Orchestrating Classrooms and Tutoring with Carnegie Learning's MATHia and LiveLab*

Stephen E. Fancsali[1,*], Michael Sandbothe[1] and Steve Ritter[1]

[1]*Carnegie Learning, Inc., Pittsburgh, PA 15219, USA*
http://www.carnegielearning.com/

Abstract
While effective for student learning, high-dosage tutoring can be costly. One promising area for achieving greater efficiency and lower costs is the augmentation and provisioning of human (possibly remote) tutoring support with data-driven guidance from artificial intelligence (AI)-driven adaptive learning software and related applications. We consider work to date on supporting classroom mathematics instructors with data from such systems, focusing on Carnegie Learning's MATHia and its LiveLab companion app for teacher classroom orchestration. We describe on-going research and a "road map" for learning analytics research on detector models and software feature development to orchestrate human tutoring in addition to more traditional classroom instruction.

Keywords
High-dosage tutoring, Intelligent tutoring systems, Orchestration

1. Introduction

1.1. Classroom Orchestration with MATHia & LiveLab

Since 2018, Carnegie Learning has provided instructors using its MATHia [1] computer-based intelligent tutoring system (ITS) with a companion application called LiveLab to help them support learners using MATHia. Applications like this have been called tools for "teacher awareness" [2] or "classroom orchestration" [3, 4, 5] because they can provide insights about how to "orchestrate" instructors' moves through their classrooms and computer labs, empowered with insights into which students may most benefit from supports like additional help or a congratulatory gesture on a recent achievement.

Holstein et al. [2] experimentally demonstrated that providing instructors with real-time analytics (in an augmented reality, "heads up" display tool) about students' usage of an ITS enhanced student learning outcomes compared to both a similar tool without real-time analytics and a "business as usual" condition without any such supporting tool. In what follows, we consider analytics and information provided to instructors using Carnegie Learning's LiveLab, which has never been detailed in the artificial intelligence in education literature.

AIED 2023 Workshop: Towards the Future of AI-Augmented Human Tutoring in Math Learning, July 07, 2023, Tokyo, Japan
*Corresponding author.
sfancsali@carnegielearning.com (S. E. Fancsali); msandbothe@carnegielearning.com (M. Sandbothe); sritter@carnegielearning.com (S. Ritter)
0000-0002-4016-8145 (S. E. Fancsali); 0000-0003-2807-9390 (S. Ritter)

Before considering LiveLab's support for classroom instructors, we consider three contexts in which learners might interact with human tutors (as opposed to classroom instructors or computer based ITSs). Our aim is to compare and contrast needs for orchestrating more traditional classrooms using an ITS like MATHia with contexts in which learners using an ITS are supported by (human) tutors. We consider promising avenues for future research investigating how to use data to support tutoring contexts and scenarios as well as a road map for new software feature development for LiveLab and related orchestration tools to support learning.

1.2. Tutoring Orchestration

The global pandemic has drastically increased the demand for approaches that address so-called "learning loss." One especially popular approach for which there is strong evidence is high-dosage tutoring (e.g., [6]). Building on its analog and digital products and services, Carnegie Learning has introduced tutoring services, delivered by state-certified math teachers, generally via remote/virtual mechanisms like video conferencing software.

In current research and development efforts, Carnegie Learning and its research partners (many represented at the present workshop), are considering three different contexts for initiating tutoring services with students:

- *Student-Initiated Tutoring*: Students individually seek out on-demand, one-on-one help from a tutor (or schedule a one-on-one session with a tutor for the near future).
- *Tutor-Initiated Tutoring (or Data-Driven Tutoring)*: Students are identified in real-time as requiring support, and a tutor initiates a one-on-one session with one (or more) student(s), empowered with students' elements of their recent ITS data to better understand how to provide just-in-time support.
- *Scheduled Tutoring*: Students are allocated to regular, (generally) small-group tutoring sessions, which typically occur twice a week for 30-60 minutes for (often) 10-week "rounds," based on criteria established by their schools to provision available tutors to students who need support.

Carnegie Learning's tutoring services, delivered to nearly 10,000 learners in the 2022-23 school year, are a combination of *student-initiated* and *scheduled tutoring*. Recent field testing with research partners at a single school site has begun to explore possibilities for data-driven, *tutor-initiated tutoring*. While tutoring initiated by a (possibly remote) tutor is a promising area for increasing efficiency and equity in tutoring delivery, we are exploring how all methods of tutoring and more traditional classroom instruction can be better "orchestrated" with AI supports and data through MATHia and LiveLab,[1].

The need for data-driven or *tutor-initiated tutoring* is motivated by at least one pre-pandemic investigation into how tutoring services are used by learners using computer-based, adaptive ITSs like MATHia, which illuminated the need for data-driven augmentation and provisioning of such services for efficient, more equitable delivery. Fancsali et al. [7] considered the extent to which remote-learning users of Cognitive Tutor (MATHia's predecessor) sought out support

[1]or related apps like PLUS (https://tutors.plus/)

from human tutors using a chat-based tutoring service. They found that this kind of "student-initiated," on-demand tutoring was used extensively by a small proportion of users, with 2.1% of 16,905 learners in the sample of university-level ITS users accounting for 55.4% of total session time with chat-based tutors. Over 80% of learners (13,585) made no use of the human, chat-based tutoring service at all. While far from conclusive, this distribution suggests that some learners are over-using resources while others are, with near certainty, not making use of resources despite having some needs for additional support. This suggests *student-initiated tutoring* alone may result in many students who would benefit from tutoring support not seeking out such support.

Whether orchestrating instructors' classrooms or tutors' tutoring (regardless of how initiated), there are at least two opportunities for AI-augmented orchestration support: identifying the appropriate learners to support and providing the instructor or tutor with helpful information that will empower them to better support identified learner(s). We now consider MATHia and ways in which LiveLab supports classroom instruction with MATHia, before we consider our road map for research and development and how supports for classroom instructors may (not) transfer to the case of tutoring.

2. MATHia & LiveLab

2.1. MATHia

MATHia (previously Cognitive Tutor [1] is an ITS for middle school and high school mathematics used by 600k+ learners each year. Typically, MATHia is used as part of a blended curriculum that has students work at their own pace in the adaptive ITS software for approximately 40% of their instructional time (often two math class periods per week), while the remaining instructional time is facilitated by instructors guided by Carnegie Learning's work-texts on collaborative problem-solving and related activities.

MATHia presents students with complex, multi-step math problems (see Figure. 1). Steps within problems are generally mapped to one or more knowledge components (KCs) [7], and context-sensitive hints and feedback are available at each step. Sets of KCs are clustered in topics or "workspaces" that present students with multiple opportunities to practice and demonstrate mastery of all KCs associated with a topic.

Student progress to KC mastery is monitored using Bayesian Knowledge Tracing (BKT) [8], and students make progress through a sequence of workspaces that comprise their curriculum by demonstrating mastery of the set of KCs associated with each topic. Within each workspace, if a student has yet to reach mastery of all associated KCs after (typically) 25 problems, the student is moved on within their curriculum sequence to the next workspace without mastery. This instructional policy, when combined with the fact that students can always access the correct answer to each problem step via "bottom out" hints that provide the answer, ensures that students are never left to endlessly "wheel spin" [9] within a particular problem or topic. Nevertheless, predicting as early as possible that non-mastery of all KCs is likely is a key target for predictive analytics (or so-called "detector" models) that are implemented in the LiveLab classroom orchestration tool [10].

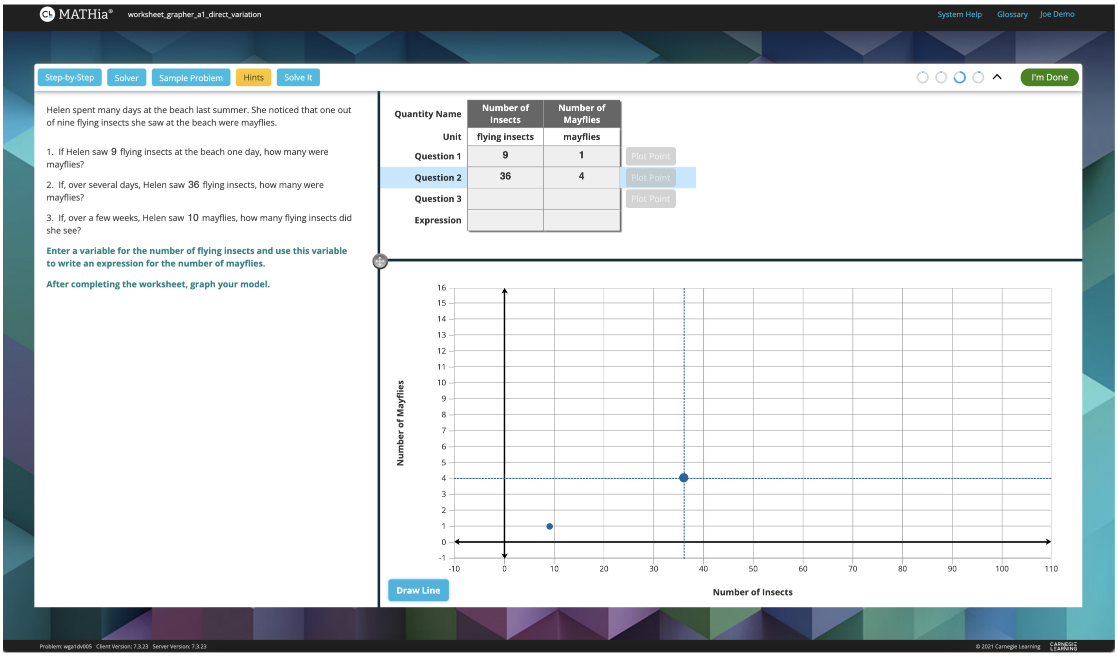

Figure 1: Screenshot of problem-solving in MATHia's "Modeling the Constant of Proportionality" workspace. A problem scenario and several questions are displayed on the left, and students provide answers via a table and interactive grapher tool on the right.

2.2. LiveLab

LiveLab's default display is a list of MATHia users (see Figure 2 with anonymous students) in the class for which they've launched the app. Each student is indicated to be in one of four activity levels (for which filtered lists also exist, showing students exclusively within a particular activity level):

- *Offline*: The student is not logged in to MATHia.
- *Idle*: The student is logged in to MATHia but has no activity for at least five minutes (with an indicator of the number of minutes of idle time).
- *Monitor*: Three or fewer of the student's last ten actions in MATHia were correct, indicating that a student may need additional support.
- *Working*: The student is active in MATHia and is not in the *Monitor* state.

Teachers within a MATHia classroom or computer lab session might first target students with *Idle* or *Monitor* status for additional support. Additionally, a notifications panel in LiveLab provides a stream of events, including that a student has completed a workspace. If a student completes a workspace having mastered all KCs, this represents a potential point of celebration. If a student completes a workspace without KC mastery, this likely represents an opportunity to provide additional support.

Unproductive Struggle Detector In addition to activity levels and notifications, LiveLab implements a "detector" of unproductive struggle, operationalized as a student's high probability

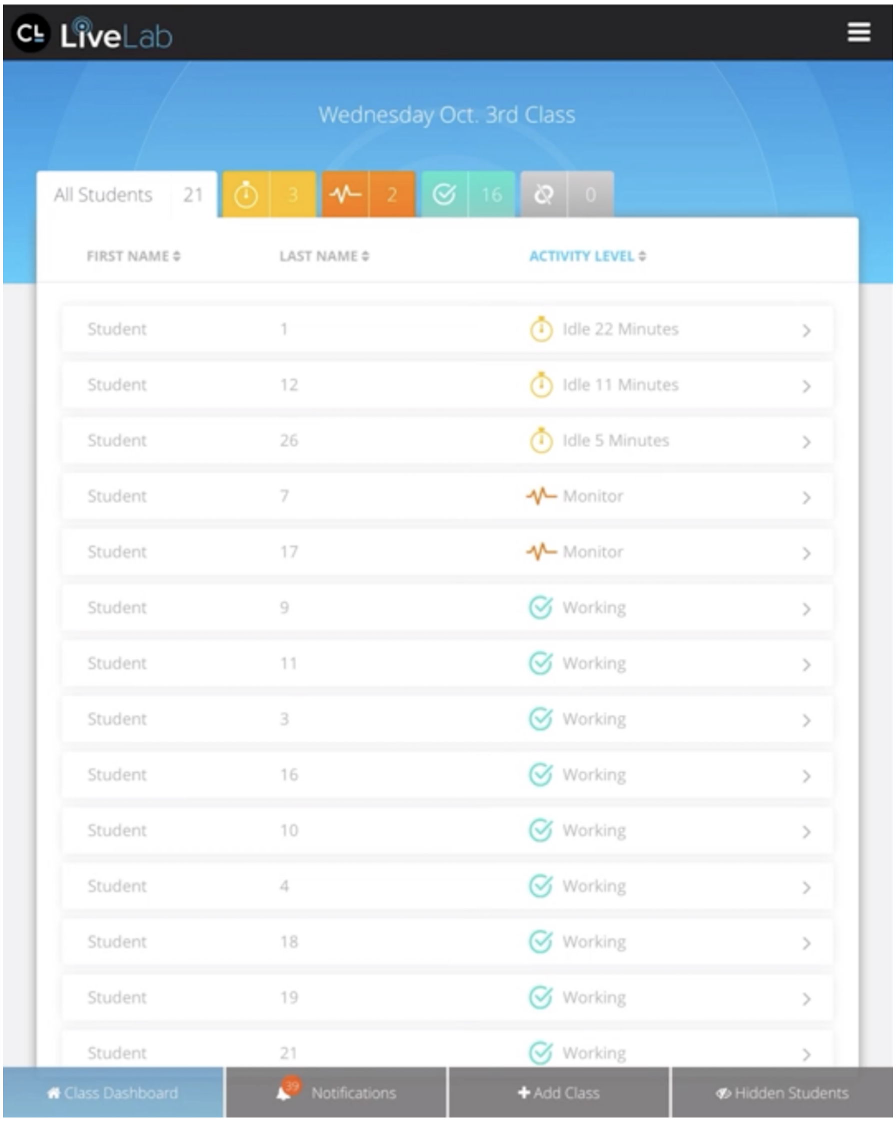

Figure 2: Screenshot of LiveLab's default display of learners in a class with their current activity level.

of failure to reach mastery of all KCs associated with a workspace before reaching the set maximum number of problems [10]. Currently, the detector is implemented as a set of univariate logistic regression models learned independently for each KC-opportunity. The single independent variable in each logistic regression model is the student's current (i.e., at the opportunity for which the model is learned) probability of mastery of the relevant KC (estimated using BKT). These models effectively establish thresholds for the probability of mastery at each KC-opportunity, below which a student is inferred to have a high probability of non-mastery. If at least one such KC-opportunity model has indicated a high probability of non-mastery, a life-

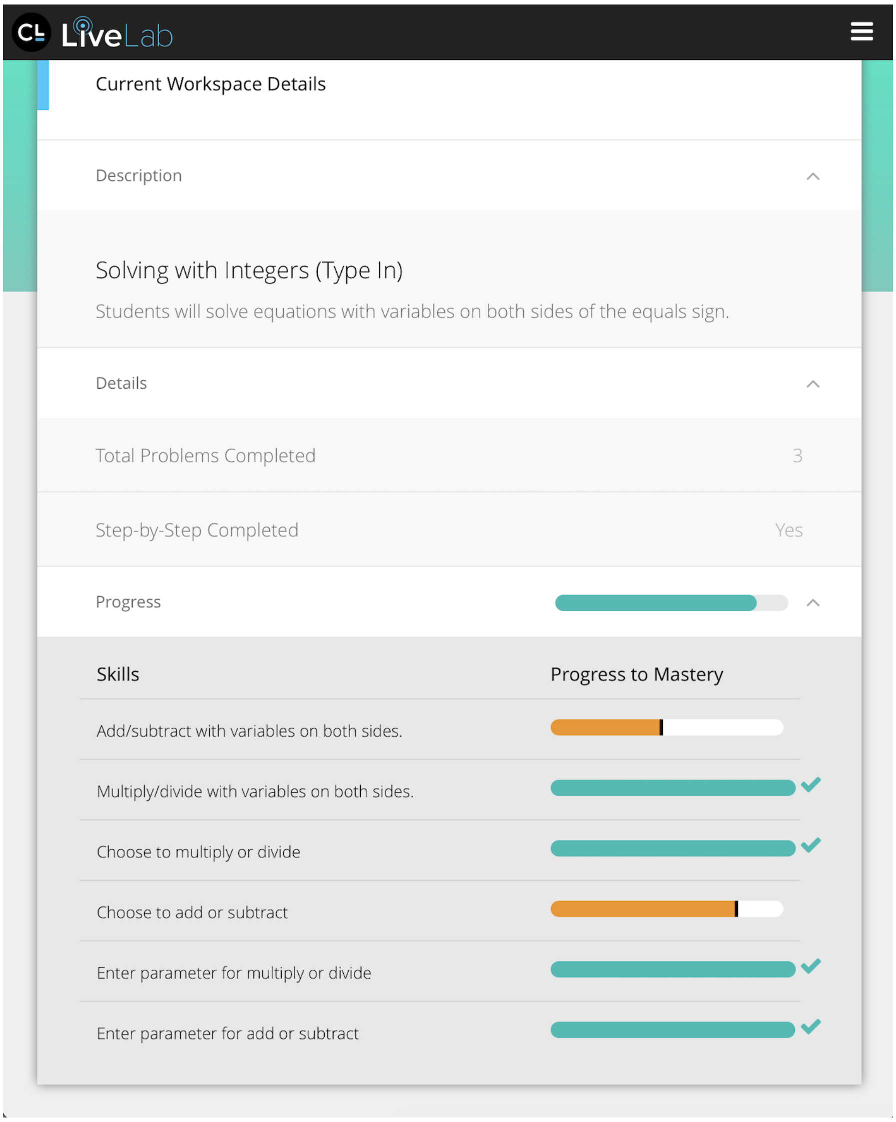

Figure 3: Screenshot of LiveLab's Workspace Details display for a particular student working in the "Solving with Integers (Type In)" workspace, including the current state of the student's "skillometer" in this workspace.

preserver icon displays on the student's row in the LiveLab class display. Such an indicator may be especially helpful in guiding an instructor's (or tutor's) attention to a student for additional support.

Content Details or "Deep Dive" To get insight into the specific content on which students are working and their progress to KC mastery, LiveLab users can select a student from the list to learn more about their current work within MATHia (see Figure 3). Details provided include the student's current workspace, a description of the workspace, the number of problems the student has completed in the workspace, whether the student has completed the step-by-step guided example provided by MATHia, as well as the current state of the student's "skillometer," which provides a visualization of BKT's current estimate of the student's KC mastery for each KC. A check-mark indicates that the student has reached the 0.95 probability threshold for KC knowledge or mastery. If the student has yet to complete the step-by-step example, this provides a natural suggestion for the instructor to provide a quick bit of assistance to the student, while considering the student's skillometer could provide more insight to have a more in-depth conversation.

On-going and future work aims to better support LiveLab users with insights about particular KCs. For example, recent work suggests we might label some KCs as especially reading-intensive [11]. We can also do more with insights from the unproductive struggle detector; some KCs that indicate likely non-mastery are not the most difficult KCs in the workspace. We call such KCs "indicator skills" because they may indicate unfinished prior learning on pre-requisite skills. We contract such "indicator" KCs with "critical" KCs, which are the KCs that are most likely to be mastered last (or never mastered) in a particular workspace, likely the most difficult KCs. Better understanding semi-or fully-automated insights we can deliver to instructors and tutors about whether and how students need support on the current topic or unfinished prior learning is an important area for future work.

3. Needs for Classroom vs. Tutoring Orchestration

LiveLab was intended to be used by a classroom instructor in the physical presence of a group of students synchronously using MATHia. However, it has been used in a variety of other contexts. For example, LiveLab was used by instructors after the shift to remote instruction made necessary by the global pandemic. Lawrence et al. [12] interviewed instructors who used LiveLab for remote instruction during the pandemic to better understand this transition. Limitations from the remote instruction use-case may inform new LiveLab features useful for the traditional classroom as well as for tutoring orchestration, including *tutor-initiated tutoring*. For example, remote instruction users of LiveLab were concerned that they were not able to identify and *immediately* correct students' perceived "problems," or conversely deliver praise and support, as they felt they could in-person. Features related to using data to initiate a tutoring session remotely with a MATHia user could facilitate more immediate interaction between instructors and their students in the context of remote instruction.

On-going piloting investigates *tutor-initiated tutoring*, using LiveLab (and a prototype app similar to LiveLab) as a way to identify learners with whom a remote, video-conferencing based tutoring session might be started as they work in a classroom. Such tutoring sessions could provide additional support to an in-classroom instructor while they work with other students in a class or be used to facilitate support in an after-school tutoring context. Pilot sessions to date have generally had tutoring facilitators physically on-site to help initiate the remote-tutoring session, linking a user of MATHia to a remote tutor in a video conference session. Preliminary feedback from use of LiveLab in these contexts echoes feedback provided by remote instructors, including, for example, the desire for step-by-step "replay" of student work in MATHia. This pilot tutoring work combined with feedback from remote and in-person instructors using LiveLab informs the road map for new feature development to better facilitate instruction and tutoring with LiveLab.

4. A Road Map for Orchestration Research & Development

We lay out a partial road map for future research and development work to support classroom and tutoring orchestration. We consider data-driven predictive models (or so-called "detector") that might help identify both who to support and how instructors/tutors might support them. We also briefly consider other MATHia and LiveLab software features that may help to facilitate new ways of orchestrating instruction and tutoring.

4.1. Detector Models

Behavior and Affective States Development of sensor-free, unobtrusive prediction models using ITS process data (from which the term "detectors" arises) has been an active area of research for at least 20 years. Baker et al. [13] describe models that detect that students are engaged in "gaming the system" or off-task behavior, and subsequent work extended this approach to affective states like boredom, confusion, frustration, and engaged concentration, including within the Cognitive Tutor ITS (now MATHia) [14, 15]. More recent work has learned such detector models for gaming the system on recent versions of MATHia [16], and we intend to build refreshed detector models of affective states in MATHia from data from the 2022-23 school year in the near future. To the extent that these models prove to be high quality predictors of their target constructs, these detectors are likely to prove valuable to instructors and tutors.

Celebration Opportunities LiveLab currently provides notifications to teachers when students complete MATHia workspaces with mastery of all KCs. However, there are a variety of more nuanced events that might serve as points of celebration with students. Lawrence et al. [12] pointed to teachers' desire to be able to tell students that they are "moving in the right direction." Fancsali et al. [17] noted relatively frequent examples of chat dialogues with human tutors in which students appeared to already have a correct solution to the problem on which they sought tutoring, but these students appeared to desire "affirmation" and encouragement from their chat-based tutor interlocutor. Detector models might be extended to predict local success within a MATHia workspace or that students have been in a state of "engaged concentration" for an extended period of time [14]. Developing more "asset-oriented" indicators or detectors, including such opportunities to celebrate student learning, is an important area for future work.

Learners with Reading Difficulties Recent efforts to predict learner outcomes on end-of-year English language arts/reading standardized tests using process data from MATHia activities suggest at least two potential detector models to indicate that learners may be having reading difficulties that ought to be addressed by instructors or tutors [11]. First, at a student-level and using only data from the first, introductory MATHia workspace, ensembled predictions of neural network models can provide accurate predictions that students are likely to have reading difficulties. This suggests the possibility of a student-level detector that could inform an instructor or tutor that a student is likely to experience reading difficulties. Second, analyses suggest that particular workspaces and KCs within workspaces may be especially difficult for learners with reading difficulties. Evidence from a recent randomized trial with 10,000+ learners, reported at the present conference [18], indicates that revising word problem content in two particular MATHia workspaces with an emphasis on readability generally led to improvements in the rate at which students master all KCs in target workspaces and that the students did so in less time. The effect was more pronounced on students inferred to have reading difficulties by the aforementioned detectors [11]. Additionally, KC-level insights, similar to those that drive the unproductive struggle detector, may provide more nuanced information to instructors and tutors. Such a detector could indicate that help on a particular KC (or perhaps particular vocabulary associated with that KC) might be the best way to provide just-in-time support to a learner.

4.2. LiveLab/MATHia Software Features

Notifications delivered via LiveLab are presently ephemeral in the sense that an instructor (or tutor) must be logged in to LiveLab to see notifications in real-time. We are exploring both the expansion of events covered by LiveLab notifications as well as more persistent delivery of such notifications (e.g., via an accumulating "feed" of events that would appear on an instructor's next login). Additionally, similar notifications could also be delivered via other modalities like email or another analytics or reporting dashboard.

Other features of LiveLab, especially those linking action(s) in LiveLab to the learner's experience in MATHia are also being explored and prototyped. These include features for data-driven or adaptive student grouping (perhaps to support more efficient and effective small-group, *scheduled tutoring*) and features that could help drive the student's attention, when necessary, toward their instructor or tutor. Such features could deliver the ability for a LiveLab instructor/tutor to "pause" MATHia interaction or to request a student's attention to join a video conferencing session for remote tutoring. The latter feature may prove especially important to realize fully remote, tutor-initiated and data-driven tutoring sessions.

Students could be given the ability in MATHia to virtually raise their hand so that an indicator would appear in LiveLab, with the caveat that some students could make excessive use of such a feature [17] without carefully considering some "guardrails." Problem "replay" of student action-level data within a math problem, viewable within LiveLab, anecdotally remains one of the most requested features across instructional and tutoring modalities. New content development efforts could provide examples to instructors and tutors that are helpful within (ideally detected) scenarios that arise (e.g., specific vocabulary support or an appropriate worked-example problem when a learner with reading difficulties is confused while working in a particular topic).

5. Conclusion

Our primary goal in the present work has been to introduce the reader to LiveLab as a classroom orchestration tool for instructors of students using Carnegie Learning's MATHia adaptive learning software. While presenting LiveLab, we've aimed to sketch out a path toward tutoring orchestration: using AI tools to augment the work of tutors in addition to more traditional classroom instructors. We've described different ways in which such tutoring is currently initiated, either by the student or by a scheduling process, as well as described the innovative goal of using data-driven, AI-augmented techniques to deliver tutoring experiences to students that are driven by real-time need for such support.

Finally, learning analytics and learning engineering researchers and educational technology providers might consider broadening the scope of stakeholders to whom we provide AI-driven assistance, supports, and/or orchestration tools as well as the types of information provided by such tools. Consequently, rather than discuss classroom orchestration versus tutoring orchestration, we might consider a notion of overall "learning orchestration." In addition to the types of data-driven insights we consider in the present work, tools for learning orchestration might consider factors like upcoming classroom topics, the availability of (possibly remote) tutors, or the extent to which groups of students require similar types of assistance with a broader audience in mind. Stakeholders for tools like this might include instructors, tutors, administrators, or caregivers, in either classroom or remote instruction and support scenarios. We look forward to continuing to investigate and develop new software features to drive improved learning for students across the gamut of potential learning experiences that include MATHia and the varied groups of stakeholders to whom we may provide valuable insights.[2]

6. Acknowledgments

The authors gratefully acknowledge their colleagues Sammy Andre, Audrea Bankston, Emily Bartko, Susan R. Berman, Courtney Lewis, Kaleb Mathieu, and Kole Norberg, as well as collaborators on preliminary piloting of *tutor-initiated* or data-driven tutoring, whose ideas about this type of tutoring and feedback informs this work: Vincent Aleven, Lee Branstetter, Danielle Chine, Erin Gatz, Shiv Gupta, and Ken Koedinger at Carnegie Mellon University and Emma Brunskill at Stanford University.

References

[1] S. Ritter, J. R. Anderson, K. R. Koedinger, A. Corbett, Cognitive tutor: Applied research in mathematics education, Psychonomic bulletin & review 14 (2007) 249–255.

[2] K. Holstein, B. M. McLaren, V. Aleven, Student learning benefits of a mixed-reality teacher awareness tool in ai-enhanced classrooms, in: Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I 19, Springer, 2018, pp. 154–168.

[3] V. Aleven, J. Blankestijn, L. Lawrence, T. Nagashima, N. Taatgen, A dashboard to support teachers during students' self-paced ai-supported problem-solving practice, in: European Conference on Technology Enhanced Learning, Springer, 2022, pp. 16–30.

[4] P. Dillenbourg, P. Jermann, Technology for classroom orchestration, New science of learning: Cognition, computers and collaboration in education (2010) 525–552.

[5] K. Holstein, B. M. McLaren, V. Aleven, Co-designing a real-time classroom orchestration tool to support teacher-ai complementarity., Grantee Submission (2019).

[2]The authors thank an anonymous reviewer for suggestions related to the idea of broader "learning orchestration."

[6] R. G. Fryer Jr, The production of human capital in developed countries: Evidence from 196 randomized field experiments, in: Handbook of economic field experiments, volume 2, Elsevier, 2017, pp. 95–322.
[7] K. R. Koedinger, A. T. Corbett, C. Perfetti, The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning, Cognitive science 36 (2012) 757–798.
[8] A. T. Corbett, J. R. Anderson, Knowledge tracing: Modeling the acquisition of procedural knowledge, User modeling and user-adapted interaction 4 (1994) 253–278.
[9] J. E. Beck, Y. Gong, Wheel-spinning: Students who fail to master a skill, in: Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16, Springer, 2013, pp. 431–440.
[10] S. E. Fancsali, K. Holstein, M. Sandbothe, S. Ritter, B. M. McLaren, V. Aleven, Towards practical detection of unproductive struggle, in: Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21, Springer, 2020, pp. 92–97.
[11] H. Almoubayyed, S. E. Fancsali, S. Ritter, Instruction-embedded assessment for reading ability in adaptive mathematics software, in: LAK23: 13th International Learning Analytics and Knowledge Conference, 2023, pp. 366–377.
[12] L. Lawrence, K. Holstein, S. R. Berman, S. Fancsali, B. M. McLaren, S. Ritter, V. Aleven, Teachers' orchestration needs during the shift to remote learning, in: Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings 16, Springer, 2021, pp. 347–351.
[13] R. S. Baker, A. T. Corbett, K. R. Koedinger, A. Z. Wagner, Off-task behavior in the cognitive tutor classroom: When students" game the system", in: Proceedings of the SIGCHI conference on Human factors in computing systems, 2004, pp. 383–390.
[14] R. S. d Baker, S. M. Gowda, M. Wixon, J. Kalka, A. Z. Wagner, A. Salvi, V. Aleven, G. W. Kusbit, J. Ocumpaugh, L. Rossi, Towards sensor-free affect detection in cognitive tutor algebra., in: Proceedings of the 5th International Conference on Educational Data Mining, International Educational Data Mining Society (2012), EDM 2012, 2012, pp. 126–133.
[15] S. Fancsali, Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra, in: Proceedings of the 7th International Conference on Educational Data Mining, International Educational Data Mining Society (2014), EDM 2014, 2014, pp. 28–35.
[16] N. Levin, R. Baker, N. Nasiar, F. Stephen, S. Hutt, Evaluating gaming detector model robustness over time, in: Proceedings of the 15th International Conference on Educational Data Mining, International Educational Data Mining Society, 2022.
[17] S. E. Fancsali, M. V. Yudelson, S. R. Berman, S. Ritter, Intelligent instructional hand offs, in: Proceedings of the 11th International Conference on Educational Data Mining, International Educational Data Mining Society (2018), EDM 2018, 2018, pp. 198–207.
[18] H. Almoubayyed, R. Bastoni, S. R. Berman, S. Galasso, M. Jensen, L. Lester, A. Murphy, M. Swartz, K. Weldon, S. E. Fancsali, et al., Rewriting math word problems to improve learning outcomes for emerging readers: A randomized field trial in carnegie learning's mathia, in: International Conference on Artificial Intelligence in Education, Springer, 2023, pp. 200–205.