When Not to Help: A Human-Teaching Lens on AI Tutoring

Emran Poh
emran.poh.2025@smu.edu.sg
School of Computing and Information Systems
Singapore Management University
Singapore, Singapore

Abstract

We invert the usual question about GenAI in education: rather than ask how to deploy new capabilities, we ask what role GenAI should play given what we know about effective human teaching. This position draws on an observational study of ten teacher-student pairs using speech, visual annotations, and remote screen control to teach Figma to novices. Teachers consistently showed restraint—avoiding high-precision modalities when it would undermine learning—which we interpret as pedagogical wisdom, not inefficiency. We argue this reveals a precision-agency tradeoff and digital territoriality as design constraints for GenAI tutors. We contribute empirical evidence of how teachers balance help with learner agency, methodological guidance for evaluation under authentic motivation, and implications for metrics and design process that treat restraint and friction as valid design goals.

1 Introduction

The discourse around Generative AI in education often proceeds from a technology-first perspective: given these new capabilities, how should we deploy them? We propose inverting this question: given what we know about effective human teaching, what role (if any) should GenAI play? This position emerges from an observational study [1] of human teachers coordinating speech, visual annotations, and remote screen control while teaching software skills to novices. Our findings reveal sophisticated adaptive strategies that challenge common assumptions about educational efficiency and suggest that resistance to automation may reflect pedagogical wisdom rather than technophobia.

The workshop's framing of student and educator opposition as "a design resource" resonates deeply with our empirical observations. Teachers in our study consistently demonstrated restraint in using high-precision instructional modalities, even when such restraint reduced immediate efficiency. One teacher explicitly avoided remote screen control, stating "if I control too much, the student
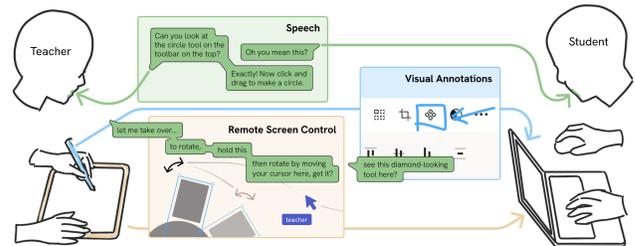
Figure 1: Teachers coordinate speech, visual annotations, and remote screen control to teach software skills—sophisticated strategies that can inform GenAI tutoring design.

don't get the experience." Another hesitated to use both annotations and control, preferring not to "touch their stuff." Rather than dismiss such reluctance as inefficiency, we argue it reveals critical design constraints: the precision-agency tradeoff and digital territoriality. These concepts, grounded in Cognitive Load Theory and joint attention research, suggest that effective instruction requires calibrating not just what information is provided, but who controls the learning process and workspace.

This paper contributes to the workshop's goals in three ways. First, we present empirical evidence of how human teachers balance instructional precision with learner agency, providing concrete design principles for GenAI tutoring systems. Second, we offer methodological guidance for evaluating educational technology under conditions of authentic motivation (a critical gap in current GenAI research). Third, we draw implications for metrics and design process that reframe resistance and friction as valid design goals rather than obstacles to overcome.

2 'Show It, Don't Just Say It': The Complementary Effects of Instruction Multimodality for Software Guidance

We conducted an observational study of ten teacher-student pairs (N=10) where experienced Figma users taught novices two progressive design tasks through videoconferencing. The study was deliberately constrained to three communication modalities (speech, visual annotations via Zoom's drawing tools, and remote screen control) to simulate capabilities that future AI tutoring systems might possess while preserving authentic human teaching dynamics (Figure 1).

2.1 Study Design for Authentic Motivation

A key methodological contribution lies in how we induced genuine learning pressure within a controlled lab environment. Teachers

Figure 2: The two Figma design tasks: Weather Icons Creation (left) and Weather Cards Design (right), showing progression from fundamental skills to advanced component usage.

were experienced Figma users (6+ months to 2+ years), and most had prior teaching experience. Students ranged from complete novices to those with 1-6 months of basic experience. Critically, students knew they would complete independent transfer tests immediately after each lesson, and teachers understood their effectiveness would be evaluated through student performance. The tasks were skill-based and procedural (students couldn't memorize their way through but had to develop actual software proficiency). This design created real stakes: teachers felt genuinely responsible for student learning, and students were intrinsically motivated to understand rather than merely complete the study. Despite physical separation by a divider (to prevent face-to-face nonverbal cues), the teacher-student relationships remained authentic, with teachers expressing real concern about intrusion and students showing genuine frustration and satisfaction.

2.2 Key Findings: Precision-Agency Tradeoffs

Our analysis revealed that each modality offered distinct advantages with corresponding costs to student agency. Speech formed the foundational modality (universal across all sessions, flexible, and preserving student control). However, it proved insufficient for conveying spatial precision ("try to make the bottom of the 16 on the same level as the bottom of the 30 and 22") and created high cognitive load as students struggled to map verbal descriptions to interface elements.

Visual annotations complemented speech by adding spatial precision (circling specific buttons, drawing layout sketches, or writing exact values next to input fields). Annotations acted as "soft" boundary crossings, marking the student's workspace without taking it over. Yet they introduced moderate agency intrusion by directing visual attention and sometimes creating clutter that teachers had to manually manage.

Remote screen control provided both spatial and temporal precision, demonstrating exact sequences of actions with interdependencies. However, it constituted a "hard" boundary crossing, temporarily reassigning workspace ownership to the teacher. Teachers used it sparingly (3 pairs heavily, 5 lightly) and strategically (primarily for complex demonstrations or critical corrections). Some teachers explicitly avoided it despite its efficiency, revealing a sophisticated understanding that over-control undermines learning.

2.3 Adaptive Teaching as Help Regulation

The most striking pattern was how teachers continuously calibrated their modality choices based on student state, task demands, and pedagogical philosophy. Teachers monitored student cursor movements as signals of cognitive processing, pausing before intervening

when students showed signs of productive struggle. They refined their communication over time (verbose explanations in early steps became concise directives as students gained familiarity). When students made errors, teachers escalated gradually: verbal correction first, then brief demonstrations if needed, then returning control for students to perform the fix themselves.

This adaptiveness reflected what help-regulation research calls the "assistance dilemma" (determining when and how much help optimizes learning). Teachers explicitly articulated this tension: "Sometimes you just have to step in and fix it, especially when they're getting frustrated" (P6-T) versus "I try to let them figure it out first, but if they're really stuck, I'll take control briefly" (P9-T). The optimal intervention point varied by student (some required multiple repetitions, others demonstrated rapid comprehension) and teachers adjusted pacing and detail accordingly.

2.4 Digital Territoriality and Social Norms

Teachers' modality choices were shaped not just by technical effectiveness but by interpersonal norms around permission and workspace ownership. The student's screen was treated as their "digital territory," and crossing into it required implicit justification. Annotations were more acceptable because they left the student in control; remote control felt intrusive because it displaced the student from their own workspace. As one teacher reflected, "I shouldn't have stepped in... I would rather annotate." These territorial boundaries help explain why teachers reached for annotations first and reserved remote control for severe breakdowns.

2.5 Learning Outcomes and Ecological Validity

All students performed well on transfer tests (10/10 completed Test 1, 9/10 completed Test 2), indicating that teachers' adaptive strategies successfully balanced guidance with agency preservation. Importantly, these outcomes emerged from teaching sessions where efficiency was not the primary goal (teachers prioritized student understanding and autonomy, sometimes at the cost of immediate progress). This validates the ecological validity of our study design: the teaching behaviors we observed reflected genuine pedagogical concerns rather than artificial compliance with study protocols.

3 Implications for GenAI Design

We focus on two areas where our findings speak directly: what metrics and methodologies best capture effectiveness, and how design processes might be adjusted so that GenAI tools for education are relevant, meaningful, and inclusive of learning (not just completion).

3.1 Metrics and Methodologies: Evaluating What Matters

Our study evaluated teaching effectiveness through immediate transfer tests where students had to perform independently, not just whether they completed the tutorial. This revealed critical patterns: students who received too much direct control struggled with transfer despite completing lessons quickly; students who experienced more struggle during lessons performed better afterward. Current GenAI evaluation often prioritizes task completion rates,

time-to-completion, or user satisfaction—metrics that can reward systems that undermine long-term learning.

Evaluating GenAI tutors by how well students perform without them could shift design incentives. Metrics could include:

- **Independence delay**: How long after tutorial completion can students perform the task unaided?
- **Transfer distance**: Can students apply learned skills to meaningfully different contexts?
- **Help-seeking calibration**: Do students request appropriate amounts of assistance for their skill level?

Methodologically, our teachers showed pedagogical restraint because the study design created real consequences: immediate transfer tests, progressive task dependencies, and procedural skills that couldn't be faked. Most GenAI education studies lack these conditions—artificial tasks, no consequences for understanding, arbitrary time limits. We argue that useful metrics require evaluation contexts where students face genuine consequences for learning, tasks require skill development rather than knowledge recall, and social dynamics reflect real pedagogical relationships (investment, responsibility, territorial boundaries). Without authentic motivation, we risk optimizing for artificial compliance: tools that look effective in the lab but fail when students actually need to learn.

3.2 Design Process: Restraint, Friction, and Territoriality as First-Class Goals

Our observations suggest that design processes for GenAI in education should treat restraint, productive friction, and territoriality as explicit design goals rather than afterthoughts.

Restraint as a core capability. When teachers declined to take remote control or hesitated to annotate, they were exercising expertise, not inefficiency. Design processes could specify not only when to help but when to hold back: e.g., deliberately withholding high-precision guidance during productive struggle, implementing "cool-down periods" after intrusive interventions, and making the AI's restraint visible ("I noticed you're exploring. I'll wait before suggesting").

Friction as a design goal. The friction of mapping verbal instructions to the interface was cognitively productive; teachers treated cursor movement and exploration as signs of learning. Design processes could target calibrated ambiguity (e.g., "the tool you need is in the left panel" vs. exact coordinates), fading precision as proficiency grows (ghost cursor → annotation → speech-only), and requiring students to explicitly request precision escalation.

Territoriality and permissions. Human teachers respected digital territory through social negotiation; GenAI can take over interfaces instantly. Territorial violations in our study were pedagogically harmful. Design processes could embed explicit permissions: students grant specific permissions for intervention types, territorial boundaries are visible (e.g., AI-controlled vs. student-controlled regions), and "suggestion mode" where the AI proposes but students execute. Consent becomes pedagogical design—the act of granting or denying permission supports learners in articulating their needs.

Together, metrics that reward independence and transfer, and design processes that institutionalize restraint, friction, and territoriality, can help ensure that GenAI tools for education serve learning rather than short-term efficiency.

4 Personal Interests and Background

I am a research-focused HCI practitioner investigating how people teach and learn complex software skills, with the goal of informing adaptive tutoring system design. My background combines design research training with deep interest in educational technology, particularly at the intersection of multimodal interaction and learning science.

This research emerged from a frustration with the technology-first framing of educational AI. As someone who has both taught software skills and struggled to learn them, I'm acutely aware of the gap between "efficient instruction" and "effective learning." I've watched colleagues provide brilliant demonstrations that students couldn't reproduce, and I've experienced the disorientation of having someone take control of my screen during troubleshooting (helpful in the moment but leaving me unable to solve the problem myself later).

What draws me to this workshop is its explicit validation of resistance and friction as design resources rather than problems to solve. The field's default posture toward GenAI often feels evangelical (celebrating capabilities without interrogating whether they serve pedagogical goals). I'm interested in convening with others who share a commitment to empirical grounding, who are willing to say "this technology shouldn't be used here" when evidence suggests it, and who recognize that effective learning often feels harder than efficient instruction.

My immediate research interests include: (1) developing design patterns for "restraint-capable" tutoring systems that know when to hold back, (2) creating evaluation methodologies that capture long-term learning outcomes rather than immediate task completion, and (3) understanding how territorial boundaries and permissions architectures might be formalized in educational technology. I'm particularly interested in cross-pollinating insights from domains beyond software learning (how do music teachers, sports coaches, or cooking instructors calibrate precision and agency? What can we learn from apprenticeship models where learners gradually take ownership?).

I hope to contribute to the workshop through: grounded empirical evidence from my study, provocations that challenge efficiency-first thinking, and genuine curiosity about how we collectively navigate the tension between GenAI's capabilities and pedagogy's complexities. I'm especially eager to learn from educators who are experiencing these tensions daily, as their situated knowledge will be essential for determining where and how GenAI can meaningfully support (rather than undermine) learning.

References

[1] Emran Poh, Yueyue Hou, Tianyi Zhang, and Jiannan Li. 2026. Show It, Don't Just Say It: The Complementary Effects of Instruction Multimodality for Software Guidance. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 1–22. doi:10.1145/3772318.3791200