

## The Impact of Real-Time AI Feedback on Students' Productive Struggle and Conceptual Understanding in Algebra

**Obi Perpetua Ngozi**

Department of Mathematics  
Benjamin Uwajumogu State College of Education, Ihitte-Uboma, Imo State, Nigeria  
zikkyn2016@gmail.com

**Iheanaetu Beatrice Ifeyinwa**

Department of Educational Foundation  
Benjamin Uwajumogu State College of Education, Ihitte-Uboma, Imo State, Nigeria  
Ifeyinwa@gmail.com

Corresponding author: zikkyn2016@gmail.com; Ifeyinwa@gmail.com

### ABSTRACT

Artificial intelligence (AI) tutors and chatbots can now give students instant feedback on every step of an algebra problem. This capability promises personalised support at scale, but it also creates a dilemma. Research in mathematics education shows that productive struggle, sustained effort on problems just beyond a learner's current competence, is a key driver of conceptual understanding. Feedback that arrives too quickly, too often or too explicitly may remove the struggle that makes learning durable, while feedback that arrives too late may leave students frustrated and stuck. This theoretical paper examines the impact of real-time AI feedback on students' productive struggle and conceptual understanding in algebra. Drawing on Vygotsky's zone of proximal development, the assistance dilemma, desirable difficulties and productive failure, it reviews evidence from intelligent tutoring systems, feedback research and recent studies of generative AI in mathematics. The review indicates that the impact of real-time AI feedback depends less on its presence than on its design: process-oriented, graduated and metacognitive feedback tends to preserve productive struggle and build understanding, while immediate answer-giving tends to encourage dependence, gaming behaviour and procedural rather than conceptual learning. The paper proposes a framework for calibrated AI feedback in algebra built on five principles: wait before helping, escalate hints gradually, target reasoning rather than answers, detect unproductive struggle, and prompt reflection. Implications for teachers, designers and Nigerian mathematics classrooms are discussed, and empirical testing of the framework is recommended.

**Keywords:** AI feedback; productive struggle; conceptual understanding; algebra; intelligent tutoring systems; assistance dilemma

### Introduction

Algebra is a gateway subject. It introduces students to abstraction, generalisation and symbolic reasoning, and success in algebra is a strong predictor of later progress in mathematics, science

and technology. Yet many students experience algebra as a set of rules to memorise, such as "move it to the other side and change the sign", rather than as a coherent system of ideas. They may solve routine equations correctly while misunderstanding what an equation, a variable or an equivalent expression actually means. Kilpatrick et al. (2001) argued that mathematical proficiency requires conceptual understanding alongside procedural fluency, and that the two must be developed together. Technology has long been used to support algebra learning. Intelligent tutoring systems such as Cognitive Tutor Algebra have provided step-by-step feedback for decades, with evidence of positive effects at scale (Pane et al., 2014). Since 2022, generative AI chatbots have made realtime feedback far more widely available. A student can photograph an equation, receive a worked solution in seconds, or ask a chatbot to check every line of their work as they write. This abundance of instant help raises a fundamental question. Mathematics education research holds that students learn deeply when they grapple with problems that are challenging but within reach, a process called productive struggle (Hiebert & Grouws, 2007; Warshauer, 2015). If AI feedback intervenes at the first sign of difficulty, it may eliminate this struggle. Conversely, well-designed feedback might help students stay in productive struggle rather than slipping into frustration. Koedinger and Aleven (2007) called this tension the assistance dilemma: how much help, and when, best promotes learning. Early evidence from generative AI suggests the stakes are real; Bastani et al. (2025) found that high school students who used an unrestricted AI tutor during mathematics practice performed better during practice but worse on later tests taken without it, whereas a version designed with guardrails largely avoided this harm. The purpose of this paper is to examine the impact of real-time AI feedback on students' productive struggle and conceptual understanding in algebra. Specifically, it (a) clarifies key concepts, (b) presents a theoretical framework, (c) reviews evidence on how AI feedback can support or undermine productive struggle and understanding, (d) proposes a framework for calibrated AI feedback in algebra, and (e) draws implications for practice and research, including in Nigeria.

## Conceptual Clarification

### Real-Time AI Feedback

Feedback is information provided to learners about their performance relative to a goal, intended to close the gap between current and desired understanding (Hattie & Timperley, 2007). Realtime AI feedback is feedback generated automatically by an AI system during the problemsolving process, rather than after a task is submitted. It may take several forms: correctness flags (right or wrong), error-specific messages, hints, worked steps, full solutions, or questions that prompt reflection. Systems range from rule-based intelligent tutors, which compare each step with a model of correct solutions, to generative AI chatbots, which respond flexibly in natural language.

### Productive Struggle

Productive struggle refers to students' effort to make sense of mathematics, to figure out something that is not immediately apparent (Hiebert & Grouws, 2007). It is productive when the challenge lies within the student's reach, the student remains engaged, and the effort leads toward understanding. Warshauer (2015) showed that how teachers respond to struggle matters: responses that maintain the cognitive demand of the task, through questioning and encouragement, support learning, whereas responses that simply tell students what to do lower the demand. Struggle becomes unproductive when students are stuck without a path forward, become frustrated, or disengage.

### Conceptual Understanding in Algebra

Conceptual understanding is comprehension of mathematical concepts, operations and relations (Kilpatrick et al., 2001). Skemp (1976) distinguished relational understanding, knowing both what to do and why, from instrumental understanding, knowing rules without reasons. In algebra, conceptual understanding includes grasping the equal sign as a statement of equivalence, seeing variables as quantities that can vary, understanding why operations performed on both sides preserve equality, and recognising structure in expressions. Rittle-Johnson et al. (2001) showed that conceptual and procedural knowledge develop iteratively, each supporting the other.

### Theoretical Framework

Zone of Proximal Development and Scaffolding Vygotsky (1978) described the zone of proximal development (ZPD) as the distance between what a learner can do independently and what the learner can achieve with guidance. Wood et al. (1976) introduced the term scaffolding for the support a tutor provides within this zone: recruiting interest, simplifying the task, highlighting critical features and, importantly, withdrawing support as the learner becomes capable. Productive struggle can be understood as work within the ZPD. Effective AI feedback acts as a scaffold that keeps the learner in the zone, whereas feedback that does the work for the learner moves the task out of the zone and removes the need to learn.

### The Assistance Dilemma

Koedinger and Alevan (2007) framed the central design question for tutoring systems as the assistance dilemma: giving information can speed performance and reduce errors, but withholding it can produce more active processing and deeper learning. Neither extreme is optimal; the best level of assistance depends on the learner's knowledge, the task and the type of help. The dilemma is directly relevant to real-time AI feedback, which by design makes assistance constantly available.

## Desirable Difficulties

Bjork and Bjork (2011) argued that some conditions that make learning slower or harder in the short term, such as spacing, retrieval practice and generating answers rather than reading them, produce better long-term retention and transfer. They called these desirable difficulties. Instant

correction can make practice feel easier and look more successful while reducing these beneficial difficulties.

## Productive Failure

Kapur (2008, 2014) showed that students who first attempt to solve complex problems on their own, often failing to reach correct solutions, and then receive instruction, develop better conceptual understanding and transfer than students who receive instruction first. Kapur (2016) explains that the initial struggle activates prior knowledge, helps students notice the critical features of the concept, and prepares them to understand the canonical solution. Productive failure suggests that feedback timing matters: intervening before students have explored the problem may cut short this preparatory process.

## A Counterpoint: Cognitive Load Theory

Cognitive load theory (Sweller, 1988) warns that novices have limited working memory and can be overwhelmed by unguided problem solving; Kirschner et al. (2006) argued that minimal guidance is less effective than explicit instruction for novices. This perspective is a necessary counterweight: struggle is not automatically productive, and novices facing unfamiliar algebra may need worked examples and timely guidance. Together, these theories suggest that the aim of AI feedback is not to eliminate struggle nor to prolong it indefinitely, but to calibrate it.

## Review of Evidence Feedback in General

Feedback is among the most powerful influences on learning, but its effects vary widely. Kluger and DeNisi (1996) found that in about a third of the studies they reviewed, feedback actually reduced performance, especially when it directed attention to the self rather than the task. Hattie and Timperley (2007) distinguished feedback about the task, the processing of the task, self-regulation and the self, concluding that process and self-regulation feedback are most effective for deep learning. A later meta-analysis by Wisniewski et al. (2020) confirmed that feedback rich in information has larger effects than simple reinforcement or correctness signals. Shute (2008) reviewed formative feedback and concluded that elaborated, specific feedback generally helps, but that its optimal timing depends on the learner and task: immediate feedback can help novices with procedural tasks, whereas delayed feedback may better support transfer.

## Intelligent Tutoring Systems in Algebra

Intelligent tutoring systems offer the most extensive evidence on real-time feedback in algebra. Cognitive Tutor Algebra provides step-level feedback and on-demand hints that escalate from general prompts to the specific next step. A large randomised trial by Pane et al. (2014) found positive effects on algebra achievement in its second year of use. VanLehn (2011) found that step-based tutors approach the effectiveness of human tutors, attributing this to the fine granularity of their feedback.

However, the same research documents risks. Baker et al. (2004) found that some students "game the system", rapidly clicking through hints until the tutor reveals the answer, and that such gaming was associated with lower learning. Aleven et al. (2006) showed that many students use help poorly, either avoiding it when needed or abusing it to obtain answers, and developed tutors

that coach help-seeking itself. These findings show that the availability of real-time help does not guarantee that it is used productively.

## Generative AI and Mathematics Learning

Generative AI raises these risks further, because chatbots can produce complete, fluent solutions instantly and on request. In a field experiment in high school mathematics, Bastani et al. (2025) found that students given unrestricted access to a generative AI tutor performed substantially better during practice, but performed worse than a control group on subsequent tests without AI, suggesting that the tool substituted for rather than supported learning. A version designed to give hints rather than answers largely mitigated the negative effect. This finding illustrates the assistance dilemma vividly: the same technology can help or harm depending on its feedback design.

## Synthesis

**Table 1. Feedback characteristics and their likely impact on productive struggle and conceptual**

| Feedback characteristic                                                  | Likely effect on productive struggle            | Likely effect on conceptual understanding                         |
|--------------------------------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------|
| Immediate full solutions on request                                      | Eliminates struggle; invites dependence         | Weak; promotes copying of procedures                              |
| Correct/incorrect flags only                                             | Maintains struggle but may frustrate            | Limited; signals errors without explaining them                   |
| Graduated hints (general to specific)                                    | Preserves struggle while preventing impasse     | Moderate to strong, if students engage with early hints           |
| Process and error-specific feedback                                      | Keeps effort focused on reasoning               | Strong; addresses misconceptions such as misuse of the equal sign |
| Metacognitive prompts ("Why does this step keep the equation balanced?") | Sustains productive effort                      | Strong; builds relational understanding                           |
| Delayed or on-demand feedback after an attempt                           | Allows exploration; supports productive failure | Strong for transfer, if novices are not overwhelmed               |

understanding Feedback characteristic Likely effect on productive struggle Likely effect on conceptual understanding Immediate full solutions on request Eliminates struggle; invites Weak; promotes copying of request dependence procedures Correct/incorrect flags only Maintains struggle but may Limited; signals errors frustrate without explaining them Graduated hints (general to Preserves struggle while Moderate to strong, if specific) preventing impasse students engage with early hints Process and error-specific Keeps effort focused on Strong; addresses feedback reasoning misconceptions such as misuse of the equal sign Metacognitive prompts Sustains productive effort Strong; builds relational ("Why does this step keep the understanding equation balanced?") Delayed or on-demand Allows exploration; supports Strong for transfer, if novices feedback after an attempt productive failure are not overwhelmed

The evidence points to a consistent conclusion: real-time AI feedback is neither inherently beneficial nor harmful to productive struggle and conceptual understanding. Its impact depends on timing, specificity, focus and the degree of control it gives students over obtaining answers.

A Framework for Calibrated AI Feedback in Algebra Drawing on the theories and evidence reviewed, this paper proposes a framework of five principles for designing and using real-time AI feedback that preserves productive struggle and builds conceptual understanding in algebra.

## 1. Wait before helping. The system should allow students a genuine first attempt before

offering help, for example by requiring a written attempt or a short period of work before hints become available, except where a student is clearly at an impasse. This protects the exploratory phase that productive failure research identifies as valuable (Kapur, 2016).

## 2. Escalate hints gradually. Help should move through levels: (a) a prompt to re-read or

restate the problem; (b) a strategic hint ("What operation would undo the multiplication by 3?"); (c) a hint focused on the specific step; (d) a worked example of a similar problem; and only then (e) the step itself. The full answer should be the last resort, not the default.

## 3. Target reasoning, not answers. Feedback should diagnose the misconception behind an

error and address the underlying concept. For example, rather than saying "Incorrect,  $x = 4$ ", the system might say, "You subtracted 5 from the left side only. What must happen to the right side for the equation to stay balanced?" Such feedback reinforces the meaning of equivalence rather than a rule.

## 4. Detect unproductive struggle. The system should monitor signals such as repeated

identical errors, long inactivity, rapid hint requests or random answer guessing, and respond differently to productive and unproductive struggle: offering encouragement and a strategic prompt for the former, and more direct support or a worked example for the latter. This balances productive struggle with the cautions of cognitive load theory.

## 5. Prompt reflection and explanation. After a problem is solved, the system should ask

the student to explain a key step, compare two solution methods, or predict what would change if a value changed. Self-explanation and reflection convert a correct procedure into conceptual understanding and develop self-regulation (Hattie & Timperley, 2007). The framework also implies a role for the teacher. AI feedback should be integrated into lessons in which teachers pose rich tasks, orchestrate discussion of different solution methods, and review data from the AI system to identify common misconceptions. The AI handles individual, moment-to-moment support; the teacher builds shared meaning and decides when struggle should continue.

## Implications for Practice

For teachers. Teachers should treat AI tools as partners in maintaining productive struggle, not as answer machines. They can set norms for AI use (for example, attempt first, ask for hints,

explain afterwards), choose or configure tools that give hints rather than full solutions, and use class discussion to consolidate conceptual understanding. For designers. Developers of AI mathematics tools should build in delayed and graduated help, misconception-specific feedback, struggle detection and reflection prompts, and should avoid defaulting to complete solutions. Generative AI tutors can be configured with instructions to guide rather than solve, as the guardrailed tutor in Bastani et al. (2025) illustrates. For Nigerian mathematics classrooms. In Nigeria, large classes and shortages of qualified mathematics teachers limit individual support, and students increasingly use free photo-solver applications and chatbots, often without guidance. Without calibration, such tools risk reinforcing the rote, rule-based algebra learning already common in examination-focused classrooms. Colleges of education and faculties of education should prepare mathematics teachers to use AI feedback in ways that preserve struggle and build understanding, and curriculum authorities should provide guidance on AI use in secondary mathematics. Offline-capable tools designed with these principles could extend high-quality feedback to schools with limited connectivity.

## Conclusion

Real-time AI feedback can either strengthen or weaken students' productive struggle and conceptual understanding in algebra. Theory and evidence agree that struggle within the zone of proximal development, supported by well-timed and well-targeted scaffolding, produces deeper understanding than either unassisted frustration or instant answers. Intelligent tutoring research shows that graduated, step-level feedback can raise algebra achievement, while research on gaming, help abuse and unrestricted generative AI shows that easy access to answers can undermine learning. The impact of real-time AI feedback therefore lies in its design and use. Feedback that waits, escalates gradually, targets reasoning, distinguishes productive from unproductive struggle and prompts reflection is most likely to turn the speed of AI into lasting mathematical understanding.

## Recommendations

### 1. Mathematics teachers should adopt norms that require students to attempt problems

before seeking AI help and to explain their reasoning afterwards.

### 2. Developers and schools should select or configure AI tools that provide graduated hints

and misconception-specific feedback rather than immediate full solutions.

### 3. Teacher education programmes should train pre-service and in-service mathematics

teachers in the theory of productive struggle and the pedagogical use of AI feedback.

### 4. Curriculum and examination bodies should issue guidance on the responsible use of AI in

mathematics learning.

### 5. Researchers should empirically test the proposed framework, comparing calibrated and

uncalibrated AI feedback on measures of conceptual understanding, retention and transfer in algebra, including in Nigerian schools.

## References

- Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101-128.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-390). ACM. <https://doi.org/10.1145/985692.985741>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). Worth.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Information Age.
- Kapur, M. (2008). Productive failure. *Cognition*

and Instruction, 26(3), 379-424. <https://doi.org/10.1080/07370000802212669> Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008-1022. <https://doi.org/10.1111/cogs.12107> Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, 51(2), 289-299. <https://doi.org/10.1080/00461520.2016.1155457> Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press. Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1) Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284. <https://doi.org/10.1037/0033-2909.119.2.254>

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239-264. <https://doi.org/10.1007/s10648-007-9049-0> Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127-144. <https://doi.org/10.3102/0162373713507480> Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93(2), 346-362. <https://doi.org/10.1037/0022-0663.93.2.346> Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795> Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20-26. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4) VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221. <https://doi.org/10.1080/00461520.2011.611369> Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18(4), 375-400. <https://doi.org/10.1007/s10857-014-9286-3> Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087> Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>