Towards Evaluating Student Agency in AI-Mediated Learning

Zaki Pauzi
UCL Knowledge Lab
University College London
London, United Kingdom
z.pauzi@ucl.ac.uk

Vidminas Vizgirda
University of Oxford
Oxford, United Kingdom
UCL Knowledge Lab
University College London
London, United Kingdom
vidminas.vizgirda@cs.ox.ac.uk

Leslye Denisse Dias Duran
Institute for Ethics in AI, Faculty of Philosophy
University of Oxford
Oxford, United Kingdom
leslye.diasduran@philosophy.ox.ac.uk

Manolis Mavrikis
UCL Knowledge Lab
University College London
London, United Kingdom
m.mavrikis@ucl.ac.uk

Abstract

AI-mediated learning systems increasingly shape how students act, choose, and engage within digital classrooms, yet designers lack concrete ways to evaluate how these systems support or constrain student agency. This work-in-progress paper proposes an initial, design-oriented framing of student agency for interactive learning environments and introduces a prototype methods suite for analysing it in practice. Synthesising insights from HCI, the learning sciences, and child-rights guidance, we conceptualise agency as an interactional and multidimensional phenomenon produced through system affordances, decision policies, feedback structures, and orchestration behaviours. We outline early methods (i.e., log-data probes, scenario-based interface walkthroughs, and mixed-methods classroom studies) that reveal how students interpret system cues, negotiate control, and enact agency socially and developmentally. Our aim is to offer designers an emerging evaluative lens that identifies when AI systems narrow or expand students' room to act, and to support the creation of more intentional, agency-sensitive learning technologies.

CCS Concepts

• **Applied computing → Interactive learning environments**; • **Computing methodologies →** *Artificial intelligence*; • **Human-centered computing → Interaction design**.

Keywords

Children, Student, Learner, Agency, AI, Data-driven systems, Learning, Education, School

ACM Reference Format:
Zaki Pauzi, Vidminas Vizgirda, Leslye Denisse Dias Duran, and Manolis Mavrikis. 2026. Towards Evaluating Student Agency in AI-Mediated Learning. In *Designing Interactive Systems Conference (DIS Companion '26), June 13–17, 2026, Singapore, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3802974.3809419

ACM ISBN 979-8-4007-2632-3/26/06
https://doi.org/10.1145/3802974.3809419

1 Introduction

AI-mediated learning environments (e.g., intelligent tutoring systems, classroom orchestration tools, and AI-driven feedback assistants) are increasingly central to how students engage with digital systems, receive guidance, and navigate their learning trajectories. We focus on systems that make or shape instructional decisions, including widely deployed systems, such as ASSISTments [1] and ALEKS [2]. Studies of AI-enhanced classrooms show that these systems actively influence classroom orchestration by shaping when students seek help, how they interpret tasks, and how teachers choose to intervene [8]. International guidance (e.g., UNESCO, UNICEF, and the OECD) emphasises the need to protect children's decision-making autonomy and uphold fairness, transparency, and explainability in educational AI [6, 12, 14], and child-rights organisations further warn that system autonomy can become harmful when it undermines or manipulates students' choices, with UNICEF's Guidance on AI for Children explicitly stating that AI systems must not persuade or manipulate children and must protect their developmental capacity for agency [15].

Within interactive systems design, agency is closely tied to long-standing HCI concerns about autonomy, sense-of-control, and ownership over action [1], as well as legibility, reciprocity, and responsiveness, which shape people's ability to act meaningfully within digital environments [7]. Contemporary discussions further highlight contestability and delegation, particularly in mixed-initiative systems where AI autonomy can redirect or override user action [1].

Despite strong theoretical foundations in HCI and clear ethical imperatives from global policy guidance, neither field provides a concrete, design-oriented framework for evaluating how agency actually manifests in interactive learning systems. This paper positions the evaluation of student agency as a foundational design challenge—one that concerns both how AI-based learning technologies are built and how their effects on agency are studied. In HCI,

[1] https://www.assistments.org/
[2] https://www.aleks.com/

evaluation is inseparable from design: both are mutually constitutive and iteratively shape one another [18].

Throughout this paper, we use the term *students* to refer to the primary actors engaging with AI-mediated learning systems. When citing international policy guidance (e.g., UNICEF), we retain their use of *children* to reflect the developmental framing of those documents. We avoid broader HCI terms such as *users* or *people* except when referring to general HCI theory. We analyse four core system attributes that shape student agency: (1) system affordances, (2) decision policies (including underlying data flows), (3) feedback structures, and (4) orchestration structures (including adaptive behaviours).

The novelty of this work lies not in proposing yet another definition of student agency, but in reframing agency as an *evaluable* interactional outcome of AI-mediated learning design. By consolidating existing theoretical accounts into a single evaluative lens and pairing this lens with concrete, design-oriented methods, we enable agency to be systematically observed, compared, and iteratively assessed across systems and over time.

2 Positioning of Student Agency

Recent work in the learning sciences highlights student agency as a central concern in AI-mediated learning, yet concrete evaluative frameworks remain limited. Vincoli et al. argue that agency is often treated in a narrow, single-dimensional way, such as contrasting "high-" versus "low-agency" task or help-seeking conditions despite AI systems influencing a far wider set of student decisions [16]. Their Student–AI Agency Framework therefore conceptualises agency as multidimensional (Content, Feedback & Help, Co-Orchestration, and Data Agency), emphasising the need for models that capture this complexity. Parallel discussions in HCI emphasise issues of contestability and delegation in mixed-initiative systems [1], while learning-sciences research stresses that agency is developmentally sensitive and shapes how students engage with and benefit from educational technologies. Brod et al. note that although agency is widely valued across psychology, education, and philosophy, current systems vary considerably in the choice, control, and self-direction they afford (often without evidence or justification) and the field still lacks systematic ways to analyse how specific design decisions (e.g., automation levels, choice structures, or personalised adaptation) shape students' opportunities for intentional action and self-regulation [3].

Complementing these system-level perspectives, Mouta et al. show that educators see AI as reshaping the subjective and relational dimensions of agency, influencing students' sense of authorship, responsibility, and ownership over decisions [10]. Teachers worry that AI-driven personalisation may narrow choices, reduce self-reflection, and promote passivity, emphasising the absence of frameworks to help safeguard agency in practice. These concerns intersect with longstanding developmental insights: students' capacity to act intentionally, regulate behaviour, and interpret choices shifts significantly across their school years, as core cognitive processes (e.g., executive function) supporting goal-directed action follow non-linear developmental trajectories from childhood into adolescence [2, 9]. Together, this highlights the need for tools that account for both the subjective and developmental dimensions through which adaptive technologies shape students' opportunities to exercise agency.

In this paper, we adopt an interactional and multidimensional view of agency grounded in HCI, learning sciences, and child-rights scholarship. Here, agency is not a fixed trait of the learner but an emergent property of interactions among students, teachers, and AI systems. Specifically, student agency refers to students' capacity to act intentionally within learning environments, make consequential choices, understand and contest system behaviour, and participate in shaping their learning trajectories [11]. This positioning draws on Brod et al.'s developmental framing [3], Vincoli et al.'s multidimensional decision spaces [16], and Mouta et al.'s subjective, relational, and collective levels of agency [10].

2.1 Resulting gap

Together, this emerging literature paints a unified picture: student agency is widely regarded as essential, increasingly complex, and deeply affected by AI-mediated interactions. In their literature review of recent HCI research on child–AI interaction, Voysey et al. [17] point out a gap: researchers and practitioners currently lack robust means to evaluate agency. There are some proposed frameworks that are relevant for evaluating student agency in AI-mediated learning (e.g., [5, 19]), but they rely on learner self-reported data. We still lack HCI studies that systematically analyse, compare, or audit agency support across different systems. This gap sets the stage for our contribution: an evaluation-oriented framing of student agency tailored to interactive learning systems, enabling designers and researchers to empirically examine how system affordances, decision policies, feedback structures (formative assessment), and orchestration structures enable or constrain students' capacity to act, choose, and shape their learning trajectories [11].

3 From Frameworks to Student Agency Evaluation

To consolidate these perspectives into a set of actionable analytical dimensions, we examined how three complementary frameworks conceptualise and structure student agency in AI-mediated learning: Brod et al.'s proposal of an agency personalisation loop, which treats agency as a dynamically assigned design variable that must be matched to students' developmental prerequisites [3]; Vincoli et al.'s Student–AI Agency Framework, which outlines the concrete decision spaces students encounter when interacting with AI systems [16]; and Mouta et al.'s account of agency as shaped through subjective, relational, and collective dynamics in educational settings [10]. Synthesising these approaches allows us to articulate the core interactional dimensions through which agency is produced in practice. Table 1 summarises these dimensions across the three frameworks and highlights the specific aspects of student–system interaction they foreground.

3.1 Observable Indicators

To make these indicators actionable for design and evaluation, we illustrate each with concrete interactional contrasts that distinguish agency-supportive from agency-constraining system behaviour.

Table 1: Synthesised dimensions of student agency across existing frameworks.

Dimension	**Brod et al. [3] (Agency Personalisation Loop)**	**Vincoli et al. [16] (Student–AI Agency)**	**Mouta et al. [10] (Agency Levels)**
Decision–making	Agency should be assigned adaptively based on student prerequisites (e.g., executive function, metacognition, prior knowledge); too much or too little student choice harms effective learning.	Content and help/feedback agency specify which learning decisions students can make.	Subjective agency: authorship, deliberation, and volition in decision processes.
System legibility & interpretability	The loop requires systems to make the basis of assigned agency legible: students must understand feedback, task options, and system cues to benefit from adaptive agency.	Data agency stresses transparency, ownership, and consent.	Teachers highlight risks of opaque automation constraining understanding and autonomy.
Actionability (responsiveness)	Adaptive agency modulates how much control students have over actions (e.g., task selection, sequencing), balancing student autonomy with scaffolding to prevent ineffective choices.	Help/feedback options and student-initiated choices affect opportunities for action.	Intersubjective scaffolding supports developing metacognitive and agentic skills.
Participation in orchestration	The personalisation loop positions agency as shared control that shifts over time: systems, teachers, and students jointly shape learning trajectories depending on student readiness.	Co-orchestration agency: student involvement in structuring collaborative activities.	Collective agency: participation and distributed decision-making.

Evaluating student agency requires translating the analytical dimensions from Table 1 into *observable, comparable* evidence at both the system and student levels.

Brod et al. propose an agency personalisation loop in which learner control (e.g., task selection, sequencing, or access to hints) is adaptively matched to students’ executive function, metacognition, and prior knowledge, since both excessive and insufficient freedom can hinder learning [3]. Evaluating decision–making therefore concerns the *choice bandwidth* exposed to students, the alignment between offered choices and the student model, and how control is negotiated through overrides between student, teacher, and AI (e.g., students can choose whether to attempt a problem independently, request a hint, or defer support, versus systems that auto-advance students without meaningful choice given). A further indicator is whether student decisions produce meaningful downstream effects on content, feedback, or learning pathways, distinguishing substantive from merely nominal agency (e.g., rejecting a system recommendation alters subsequent task difficulty or feedback, rather than only dismissing a message while system behaviour remains unchanged).

Agency also depends on the legibility and interpretability of system behaviour. Students and teachers must understand what the system is doing and why, as opaque automation undermines autonomy and trust [7]. Indicators of legibility include the clarity of explanations accompanying recommendations (e.g., explicit rationales vs. generic prompts), transparency around feedback and progression logic, and the availability and actual use of contestation mechanisms (e.g., visible overrides or appeal options vs. hidden or absent controls). These features jointly determine whether students can make sense of how and why their agency is granted or constrained.

Actionability concerns system responsiveness to learner-initiated actions. Brod et al. emphasise that underscaffolding can overwhelm learners while overscaffolding restricts intentional action [3]. In line with Vincoli et al.’s account of Feedback & Help Agency [16], actionability can be evaluated by examining the proportion of learner–versus system-initiated moves, whether the system preserves or redirects learner actions (e.g., amplifying learner-initiated exploration versus frequently correcting or overriding it), and the action horizon (i.e., whether students can see and meaningfully influence next steps).

Participation in orchestration captures how roles and authority shift across students, teachers, and the AI system. Building on Mouta et al.’s account of agency as distributed across subjective, relational, and collective levels [10], Key indicators include the extent to which students participate in structuring activities such as grouping or pacing (e.g., students can negotiate groupings or timelines versus assignments fixed by the system), how easily teachers can mediate or override AI decisions, and opportunities for collective influence such as joint task selection. These dynamics reveal how agency is enacted not only individually but relationally within AI-mediated learning. Because agency is relational, future work will also explore participatory evaluation approaches that involve students and teachers as co-evaluators of agency rather than passive subjects.

While this work-in-progress paper focuses on affordances and orchestration structures, decision policies and underlying data flows are not excluded from the framework. Rather, we treat them as system attributes whose effects on student agency often manifest indirectly—through visible choices, feedback consequences, and opportunities for contestation. Even when models are partially opaque, their agency implications remain evaluable through interactional traces, such as whether personalisation narrows choice, whether overrides are possible, and whether students can understand or challenge system behaviour (contestability). Future work will extend our evaluative methods to explicitly probe these attributes in deployed systems.

3.2 Methods Suite

To capture the indicators outlined above, we propose a methods suite combining log-data analysis, scenario-based interface walkthroughs, and mixed-methods classroom studies. The core premise is that student agency is visible through patterned interactions between students, teachers, and AI systems over time, and not just through static self-reports.

Log-data (with timestamps) probes. Logs allow us to quantify patterns such as the rate of student-initiated versus system-initiated actions, the frequency and context of overrides, and the meaningfulness of decision trajectories, whether student choices produce downstream changes in content or sequencing. We additionally draw on Brooks et al.'s time-series interaction analysis method, which converts clickstream logs into temporally structured features (including daily/weekly indicators and n-gram behavioural patterns) to predict outcomes at scale [4]. Although their aim was prediction of student success, the same feature-engineering approach can be repurposed to evaluate agency by capturing indicators such as initiative (learner-initiated vs system-initiated actions), persistence of self-directed behaviour, and shifts in action patterns over time.

Scenario-based interface walkthroughs. While logs reveal what students do, scenario-based walkthroughs show what *they can do*. Researchers construct tasks that foreground key decision points, legibility demands, or scaffolding behaviours, drawing on cognitive walkthroughs to identify where users notice actions, interpret cues, or understand consequences [20]. Although originally a usability method, it effectively surfaces agency-relevant breakdowns (e.g., when students fail to recognise choices or cannot act on their intentions). Participants "think aloud" while navigating scenarios, making explicit how they interpret system cues and why they follow or reject system suggestions. Walkthroughs also expose hidden constraints by revealing whether ostensibly open actions are practically achievable. Because they can be repeated with alternative designs or configurations, they support controlled comparisons of agency-sensitive features.

Mixed-methods classroom studies. Finally, to understand how agency operates within the social ecology of a classroom, mixed-methods studies are essential. Classroom contexts make visible how teachers mediate AI decisions, how students negotiate choices collaboratively, and how institutional norms or expectations shape the enactment of agency. Our approach to mixed-methods classroom studies aligns with Small's account of how mixed-method designs can effectively combine heterogeneous data sources to capture mechanisms that neither qualitative nor quantitative approaches can reveal alone [13]. Mixed-methods allows us to link log-level interaction patterns with ethnographic observations and teacher interviews, providing both distributional evidence (e.g., how agency varies across students) and process-level insight (e.g., how control is negotiated moment-to-moment). Following Small's emphasis on integrating multiple forms of data to explain both patterns and mechanisms, this design enables us to trace how student agency is enacted socially, individually, and collectively within AI-mediated classrooms.

Together, these methods allow us to examine agency from complementary angles. Log data surfaces structural patterns in how control and choice unfold over time; scenario walkthroughs reveal how students interpret those patterns, where they feel supported or blocked, and how they make sense of system cues; and classroom studies expose the social dynamics through which teachers mediate system decisions and students negotiate choices.

An applied example. To illustrate how the proposed agency evaluation methods suite might work, consider a pilot study of an AI-supported mathematics learning platform that recommends new problems for learners to solve. Log data probes could track how often students accept, defer, or ignore recommendations, whether student overrides lead to meaningful changes in task sequence or difficulty, and whether students make self-directed choices consistently or only sporadically. Scenario-based walkthroughs could then ask students to complete a task, such as revising their learning path after receiving a recommendation, while thinking aloud, to reveal whether they understand how the system produces recommendations, notice alternative options, recognise consequences of each choice, and feel able to make such decisions. Finally, classroom observation and teacher interviews could examine whether teachers set paths for students or encourage them to exercise discretion, how teachers support students to self-assess own progress and make learning path decisions, and what teachers think about their students' genuine ability to make choices and enact them. In combination, these approaches move beyond learner self-reported agency evaluation. Observations triangulated across data sources could better inform design, highlighting where systems quietly narrow choice, over-automate learning flow, or obscure their reasoning. The suite therefore provides concrete signals for iteratively calibrating control, improving legibility, and supporting intentional action as system behaviour adapts to students' developing capacities.

4 Conclusion

This work-in-progress paper outlines an initial framing for evaluating student agency in AI-mediated learning and introduces a prototype methods suite. By synthesising how agency is assigned, enacted in decision spaces, and experienced subjectively and relationally, we identify dimensions that reveal where interaction design shapes students' room to act. Collectively, this work enables designers and researchers to (1) compare agency profiles across AI-mediated learning systems, (2) identify tradeoffs between control, legibility, and orchestration, and (3) track how agency shifts as systems and learners co-develop over time. As we refine early methods, our goal is to offer a practical, design-led way to identify when systems narrow or expand agency, and how adjustments in scaffolding, legibility, or orchestration can better align AI behaviour with students' developing capacities. For the community, this work provides an emerging evaluative lens to support more intentional, agency-sensitive learning technologies, with next steps involving iterative application across diverse AI-mediated tools.

Acknowledgments

This work was supported by the UK Research and Innovation (UKRI) Cross Research Council Responsive Mode project CHAILD - Children's Agency In the age of AI: Leveraging InterDisciplinarity (MR/Z505882/1).

References

[1] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 375, 18 pages. doi:10.1145/3544548.3580651

[2] John R Best and Patricia H Miller. 2010. A Developmental Perspective on Executive Function. *Child Development* 81, 6 (11 2010), 1641–1660. arXiv:https://academic.oup.com/chidev/article-pdf/81/6/1641/64423329/j.1467-8624.2010.01499.x.pdf doi:10.1111/j.1467-8624.2010.01499.x

[3] Garvin Brod, Natalia Kucirkova, Joshua Shepherd, Dietsje Jolles, and Inge Molenaar. 2023. Agency in Educational Technology: Interdisciplinary Perspectives and Implications for Learning Design. *Educational Psychology Review* 35, 1 (Feb. 2023). doi:10.1007/s10648-023-09749-x

[4] Christopher Brooks, Craig Thompson, and Stephanie Teasley. 2015. A time series interaction analysis method for building predictive models of learners using log data. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York) *(LAK '15)*. Association for Computing Machinery, New York, NY, USA, 126–135. doi:10.1145/2723576.2723581

[5] Federica Cavazzoni, Alec Fiorini, and Guido Veronese. 2022. How Do We Assess How Agentic We Are? A Literature Review of Existing Instruments to Evaluate and Measure Individuals' Agency. *Social Indicators Research* 159, 3 (Feb. 2022), 1125–1153. doi:10.1007/s11205-021-02791-8

[6] Virginia Dignum, Melanie Penagos, Klara Pigmans, and Steven Vosloo. 2021. Policy guidance on AI for children. *Report, UNICEF, Office of Global Insight & Policy, United Nations Children's Fund* (2021), 10017.

[7] Romualdo Gondomar and Enric Mor. 2021. Understanding Agency in Human-Computer Interaction Design. In *Human-Computer Interaction. Theory, Methods and Tools*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 137–149.

[8] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2019. Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In *Artificial Intelligence in Education*, Seiji Isotani, Eva Millán, Amy Ogan, Peter Hastings, Bruce McLaren, and Rose Luckin (Eds.). Springer International Publishing, Cham, 157–171.

[9] Jianping Ma, Xiaoxi Liu, Teng Pan, Linqin Ji, and Wenxin Zhang. 2025. Developmental Changes in the Structure of Executive Function from Early To Late Adolescence. *Journal of Youth and Adolescence* (Dec. 2025). doi:10.1007/s10964-025-02293-7

[10] Ana Mouta, Ana María Pinto-Llorente, and Eva María Torrecilla-Sánchez. 2025. "Where is Agency Moving to?": Exploring the Interplay between AI Technologies in Education and Human Agency. *Digital Society* 4, 2 (June 2025). doi:10.1007/s44206-025-00203-9

[11] OECD. 2019. OECD Future of Education and Skills 2030. In *OECD Learning Compass 2030. A Series of Concept Notes.* https://www.oecd.org/education/2030-project/

[12] OECD. 2023. *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem.* OECD Publishing. doi:10.1787/c74f03de-en

[13] Mario Luis Small. 2011. How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature. *Annual Review of Sociology* 37, 1 (Aug. 2011), 57–86. doi:10.1146/annurev.soc.012809.102657

[14] Unesco. 2022. *Recommendation on the ethics of artificial intelligence.* United Nations Educational, Scientific and Cultural Organization.

[15] UNICEF Innocenti. 2025. *Guidance on AI and Children: Version 3.0.* Technical Report. UNICEF Office of Global Insight and Policy, Florence, Italy. https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf Recommendations for AI policies and systems that uphold child rights.

[16] Martina Vincoli, Niklas Scholz, and Tomohiro Nagashima. 2025. Multidimensional Student Agency in Learning with AI: A Conceptual Framework and Design Implications. In *Artificial Intelligence in Education*, Alexandra I. Cristea, Erin Walker, Yu Lu, Olga C. Santos, and Seiji Isotani (Eds.). Springer Nature Switzerland, Cham, 218–232.

[17] Isobel Voysey, Vidminas Vizgirda, Sarah Turner, Leslye Denisse Dias Duran, Zaki Pauzi, Manolis Mavrikis, Carina Prunkl, and Jun Zhao. 2026. Agency in Child-AI Interaction: A Review of How It Is Conceptualised, Studied, and Supported in HCI. (Feb. 2026). https://hal.science/hal-05500673

[18] Christine E. Wania, Michael E. Atwood, and Katherine W. McCain. 2006. How do design and evaluation interrelate in HCI research?. In *Proceedings of the 6th Conference on Designing Interactive Systems* (University Park, PA, USA) *(DIS '06)*. Association for Computing Machinery, New York, NY, USA, 90–98. doi:10.1145/1142405.1142421

[19] Liangliang Xia, Kexin Shen, Herui Sun, Xin An, and Yan Dong. 2025. Developing and Validating the Student Learning Agency Scale in Generative Artificial Intelligence (AI)-Supported Contexts. *Education and Information Technologies* 30, 10 (July 2025), 13999–14021. doi:10.1007/s10639-024-13137-5

[20] Seyma N. Yildirim-Erbasli, Okan Bulut, Carrie Demmans Epp, and Ying Cui. 2023. Conversation-Based Assessments in Education: Design, Implementation, and Cognitive Walkthroughs for Usability Testing. *Journal of Educational Technology Systems* 52, 1 (2023), 27–51. doi:10.1177/00472395231178943