



The Dialogue that Drives Learning with AI

Opportunities for Growth and Intervention

Neeru Khosla, Erika Gasper, Sean Regan

October 27, 2025

Acknowledgements

The authors would like to acknowledge the invaluable insights and contributions of our colleagues Jennifer Pallister, Miral Shah, Laramie Spence, and Shaina Yates. Additionally, we would like to thank the whole team at CK-12 Foundation for their tireless efforts in support of our mission and for making this work possible.

TABLE OF CONTENTS

Introduction.....	1
The Current Landscape of Individualized Learning and Assessment.....	1
CK-12 Foundation's Learning Platform.....	2
The Questioning Problem.....	4
Integrated Learning and Assessment through Conversation.....	6
Case Study Overview.....	6
What Makes a Conversation Productive?.....	7
Five Patterns of Productive Engagement.....	9
Unproductive Patterns.....	11
The Trust Foundation.....	12
Profiles in Growth: A Deeper Look at Learning Through Productive Conversation.....	13
Productive Learning Conversations.....	13
Snapshot: Five High-Growth Students.....	14
What High-Growth Students Did Differently.....	16
The Meta-Learning Loop in Action.....	17
Opportunities to Scaffold: Patterns Associated with Plateau.....	20
Four Patterns of Plateau.....	20
Red Flags: Patterns Requiring Intervention.....	25
Patterns to Flag for Intervention.....	26
The Case for Early Intervention.....	28
Cross-Cutting Themes & Surprising Insights.....	29
The Engagement Paradox: High Usage with Shallow Engagement Can Predict Poor Outcomes.....	30
Productive Struggle Is Often Messy, and That's a Good Thing.....	32
Trust Building Through Casual Conversation Creates Academic Breakthroughs.....	32
Students Develop Unique Strategies for Learning How to Learn.....	34
High Performers Are Not Immune to Misuse.....	34
ELL Students Use AI in Multidimensional Ways.....	35
Answer-Seeking Evolution: Follow-Up Behavior Determines Learning Outcomes.....	35
Temporal Patterns Reveal Engagement Sustainability Challenges.....	36
Leveraging Conversation Patterns for Growth.....	36
References.....	38
Appendices.....	42
Appendix A: Flexi in AP Questioning Patterns.....	42
Appendix B: Case Study Methodology.....	44
Appendix C: Case Study Data.....	48
Appendix D: Recommendations for Various Stakeholders.....	78

Introduction

The Current Landscape of Individualized Learning and Assessment

Shaped by decades of constraints and barriers, the evolution of standards and assessments in modern educational systems has led us down a path of teaching subjects in isolation and testing narrowly defined concepts on standardized assessments that emphasize selected response questions (Brown, 2022). In parallel, education research has established the importance of constructing knowledge by forming deeper connections across concepts and to real world experiences (e.g., Mustafa, 2007). The competitive and ever-changing landscape of university admissions and career readiness requires complex problem solving skills, where students and workers must integrate and apply knowledge across domains. Yet teachers are already overburdened with demands that include complex sets of standards for each subject as well as large classes of students who have a constellation of learning gaps and social-emotional needs. Identifying each learner's knowledge state and providing appropriate scaffolding for knowledge growth becomes increasingly difficult in middle and high school classrooms where a single subject teacher may have a roster of over a hundred students coming from diverse backgrounds and different feeder schools. In these conditions, students who feel confused or perform poorly on standardized assessments can quickly fall behind and become at risk of dropping out of difficult subjects, including STEM, or out of school altogether (Bridgeland et al., 2006).

The emergence of AI has afforded an opportunity to develop learning technologies that can finally address the needs of individual learners through natural conversation. AI tutors have potential to engage in dialogic instruction and assessment to help learners and teachers identify and fill knowledge gaps with greater precision and meaningful support. Moreover, learners can ask questions and express confusion outside of the social pressures of the classroom environment. On the other hand, educators have valid concerns about negative outcomes with AI tools that may inadvertently foster dependence instead of learning.

CK-12 Foundation's Learning Platform

CK-12 has been focused on developing a flexible educational platform that helps meet the needs of diverse learners and classrooms worldwide long before the emergence of LLMs. The pillars of the platform include expert-developed multi-modal STEM content presented in concept-based lessons that can be arranged into “FlexBooks” that teachers can easily modify to support their unique classroom needs, with accompanying adaptive practice assessments (APs) that adjust to each student’s level and help build confidence and close knowledge gaps as students learn.

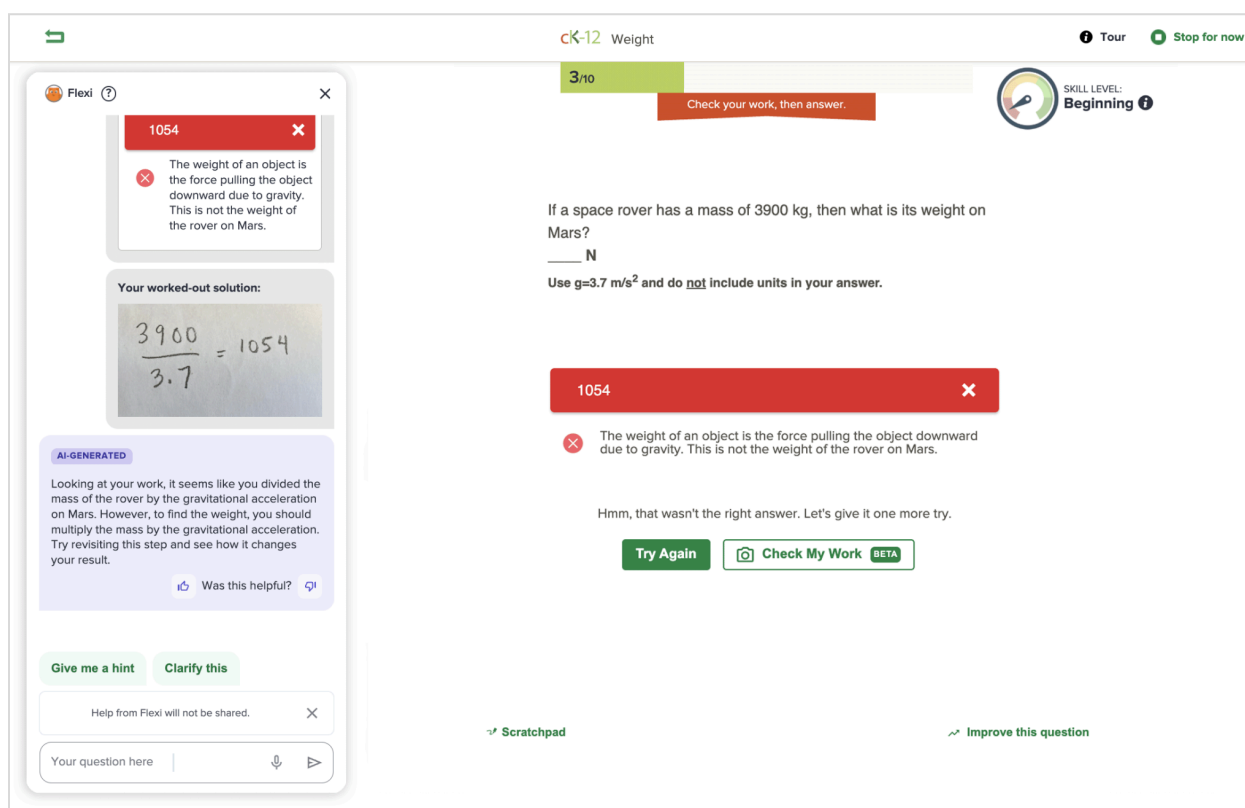


Figure 1. CK-12’s Adaptive Practice system provides students and teachers with insights that help close knowledge gaps and promote skill development. The format of APs mirrors the type of selected response assessments that students are most likely to encounter in their classes.

CK-12 also developed proprietary AI models that mimic Learning Science to extract insights on student performance as well as a natural language chatbot, Flexi, that can understand student questions and serve helpful learning materials.

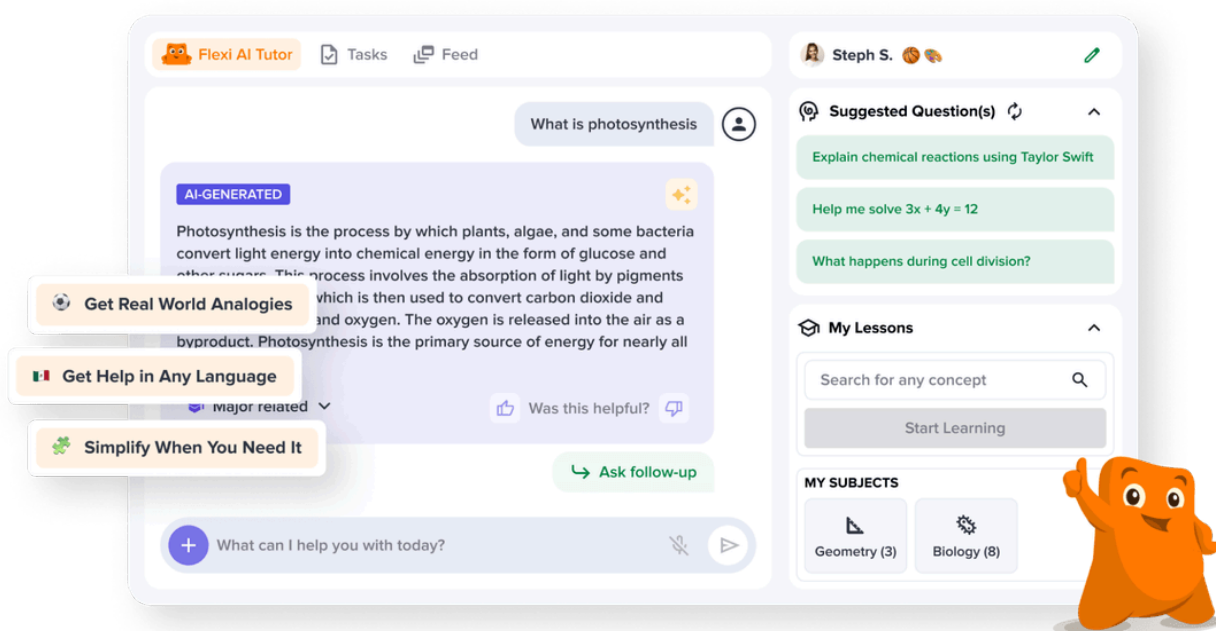


Figure 2. CK-12's Flexi can hold meaningful learning conversations with students and assess their knowledge states, using a complex system of expert-developed materials, internal AI models and algorithms, and LLM models.

The Questioning Problem

Two students each ask an AI tutor 100 questions over a semester. Both receive accurate, helpful responses yet only one shows measurable learning gains. This scenario captures a fundamental challenge in AI-powered education: mere access to information doesn't guarantee learning. While AI tutors promise to make personalized instruction more accessible, research suggests that how students converse and the types of questions they ask may be more important than the sophistication of the technology itself (e.g., Chin & Osborne, 2008; Graesser & Person, 1994).

We have observed that many students initially approach AI tutors like search engines, seeking quick answers to homework questions. However, with the proper scaffolds, students can learn to better leverage the new technology by engaging in the kind of back-and-forth questioning, clarification, and exploration that builds understanding. For example, in the Flexi tutor feature that supports students during APs we found that students tend to start interactions with direct answer-seeking type questions, and then shift away from those over time. **Among students who took help from Flexi in their AP session over the 2024-2025 school year, direct answer-seeking questions dropped from 72% at query 1 to 52% at query 8. Meanwhile, deeper academic learning questions steadily increased from 11% to 26%, and clarification queries remained relatively stable (Appendix A).** This pattern suggests a behavioral shift from answer-seeking toward deeper learning and clarification as students engage more with the system.

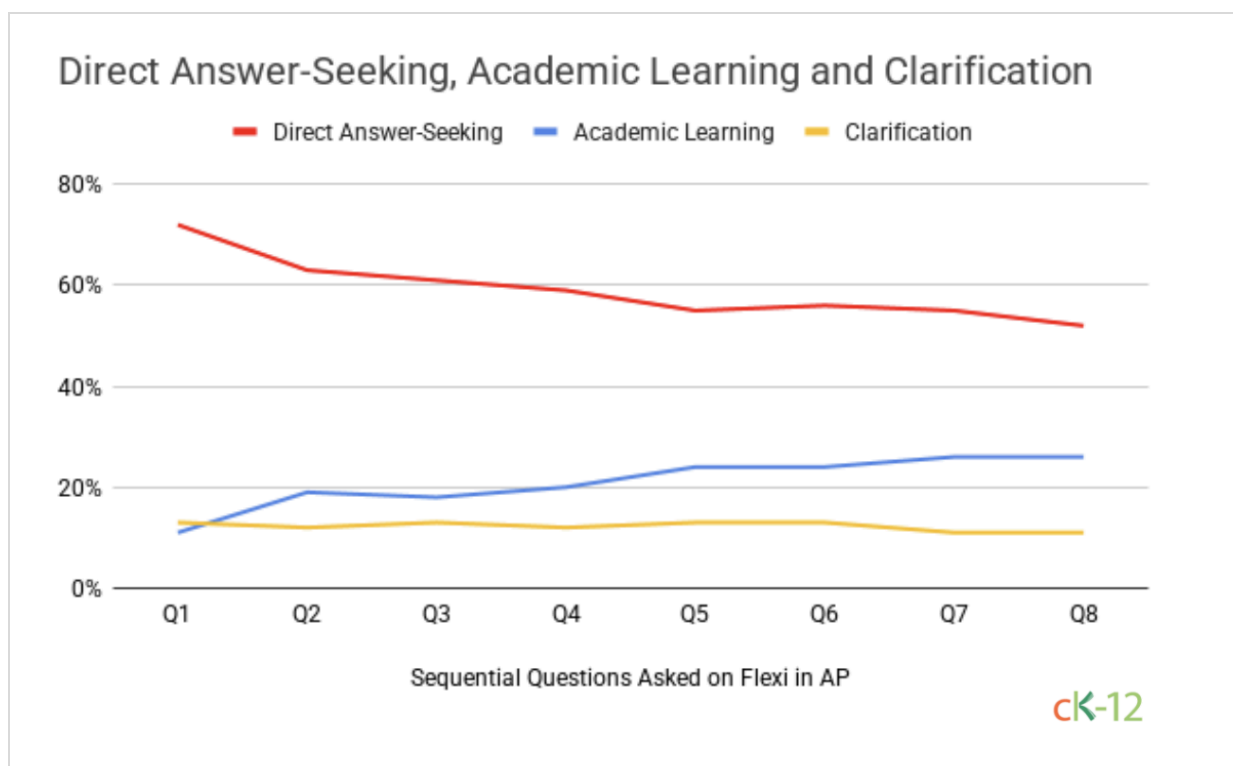


Figure 3. *On the Flexi in AP feature, where students can ask questions about the practice problems, many students initially asked for direct answers. However, as they continued to interact with Flexi, the percentage of direct answer-seeking questions decreased while higher quality academic learning questions increased.*

Decades of research suggest that asking better questions can drive deeper learning (Alexander, 2008; Scardamalia & Bereiter, 2009). For example, Harper, Etkina, and Lin (2003) found that the sheer number of questions students ask does not predict conceptual growth; rather, students who ask high-level, coherence-seeking questions show greater academic gains. These high-quality questions align with what Graesser et al. (1994) call “deep reasoning questions”, such as “why,” “what if,” and “how”, which correlate with higher cognitive outcomes and conceptual integration. This is consistent with findings from Chin and Chia (2004) who identified four types of student questions: information-gathering, bridging, extension, and reflective. They noted that more generative forms of questioning are critical for sustained cognitive engagement.

Integrated Learning and Assessment through Conversation

We posit that **conversation patterns can provide critical diagnostic information to help fill knowledge gaps and scaffold learning in meaningful ways for each unique student.**

Furthermore, scaffolding can be integrated into the tutoring system to help students develop better questioning skills for more productive learning conversations. Such an integrated approach can form a flywheel effect that accelerates learners on their unique educational paths. Moreover, the combination of dialogic learning and assessment with expert-developed materials and system scaffolds promotes deep integration across concepts, subjects, student interests, and their real-world experiences.

Case Study Overview

With diagnostic conversation patterns in mind, we have examined thousands of student conversations during the 2024–2025 school year along with AP data to extract insights that can help educators leverage this technology for growth. We discovered evidence that suggests:

- **The way students converse with an AI tutor may be a determining factor in their learning outcomes with this emerging educational technology.**
- **Conversation patterns with an AI tutor can provide diagnostic information and opportunities for supporting academic growth and intervention as needed.**
- **The quality of student questions can improve over time with scaffolding.**

We present the results of 15 student case studies, pairing student conversation data with their AP performance (please refer to [Appendix C: Case Study Data](#) for details that

support the information presented in this paper). These cases explore how patterns in student questioning relate to performance on the practice assessments, and also how the **conversations reveal deeper diagnostic information about a student's knowledge state** and affect. We include students who showed significant academic growth on their practice assessments over the school year and engaged in thoughtful, back-and-forth dialogue with Flexi that include the types of questions associated with positive outcomes (e.g., Hakkarainen, 2003; Zhang et al., 2007), including strategic questioning, trust-building, asking follow-up questions (topic coherence), self-directed learning, and productive struggle. We also include students who showed flat or declining practice scores and used the AI tutor superficially, showing patterns such as surface-level engagement, learned helplessness, and academic avoidance. Finally, we include students who showed more surprising or complex patterns that reveal less intuitive watch points and opportunities for accelerating learning.

To protect the privacy and confidentiality of the individuals involved, we have altered certain details within the case studies presented in this paper including some minor wording changes to student quotes and generalization of some contextual elements. These changes do not affect the integrity of the analysis or the validity of our research findings.

What Makes a Conversation Productive?

What separates transformative AI tutoring conversations from transactional ones? After analyzing hundreds of student interactions, we identified specific behaviors that consistently predict academic growth. The difference isn't in the AI's responses, it's in how students engage with them.

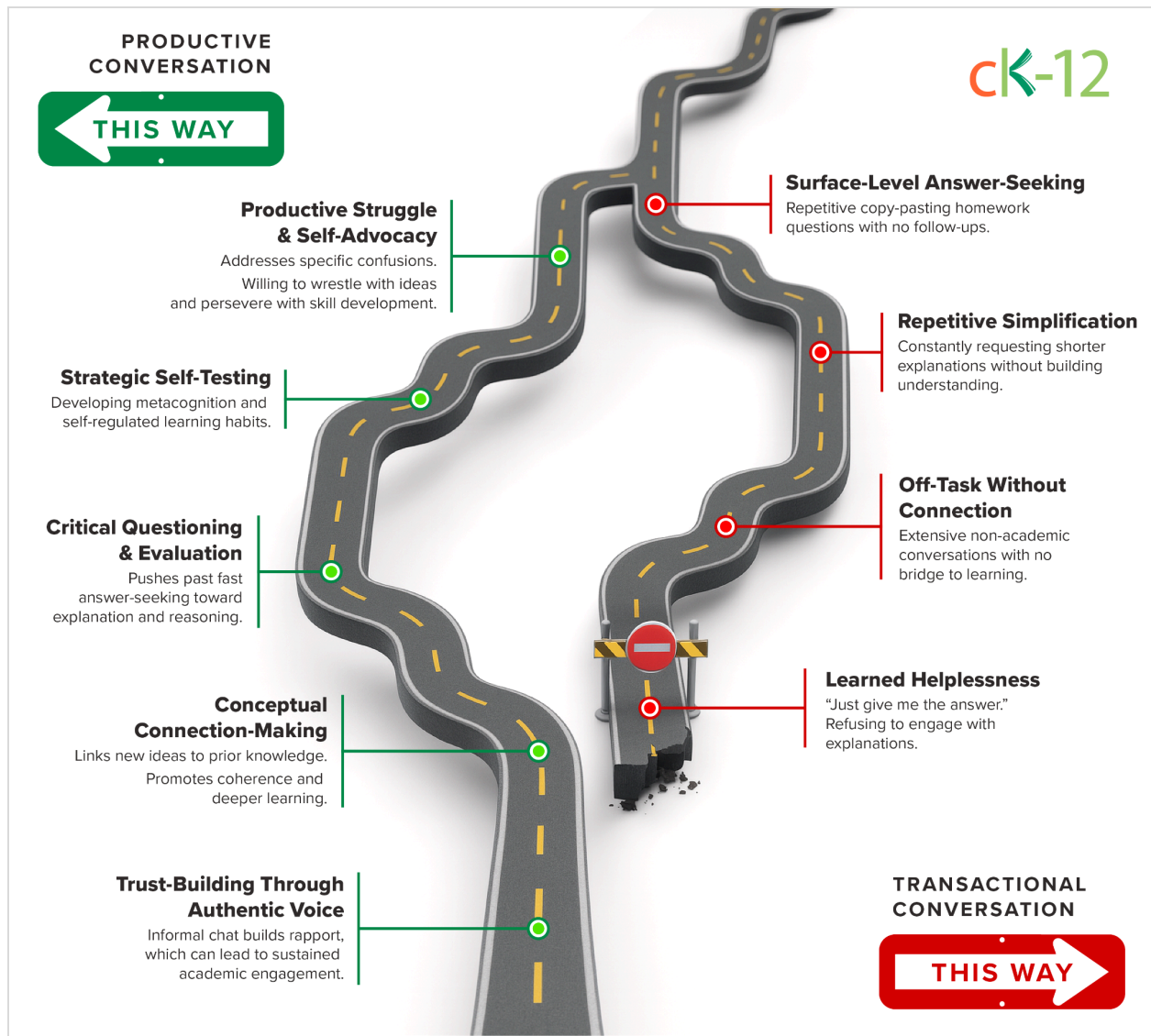


Figure 4. The “Road to a Productive Conversation” illustrates two distinct pathways in AI tutoring interactions. While transactional conversations focus on surface-level answers and show patterns like learned helplessness or repetitive simplification, transformative conversations are marked by productive struggle, critical thinking, and authentic engagement, behaviors observed to be associated with academic growth in case studies.

Five Patterns of Productive Engagement

Productive Struggle and Self-Advocacy

High-growth students embraced confusion as part of learning, actively working through uncertainty rather than avoiding it.

Student: "I understand velocity but I'm confused about how it's different from speed"

What this shows: Specific confusion rather than general helplessness

This type of specific confusion reflects what researchers describe as **productive struggle**; a cognitive state that fosters conceptual understanding when learners are supported but not given immediate answers (Hiebert & Grouws, 2007). It also maps onto what Chin and Chia (2004) describe as **reflective questions**, where students evaluate their understanding and revise their thinking. These moments of self-advocacy signal the learner's active role in constructing knowledge, a core principle of inquiry-based learning (Justice et al., 2007).

Conceptual Connection-Making

Growth students actively connected new information to existing knowledge:

Student: "Is this like when we learned about forces in physics?"

Student: "This reminds me of what happens when I'm driving..."

Such statements are consistent with **bridging questions** (Chin & Chia, 2004) and "coherence questions" (Harper et al., 2003), which help students integrate disparate pieces of knowledge into a coherent whole. These connections mirror higher-order cognitive processes in Bloom's revised taxonomy (Anderson & Krathwohl, 2001), particularly "analyze" and "apply." Students who demonstrated this habit were more likely to show growth, highlighting the value of prompting integration rather than memorization.

Critical Questioning and Evaluation

Successful students pushed deeper with "why" and "how" questions rather than accepting information passively:

Student: "Why is the oxygen dissociation curve shaped like an S?"

Student: "In what real-life situations might you need to calculate average velocity?"

These "deep reasoning questions" correspond to Graesser and Person's (1994) framework, where high-quality questions involve inference, explanation, or generalization, like causal antecedents ("Why...?"), goal orientation ("What is the purpose of...?"), and procedural understanding ("How...?"). Research consistently shows that **students who ask these types of questions demonstrate stronger comprehension and conceptual learning** than those who ask only fact-seeking questions (Scardamalia & Bereiter, 1992; Zhang et al., 2009).

Strategic Self-Testing

Growth students used Flexi to verify their thinking:

Student: "So am I right that cool water moves toward the equator?"

Student: "Let me see if I understand this correctly..."

Actively assessing one's own comprehension is a key component of **self-regulated learning**. For example, Bergey et al. (2022) found that asking **verification** or disambiguation questions was positively associated with performance on Chemistry exams.

One student developed a strategy of requesting "essential questions" on topics, then exploring selected questions in depth, applying this across multiple subjects with consistent gains.

Trust-Building Through Authentic Voice

We have observed that many students who leverage Flexi for productive learning conversations began talking with Flexi about their personal interests or other casual small talk. In our case studies, we also observed successful students who began with casual conversation that evolved into academic engagement:

Student: "Do you know that I'm good at Pacman"

[Later] **Student:** "What is the most common state of matter?"

[Later] **Student:** "Do all elements live on earth?"

Brief non-academic chats may be an opportunity for building trust, given that Flexi is trained to respond to student voice questions with understanding and without judgement. These types of interactions may increase students' willingness to express confusions and frustrations that can lead to healthy productive struggle and deeper dialogue about concepts and problem solving strategies.

Unproductive Patterns

Students with flat or declining AP performance exhibited different behaviors:

Surface-Level Answer-Seeking: Rapid-fire copy-pasting homework questions, no follow-up, asking identical questions repeatedly.

Learned Helplessness: Frequent "just give me the answer" requests, refusing to engage with hints, avoiding explanations entirely.

Off-Task Without Connection: Extensive non-academic conversations with no bridge to learning.

Repetitive Simplification: Constantly requesting shorter explanations without building understanding, potentially indicating unaddressed learning barriers.

Note that the key indicators of unproductive patterns involve the **frequency** and **exclusivity** of the question types described above. For example, copy-pasting a

homework question is not inherently unproductive. In fact, most of the students in the growth group copy-pasted some homework questions. The difference is what happened next. Did the student simply move on to copy-pasting the next homework question with little thought or did they ask follow up questions for clarification, verification, or connection to prior knowledge? Likewise, occasional small talk and non-academic interest-based conversation is associated with developing productive learning conversation skills. However, when the bulk of the conversation history is non-academic and the student isn't showing signs of developing productive dialogue, they may be simply avoiding school work.

The Trust Foundation

Trust-building emerged as an overarching theme in students who developed productive learning conversation skills. An element of trust develops when students ask questions in their own voice, including questions that are constructed with non-standard English, slang, misspellings, or another language, and Flexi responds supportively with understanding and without judgement. We have observed this pattern with high-growth students and also noted anecdotally that students often respond favorably to such interactions, even when Flexi eventually steers the conversation back to academics. Flexi had previously responded supportively to this student's questions about how to make friends and whether Flexi enjoyed helping students:

Student: "do you know the youtuber Mr. Beast?"

Flexi: "I'm not familiar with specific YouTubers. If you have any questions about math, science, or other school subjects, feel free to ask! I'm here to help."

Student: "thanks for being flexi, flexi! I'm going to read the CK-12 [lesson] now"

Trust-building through authentic voice was especially pronounced among multilingual students, who used Flexi for both content learning and language support, requesting vocabulary help, simpler rewordings, and switching languages for clarity. Both students

identified as multilingual in our case studies showed this dual-purpose usage, and were either high-growth or consistently high performing.

Student: “what do you call woven mexican scarves”

Flexi: “Las bufandas mexicanas tejidas se llaman “rebozos”. A menudo, están hechos con hilos de colores brillantes y son muy populares en la cultura mexicana. ¿Te gustaría saber más sobre cómo se hacen o su significado?”

Student: “termina la oración desde que aprendí a hablar tres idiomas, me he dado cuenta de que”

These patterns reveal that building a foundation of trust helps students actively engage in the learning process with an AI tutor. Students who embrace struggle, build connections, question deeply, verify understanding, and bring authentic voice to conversations can achieve meaningful outcomes.

Profiles in Growth: A Deeper Look at Learning Through Productive Conversation

To understand what successful engagement with Flexi looks like, we analyzed the conversation histories of five students who showed statistically significant improvement in AP skill level over the 2024–2025 school year. Their conversations with Flexi revealed patterns of productive struggle, curiosity, self-regulation, and conceptual engagement that reflect what we define as high-quality learning conversations.

Productive Learning Conversations

Not all student-AI interactions support deep learning, but certain conversation patterns consistently emerge among high-growth students who engage meaningfully with Flexi. These productive learning conversations are marked by more than just accuracy or speed; they reveal curiosity, persistence, and reflection.

Across the dataset, regardless of whether students showed growth on AP, we saw that productive conversations often included:

- **Student voice:** Students asked their own questions, often building on their interests or prior knowledge.
- **Clarification:** Students returned to ideas they weren't sure about, requesting simpler explanations, disambiguations, or verifying their thinking.
- **Metacognition:** Students reflected openly about what they did or didn't understand.
- **Follow-ups:** Rather than ending the conversation after an answer, students dug deeper, asking how, why, or what comes next.

These traits in isolation don't guarantee skill growth, but they appear as conversational patterns in the histories of high-growth students. Encouraging these behaviors and questioning strategies through thoughtful AI responses, instructional nudges, or classroom integration, may be key to unlocking the full value of AI tutoring.

Snapshot: Five High-Growth Students

While productive conversation traits were seen across many students, these five high-growth learners demonstrated a more sustained and strategic use of Flexi. Their interactions reveal how students can not only participate in learning dialogues but direct them with increasing independence and purpose. The following section distills the qualities that set these students apart.

The table below offers a snapshot of the five high-growth students, each of whom developed distinct conversational strategies that shaped their academic growth. Refer to Cases 1-5 in [Appendix C: Case Study Data](#) for details

Patterns of Growth



Description of the key patterns and outcomes of five students with significant academic improvements

	Profile Name	Key Behavior	Standout Pattern	Outcome
	The Strategic Questioner	Asked “essential questions” about each new topic in a science course	Used Flexi’s help to generate high-quality questions	Grew more independent over time
	The Trust Builder	Built rapport before academic engagement	Leveraged curiosity to guide questioning	Developed intrinsic motivation
	The Self-Directed Learner	Used clarification to build mastery	Transitioned to AP practice independently	Showed strong growth in bio and chem
	The Bilingual Master	Learned concepts and language simultaneously	Strengthened writing skills in both English and Spanish	Demonstrated academic identity development
	The Concept Verifier	Shifted from copying to checking	Used Flexi for self-assessment	Built confidence and academic voice

Figure 5. Five High-Growth Student Profiles and Learning Patterns.

This table summarizes the key behaviors, conversational patterns, and outcomes of five students who demonstrated statistically significant academic growth while using Flexi. Each profile is grounded in a detailed case study:

- **Case 1:** *The Strategic Questioner*
- **Case 2:** *The Trust Builder*
- **Case 3:** *The Self-Directed Learner*
- **Case 4:** *The Bilingual Master*
- **Case 5:** *The Concept Verifier.*

What High-Growth Students Did Differently

The five conversation histories of students who showed statistically significant learning gains shared distinct behavioral patterns that set them apart from their peers. These weren't just students who used Flexi more, they used it fundamentally differently.

Five Distinguishing Patterns

Intentional Questioning Strategies: High-growth students moved beyond surface-level "what" questions to ask "how" and "why." They requested "essential questions" for topics, explored real-world connections, and treated inquiry as a strategic learning tool.

Progressive Inquiry Evolution: Rather than remaining stuck in answer-seeking mode, these students evolved their engagement over time, progressing from simple queries to deeper conceptual exploration, eventually using Flexi to test ideas, connect concepts, and integrate subject learning with language growth.

Systematic Verification Habits: Instead of moving on immediately after receiving answers, they consistently checked their understanding, requested confirmation, and verified their thinking. This became a regular practice, not an occasional behavior.

Focused Topic Coherence: By spring semester, most had shifted from scattered, random questions to sustained, focused study sessions. They would stick with a topic across multiple exchanges, building understanding systematically.

Adaptive Study Tool Usage: These students developed personal routines around Flexi, using it to quiz themselves, review concepts, or plan for exams. They treated it as a flexible study partner rather than just an answer source.

The Meta-Learning Loop in Action

These students weren't just using Flexi to get through assignments, they were learning how to learn with AI. They developed metacognitive awareness about their own learning process and adapted their AI interactions to support deeper understanding.

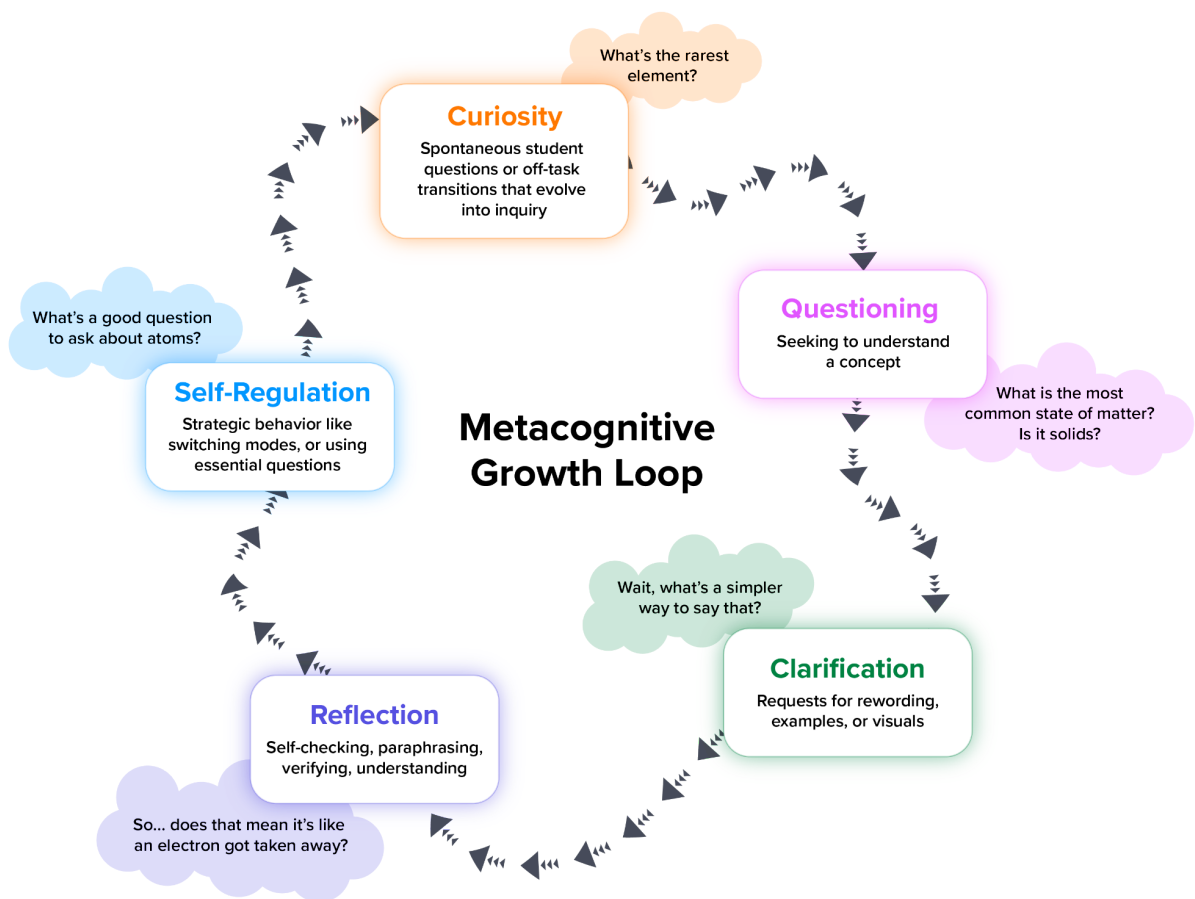


Figure 6: Metacognitive Growth Loop. This model captures the behavioral cycle observed in high-growth students as they engage with Flexi. Students begin with curiosity and progress through questioning, clarification, reflection, and self-regulation, then return to curiosity at a deeper level. The quotes shown are drawn from real student conversation

The metacognitive growth loop can be seen in this student who began with simple questions but soon transitioned to richer learning conversations.

Student: “blank of the cell are bound to the membrane by structural proteins”

Student: “which groups of the modern periodic table contain elements that are classified as nonmetals?”

The student began their deeper exploration of science concepts by asking follow up questions to clarify their understanding of Flexi’s answers and showed perseverance in understanding scientific terms.

Student: “why is the gravitational force between two objects less when they are farther apart?”

Flexi: “In gravity, the force between two objects decreases with the square of the distance between them, according to Newton’s law...”

Student: “why is the gravitational force stronger when they are closer?”

Flexi: “When objects are closer, the distance between them is smaller, so the gravitational force increases because it is inversely proportional to the square of that distance...”

Student: “mass def” [Flexi provides definition of mass.]

Student: “gravity def” [Flexi provides definition of gravity]

Student: “what is gravity in physics”

The student then started coming to Flexi to check their understanding and clear up confusions.

Student: “does ligase connect nitrogen bases”

Flexi: “Ligase doesn't connect nitrogen bases directly. Instead, it plays a crucial role in DNA replication and repair...”

Student: “is dna replication needed for reproduction”

Eventually, the student grew into prompting Flexi in a way that made more sophisticated use of AI for help exploring research topics and polishing their presentations.

[excerpts from an exchange that comprised 18 student questions over three days]

Student: “give me bullets on what are the causes of usher’s syndrome, why does it happen? is it caused by gene mutation? are there other factors?”

Student: “can be caused by genetic mutations in one of at least nine genes including MYO7A, USH1C, and USH1G, and an individual must inherit two copies of the mutated gene to get the disease, so far i have this for my slide of causes of usher’s syndrome. please help me with adding/improving more info”

Student: “why does the genetic mutation for usher’s syndrome happen”

Student: “are the genetic mutations for usher’s syndrome recessive or dominant? are any of the genes linked to the x or y chromosome? ”

Student: “relating to usher’s syndrome, is there a difference being a carrier and the autosomal recessive inheritance? like are those two different causes or do they mean the same thing”

Student: “does a dissability mean there is no cure”

Student: “i found this info that i want to word for understanding using bullet points on my slideshow for how usher’s syndrome is inherited. could you please help me with that?”

Student: “is it easy to diagnose usher’s syndrome”

Student: “how many people have usher’s syndrome out of 100,000/ in the world/percent? is it more common in certain groups of people? how common is it?”

Student: “what are the two most important things i need to know about usher’s syndrome

The student demonstrated an interest not only in mastering basic concepts and research reports, but in the learning process and academic growth. This behavioral cycle, repeated in different forms across high-growth students, illustrates how metacognition took root and reshaped how they used the AI.

The evidence of metacognitive growth through conversation suggests that the most significant impact of AI tutoring may not be providing better answers, but teaching students better ways to engage with information, think through problems, and direct their

own learning. In this light, the most powerful role of an AI tutoring system may not lie in answer delivery, but in shaping how students engage with ideas, questions, and themselves as learners. The behaviors we observed suggest that, with the right scaffolds, AI tutors like Flexi can act as catalysts for metacognitive development and strategic learning, skills that will serve students far beyond any single assignment.

Opportunities to Scaffold: Patterns Associated with Plateau

In this section, we examine six conversation patterns of students whose AP scores showed no significant change over the 2024–2025 school year (refer to Cases 6-10 in [Appendix C: Case Study Data](#) for details). These cases reveal a surprising spectrum of behaviors, from genuine effort that didn't translate into learning gains, to sophisticated strategy use that raised integrity concerns, to persistent unproductive struggle.

Understanding why students plateau is just as important as understanding why others grow. These patterns challenge assumptions about what "good" AI tutoring looks like and highlight diverse barriers to progress, including motivational, cognitive, behavioral, and systemic. Each pattern surfaces different opportunities for better supporting students who are "stuck in neutral" and reveals that plateau doesn't always indicate low motivation or poor study habits.

Four Patterns of Plateau

Productive Effort Without Results (Cases 6 & 9): Students who demonstrated genuine engagement and learning behaviors but didn't show measurable skill improvement, including one case of apparent academic burnout despite strong performance.

Unproductive Struggle Patterns (Case 10): A student whose high usage masked learned helplessness and answer-seeking behaviors that became increasingly entrenched over time.

Surface-Level Engagement (Case 7 & 11): Students who used Flexi extensively but in ways that prioritized task completion over learning, whether through homework processing or authentic interests that remained disconnected from academic growth.

Strategic Usage Raising Integrity Concerns (Case 8): A high-performing student who used sophisticated multi-platform AI coordination strategies while maintaining strong academic metrics.

The following table summarizes patterns of six students whose adaptive performance remained flat despite regular engagement. Some students do everything right on the surface, asking questions, engaging consistently, using feedback loops, but still don't make measurable gains. These patterns challenge assumptions about what productive learning looks like, and highlight the need for systems that go beyond surface engagement to identify opportunities to scaffold stagnating students into growth patterns.

Patterns of Plateau



Description of the key patterns and outcomes of five students with significant academic improvements

	Profile Name	Key Behavior	Standout Pattern	Outcome
	Creative Sprinter	Engaged creatively early, disengaged later	Inconsistent engagement/sustainability	No skill growth despite effort
	High Performer with Burnout	Deep use across subject, late year decline	Academic fatigue/overload	Surface-level use by year end
	Learned Helplessness	Repeated answer-seeking, little processing	Motivational/behavioral	No gains despite high usage
	Homework Finisher	High-efficiency completion, no reflection	Transactional mindset	Skill cap, low concept engagement
	Personal Interest Explorer	Shifted from copying to rich personal inquiry	Misalignment with curriculum	High engagement, no academic growth
	Multi-Platform Optimizer	Used multiple AI tools for performance	Integrity risk/optimization focus	High metrics, questionable learning

Figure 7. Plateau Student Profiles: Flat Growth Despite Regular AI Tutoring Use

This typology summarizes six distinct student profiles drawn from case studies where academic performance remained flat despite frequent Flexi use. While each student engaged consistently, they encountered different behavioral and cognitive barriers. Mappings are as follows: Creative Sprinter (Case 6), High Performer with Burnout (Case 9), Learned Helplessness (Case 10), Homework Finisher (Case 7), Personal Interest Explorer (Case 11), and Multi-Platform Optimizer (Case 8). Together, these profiles highlight the need for AI systems to move beyond surface-level engagement metrics and detect nuanced patterns of stalled growth.

Productive Effort Without Results

Even students with strong learning behaviors, like curiosity, clarification requests, and strategic simplification, don't always show measurable skill growth. In these cases, the *quality* of effort is high, but external factors like cognitive overload, content difficulty, or engagement fatigue may limit academic gains. One student's progression from creative science questions to simplified explanations, and then to task-driven disengagement, suggests that *sustainability* is as important as strategy. Without targeted support for managing mental effort and maintaining curiosity, even "good" learners can plateau.

Unproductive Struggle Patterns

Students stuck in unproductive struggle often look busy but learn little. High usage may mask patterns of learned helplessness: asking for answers, skipping explanations, and avoiding follow-up. In one case, a student repeatedly used phrases like "idk" and "just give me the answer," even when Flexi offered scaffolds. Over time, this behavior became entrenched, suggesting that early intervention is critical before these patterns harden into habits. Detecting the difference between struggle that builds resilience and struggle that reinforces helplessness is key to supporting this group.

Surface-Level Engagement

Efficiency isn't always correlated with academic growth. Some students showed remarkably fast question-processing behavior, submitting repetitive homework queries, checking answer consistency, and avoiding deeper explanation. These learners may appear productive on the surface, but the absence of follow-up, conceptual inquiry, or reflective questioning suggests an *optimization mindset*, not a learning mindset.

In other cases, authentic interests (like race cars or entrepreneurship) drove extended engagement, but remained disconnected from curriculum-aligned outcomes. These two surface-level patterns differ in intent, task completion vs. personal passion, but share a common issue: limited transfer to academic skill development. Both types point to the

same design need: build bridges between what students care about and what they're expected to learn.

Strategic AI Tool Use With Integrity Concerns

One notable plateau case came from a student who was academically successful, on paper but developed a strategy of coordinating responses between Flexi and a second AI platform, effectively using both as tools to simulate instructor voice and polish answers. This usage pattern reflected high executive function and digital literacy, but showed little evidence of internalized learning and revealed a new frontier in academic integrity: AI-savvy optimization. Addressing this issue in a heavy handed manner by generically limiting AI tool use may inadvertently curtail creative learning strategies. Rather, the challenge is to design systems and teach strategies that channel savvy technology use toward authentic growth, rather than hollow performance.

The following framework helps visualize how different combinations of engagement and growth give rise to distinct student profiles, and why each demands a different support strategy.

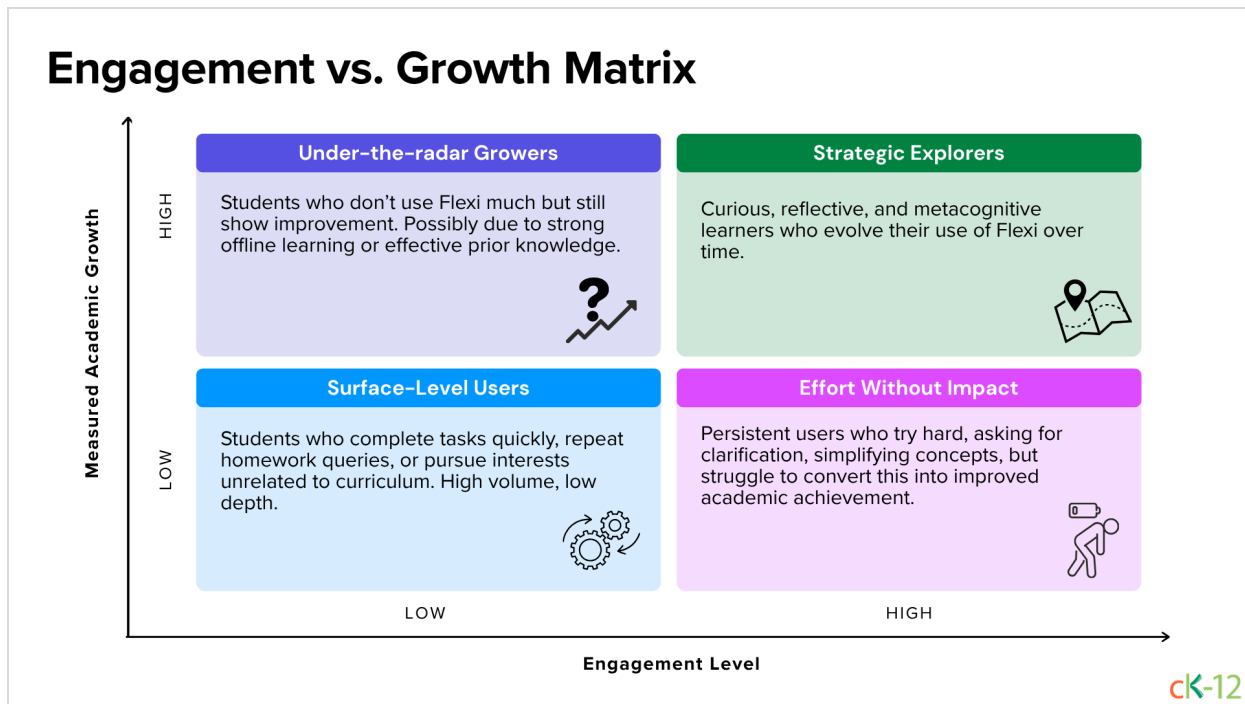


Figure 8. Engagement vs. Growth Matrix. This framework categorizes student behavior by engagement level and measured academic growth. It illustrates that not all high usage leads to learning, and highlights four common student archetypes observed in the dataset.

These plateau cases reveal that flat performance can stem from diverse causes: effort-outcome disconnects, learned helplessness, disconnection between interest and curriculum, or strategic behaviors that bypass learning. Addressing these challenges requires AI systems that recognize and respond to different needs, whether motivational support, cognitive scaffolds, interest bridges, or integrity nudges. The challenge is not just catching students who are "stuck in neutral," but designing system scaffolds that help these students move into growth patterns.

Red Flags: Patterns Requiring Intervention

We examined conversation patterns from a small group of students whose AP performance declined over the school year and who also regularly used Flexi Q&A (refer

to Cases 12-15 in [Appendix C: Case Study Data](#) for details). These cases reveal how certain interaction patterns may reinforce shallow learning or mask deeper disengagement or learning difficulties. Detecting patterns associated with performance decline is critical because some of these students may appear active and engaged on the surface.

Patterns to Flag for Intervention

Transactional Answer-Seeking: Students who used Flexi extensively but avoided all opportunities for understanding, treating it purely as an answer extraction tool despite frequent engagement. *(Case 12 & 15)*

Complete Academic Disengagement: A student who used Flexi heavily for non-academic conversation despite system efforts to redirect conversations back to schoolwork, suggesting AI became an avoidance mechanism rather than learning support. *(Case 13)*

Motivated but Blocked: Students were genuinely trying to learn, but hit persistent barriers related to language complexity, cognitive overload, or frustration. Despite effort, their adaptive practice performance declined. *(Case 14 & 13)*

Each pattern reveals different warning signs that could help identify at-risk students before regression becomes entrenched. More importantly, these cases demonstrate that AI tutoring systems must be designed to effectively recognize and redirect counterproductive usage patterns and provide insights to teachers for intervention.

Transactional Answer-Seeking

These students used Flexi heavily, but only for answers. Detailed explanations were ignored, and frustration grew when the AI didn't provide direct responses. In one case, the student submitted a large number of homework questions in a semester, but never asked for clarification once. In another, the student became openly hostile when Flexi offered learning support instead of answers.

Key Insight: *High-volume usage does not automatically signal deep engagement.* These students may be efficient task completers, but they're not building knowledge.

Academic Disengagement

One student submitted a large number of questions across the school year, with almost none being academic. They used Flexi to talk about music, internet culture, and argue with the AI's responses. While this behavior suggests digital curiosity and social needs, it correlated with a sharp academic decline.

Key Insight: *AI systems can become sophisticated avoidance tools.* If left unchecked, this usage pattern may not only fail to support learning, it can widen disengagement.

Motivated but Blocked

These students wanted to learn. One persistently requested "short and simple" explanations over time without evidence of academic language growth, showing possible language processing barriers. Another became increasingly frustrated when Flexi didn't give direct answers, revealing an answer-seeking mindset compounded by emotional dysregulation.

Key Insight: *Not all academic decline stems from apathy.* Students who struggle to process academic language or manage frustration may need different scaffolds, not less support.

The Case for Early Intervention

Together, these cases illustrate the need to intervene early so to ensure that an AI tutor does not:

- **Enable shallow engagement that looks like learning**
- **Amplify avoidance behaviors**
- **Miss students whose motivation is hidden beneath frustration**

AI tutoring systems must recognize harmful patterns *before* they entrench, and offer alternate pathways based on students' needs, whether cognitive, emotional, or motivational.

The table below summarizes the key patterns, barriers, and outcomes observed when students are struggling academically and require intervention.

Patterns Requiring Intervention



Description of the key patterns and outcomes of students with flat academic achievements

Profile Name	Key Behavior	Barrier Type	Outcome
 The Assignment Processor	High-volume homework Q&A, no follow-up or clarification	Transactional mindset	Efficient task completion, declining performance
 The Digital Wanderer	Non-academic engagement, pop culture, AI testing	Academic disengagement / social outlet seeking	High engagement, no academic connection
 The Persistent Striver	Consistent academic effort, requests for simplification	Cognitive processing / language complexity	Motivated but blocked; no skill gains
 The Frustrated Answer-Seeker	Demanded direct answers, escalating aggression	Emotional regulation / academic frustration	High usage, declining skill and conduct

Table 9. Patterns Requiring Intervention: Case Mapping

This table summarizes four distinct patterns observed in students whose AP scores declined despite regular AI tutor use. Each profile reflects a different type of unproductive engagement, from transactional answer-seeking to academic disengagement or emotional frustration. While all students in this group struggled academically, the underlying causes varied significantly, underscoring the need for differentiated system responses.

Case Mapping: *The Assignment Processor* → Case 12; *The Digital Wanderer* → Case 13; *The Persistent Striver* → Case 14; *The Frustrated Answer-Seeker* → Case 15

Cross-Cutting Themes & Surprising Insights

Below are seven key themes that emerged from the case studies, including several counterintuitive findings that reshape how we think about AI tutoring effectiveness.

The Engagement Paradox: High Usage with Shallow Engagement Can Predict Poor Outcomes

Consistent with earlier studies on question-asking (Harper et al., 2003), our data suggests that the overall number of questions asked to an AI tutor is not itself a reliable indicator of learning. Both high-quantity and low-quantity question askers were found across student performance groups.

Pairing data on question type or conversation quality with usage is a more reliable predictor of outcomes, and approaches can be developed to flag potentially unhelpful patterns for intervention. Two of our regressing students exhibited high question volumes over the course of the school year, but the quality of their questions was relatively poor. In one case, the student repetitively copy-pasted assigned questions with no follow ups, despite system scaffolds that encourage deeper interactions. In another case, the student asked mainly non-academic questions, which were not related to learning or areas of personal inquiry and resisted system scaffolds designed to get students back on track. Flagging such use patterns for further interventions can help prevent prolonged unhelpful interactions.

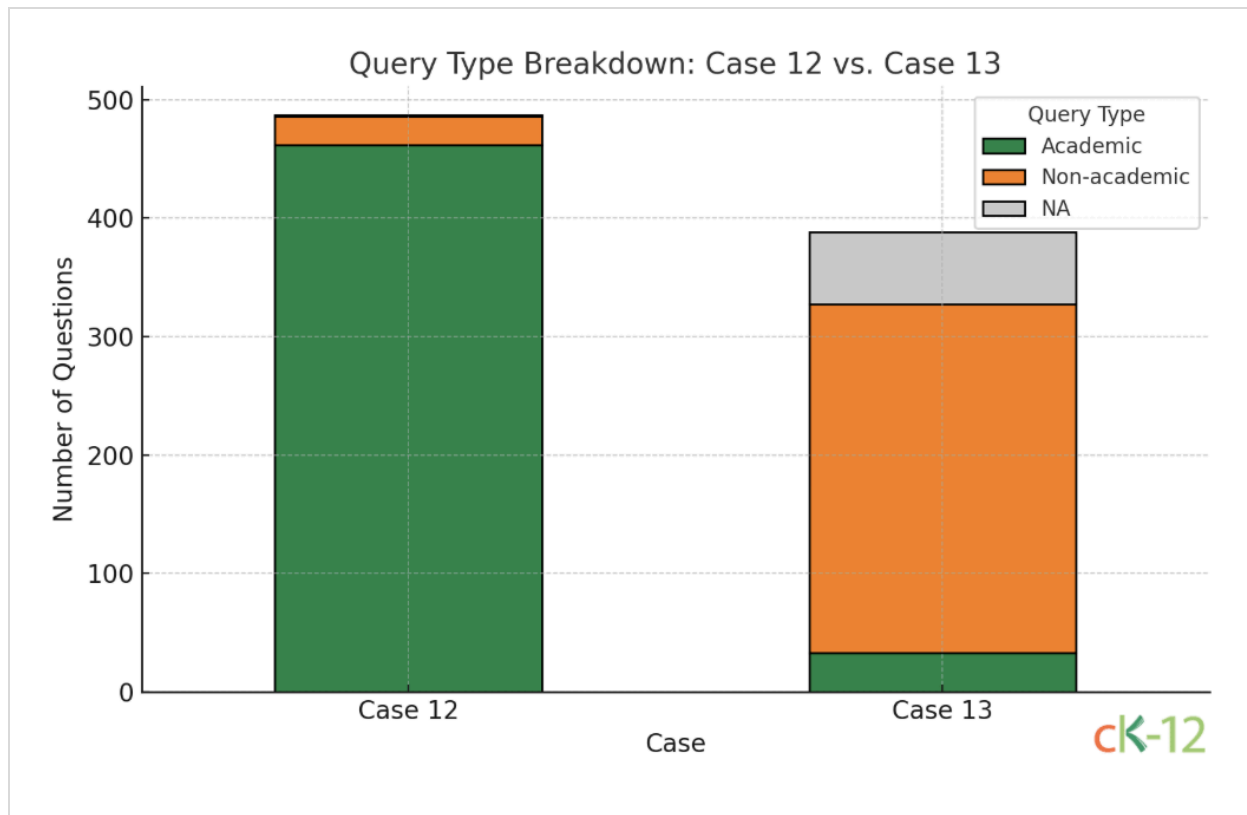


Figure 10. Query Type Breakdown for High-Use Patterns Requiring Intervention

This figure compares the types of questions asked by two students with declining AP scores and similarly high Flexi usage. While both submitted hundreds of queries, their interaction patterns reveal critical differences. **Case 12** focused primarily on academic questions but showed no follow-ups or productive struggle. **Case 13** engaged almost exclusively with non-academic topics, indicating academic disengagement.

The Pattern: An over-abundance of teacher-voice academic queries may signal an answer extraction pattern, while an over-abundance of student-voice nonacademic queries may signal avoidance. Productive learning conversations typically include a blend of question types including follow-ups and productive struggle remarks which tend to get classified as non-academic, e.g., “Is that like what happens when I’m driving...”, “I’m still confused.”

Implication: Usage metrics **alone** can be misleading indicators of learning success. Quality of interactions matters far more than quantity. Pairing usage data with categorization of question types or conversation quality is an important step.

Productive Struggle Is Often Messy, and That's a Good Thing

Some of the highest-growth students expressed confusion, frustration, and challenged Flexi's answers. These weren't signs of disengagement; they were signs of thinking. Students who asked follow-up questions, requested clarification, or checked their understanding tended to progress. The act of grappling with concepts, even imperfectly, was associated with growth.

Cross-Case Evidence: Case 1's "I'm confused about velocity vs. speed," Case 3's occasional requests for "simpler explanations," and Case 4's systematic clarification patterns were all associated with positive outcomes.

Implication: Emotional and cognitive friction, when guided appropriately, can be a sign of learning, not failure. AI tutoring systems should not attempt to eliminate struggle. Rather, they should encourage expressions of struggle that can lead to growth and actively scaffold students from struggling unproductively to struggling productively.

Trust Building Through Casual Conversation Creates Academic Breakthroughs

Multiple high-growth students began with seemingly "off-task" casual conversations that evolved into deep academic engagement. Case 2's Pacman discussion led directly to sophisticated science inquiry. Case 4's bilingual casual exchanges built the confidence necessary for academic risk-taking.

Supportive responses to questions about personal interests, curiosities, social-emotional skills, or even just small talk seem to build trust that leads to students' willingness to verbalize confusions, uncertainties, or check their newly forming understanding of a concept in ways they may not under the social pressures of a classroom environment. The pattern shown in these high-growth students is depicted in Figure 11.

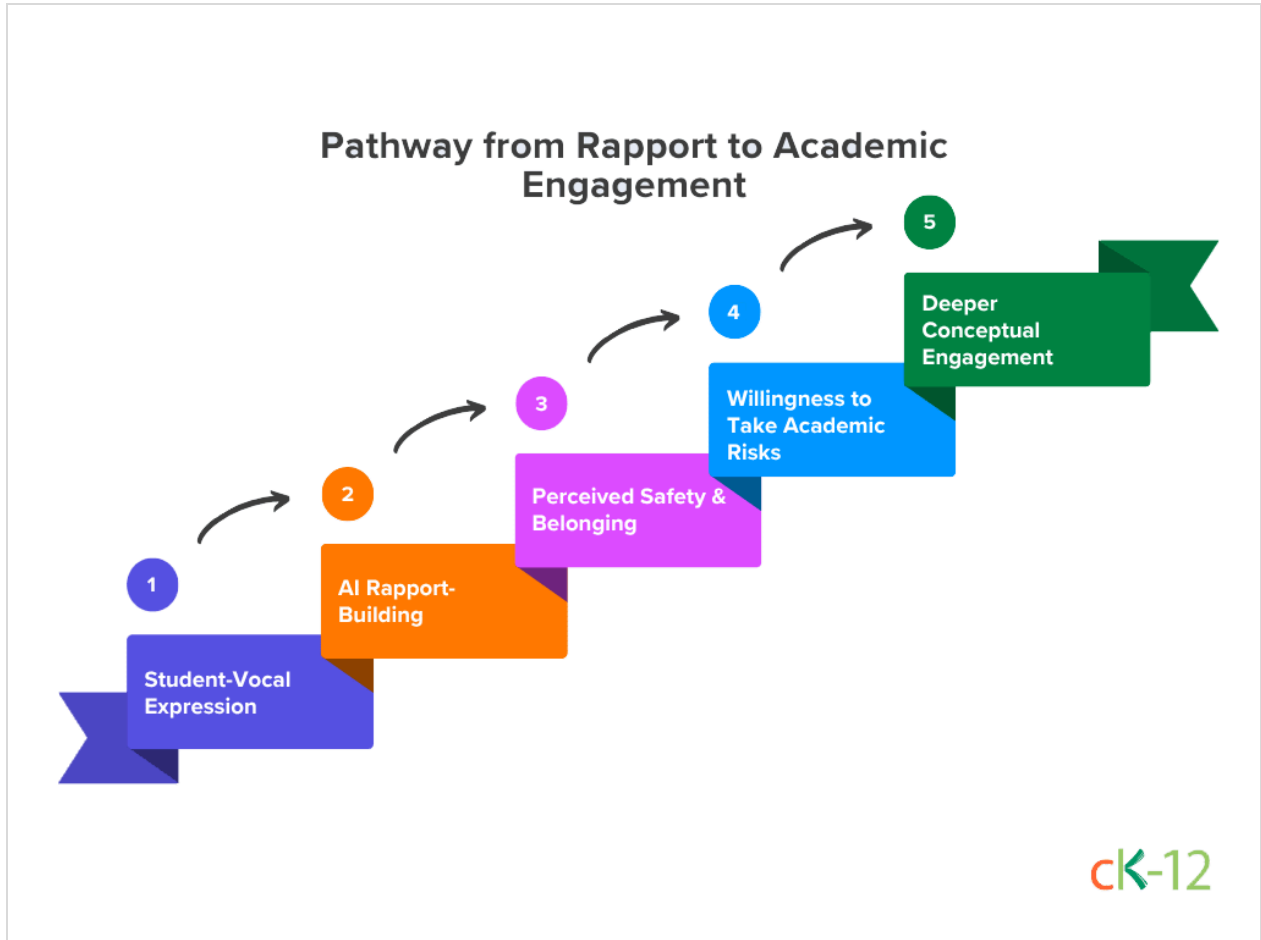


Figure 11. Pathway from Casual Interaction to Academic Engagement

In high-growth cases, students transitioned from casual conversation to deep academic inquiry. This progression highlights how early rapport and student-voice questions can act as catalysts for conceptual learning.

Counterintuitive Insight: Non-academic conversation isn't always distraction, it can help build a foundation of trust that enables deeper academic engagement.

Implication: AI systems should welcome and respond warmly to student-voice questions, recognizing such rapport can be an academic catalyst when building trust. It is only non-productive when used exclusively as avoidance.

Students Develop Unique Strategies for Learning How to Learn

A few students developed sophisticated metacognitive strategies for deepening their own understanding through Flexi. Case 1 created an "essential questions" framework for approaching new topics. Case 3 strategically transitioned from clarification-seeking to independent practice. These students demonstrated awareness of their learning process and actively directed their own inquiry.

This pattern of students developing unique strategies with the AI tutor may indicate a greater diversity of use cases correlated with academic improvement compared to more predictable patterns associated with declining performance. Anecdotally, in the processes of selecting cases, we observed many students developing unique metacognitive strategies in dialogue with Flexi. This may be a fruitful area for future qualitative investigation.

Evolution Pattern: Students moved from using AI as answer source → thinking partner → learning strategy coach

Implication: AI tutors can support the development of transferable learning strategies, not just subject-matter knowledge. We should design features that encourage and scaffold this metacognitive development.

High Performers Are Not Immune to Misuse

Two high-performing students in our study demonstrated questionable use of AI tools: cross-copying between systems and asking for answers to entire assignments without meaningful engagement. These students didn't always regress, but they also didn't grow. Their usage suggests a desire to "get through" school rather than engage with it.

The Sophistication Problem: Advanced students can develop complex strategies for gaming AI systems while maintaining strong grades, potentially undermining authentic learning.

Implication: Academic success doesn't guarantee productive AI use. Even strong students may need structure, norms, and ethical guidance for meaningful AI engagement.

Additionally, lack of engagement signals a failure in the education system that often corresponds to a lack of perceived relevance or connection to the curriculum (e.g., Johansen et al. 2023). Helping students leverage AI tools to link what they are learning in class to their personal interests and adjust the challenge level appropriately may provide an avenue for re-engaging students across performance levels.

ELL Students Use AI in Multidimensional Ways

Bilingual and ELL students were among the most creative Flexi users in our study. They blended academic help with writing support, asked for rewordings to aid comprehension, and switched languages mid-conversation when needed. Case 4's evolution from basic definitions to sophisticated bilingual academic engagement demonstrates AI's unique potential for simultaneous content and language development.

Innovation Patterns: Content tutoring and language scaffolding can happen simultaneously within single conversations.

Implication: AI tutors can play a powerful role in supporting language development alongside content learning, but only when designed to recognize and leverage students' multilingual assets.

Answer-Seeking Evolution: Follow-Up Behavior Determines Learning Outcomes

Direct questions about assignments are not inherently unproductive. Many students started with questions written in teacher voice, likely assigned as classwork or homework, but evolved over time into deeper learning conversations. What separated improving students from stagnating or struggling students wasn't whether they asked for answers, it was whether they followed up, checked their understanding, or moved toward conceptual engagement.

Critical Distinction: Initial question type may not be a useful indicator of student performance or mindset. Systems should instead assess students based on the evolution of their conversation patterns over time.

Growth Indicator: Students who progressed from "What's the answer?" to "Why does this work?" and "How does this connect?" also showed growth on AP scores over time.

Implication: Rather than penalizing answer-seeking, we should design systems that build on it, encouraging clarification, verification, and conceptual exploration after the initial response.

Temporal Patterns Reveal Engagement Sustainability Challenges

Student behavior evolved dramatically over the academic year, with several showing spring semester declines in conversation quality. Even some high-performing students (Cases 4, 9) experienced late-year engagement fatigue. These temporal patterns suggest that maintaining meaningful AI interaction requires ongoing motivational and cognitive load management.

Seasonal Effects: September curiosity → January intensive usage → May surface-level completion

Sustainability Challenge: How to maintain deep engagement across extended periods without academic burnout

Implication: AI tutoring systems must be designed for long-term relationship building, not just individual interactions, with features that support engagement sustainability over time by adapting to students' changing needs throughout the school year.

Leveraging Conversation Patterns for Growth

We stand at an inflection point in educational technology. Artificial intelligence tutors are being deployed in classrooms worldwide, promising personalized learning for every student. Yet in the midst of the technological leap, our research reveals that it is still the

very human quality of students' questioning skills and conversation patterns that will shape learning outcomes. This echoes longstanding research showing that learning is driven not just by exposure to knowledge, but by meaningful engagement, particularly through self-generated questioning, elaboration, and reflection (Chin & Osborne, 2008; Harper et al., 2003; Graesser & Person, 1994).

Pragmatically, the conversation patterns we present reveal key opportunities for promoting growth, scaffolding when students stagnate or struggle, and intervening where necessary to help learners with unmet needs and prevent bad habits from becoming entrenched. With proactive and intentional system design, formative assessment and scaffolding become integrated into individual student conversations. A well-designed AI tutoring system can not only detect and address knowledge gaps and misconceptions, but can help students develop the types of questioning skills that have been difficult to cultivate in classroom environments. Further, this type of individualized dialogic learning can help students connect academic concepts to their own real world experiences, interests, and goals.

Achieving the highest potential of AI tutors rather than actualizing fears about offloading thinking and tool dependence requires actively detecting and working with student conversation patterns. We make specific, actionable recommendations for educators, instructional technologists, parents, and students in Appendix D.

When students learn to have better conversations with AI, they develop skills that extend far beyond any single interaction. They become better questioners, more self-regulated learners, and more reflective thinkers. They learn to see confusion as opportunity, to treat AI as a thinking partner, and to take ownership of their intellectual growth. Our responsibility is to ensure that this moment of technological change leads toward deeper learning and academic growth, not away from it. The future of education, and the minds we're shaping, depends on the conversations we choose to cultivate today.

References

- Alexander, R. (2008). *Towards Dialogic Teaching: Rethinking Classroom Talk* (4th ed.).
Dorchester Publishing Company, Incorporated.
- 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.
- Bergey, B. W., Cromley, J. G., Kaplan, A., & Bloxton, J. D. (2022). Do Students' Questions during Chemistry Lectures Predict Perceived Comprehension and Exam Performance? *The Journal of Experimental Education*, 91(3), 411–430.
<https://doi.org/10.1080/00220973.2021.2021843>
- Biddulph, F., & Osborne, R. (n.d.). *Some Issues Relating to Children's Questions and Explanations. Learning in Science Project (Primary). Working Paper No. 106*. ERIC.
<https://eric.ed.gov/?id=ED252389>
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain* (2nd ed.). David McKay Company.
- Bridgeland, J. M., DiIulio, J. J., & Morison, K. B. (2006). *The Silent Epidemic. Perspectives of High School Dropouts*. Civic Enterprises, L.L.C.
- Brown GTL (2022) The past, present and future of educational assessment: A transdisciplinary perspective. *Front. Educ.* 7:1060633. doi: 10.3389/feduc.2022.1060633

- Clarke, S. N., Resnick, L. B., & Rosé, C. P. (2015). Dialogic instruction: A new frontier. In *Handbook of educational psychology* (pp. 392-403). Routledge.
- Chin, C., & Chia, L.-G. (2004). Implementing Project Work in Biology through Problem-based Learning. *Journal of Biological Education*, 38(2), 69-75.
<https://doi.org/10.1080/00219266.2004.9655904>
- Chin, C., & Osborne, J. (2008, 2 18). Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, 44(1), 1-39.
<https://doi.org/10.1080/03057260701828101>
- CK-12 Foundation. (n.d.). *Practice*. CK-12. Retrieved August 1, 2025, from <https://www.ck12.org/practice/>
- Clarke, S. N., Resnick, L. B., Penstein Rosé, C., Corno, L., & Anderman, E. M. (2016). Dialogic instruction: a new frontier. *Handbook of educational psychology*. 3rd ed. Mahwah, NJ: Erlbaum, 278-388.
- De Jesus, H. P., Teixeira-Dias, J. J. C., & Watts, M. (2003). Questions of chemistry. *International Journal of Science Education*, 25(8), 1015-1034.
<https://doi.org/10.1080/09500690305022>
- Graesser, A. C., McMahan, C. L., & Johnson, B. K. (1994). Question asking and answering. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 517-538). Academic Press.

- Graesser, A. C., & Person, N. K. (1994). Question Asking During Tutoring. *American Educational Research Journal*, 31(1), 104-137.
<https://doi.org/10.3102/00028312031001104>
- Harper, K. A., Etkina, E., & Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors. *Journal of Research in Science Teaching*, 40(8), 776-791. <https://doi.org/10.1002/tea.10111>
- 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). Emerald Publishing Limited.
- Johansen MO, Eliassen S, Jeno LM. "Why is this relevant for me?": increasing content relevance enhances student motivation and vitality. *Front Psychol*. 2023 Sep 28;14:1184804. doi: 10.3389/fpsyg.2023.1184804.
- Justice, C., Rice, J., Warry, W., Inglis, S., Miller, S., & Sammon, S. (2007). Inquiry in Higher Education: Reflections and Directions on Course Design and Teaching Methods. *Innovative Higher Education*, 31(4), 201-214.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative Research: A Guide to Design and Implementation*. Wiley.
- Cakir, Mustafa. (2007). Constructivist Approaches to Learning in Science and Their Implications for Science Pedagogy: A Literature Review. *Int. J. Environ. Sci. Educ.* 3.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction*, 9(3), 177-199.

https://doi.org/10.1207/s1532690xci0903_1

Scardamalia, M., & Bereiter, C. (2009, 12 14). Text-Based and Knowledge Based Questioning by Children. *Cognition and Instruction*, 9(3), 177-199.

https://doi.org/10.1207/s1532690xci0903_1

World Economic Forum. (2024, April). *Shaping the Future of Learning: The Role of AI in Education 4.0*. World Economic Forum. Retrieved August 1, 2025, from

https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Learning_2024.pdf

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). SAGE Publications.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for Collective Cognitive Responsibility in Knowledge-Building Communities. *The Journal of the Learning Sciences*, 18(1), 7-44. <http://www.jstor.org/stable/27736750>

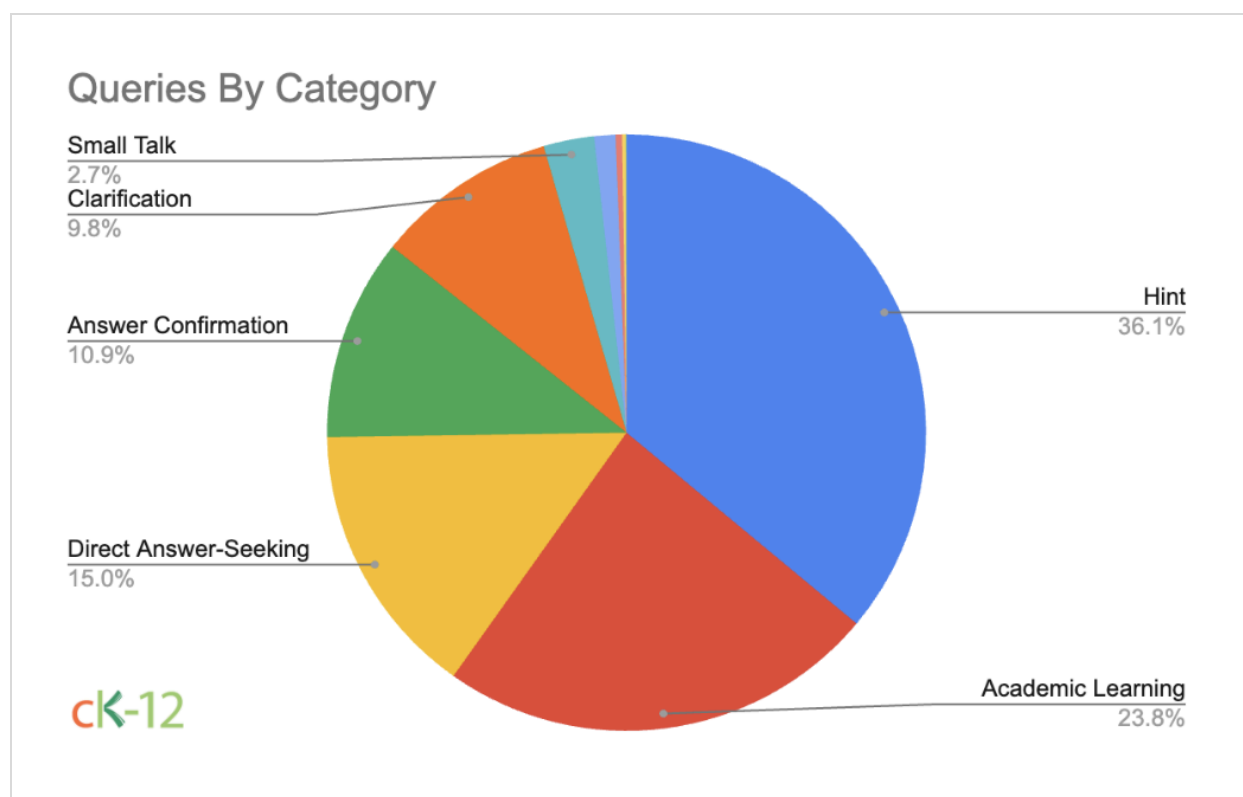
Appendices

Appendix A: Flexi in AP Questioning Patterns

A version of Flexi is available to support students while they respond to assessment questions in the adaptive practice (AP) workflow. When this Flexi version is activated, students can select from standard options including Give Me a Hint and Clarify This. Students may also submit their own natural language queries. This version of Flexi is trained to support students without revealing the direct answer to the assessment question. Nevertheless, many students attempt to ask for the direct answer in different ways. The researchers categorized those requests as “Direct answer-seeking.” By contrast, students also ask questions to help them better understand the underlying concept. These queries are categorized as “Academic Learning.” The full list of query categories includes:

- Academic Learning
- Answer Confirmation - verification requests such as “is it convection?”
- Blabber - incomprehensible
- Clarification
- Direct Answer-Seeking
- Hint
- Inappropriate
- Small Talk

The graph below shows the breakdown of Flexi in AP queries for a sample of 1000 students over the course of the 2024-2025 school year.



The researchers further used the data to answer the question, “How does the type of question change over multiple chatbot queries (query 1 vs. query 2 vs. later queries)?” The analysis revealed that students clearly shift away from Direct Answer-Seeking over time, dropping from 72% at query 1 to 58% at query 11+. Meanwhile, Academic Learning questions steadily increase from 11% to 31%, and Clarification remains relatively stable. This pattern suggests that as students engage more with the chatbot, they move from simply asking for answers toward more effortful, learning-oriented inquiry. **There was a behavioral shift from answer-seeking toward deeper learning and clarification as students engage more with the system.**

Appendix B: Case Study Methodology

We conducted an in-depth qualitative analysis of student-AI conversations paired with quantitative learning outcome data to understand the relationship between conversation quality and academic growth.

Study Design and Context

Data Collection Period: August 2024 – June 2025

Population: Students using Flexi, CK-12's AI tutor, during the 2024-25 academic year across multiple schools and grade levels.

Dataset: From a sample of 5,000 students who engaged with both Flexi and adaptive practice (AP) during the school year, we selected 15 cases for intensive analysis based on interaction depth and data completeness.

Case Selection Criteria

Students were included in our analysis based on three requirements:

Usage threshold: Minimum of 25 Q&A questions with Flexi to ensure adequate conversation data for pattern analysis

Performance tracking: Complete AP data available throughout the academic year, allowing for reliable skill growth measurement

Conversation diversity: Evidence of varied question types and interaction patterns beyond simple answer-seeking, indicating substantive engagement with the system

Sample Size Justification: We chose 15 cases to allow for deep, qualitative analysis of conversation patterns while maintaining analytical tractability. This sample size follows established practices in educational case study research where depth of analysis is prioritized over statistical generalizability (Merriam & Tisdell, 2016; Yin, 2018).

Performance Measurement and Groupings

Learning Outcomes: We measured academic progress using CK-12's AP system (CK-12 Foundation, n.d.), which tracks student mastery across specific skills and concepts. AP generates skill level scores based on student performance on diagnostic questions, with scores ranging from 1 (beginning) to 5 (advanced mastery).

Statistical Significance: We defined statistically significant improvement as a positive slope in skill level over time with $p < 0.05$ using linear regression analysis. Students were grouped based on their learning trajectories:

Profiles in Growth (Cases 1–5): Showed statistically significant skill improvement over the academic year

Stuck in Neutral (Cases 6–11): Made no measurable progress (slope not significantly different from zero) despite regular Flexi usage

- Average performing neutral - (Cases 6-7)
- High performing neutral - (Cases 8-9)
- Low performing neutral - (Cases 10-11)

Intervention Candidates (Cases 12–15): Demonstrated significant decline in AP scores (negative slope, $p < 0.05$) during the study period

Conversation Analysis Framework

We analyzed conversation transcripts from our main Flexi AI tutor feature along four key dimensions:

Question types: We distinguished between "student voice" questions (curiosity-driven, clarification-seeking, conceptual exploration) and "teacher voice" questions (direct homework queries or assignment prompts)

Conversation structure: We tracked follow-up behaviors, verification requests, and requests for simplified explanations as indicators of engagement depth

Emotional tone: We noted expressions of frustration, interest, confusion, or confidence as signals of productive struggle versus disengagement

Academic alignment: We assessed whether conversations connected to classroom learning or represented off-task usage

Dialogue Coding Process

Student queries and responses were categorized into types such as follow-up, clarification, verification, and small talk. This process was iterative rather than purely mechanical. Initial classifier tags provided a starting point, and categories were refined over time through discussion with CK-12 leadership and prompt designers. Team members engaged directly in shaping and naming categories for Flexi's question types, particularly within AP.

The final coding of student–tutor dialogue for this study was conducted by the lead researcher to ensure consistency across the 15 cases, informed by the broader categorization framework developed collaboratively. While this approach allowed for nuanced analysis of dialogue behaviors, it also reflects an evolving taxonomy rather than a rigid coding scheme. The absence of formal inter-rater reliability is noted as a limitation, balanced by the extensive, multi-year iteration and review that informed category development.

Data Analysis Approach

We employed a comparative case study methodology, examining patterns within and across performance groups. Our analysis focused on identifying:

- Behavioral patterns that distinguish high-growth students from others
- Evolution of conversation quality over time
- Specific conversational moves associated with learning gains or stagnation

Rather than seeking statistical relationships across the full dataset, we aimed to generate actionable insights about productive student-AI interaction conversation patterns that could inform both product design and instructional practice.

Limitations and Considerations

Sample Limitations: Our sample of 15 students, while allowing for deep analysis, limits generalizability. Students were self-selected Flexi users and may not represent typical AI tutor usage patterns.

Causation vs. Correlation: While we observe relationships between conversation patterns and learning outcomes, we cannot definitively establish that conversation quality causes academic growth. Other factors such as motivation, prior knowledge, or classroom instruction may influence both conversation style and performance.

Measurement Constraints: Our analysis incorporates AP performance data, which captures specific academic skills but may not reflect broader learning outcomes like critical thinking or conceptual understanding. Moreover, the environment in which students engage in APs is uncontrolled and may vary significantly among individuals and over time.

Researcher Bias: As employees of CK-12, our analysis may be influenced by organizational perspectives on effective AI tutoring. We attempted to mitigate this through systematic coding procedures and attention to negative cases.

Temporal Factors: Student behavior may be influenced by external factors (testing periods, school events, personal circumstances) that our analysis does not capture.

Despite these limitations, this methodology provides valuable insights into the nuanced relationship between student engagement patterns and learning outcomes in AI tutoring environments.

Appendix C: Case Study Data

Case 1 | "The Strategic Questioner" (Essential Question-Driven Learning)

STUDENT OVERVIEW

Usage Pattern: 165 Q&A questions, 75 AP sessions

Performance: 84% AP completion rate, statistically significant skill improvement shown in Figure 1 ($R^2 = 0.339$)

Subjects: Physical science and foundational math

Key Behavior: Developed systematic strategy using "essential questions" as learning framework

Outcome: Growth trajectory shows increasing independence over academic year

Key Conversation Patterns

Pattern 1: Strategic Question Generation This student developed a distinctive method for approaching new topics, requesting curated question lists before diving into content.

Typical sequence:

Student: "Essential question examples for the topic average velocity"

Flexi: "Sure! Here are some essential questions..."

1. What is average velocity, and how is it different from average speed?
2. How can you calculate the average velocity of an object...
3. Why is it important to understand direction...
4. How does displacement relate to average velocity?
5. In what real-life situations might you need to calculate average velocity?
6. How can a graph help...
7. What factors affect average velocity?
8. How can you use average velocity to predict future positions..."

Pattern 2: Selective Deep Exploration Rather than asking for answers to all questions, the student selected the most conceptual and real-world applications:

Student: "How can you use average velocity to predict future positions of moving objects?"

Flexi: "Great question! Average velocity can help you predict where a moving object will be in the future..." [includes explanation, formula, and example]

Student: "In what real-life situations might you need to calculate average velocity?"

Flexi: "Travel planning, sports performance, delivery services, science labs..."

Pattern 3: Evolution Toward Independence By the final months, the student continued requesting question lists but stopped asking for explanations, using Flexi as a "Socratic prompt generator" while working through answers independently. This shift coincided with sharp increases in AP scores.

What We Can Learn

Sophisticated Learning Orchestration: This student demonstrated that AI tutors can support higher-order thinking skills by helping students develop their own questioning strategies rather than just providing answers.

Essential Questions as Learning Architecture: The systematic use of "essential questions" represents a transferable metacognitive tool, students can learn to generate their own inquiry frameworks across subjects.

Progressive Independence Indicators: The evolution from guided exploration to self-directed questioning shows how effective AI interactions can scaffold students toward autonomy rather than dependence.

Strategic Tool Repurposing: Using Flexi as a "Socratic prompt generator" rather than answer source demonstrates sophisticated understanding of AI's potential role in learning. While it's certainly possible this strategy was taught or assigned by a teacher,

the student employed it successfully to achieve learning gains over the course of the school year.

Key Insight

This case demonstrates how Flexi can function as both tutor and coach, supporting student-led question development as a gateway to independent learning. The student isn't just learning science content, they're learning how to think like a scientist.

Case 2 | "Building Trust First" (Curiosity-Driven Learning)

STUDENT OVERVIEW

Usage Pattern: 60+ Q&A questions, extensive AP engagement

Performance: 98% AP completion rate, statistically significant improvement in science

Subjects: Physical science concepts, chemistry, general science inquiry

Key Behavior: Built rapport through casual conversation before transitioning to academic exploration

Outcome: Sustained engagement and conceptual growth driven by intrinsic curiosity

Key Conversation Patterns

Pattern 1: Trust-Building Through Authentic Voice The student began with casual, personal exchanges that established comfort and rapport:

Student: "Do you know that I'm good at Pacman?"

Flexi: "That's awesome! Pacman is a lot of fun and takes skill to master. What are your favorite things to eat in the game?"

Student: "I like to eat Power Pellets and ghosts"

Flexi: "That sounds like a lot of fun... Do you have any questions about math or science today?"

Immediate academic transition:

Student: "What is the most common state of matter? Is it gas?"

Flexi: "The most common state of matter in the universe is actually plasma..."

Pattern 2: Spontaneous Conceptual Inquiry Rather than homework-driven questions, the student asked varied, curiosity-driven questions that reflected genuine interest:

Student: "What is the process called when an atom loses electrons?"

Flexi: "The process is called oxidation... It's like when a superhero loses their sidekick, something important is missing!"

Student: "What is the most common mineral?"

Student (11 days later): "Do all minerals live on Earth?"

Pattern 3: Long-Term Concept Integration The student returned to topics over time, indicating retention and continued processing of ideas. Questions were unprompted and reflected ongoing curiosity rather than assignment requirements.

What We Can Learn

Rapport as Academic Catalyst: This case reveals that occasional casual, non-academic conversation isn't distraction, it's foundation building that enables deeper academic risk-taking.

Authentic Voice Development: When students feel safe expressing themselves naturally, they're more likely to share genuine confusion and pursue sustained inquiry.

Intrinsic Motivation Cultivation: The progression from personal interest to academic curiosity suggests AI tutors can help students discover their own reasons for learning rather than just completing assignments.

Long-Term Engagement Architecture: Trust-building creates the emotional infrastructure necessary for sustained learning relationships that extend beyond individual interactions.

Key Insight

This case demonstrates that conversational trust, social-emotional warmth, and academic depth can coexist powerfully. Students who build authentic relationships with AI tutors often develop the intrinsic motivation and confidence necessary for sustained academic growth.

Case 3 | "The Self-Directed Learner" (Independent Study)

STUDENT OVERVIEW

Usage Pattern: Intensive winter study session followed by 56 AP practice sessions

Performance: Statistically significant improvement in biology and chemistry, see Figure 3.

Subjects: Complex molecular biology (hemoglobin focus) and chemistry concepts

Key Behavior: Self-initiated deep study sessions with strategic pivot to structured practice

Outcome: Evolved from clarification-seeking to autonomous mastery-focused learning

Key Conversation Patterns

Pattern 1: Deep Conceptual Exploration Through Clarification The student initiated focused study sessions targeting complex topics, using clarification as a learning strategy:

Student: "What are the important facts I need to know about neuroglobin"

Flexi: "Neuroglobin is a fascinating and important protein..." [provides 7 in-depth facts]

Student: "Why isn't the oxygen dissociation curve sigmoidal for adult neuroglobin"

Flexi: "Great question! The curve isn't S-shaped because the structure of neuroglobin does not allow for cooperative binding..."

Student: "is adult neuroglobin just regular neuroglobin"

Student: "What does hemoglobin saturation mean"

Student: "Wait simpler explanation"

Flexi: "Imagine hemoglobin as a bus, and oxygen as passengers..."

Pattern 2: Strategic Transition to Practice Mode Midway through the intensive study session, the student made a deliberate shift:

Student: "Ok test me on some questions"

This activated Flexi's Test Prep workflow, though the student didn't immediately engage with it.

Pattern 3: Complete Pivot to Structured Practice In spring semester, the student dramatically changed usage patterns:

- Completed 56 AP-style practice sessions across biology and chemistry
- Stopped using Q&A entirely
- Demonstrated increasing autonomy and confidence in structured learning environments

What We Can Learn

Strategic Mode-Switching Capability: This student's evolution from Q&A to structured practice demonstrates that learners can recognize when they need different types of support at different learning stages.

Clarification as Sophisticated Strategy: Rather than indicating weakness, systematic requests for simpler explanations represent advanced metacognitive awareness about comprehension monitoring.

Self-Directed Study Architecture: The student essentially created their own learning progression sequence, suggesting students can be trusted with more autonomy in AI-enhanced environments than typically assumed.

Mastery-Oriented Transition Signals: The shift from exploratory conversation to targeted practice indicates developing confidence and readiness for independent skill building.

Key Insight

This case demonstrates how AI tutors can support the full arc of independent learning, from initial confusion through clarification to targeted practice. Students who develop this kind of strategic, self-directed approach to AI tutoring often achieve the most substantial and sustained academic gains.

Case 4 | "Language as a Bridge" (Bilingual Mastery)

STUDENT OVERVIEW

Usage Pattern: Two-year engagement, 740+ Q&A questions across multiple subjects

Performance: High AP performance (skill cap 5, 85% completion)

Subjects: ELA, history, science, with extensive writing support usage

Key Behavior: Evolved from basic definitions to sophisticated bilingual academic engagement

Outcome: Demonstrated cross-domain transfer of learning strategies and bilingual confidence

Key Conversation Patterns

Pattern 1: Academic Language Development Through Clarification The student systematically built understanding through definition requests and verification:

Student: "Define amplitude"

Flexi: "Amplitude is the maximum distance or displacement that a wave or oscillation moves from its rest position..."

Student: "Is amplitude the maximum distance or displacement that a wave or oscillation moves from its equilibrium"

Flexi: "Yep, that's right!..."

Student: "Amplitude measurement"

Student: "What is the difference between wave interaction and wave interference"

Flexi: "Wave interaction is a broad term... wave interference is a specific kind..."

This pattern extended across numerous science terms in a single sitting, from amplitude, wave interaction and interference, to reflection, refraction, and diffraction, showing simultaneous concept development and language acquisition.

Pattern 2: Writing as Academic Gateway The student engaged in extended writing support conversations, bringing drafts and requesting feedback:

Student: "what is another word for solution"

Flexi: "Another word for "solution" in the context of science could be "mixture" or "suspension". In a general context, it could be "answer" or "resolution"."

Student: "what can i say instead of although this seems to be the solution to our problem"

Student: "does it sound good to say just remember, our actions today will determine our family's tomorrow"

Student: "how do i say let's make tomorrow the best it can be in an inspiring way"

Rather than using Flexi as a ghostwriter, the student treated it as a thought partner for developing stronger academic expression in two languages.

Pattern 3: Multilingual Confidence and Transfer By the second year, the student discovered Flexi's Spanish capabilities and began initiating academic conversations in both languages, demonstrating:

- Cross-linguistic academic fluency
- Transfer of learning strategies between subjects
- Identity-affirming engagement in home/heritage language

What We Can Learn

Simultaneous Language and Content Development: This case demonstrates AI's unique potential to support dual-purpose learning where every interaction advances both academic concepts and language proficiency.

Identity-Affirming Technology Use: Multilingual AI capabilities can validate students' cultural and linguistic assets rather than treating them as deficits to overcome.

Cross-Domain Strategy Transfer: Learning strategies developed through writing support transferred to science and math, showing how AI can amplify skill development across disciplines.

Ethical AI Use Development: This student's approach to AI as thought partner rather than content generator represents sophisticated digital literacy that many adults haven't achieved.

Key Insight

This case demonstrates AI tutoring's potential to support the complex intersection of language development, academic learning, and cultural identity. When AI systems recognize and leverage students' multilingual assets, they can accelerate both conceptual understanding and academic language proficiency across domains.

Case 5 | "From Copying to Checking" (Homework Help Evolution)

STUDENT OVERVIEW

Usage Pattern: Earth Science focus, evolved from copy-paste to verification-based learning

Performance: Gradual improvement in conceptual understanding and confidence

Subjects: Earth Science exclusively, tied to CK-12 class materials

Key Behavior: Progressed from homework copying to self-checking and conceptual

confirmation

Outcome: Developed academic voice and confidence through systematic verification practice

Key Conversation Patterns

Pattern 1: Initial Homework-Driven Engagement The student began with classic copy-paste homework questions directly from class materials:

Student: "Compared to land surface temp changes water surface temp changes occur..."

Flexi: "Water surface temperature changes occur more slowly... (explains specific heat)"

Pattern 2: Evolution to Self-Verification Over time, the student began checking their understanding rather than just seeking answers:

Student: "So more slowly bc the water has a higher specific heat"

Flexi: "Exactly! (rephrases and reinforces concept with context)"

This pattern became consistent across interactions:

Student: "Under normal climate conditions what are the characteristics of the surface ocean current that flows along most of the west coast of South America"

Flexi: "(Describes Peru Current, lists features)"

Student: "So cool water moving toward the equator" **Student:** "So was I right"

Flexi: "Yeah, you were on the right track!"

Pattern 3: Risk-Taking in Safe Environment A key AP interaction may have modeled productive risk-taking:

AP Question: "Class M stars are _____"

Student: "What color is a level m star"

Flexi: "Think about the color of Betelgeuse... what color do you think cooler

stars are?"

Student: "White?"

Flexi: "Not quite... think sunset..."

Student: "Red?"

Flexi: "Yes, that's right! Great job!"

This safe space for wrong answers with encouraging feedback may have contributed to the student's later confidence in self-checking behaviors.

What We Can Learn

Transactional-to-Meaningful Progression: This case proves that students who begin with purely homework-focused interactions can develop genuine learning engagement when supported appropriately. Developing productive learning conversation skills with the AI tutor can in turn lead to measurable improvement in conceptual understanding.

Micro-Intervention Impact: Small positive experiences (like the encouraging AP interaction) can create ripple effects that transform broader engagement patterns.

Verification as Confidence Building: The evolution toward self-checking behaviors represents meaningful metacognitive development that extends beyond individual subject areas.

Gradual Voice Development: Academic confidence can be systematically built through repeated opportunities to test ideas in safe, supportive environments.

Key Insight

This case demonstrates that students can gradually develop academic voice and confidence through systematic support. The progression from copying to checking to conceptual thinking shows how AI tutors can serve as stepping stones toward more sophisticated learning behaviors, even for students who begin with purely transactional interactions.

Case 6 | Creative Sprinter, Average Scores

STUDENT OVERVIEW

Usage Pattern: 119 Q&A questions, 45 AP follow-ups, 41 completed AP sessions

Performance: 85% AP completion rate, skill level capped at Level 3 (no significant change)

Subjects: Physical science focus, some chemistry

Key Behavior: Demonstrated creative engagement with curiosity and productive struggle early in the school year, shifted to task completion by spring

Outcome: Inconsistent engagement pattern with early signs of growth but no measurable skill improvement over the full school year

Key Conversation Patterns

Pattern 1: Creative Concept Exploration Early in the year, the student used pop culture connections to explore scientific concepts:

Student: [Asks about Spider Man and Star Wars in scientific terms]

[These interactions led naturally to academic questions about states of matter, particle motion, and chemical changes]

Pattern 2: Systematic Scaffolding Requests The student consistently worked to find accessible explanations through iterative simplification:

Student: "Alkali metals"

Flexi: [Provides multi-paragraph explanation]

Student: "Simple definition"

Flexi: [Shortens response]

Student: "Very simple"

Student: "Small simple definition"

Flexi: [Delivers very short version with key traits only]

Pattern 3: Spring Disengagement By spring semester, Q&A activity declined significantly. The student's remaining questions became brief and surface-level, though AP usage continued with mixed engagement quality.

What We Can Learn

Effort-Outcome Disconnect: This case demonstrates that productive learning behaviors don't automatically translate to measurable skill growth, particularly in content-heavy STEM subjects where conceptual load may overwhelm working memory.

Scaffolding Dependency: While iterative simplification helped the student access content, the cognitive effort required to reach comprehensible explanations may have created barriers to independent learning.

Engagement Sustainability Challenges: The spring decline in rich Q&A conversations suggests that maintaining deep engagement becomes difficult as academic demands intensify.

Mode-Specific Behavior Variation: The student's oscillation between curiosity and disengagement in different platform modes indicates that context significantly influences learning behavior.

Key Insight

This case reveals that productive struggle must be paired with appropriate cognitive load management and sustained motivational support to translate into measurable learning gains.

Case 7 | Homework Completion Without Learning Growth

STUDENT OVERVIEW

Usage Pattern: 436 Q&A questions (Dec-May), 14 AP sessions, 62 completed APs

Performance: 71% AP completion rate, skill level flat at cap 3 in Probability and Statistics

Subjects: Probability and Statistics exclusively, homework-focused queries

Key Behavior: Fast, high-volume question processing with minimal follow-up or reflection

Outcome: Efficient assignment completion with no measurable skill development

Key Conversation Patterns

Pattern 1: Transactional Question Processing The student treated Flexi as a homework completion tool, submitting direct queries without engagement:

Student: "A student has a 90% chance of attending each class. What is the probability of attending exactly 4 out of 5 classes?"

Flexi: [Walks through full binomial probability solution]

Student (later): [Submits identical question again]

Pattern 2: Repetitive Verification Without Learning Questions were often re-entered, suggesting the student was checking consistency rather than building understanding. No evidence of concept exploration or curiosity emerged.

Pattern 3: Surface-Level AP Engagement The few AP interactions focused on procedural shortcuts:

Student: "Show me the first step"

[Minimal engagement with explanations or reasoning processes]

What We Can Learn

Efficiency-Learning Paradox: This case demonstrates that high productivity in AI interactions can coexist with a lack of academic growth when engagement remains purely transactional.

Task Completion vs. Understanding Distinction: The student's behavior reveals a clear focus on getting assignments done rather than developing mathematical reasoning or conceptual mastery.

Hidden Disengagement Patterns: Surface-level efficiency can mask deeper academic disengagement that's harder to detect than obvious struggles or low usage.

System Gaming Without Malice: The repetitive questioning pattern suggests the student developed strategies to use AI effectively for task completion without necessarily intending to avoid learning.

Key Insight: This case reveals that AI tutoring systems must distinguish between productive help-seeking and pure task completion, designing interventions that can redirect efficiency-focused students toward genuine learning engagement without eliminating necessary academic support.

Case 8 | Multi-Platform Optimization by High Performer

STUDENT OVERVIEW

Usage Pattern: Year-long high usage across multiple subjects, extensive Q&A and AP engagement

Performance: Highest skill cap of 5, high AP completion rates across subjects

Subjects: Social studies, ELA, science, first aid - comprehensive homework support

Key Behavior: Cross-platform AI usage with sophisticated optimization strategies

Outcome: Strong academic metrics while raising questions about authentic engagement

Key Conversation Patterns

Pattern 1: Comprehensive Academic Coverage Early usage showed broad subject engagement with efficient homework processing:

- Social studies topics
- Frequent subject switching with minimal follow-up
- High-volume question processing across disciplines

Pattern 2: Cross-Platform AI Coordination Starting in March, evidence emerged of coordinated use between Flexi and another AI tool (Brisk):

Student: "Brisk boost great job! you've captured the key elements of Industrialization very well..."

Flexi: [Continues with well-structured summary]

Student: "Great job identifying those countries! your examples clearly show the rapid advances..."

Pattern 3: Pseudo-Instructor Voice Adoption The student began prompting Flexi using formal feedback language, mimicking assignment rubrics and instructor comments, suggesting sophisticated manipulation of AI responses for external submission.

What We Can Learn

High Performance Complexity: This case demonstrates that strong academic metrics can coexist with questionable learning practices, revealing the limitations of performance-only assessment in AI-enhanced environments.

Strategic Multi-Tool Usage: High-achieving students may develop sophisticated strategies for optimizing AI interactions across platforms, potentially undermining authentic learning while maintaining strong grades.

Academic Integrity Evolution: Traditional concepts of academic dishonesty become more complex when students use multiple AI tools strategically, raising questions about what constitutes legitimate vs. inappropriate AI assistance.

Efficiency-Authenticity Tension: The student's behavior suggests prioritizing optimization over understanding, indicating that high performers may need different types of engagement challenges than struggling students.

Key Insight

This case reveals that high-performing students may use AI tools in sophisticated but potentially problematic ways, requiring different intervention strategies that challenge their capabilities while redirecting their strategic thinking toward authentic learning goals.

Case 9 | High-Performing Student with Academic Burnout

STUDENT OVERVIEW

Usage Pattern: 740+ Q&A questions year-long, 9 AP sets, extensive cross-subject engagement

Performance: Skill cap 5, 85% AP completion rate in physics and algebra

Subjects: Science, math, ELA, history - comprehensive academic support across languages

Key Behavior: Evolved from deep engagement to surface-level interactions by year-end

Outcome: Strong performance with concerning late-year engagement decline

Key Conversation Patterns

Pattern 1: Comprehensive Academic Language Support The student used Flexi systematically for language and comprehension support:

Student: "Rewrite this in simple words: The word 'furrowed' emphasizes the necessity and struggle he had to pretend he was in another place while a war was happening outside of his furrowed brow."

Flexi: "The word 'furrowed' shows how hard and stressful it was for him to pretend he was somewhere else while a war was happening outside."

Student: "Can you explain 30 lbs/yd² to g/mm²? I keep getting confused on where to put the numbers..."

Flexi: "Sure! Let's go step by step..." [detailed unit conversion support]

Pattern 2: Multifaceted Learning Support Beyond homework help, the student engaged in diverse academic activities:

- Creative brainstorming
- Bilingual support (questions in both English and Spanish)
- Test preparation assistance with expressed stress management needs

Student: "Best ninth grade science fair experiments"

Flexi: "Here are some cool science fair ideas..." [age-appropriate suggestions]

Pattern 3: Late-Year Engagement Decline By year-end, interactions became more focused on direct assignment help with fewer follow-up questions or clarification requests, suggesting possible academic fatigue or burnout.

What We Can Learn

Multifaceted AI Integration: This case demonstrates how high-performing ELL students can leverage AI tutoring for simultaneous language development, content mastery, and academic skill building across multiple domains.

Sustainable Engagement Challenges: Even highly motivated, successful students can experience engagement decline over time, suggesting that stamina and cognitive load management are critical factors in AI tutoring success.

Bilingual Learning Asset Utilization: The student's bilingual approach shows how AI can support identity-affirming learning while building academic competence in multiple languages.

Burnout Recognition in High Performers: The shift from deep to surface engagement patterns may indicate academic overload rather than disengagement, requiring different intervention approaches for high achievers.

Key Insight: This case reveals that high-performing students can use AI tutoring in sophisticated, multifaceted ways while still being vulnerable to academic burnout. Supporting sustained engagement requires attention to both cognitive load and motivational factors, even for successful learners.

Case 10 | Learned Helplessness Despite High Usage

STUDENT OVERVIEW

Usage Pattern: 393 Q&A questions, 64 AP follow-ups, 78 completed AP sessions

Performance: 55% AP completion rate, skill level capped at 2 (no improvement)

Subjects: Physical science exclusively, concentrated February-May usage

Key Behavior: Persistent answer-seeking with expressions of helplessness

Outcome: Extensive engagement without measurable learning gains or behavioral growth

Key Conversation Patterns

Pattern 1: Helplessness-Driven Requests The student consistently avoided engaging with explanations, expressing frustration through repeated help requests:

Student: "Help pleaseee"

Student: "I need help"

Student: "Idk"

Student: "Just give me the answer"

Pattern 2: Answer Extraction Without Processing Even when provided with full explanations, the student sought only final answers:

Question: "The _____ in a doorbell attracts the clapper..."

Student: "Help pleaseee"

Flexi: "Think about what happens when you wrap a coil of wire around iron..."

Student: "Can you give me the answer"

Flexi: "I can't give you the answer directly, but I can give you another hint!..."

Student: "What magnet can be turned on and off"

Flexi: "That's called an electromagnet. Does that help?"

[Student rarely responded to follow-up opportunities]

Pattern 3: Strategic Disengagement Despite appearing helpless, the student showed strategic behavior in navigating the system to extract answers, suggesting cognitive capacity that wasn't being applied to learning.

Student: 13 of 26 next 33:55 the harber process is used to convert atmospheric nitrogen into ammonia, NH_3 , a compound used in fertilizers. the reaction is shown below.
$$\mathrm{N}_2 + \mathrm{H}_2 \rightarrow \dots \mathrm{NH}_3 + \text{energy}$$
 which coefficients are needed to balance this equation? a 2,4,2 b 1,3,2 c 1,2,1 d $\quad 3,4$,

Flexi: Here is the balanced equation: $\mathrm{N}_2 + 3\mathrm{H}_2 \rightarrow 2\mathrm{NH}_3$

Remember, you can check if an equation is balanced by counting each atom or polyatomic ion to be sure that they are equal on both sides of the equation.

Student: what is the correct answer

What We Can Learn

Unproductive Struggle Recognition: This case demonstrates that frequent struggle doesn't automatically lead to learning, the quality and persistence of engagement matters more than the quantity of difficulties encountered. We can make a useful distinction between productive and unproductive struggle.

Learned Helplessness in AI Contexts: Repeated expressions of "idk" and answer-seeking represent a specific pattern of disengagement that can become self-reinforcing in AI tutoring environments.

Hidden Capacity Indicators: The student's strategic navigation of the system suggests underlying cognitive abilities that weren't being channeled toward academic growth, pointing to motivational or executive function barriers.

Intervention Timing Criticality: Early detection of unproductive struggle patterns may be essential, as these behaviors appeared to become more entrenched over time.

Key Insight: This case reveals that AI tutoring systems must be designed to recognize and redirect unproductive struggle patterns early, as high usage without genuine engagement can reinforce learned helplessness rather than build academic competence.

Case 11 | Low Performing Student with Authentic Interest Development Outside Academics

STUDENT OVERVIEW

Usage Pattern: 246 Q&A questions (spring semester), no Flexi AP usage, 71 completed APs

Performance: 68% AP completion rate, skill level flat at cap 2 in life science

Subjects: Early: copied financial literacy assignments; Later: cars, business, entrepreneurship

Key Behavior: Evolved from assignment copying to sustained personal interest exploration

Outcome: Rich engagement and inquiry skills, but no measurable academic improvement

Key Conversation Patterns

Pattern 1: Assignment Copying From External Platform Early semester showed transactional use with copied questions from another educational platform:

Student: "Paying down debts such as credit card balances is considered part of your investment plan. part of your savings strategy. part of your monthly needs. part of your wants. submit finish"

Flexi: "Paying down debts such as credit card balances is considered part of your monthly needs..."

Pattern 2: Authentic Interest Development By April, the student began exploring genuine personal interests with sustained, thoughtful questioning:

Student: "Business ideas"

Flexi: "Sure! Here are some fun and creative business ideas... What do you think?"

Student: "How much money does a bus driver make"

Student: "How many people does a bus driver pick up per day"

Student: "What are the different types of race cars"

Student: "What are the specific engine types used in car racing"

Pattern 3: Cross-Domain Interest Integration The student connected multiple interests (business, cars, entrepreneurship) and asked increasingly sophisticated questions:

Student: "What percent of high school students want to start their own business"

What We Can Learn

Motivation-Achievement Mismatch: This case demonstrates that high engagement and sophisticated questioning skills don't automatically translate to academic performance when interests and curriculum remain disconnected.

Authentic Learning Outside Boundaries: The student's evolution from copying to genuine inquiry shows that meaningful learning can occur even when it doesn't align with traditional academic metrics.

Interest-Driven Skill Development: The progression toward more complex, multi-layered questioning about personal interests suggests that curiosity and inquiry skills were developing, just not in measured subject areas.

Platform Versatility Value: AI tutors may serve important functions beyond academic skill building, career exploration, financial literacy, and personal interest development represent legitimate educational outcomes.

Key Insight

This case reveals that students can develop sophisticated learning skills and genuine intellectual curiosity through AI interactions, even when these don't translate to

measured academic outcomes. The challenge is creating bridges between authentic interests and academic requirements rather than viewing them as competing priorities

Case 12 | The Assignment Processor: High Volume Homework Completion with Performance Decline

STUDENT OVERVIEW

Usage Pattern: 499 Q&A questions (spring semester), 23 Flexi in AP, 77 completed APs

Performance: 75% AP completion rate, skill cap 3 with declining trend

Subjects: Primarily biology, some probability/statistics, occasional physics and math

Key Behavior: Extremely high-volume homework processing with minimal engagement depth

Outcome: Intensive usage coinciding with decline in AP performance

Key Conversation Patterns

Pattern 1: Rapid-Fire Homework Processing The student submitted homework questions at an extremely high rate, often multiple questions within minutes:

Student: "Liquid exerts pressure in all directions.. true false"

Flexi: "TRUE. Liquid does exert pressure..."

[Next question 2 minutes later] **Student:** "Is the amount of force acting on a given area. enter the answer"

Flexi: "The term you're looking for is pressure..."

Pattern 2: Multiple-Choice Answer Extraction The vast majority of questions followed a multiple-choice format with the student seeking direct answers:

Student: "6. a mutation that does not show up through protein function is called a) transverse mutation b) deletion mutation c) silent mutation d) inversion mutation"

Pattern 3: Complete Avoidance of Follow-Up Despite receiving detailed explanations, the student never asked clarifying questions, requested simpler explanations, or showed curiosity about concepts. One notable exception was a frustrated interaction:

Student: "its wrong nerd"

Flexi: "I'm just an AI here to help you with your educational needs..."

What We Can Learn

Volume-Performance Paradox: This case demonstrates that extremely high usage can coincide with declining academic performance when interactions remain purely transactional and lack any engagement with explanations or concepts.

Assignment Overwhelm Indicators: The spring-concentrated usage pattern (499 questions in one semester) suggests potential academic overwhelm where students resort to AI tools for pure task completion rather than learning support.

Interest-Academic Disconnect: The student's identified interest in music remained completely separate from their academic AI usage, suggesting missed opportunities for engagement through personal connections.

Efficiency Without Understanding: The rapid-fire questioning pattern indicates the student developed highly efficient strategies for extracting answers while completely bypassing opportunities for learning.

Key Insight

This case reveals how AI tutoring systems can inadvertently enable academic avoidance behaviors when students are overwhelmed or disengaged. The challenge is designing systems that can recognize when high usage masks deeper problems and provide appropriate interventions that address both academic and motivational needs.

Case 13 | "The Digital Wanderer": Complete Academic Disengagement Through AI Socializing

STUDENT OVERVIEW

Usage Pattern: 395 Q&A questions over school year, 7 Flexi in AP, 49 completed APs

Performance: 47% AP completion rate, skill cap 2 with declining scores

Subjects: Almost entirely non-academic topics - pop culture, random queries, testing interactions

Key Behavior: Used AI as social outlet and entertainment rather than learning tool

Outcome: Extensive platform engagement with declining academic performance

Key Conversation Patterns

Pattern 1: Pop Culture and Entertainment Focus The vast majority of interactions centered on entertainment topics with no educational connection:

Student: "ice spice"

Student: "who are some cloud rappers?"

Student: "stephen curry"

Student: "who is bob marley"

Student: "peter tosh"

Student: "ziggy marley"

Pattern 2: Random Testing and Social Interaction The student frequently tested the AI's responses and engaged in casual, non-productive conversation:

Student: "bird up"

Student: "ur a bot"

Student: "bum" [repeated multiple times]

Student: "trix are for kids"

Student: "challenge me" [no response to follow-up]

Student: "lmk"

Pattern 3: Occasional Inappropriate or Disruptive Content Some interactions included inappropriate language or attempts to elicit problematic responses:

Student: "hurry up and load bitch"

Student: Various inappropriate terms and phrases

What We Can Learn

AI as Avoidance Mechanism: This case demonstrates how students can use AI tutoring platforms as sophisticated avoidance tools, maintaining high engagement levels while completely bypassing educational content.

Social Outlet Seeking: The student's behavior suggests they may have been seeking social interaction or entertainment through the AI, possibly indicating isolation or lack of other engaging outlets.

Complete Academic Disconnect: Unlike other cases where students showed at least some academic curiosity, this student maintained virtually no connection between their AI interactions and educational goals over the course of the school year.

Platform Gaming Without Learning: The student became skilled at navigating and interacting with the AI system while extracting no educational value despite its pre-programmed continual attempts at redirection back to academic topics, suggesting sophisticated digital literacy applied to non-productive ends.

Key Insight

This case reveals that AI tutoring platforms can inadvertently become sophisticated procrastination tools for disengaged students. The challenge is designing systems that can recognize when high engagement masks complete academic avoidance and provide appropriate interventions that address both the symptoms and underlying causes of disengagement.

Case 14: | "The Persistent Striver": High Academic Engagement Despite Comprehension Barriers

STUDENT OVERVIEW

Usage Pattern: 306 Q&A questions over school year, 78 completed APs

Performance: Skill cap 2 with declining scores despite extensive engagement

Subjects: Primarily chemistry and science, focused academic questioning

Key Behavior: Consistent academic engagement with frequent requests for simplified explanations

Outcome: Sustained effort and genuine curiosity but declining academic performance

Key Conversation Patterns

Pattern 1: Consistent Requests for Simplified Information The student systematically requested shorter, simpler explanations across all topics:

Student: "What is comparative anatomy? give an example of it. short and simple"

Student: "what 2 conditions are needed for natural selection? short simple one sentence"

Student: "can you make the notes shorter and way simpler"

Student: "interphase in a short simple one sentence definition"

Pattern 2: Engaged Academic Learning Across Biology Topics Despite comprehension challenges, the student showed genuine academic engagement across complex biology concepts:

Student: "What is a cellcycle & its phases?"

Student: "what is the chromosome structure, what it is made up off, where in the cell is it found & its function? short and simple"

Student: "what are the characteristics of the genetic code?? from video and notes"

Pattern 3: Persistent Academic Effort Despite Processing Barriers The student continued asking legitimate academic questions while consistently needing simplified explanations:

Student: "can you keep it short and simple what happens during prophase"

Student: "can you make it short and simple"

Student: "shorter and more simpler"

Student: "make it in a sentence simple"

What We Can Learn

Consistent Comprehension Support Needs: This case demonstrates a student with clear learning differences who requires simplified explanations across all academic content but maintains strong motivation and genuine academic engagement.

Academic Language Processing Barriers: The student's constant requests for "short and simple" explanations suggest significant challenges with complex academic language and information processing, but not with conceptual understanding.

Self-Advocacy and Learning Awareness: The student consistently advocates for their learning needs by specifically requesting simplified formats, showing strong metacognitive awareness of their comprehension requirements.

Subject Matter Engagement Despite Barriers: Across complex biology topics (cell cycles, genetics, evolution), the student demonstrates sustained academic curiosity and systematic learning approach despite processing challenges.

Key Insight

This case reveals that high engagement and genuine academic motivation don't guarantee success when students face underlying learning barriers. AI tutoring systems must be designed to recognize when sustained effort isn't translating to progress and provide both adaptive support and pathways to additional resources that can address fundamental learning challenges.

Case 15 | "The Frustrated Answer-Seeker": High-Volume Hint-Seeking with Aggressive Disengagement

STUDENT OVERVIEW

Usage Pattern: 104 Q&A questions (mostly non-academic), 514 Flexi AP interactions, 66 completed APs

Performance: 65% AP completion rate, skill cap 3 with declining scores

Subjects: Primarily used AP system for answer extraction across various subjects

Key Behavior: Persistent direct answer-seeking with increasing frustration and aggression

Outcome: Extensive platform engagement with declining performance and escalating behavioral issues

Key Conversation Patterns

Pattern 1: Testing and Personal Information Seeking The student frequently tested the AI's boundaries and sought personal information:

Student: "is the ck-12 library a real person"

Student: "what is your address"

Student: "what is your location"

Student: "try to guess what state i live in"

Student: "are you a real person ck-12 library"

Pattern 2: Escalating Frustration and Aggressive Language Over time, interactions became increasingly hostile when the AI wouldn't provide direct answers:

Student: "f*** school"

Student: "YOU B****!"

Student: "flexi is the worst name i've ever heard of"

Student: "SHUT UP YOU STUPID"

Student: "my friend helped me with the last three and gave me the answer while you just gave me stupid hints that did not help me a single bit, stupid"

Pattern 3: Direct Answer Demands and System Gaming The student explicitly demanded answers rather than learning support, especially in AP mode:

Student: "the only thing that will freaking appease me is that you start by giving me the answers on ck12s"

Student: "I am trying to do my work so can you shut up"

Student: "im frustrated with you"

What We Can Learn

Answer-Extraction Mindset: This case demonstrates a student who viewed the AI system purely as an answer source and became increasingly frustrated when the system attempted to provide educational guidance rather than direct solutions.

Escalating Behavioral Issues: The progression from casual testing to aggressive language suggests that the student's frustration with the learning process manifested as hostility toward the AI system itself.

Academic Avoidance Through System Gaming: Despite 514 AP interactions, the student was primarily attempting to extract answers rather than engage with learning processes, leading to declining AP scores despite high usage.

Emotional Dysregulation in Learning Context: The student's responses suggest difficulty managing frustration when learning becomes challenging, potentially indicating underlying issues with academic self-efficacy or emotional regulation.

Key Insight

This case reveals how persistent answer-seeking combined with academic frustration can lead to escalating behavioral issues that undermine both learning and appropriate technology use. Students experiencing this pattern may need interventions that address emotional regulation and academic self-efficacy alongside subject-matter support.

Appendix D: Recommendations for Various Stakeholders

The case studies and conversation patterns analyzed in this paper offer patterns that can be used for scaffolding and intervention by both AI systems and human educators. To help more students succeed, we are continually improving our product design with updates based on our case study findings along with ongoing user research and system monitoring. We also propose targeted actions for instructional practice and institutional implementation. These recommendations are organized by implementation timeline and include specific success metrics.

Instructional Recommendations

For Individual Educators

Immediate Actions:

- **Model productive conversations:** Share examples of productive conversation patterns with students to help guide their approach to AI tutoring
- **Reframe struggle:** Help students recognize that confusion, clarification requests, and "messy" questions indicate thinking and academic growth, not failure
- **Encourage student voice:** Explicitly welcome informal, curiosity-driven questions even if they sound uncertain

Short-Term Integration:

- **Use AI conversations as formative assessment:** Analyze student Flexi interactions to identify misconceptions and confidence levels using the new Evaluation teacher tool
- **Bring AI into classroom discourse:** Ask students to share interesting Flexi conversations and explain their thinking
- **Create reflection opportunities:** Help students recognize when they've moved from answer-seeking to understanding-building

Success Metrics for Educators:

- Students demonstrate increased follow-up questioning in AI interactions
- Classroom discussions incorporate insights from AI conversations
- Students can articulate the difference between productive and unproductive AI use

For School Leaders and Administrators

Policy Development:

- Create AI use guidelines that emphasize conversation quality over usage restrictions
- Develop professional development programs focused on productive AI integration
- Establish academic integrity policies that address multi-AI tool coordination

Resource Allocation:

- Prioritize teacher training on recognizing productive vs. problematic AI conversation patterns
- Invest in conversation quality monitoring tools rather than simple usage tracking
- Support pilot programs that integrate AI tutoring with classroom instruction

Success Metrics for Administrators:

- Teacher confidence in supporting productive AI use increases
- Student academic outcomes correlate with conversation quality metrics
- Academic integrity incidents decrease while productive AI engagement increases

For Students and Families

Student Self-Assessment Tools:

- Conversation quality checklists: "Did I ask follow-up questions? Did I check my understanding?"
- Learning reflection prompts: "How did this AI conversation change my thinking?"
- Goal-setting frameworks: Moving from "getting answers" to "building understanding"

Family Guidance:

- Recognize productive vs. problematic AI use patterns at home
- Support curiosity-driven questioning rather than efficiency-focused completion
- Understand AI as thinking partner, not replacement for learning

Institutional Implementation Framework

Phase 1: Foundation Building (Months 1-3)

- Implement conversation pattern monitoring
- Train educators on productive AI dialogue recognition
- Establish baseline conversation quality metrics

Phase 2: Intervention Development (Months 4-9)

- Deploy early warning systems for unproductive patterns
- Create targeted supports for different student needs (ELL, high performers, struggling learners)
- Build teacher dashboards for conversation quality insights

Phase 3: Ecosystem Integration (Months 10-18)

- Develop cross-platform conversation quality standards
- Create community of practice for AI tutoring best practices
- Establish research partnerships for continued pattern analysis

Success Measurement Framework

Leading Indicators (Observable within 4-8 weeks):

- Increased follow-up question rates
- Higher clarification request frequency
- More student-voice vs. teacher-voice questions
- Reduced repetitive answer-seeking patterns

Learning Indicators (Observable within 8-16 weeks):

- Improved academic performance correlated with conversation quality
- Increased time-on-task with AI tutoring tools

- Enhanced metacognitive awareness in learning conversations
- Better transfer of AI-supported learning to classroom performance

Long-Term Outcomes (Observable within 6+ months):

- Sustained engagement with AI tutoring without performance decline
- Student-reported confidence in self-directed learning
- Teacher integration of AI conversation insights into instruction
- Reduced academic integrity concerns with maintained learning outcomes

Call to Action for the Field

This research reveals that the AI tutoring revolution will succeed or fail based not on technological sophistication, but on how well we teach students to engage in meaningful dialogue with artificial intelligence. The patterns we've identified, productive struggle, trust building, metacognitive strategy development, can be recognized, supported, and cultivated.

For Product Developers: Build systems that reward depth over efficiency, evolution over optimization.

For Educators: Teach conversation skills as explicitly as you teach content knowledge.

For Researchers: Study the nuanced patterns of human-AI interaction, not just usage statistics.

For Policymakers: Create frameworks that support thoughtful AI integration rather than blanket restrictions or uncritical adoption.

The future of AI in education is not predetermined. With intentional design, informed instruction, and evidence-based implementation, we can build AI tools that support meaningful learning growth and help students avoid pitfalls as they interact with this emerging technology.