The Autonomy Paradox: Correlation Analysis

The most striking finding lies in the intercorrelations between modes (Table 5). A significant negative correlation was found between Quiz 3 scores and earlier performance (Q1 & Q2). This implies that students who excelled under teacher-led instruction experienced the greatest difficulty when granted full autonomy. However, the positive correlation between Q3 performance and Resilience Efficacy (r = .180) indicates that those who maintained confidence in "handling trouble" were more successful in self-selecting appropriate tasks (Zimmerman, 2002).

Table 5. *Correlations between Performance and Self-Efficacy Measures.*

Variables	1	2	3	4	5
1. Quiz 1 Score	-				
2. Quiz 2 Score	.779***	-			
3. Quiz 3 Score	-.660***	-.578***	-		
4. Self-Efficacy (Overall)	-.091	-.072	.074	-	
5. Resilience Efficacy	-.207*	-.157	.180*	.804***	-

**p < .05, ***p < .001*

4.4. Synthesis and Discussion

The findings highlight the "Autonomy Paradox" in STEM assessment. The significant performance drop in Q3 supports Cognitive Load Theory (Sweller, 1988), showing that unguided choice creates an extraneous load that exceeds students' germane resources.However, AI-assisted selection (Q2) serves as a critical cognitive scaffold. It mitigates the performance decline seen in Q3 while maintaining the motivational benefits of choice. This study suggests that AI-powered task cycles can bridge the gap between teacher-led precision and student agency, fostering a

Shared-Regulation environment where students remain confident even as they navigate the complexities of autonomous learning.

5. Implications & Limitations

5.1. Pedagogical Implications

The results of this study offer profound theoretical and practical insights for the integration of AI-assisted assessment in STEM education. First, the sharp decline in performance from teacher-led (Q1) to fully autonomous selection (Q3) validates the perspective of Jones and Saville (2016), who argue that the traditional Task Cycle in Learning Oriented Assessment (LOA) relies heavily on professional teacher decision-making to ensure task difficulty is appropriately calibrated to student ability. However, the significant mitigation of this performance drop in the AI-assisted mode (Q2) supports the contentions suggesting that AI-driven adaptive systems can effectively automate certain clinical judgments by identifying knowledge gaps and providing precise task recommendations. This implies that STEM educators should not strive for an extreme, unguided form of student autonomy; instead, they should adopt a "Shared-Regulation" model. In this framework, AI serves as a digital scaffold that preserves learner agency while ensuring that assessment tasks remain within the student’s Zone of Proximal Development (ZPD).

Furthermore, this research identifies a crucial psychological buffering effect: despite the objective decline in academic scores, students’ STEM self-efficacy remained remarkably stable across all modes. This aligns with Self-Determination Theory (SDT), indicating that the sense of choice provided by AI-assisted interfaces can compensate for the frustration of lower performance. In high-stakes STEM tasks such as coding or circuit troubleshooting, this agency-driven psychological protection is vital to prevent students from falling into learned helplessness. Most notably, the "Autonomy Paradox" revealed by the correlation analysis—where high-achievers under teacher-led conditions struggled most when granted full autonomy—suggests that for these students, the instructional focus should shift from pure content mastery to the development of metacognitive selection strategies. This ensures that they can eventually internalize the decision-making logic provided by AI and apply their technical expertise in unstructured, real-world problem-solving scenarios.

5.2. Research Limitations

While this study provides valuable empirical evidence, several limitations warrant consideration for future research. First, the sample was confined to junior secondary students (Form 1 and Form 2) within a single institutional context. Given that metacognitive maturity evolves with age, senior secondary or university students might exhibit different behavioural patterns when interacting with AI recommendations and exercising autonomy. Second, the short-term experimental design of this study does not account for long-term "algorithm dependency." It remains unclear whether prolonged exposure to AI-assisted choice will eventually allow learners to internalize AI logic and improve independent self-calibration, or if it will lead to a reliance on automated suggestions. Finally, the specific diagnostic logic of the AI model used in this experiment has inherent limitations. Future studies should investigate how different types of AI interfaces—such as those based on Large Language Models (LLMs) versus those based on rigid knowledge-tracing models—differentially impact student resilience and academic outcomes.

6. Conclusion

The This study investigated the intricate relationship between student agency, academic performance, and self-efficacy within an AI-supported STEM assessment framework. By transitioning students from teacher-led tasks to AI-assisted and eventually unguided selection, the research captured a phenomenon termed the "Autonomy Paradox."

The results demonstrate that while increased learner agency is a significant driver of intrinsic motivation and psychological resilience, it also imposes a substantial cognitive burden that can hinder objective performance if not properly scaffolded.

The empirical evidence suggests that AI-assisted selection (Phase 2) serves as a critical bridge between traditional instruction and independent learning. While performance naturally declined as students took on the responsibility of task selection, the AI's "Learning Mentor" role provided a necessary scaffold that prevented the performance collapse observed in the unguided phase. Furthermore, the stability of STEM self-efficacy despite fluctuating scores indicates that the presence of choice—even when supported by an algorithm—protects students' confidence. This finding challenges the traditional "failure-frustration" cycle, suggesting that agency allows students to internalize their learning process and maintain belief in their technical potential.

Perhaps the most significant finding for STEM educators is the vulnerability of high-achieving students in autonomous settings. The negative correlation between traditional performance and unguided selection performance ($r = -.660$) highlights a critical gap in current pedagogical practices: students who master content under direction may lack the metacognitive skills required for self-regulation. Therefore, the goal of integrating AI into STEM education should not be to replace the teacher or simply automate grading, but to facilitate a Shared-Regulation model. By utilizing AI to navigate the Zone of Proximal Development, educators can empower students to exercise agency responsibly, ensuring they develop both the technical competence and the self-regulatory maturity needed for the complex, unstructured challenges of the modern technological landscape.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W. H. Freeman. https://archive.org/details/selfefficacyexer0000band

Black, P., & Wiliam, D. (1998). *Assessment and classroom learning*. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

Dasgupta, N., & Stout, J. G. (2014). *Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers*. Policy Insights from the Behavioral and Brain Sciences, 1(1), 21–29. https://doi.org/10.1177/2372732214549471

Deci, E. L., & Ryan, R. M. (2000). *The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior*. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). *Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and families*. American Psychologist, 48(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Jones, N., & Saville, N. (2016). *Learning oriented assessment: A systemic approach.* Cambridge University Press. https://www.cambridge.org/gb/cambridgeenglish/catalog/teacher-training-development-and-research/learning-oriented-assessment

Kornell, N., & Bjork, R. A. (2008). *Learning concepts and categories: Is spacing the "enemy of induction"?.* Psychological Science, 19(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Leung, D. Y., & Leung, A. Y. (2011). Factor structure and gender invariance of the Chinese General Self-Efficacy Scale among soon-to-be-aged adults. Journal of advanced nursing, 67(6), 1383–1392. https://doi.org/10.1111/j.1365-2648.2010.05529.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* Pearson.https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/Intelligence-Unleashed-Publication.pdf

Paris, S. G., & Newman, R. S. (1990). *Developmental aspects of self-regulated learning.* Educational Psychologist, 25(1), 87–102. https://doi.org/10.1207/S15326985EP2501_7

Schunk, D. H. (1991). *Self-efficacy and academic motivation.* Educational Psychologist, 26(3-4), 207–231. https://doi.org/10.1080/00461520.1991.9653133

Sweller, J. (1988). *Cognitive load during problem solving: Effects on learning.* Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Zimmerman, B. J. (2002). *Becoming a self-regulated learner: An overview.* Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2