Virtual Agents for Real Teachers:
Applying AI to Support Professional Development of Proportional Reasoning

Benjamin D. Nye[1], Aaron Shiel[1], Ibrahim Burak Olmez[2], Anirudh Mittal[1]
Jason Latta[1], Daniel Auerbach[1], Yasemin Copur-Gencturk[2]

[1]Institute for Creative Technologies, University of Southern California
[2]Rossier School of Education, University of Southern California
nye@ict.usc.edu

Abstract

Despite the critical role of teachers in the educational process, few advanced learning technologies have been developed to support teacher-instruction or professional development. This lack of support is particularly acute for middle school math teachers, where only 37% felt well prepared to scaffold instruction to address the needs of diverse students in a national sample. To address this gap, the Advancing Middle School Teachers' Understanding of Proportional Reasoning project is researching techniques to apply pedagogical virtual agents and dialog-based tutoring to enhance teachers' content knowledge and pedagogical content knowledge. This paper describes the design of a conversational, agent-based intelligent tutoring system to support teachers' professional development. Pedagogical strategies are presented that leverage a virtual human facilitator to tutor pedagogical content knowledge (how to teach proportions to students), as opposed to content knowledge (understanding proportions). The roles for different virtual facilitator capabilities are presented, including embedding actions into virtual agent dialog, open-response versus choice-based tutoring, ungraded pop-up sub-activities (e.g. whiteboard, calculator, note-taking). Usability feedback for a small cohort of instructors pursuing graduate studies was collected. In this feedback, teachers rated the system ease of use and perceived usefulness moderately well, but also reported confusion about what to expect from the system in terms of flow between lessons and support by the facilitator.

A key strategy for addressing students' poor mathematics achievement is to improve professional learning opportunities for middle school teachers. Research indicates that teachers' content and pedagogical content knowledge could be improved through professional development (Copur-Gencturk, Plowman, and Bai 2019) and that targeting both content knowledge and pedagogical content knowledge is more successful than either of them in isolation (Scher and O'Reilly 2009). Furthermore, gains in teachers' content and pedagogical content knowledge have been linked to improvements in mathematics instruction that are associated with students' mathematics achievement (Blazar 2015). Thus, professional development (PD) designed to enhance both content and pedagogical content knowledge has the po-

tential to improve the quality of mathematics instruction, which will improve students' mathematics understanding. To research this issue, the Advancing Teachers Understanding of Proportional Reasoning (ATProportion) project is developing and testing a personalized, online professional development program for middle school teachers.

As adult learners and as experts in education, teachers may react differently to virtual facilitator pedagogy which means that usability factors are still being established. In this paper, we describe the design and preliminary formative testing of the ATProportion system for professional development of proportional reasoning. When presenting the design, we focus on two aspects. First, iterating on the virtual facilitator pedagogy has been a central effort of the project, building capabilities that can be composed into distinct pedagogical patterns that could generalize to other domains. Second, developing content has also been a core focus, because math educational experts on our team require a high degree of control over content to make effective professional development training. As such, the content development process has been refined to help math experts directly author content wherever possible. Next, we present data from the formative testing. As this work is over a small sample and teachers are a diverse population, the results are not conclusive but they identify areas to explore further in the future.

Background and Rationale

The scalability of successful online professional development is constrained by the availability of quality interaction between users and facilitators (Pianta et al. 2008). These constraints can result in inadequate feedback and shallow discussion. In this work, we investigate virtual agents as a solution to overcome these problems. To support online professional development, a virtual facilitator must play both a pedagogical and a motivational role.

Research indicates that virtual agents have been effective in both roles in various contexts (Schroeder, Adesope, and Gilbert 2013), but to our knowledge it has not yet been assessed in the context of a computer-based professional development program. This gap is important, because teachers are a special category of learner for two reasons. First, teachers must learn master both domain knowledge (the skills themselves) and also pedagogical domain knowledge (how to teach the skills, how different kinds of students under-

stand a topic, etc.). Second, in-service teachers are adult professional learners, which is a stage that is overall not well-studied by research on virtual agents or tutoring systems.

Research on virtual agents to support learning, and mathematics learning in particular, have shown a number of key advantages over a more traditional "faceless" online system. First, virtual facilitators have been used to optimize learning from examples (Atkinson 2002), and professional development is taught through examples or case studies across many domains. Second, multimedia agents who communicate with learners via speech can increase learning gains (Schroeder, Adesope, and Gilbert 2013). This is relevant because high-quality mathematics instruction often talks through multiple steps of worked example (e.g., solving a word problem). Third, virtual agents can provide continuity and personalized support for learners as they learn from many different objects of joint attention, such as concept maps, equations, videos, and figures (Nye et al. 2018). An agent to provide continuity is especially relevant for mathematics instruction, where deep conceptual understanding often requires teaching multiple representations that may initially appear unrelated.

This work also leverages virtual agents to deliver a dialog-based intelligent tutoring system (ITS). Open-response dialog-based ITS such as AutoTutor have shown learning gains on the order of 0.8σ (Nye, Graesser, and Hu 2014). While such ITS content is more time-consuming to develop, these resources have been particularly effective for learning deep, conceptual knowledge (Graesser, Lippert, and Hampton 2017). With that said, open-ended tutoring dialogs are not easily paired with procedural problem solving, due to the conversation and the task potentially having a different state, so dialog-based ITS for math often focus on conceptual knowledge or break the problem into multiple stages (Nye et al. 2018).

Research on the impact of virtual agents on engagement and motivation is also relatively strong, but requires further investigation for teachers as learners. Overall, disengagement is often a problem for online learning (Feild et al. 2018) and virtual agents have shown the ability to increase motivation in computer-based learning (Sträfling et al. 2010). However, research has not studied how teachers specifically react to pedagogical agents and research with other adult professionals limited. For example, research on military training has shown some evidence for strong engagement during agent-based scenarios, but without a comparable traditional online learning control condition (Lane et al. 2013). Research on agents to support long-term engagement in medical interventions has shown benefits and give some evidence that agents engage adults, but the agents were not tutors (Bickmore, Schulman, and Yin 2010). One of the most popular voluntary learning apps (DuoLingo) employs an animated pedagogical agent, but the agent is only one of many engagement mechanisms in the app. Perhaps most important, learners' reactions to agents are dependent on the type of agent (Baylor 2011) and their reactions to agents can impact learning outcomes (Schroeder et al. 2018). As such, the relevant question may not be if teachers can be motivated by a pedagogical agent but what kind of agent appearance and role is appropriate to engage mathematics teachers.

System Design: ATProportion

Based on this background work, the number of design criteria were identified as being essential for the success of the ATProportion professional development for proportional reasoning:

1. Virtual Facilitator: A virtual agent, whose interactions and appearance must be engaging and accepted by teachers.

2. Multi-Stage Examples: Walking a learner through multiple stages of a problem or concept, such as different steps, comparing examples, or multiple representations.

3. Conceptual Tutoring: Guiding a learner to discuss and explain their understanding of proportional reasoning.

4. Virtual Manipulatives: Multiple categories for objects of joint attention, such as images, videos, note-taking, whiteboards, and interactive objects.

These elements represent the core capabilities of the system, which are the lessons that the teachers learn from. A number of other capabilities support this functionality, which will be described briefly. The system is structured in terms of two modules: Content Knowledge and Pedagogical Content Knowledge. Inside each module is a list of submodules, which train specific skills as shown in Figure 1. Each submodule contains a finite set of lessons, where completed lessons show a traffic light for their completion status and attemped lessons have a notepad link to review and edit any note-taking from the user. Lessons are adaptively recommended to the learner through a simple algorithm: each lessons is completed once in-order and then lessons with scores below passing are recommended up to once more if the submodule performance is below passing. While submodules may be completed out of order, the system will always recommend the next uncompleted submodule starting at the top. This simple recommendation approach is intended to help learners resume progress quickly, revisit activities for modules they struggled with, but also prevent repeating content too many times (which would be uncharacteristic of professional development).

As shown in Figure 2, each lesson is displayed with three panels: the virtual facilitator on the upper-right, the chat window on the middle/bottom right, and the content panel on the left side. The virtual facilitator agent speaks question prompts and may have limited gestures/facial responses. The chat area shows a transcript with text equivalents for each statement from the agent and it records each input by the user. It also has a glossary functionality, with automatically embeds lookup links for key definitions (e.g., "ratio table"). At the bottom of the chat area is an input box, which is enabled for open-response answers but is disabled if answers are input by other means (e.g., multiple choice, tables). The content panel shows the primary question prompt at the top, an object of joint attention in the middle (e.g., some media), and any structured input below that (choices, table). In the example given, multiple-choice options are displayed.
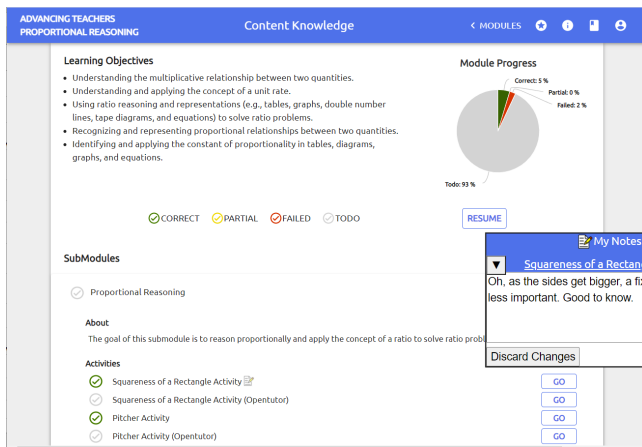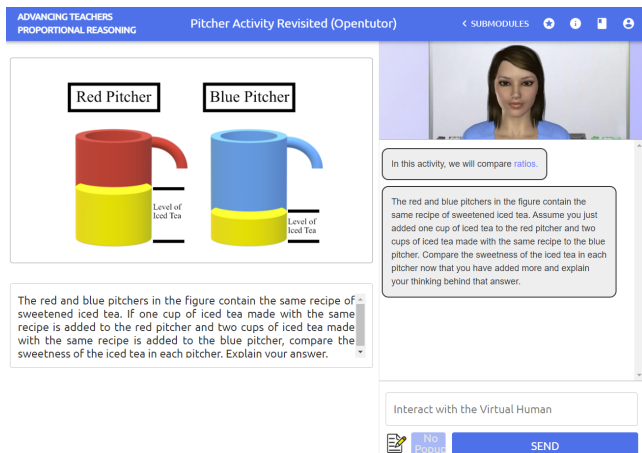
Figure 1: Submodules and Lesson List

Figure 2: Lesson Activity User Interface

Virtual Facilitator

While the avatar of the virtual facilitator acts as a fairly standard animated virtual agent, focus was placed on its appearance and interaction style. Specifically, the agent provides redundant speech (text plus chat). While this is not typically ideal as per multimedia principles (Schroeder, Adesope, and Gilbert 2013), a transcript is essential to help users proceed after a lapse in attention (or lack of speakers) and a lack of agent voice would detract from engagement. As such, both were retained. For the agents' appearance, this work has been guided by research that indicates that while less-represented peoples are typically more positive toward someone of similar demographics, majority group members may be neutral toward demographic differences (Nye et al. 2020). Given that the expected population of math teachers is majority-female and ethnically diverse, this meant that all versions of the agent were female and, to the degree possible with available agent configurations, aimed for a degree of ethnic ambiguity, with somewhat tan skin and brown hair.

Multi-Stage Examples

Within each module, the system uses two types of lessons: multi-stage lessons and open-response tutoring. Multi-stage tutoring can present a number of activities in a sequence,

which include multiple choice, table completion, and open response tutoring (described later). Multiple choice and table completion work similarly to each other, where a user's input pattern (item for multiple choice, pattern of values for a table) is scored as correct or incorrect. For each pattern or option, the agent can provide a series of feedback messages and virtual manipulative actions. After the agent's response, the system may either allow the user to try another answer, proceed to the next stage (if correct), or end the current activity and proceed to a remediation/drill-down activity. As such, each stage has limited branching capability.

Conceptual Tutoring (Open Response)

The second type of activity leverages expectation-misconception tutoring, derived from AutoTutor (Graesser, Lippert, and Hampton 2017). In an AutoTutor dialog, the agent asks an open-ended question with a number of expectations (concepts the user is expected to state/explain). As the user answers, the dialog-based tutor analyzes their inputs to identify the parts of the answer they have already covered and to give hints as leading questions to help the user explain each of the remaining concepts. The structure and natural language processing for these dialogs is well-explained in other work, so it will not be covered in detail (Nye, Graesser, and Hu 2014). As with the multi-stage lessons, any response by the virtual facilitator may include a series of statements and actions.

Virtual Manipulatives

As noted, any response by the virtual facilitator can include a series of speech and virtual manipulative actions. These actions are critical to the pedagogy of the system, as they serve two roles. First, they enable just-in-time media or interactive objects to be presented that accompany the agent's dialog. Second, these injected actions offer a way to break up larger explanations by the tutor by engaging the learner in actions with the manipulatives.

The current set of manipulatives in the system fall into three categories: content changes, ungraded input, and just-in-time activities. Content changes affect the content panel, such as changing the displayed question prompt, changing which picture/video is shown, adding highlighted regions to a picture/video, or playing a segment of a video. Ungraded input means that the user can be prompted to speak or type an open-response answer (e.g., explain their reasoning) at any point, which is entered in the transcript but not given conditional feedback or graded. While these first two actions may seem mundane, they can be powerful when used together. For example, after a multiple choice wrong answer the agent can offer brief feedback, replay the relevant part of a video with a box to highlight the important elements, prompt the user to explain their thinking, and ask the user to try another option.

Just-in-time virtual manipulatives pop-over the current activity, so they can in principle be nearly any ungraded activity, including embedded views of third party websites. At present, the following activities can be triggered: whiteboard, note-taking, calculator, and interactive proportion visualizations (e.g., a rectangle that maintains constant a con-

stant difference between its sides as it is dragged larger). For a whiteboard virtual manipulative, a user is given a prompt to drawing shapes, simple graphs, or tables. Note-taking pops up a pad with the ability to take and review notes on any in-progress or completed activity (also available in the list of lessons). Notes are typically prompted during the ending summary of a problem. The calculator popup is self-explanatory, and used when prompting the user to extrapolate potential proportional relationships. Finally, specialized popups exist for specific problems which focus on dragging or comparing rectangles as proportional relationships.

Pedagogical Strategy Development

While each system capability in isolation is useful, the impact of the system relies on math education experts being able to coordinate these capabilities together. Based on collaborations between the ITS development team and education experts, this was supported through two techniques: pedagogy patterns and authoring capabilities.

Pedagogy Patterns

Based on the domain content and the system capabilities, a set of pedagogical patterns were developed as "recipes" for authors to consider when developing lessons. These are used inside lessons, where the virtual facilitator combines feedback with one or more virtual actions. The most common combined tactics used by the system were:

1. Brainstorm and Confirm: Before presenting multiple choice items, ask for the learner's open response guess and then prompt them to pick the most relevant choice. This was done to encourage self-explanation, to break up long introductions, and to collect open response answers.

2. Salience Cues: During a hint or feedback, use content change actions to draw attention to an important section (position or video segment).

3. Solve a Sub-problem: Prior to answering, or as feedback for an incorrect/incomplete answer, break out a smaller task to complete using a whiteboard or calculator.

4. Diagnostic Remediation: Present a brief initial problem choice, but if the user answers wrong then they must complete a more extensive remedial task (conceptual tutoring or multiple choice) to understand the concept.

5. Closing Summarization: For dialogs that were not conceptual tutoring, this tactic prompts the user with a notepad so they can record their understanding at the end of the problem. The notes are retained to review later.

Teacher Authoring

These pedagogical tactics were used extensively within the system, with about a third of responses to wrong answers involving a tactic that combined both feedback and virtual manipulatives. However, these tactics are fine-grained and vary by problem, so they needed to be controlled by expert math educators. In the first rounds of content development, spreadsheet templates were designed and filled by expert authors. However, this had the major drawback that

they could not immediately see the system change and test their changes in real time. As a result, the system content was moved to a headless (cloud-hosted) content management system (Contentful), where each author could log in and collaboratively edit and publish content.

Being able to directly edit the virtual agent and interactions, as opposed to working through programmers, appeared to have a significant impact on improving the AI system in two ways. First, the content team was able to review and align to current curriculum standards for the teachers using the PD. For example, the ATProportion lessons must consider the development of concepts across grade levels, which is directed through guiding questions in those activities. Directly editing these items mean the ability to refine the items based on educational principles such as asking questions starting from general to specific and injecting open-ended questions even during multiple choice stages. Second, by directly editing content, the content team did not limit their changes to the system's initial technical boundaries. This greater flexibly meant that the content team was able to request improvements that made new features possible and which were not obvious otherwise (e.g., options to hide multiple choice answers until after the facilitator's initial dialog and actions are complete). This process was also substantially more efficient in the long term, allowing quicker editing and review.

Formative Testing Results

Many of the features for this system have been developed due to a combination of internal testing and also a small sample (N=6) of formative testing results. While many of the capabilities presented were included in these results, a number of them were also developed due to the feedback from this formative test [1]. The participating math educators had prior experience as mathematics teachers, and were involved in graduate or postdoctoral studies in the graduate school of education but were not involved in the project.

Table 1 shows the mean and standard deviation of the usability study for different across all lessons and for categories of Lesson Types (ranked from 1=Not at All to 5=Very). As shown, Ease of Use items and the two Expected Benefit items were both moderate (around 4 out of 5). Across different lesson types, Conceptual Tutoring lessons were rated with a higher Expected Benefit but lower Ease of Use, while Single One-Shot Multiple Choice were rated as easiest to use but lower Expected Benefit. Ratings of Engagement with the system was fairly high (4.3) and did not vary by activity type. The three three final items were about the content quality and appropriateness (potential value) and were also fairly high (4.4, for each). However, as with the Expected Benefit, Conceptual Tutoring was rated slightly higher for the content quality and appropriateness. This indicates that the users felt the Conceptual Tutoring was considered valuable, despite problems during the interaction.

---

[1] Specifically, the initial testing session contained only 9 lessons (out of a current set of 46) and did not yet have note-taking, a calculator, a glossary, table-based problems, ungraded open response prompts, or the revised agent design.

Question	All Items (N=4)	Conceptual Tutoring (N=2)	Single MC (N=4)	Multi-Stage Example (N=4)
How clear and easy to use was this activity?	4.0 (0.8)	3.7 (1.1)	4.3 (0.7)	3.9 (0.8)
How helpful was this activity to understanding the concept?	4.1 (0.9)	4.3 (1.0)	3.9 (1.0)	4.2 (0.9)
How much will this help a teacher learn about proportions?	4.0 (0.8)	4.1 (0.8)	3.9 (0.9)	4.0 (0.7)
How engaging was the system?	4.3 (0.8)	4.3 (0.7)	4.3 (0.8)	4.3 (0.7)
How clear were the questions in the activity?	4.4 (0.8)	4.6 (0.4)	4.4 (0.8)	4.3 (0.9)
How important was the content in this activity?	4.4 (0.7)	4.5 (0.8)	4.4 (0.7)	4.5 (0.7)
How useful was the content in this activity?	4.4 (0.7)	4.4 (0.8)	4.4 (0.7)	4.4 (0.7)

Table 1: User Ratings for Lesson Types in AT Proportion

Talk-aloud data and open response items confirmed these intuitions and raised additional areas to improve in each of the four design criteria noted earlier.

1. Multi-Stage Examples: Positive feedback, with 3 users noting those as their favorite part of the system ("More activities like Plane Travel activity consisting of multiple parts/stages"), as they appreciated the broad range of media that was coordinated by the agent ("classroom video, teacher slides, drawings etc.").

2. Conceptual Tutoring: Positive about the ability to receive feedback ("liked that I typed the answer and received immediate feedback"), but did not like the facilitator sometimes told them they were wrong when they were right ("agent did not understand his answers").

3. Virtual Manipulatives: Positive about using varied activities (e.g., "All the activities with cool technology were engaging: strip diagram, dragging thing", but were negative about inability to move the popup and they also found it time-consuming to draw tables on a whiteboard.

4. Virtual Facilitator: Positive in terms of its overall interaction style ("pace of speech was good. Its comments were helpful") but feedback on its appearance was negative ("taking much space", "dressed a little like an FBI agent and not like a teacher").

Discussion: Applying Formative Insights

Overall, this first formative test showed promising results, but the results must be taken with a grain of salt because the users were moderately advanced (e.g., pursuing graduate studies). As such, their interactions are not necessarily representative of teachers overall. However, their feedback has been significantly influenced recent changes to the system. Since the Multi-Stage problems were highly regarded, more in that form were created, ultimately representing 24 out of 46 lessons developed. Future systems using animated virtual agents should consider these items, as the most basic versions can be developed with sequences of multiple choice items that contain feedback and where the feedback includes simple markup to start actions (e.g., "type : 'FREE_TEXT'"). These multi-stage lessons resemble a constrained example-tracing tutor (Heffernan and Heffernan 2014), due to their ability to interleave feedback messages, content changes, and sub-activities.

Likewise, the use of Virtual Manipulative actions were expanded significantly across problems, and new ones such as note-taking and voice open response were added. The usability of these activities were also improved, by adding functions to drag and hide/reopen just-in-time activities. Content authors also learned the simple markup to insert their own virtual manipulative action triggers, which further expanded their use. While the virtual agent is less interactive in this case, the continuity and ability to guide learners through multiple types of interactions was valued strongly by teachers, especially if those interactions showed them genuine examples of classroom artifacts and instruction.

The Virtual Faciliator was substantially revised based on feedback. Three aspects of the agent were optimized: the frame, the appearance, and the background. In terms of formatting, the first iterations used a standing, waist-up avatar to enable greater movement of the hands (e.g., gestures, pointing). However, this reduced the transcript size which users found more important than greater agent presence. As a result, the later iterations were cropped at the shoulders for a close-up view. The clothing and scene also changed significantly. Initial agent avatars had clothing that was too elaborate (multiple layers) or too austere (suits, black top).

To resolve this issue, a selection of videos from real life middle school math teachers were reviewed, to identify common clothing styles. This indicated that neutral-to-bright one-piece long sleeve tops appeared common, without layering, which replaced the more formal style. Finally, the background evolved over time to increasingly resemble a classroom. Early styles had a light green wall and a simple whiteboard. Based on feedback, it was identified that blue or white walls were most common, so the wall was updated to blue. The whiteboard was also notable as "too empty" so an eraser and markers were added. These small changes to normalize the scene were conducted to reduce distraction. As a takeaway for future projects, starting with a library of comparison videos for humans similar to the agent under development was an efficient technique to identify a more authentic look and feel, as compared to a more abstract discussion.

Finally, improvements to Conceptual Tutoring are ongoing, but the core problem is a common one for new dialog-based tutors (Nye, Graesser, and Hu 2014). When acting as learners, teachers (like most experts) are sensitive to the system treating them as incorrect when they are answering correctly. To address this, we have been investigating methods to apply supervised learning to improve dialog classification, but these efforts have not yet significantly improved dialog performance. Research is ongoing on techniques to improve dialog assessment quality. This is a key area, because teach-

ers found the dialogs to be valuable (even in their current state) but found them hard to use. These issues will likely be amplified in a broader population of teachers.

Conclusions and Next Steps

This work has produced a number of lessons-learned for developing a virtual facilitator for online professional development. First, at least among math educators, drilling down into multiple facets of an example was appealing: there was clear engagement in seeing a concept, analyzing how students reacted to it, and discussing real videos of classroom instruction and artifacts. One teacher reported that early in her career, she "did not get any feedback like that" until she received a special grant for teacher quality. Additionally, teachers appear motivated by considering real students. Second, despite being educators, they were not qualitatively more demanding about the pedagogy for Conceptual Tutoring. Their reactions were fairly typical for AutoTutor-style dialogs by knowledgeable learners, which were that they appreciated immediate feedback/guidance but were averse to having good answers rejected.

Further user testing is required to see how effectively the formative feedback has improved the system design. Substantial improvements were implemented in each of these areas since the formative testing, so a larger usability pilot should help establish how effectively these changes improve teachers' usability and expected benefit from the system.. Additionally, the ATProportion project has been developing assessments to measure pre-post learning gains, which should give a better understanding of the content submodules and associated item types that demonstrate greater learning gains. These studies are upcoming and represent the next data points for this work.

Acknowledgments

References

Atkinson, R. K. 2002. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology* 94(2):416–427.

Baylor, A. L. 2011. The design of motivational agents and avatars. *Educational Technology Research and Development* 59(2):291–300.

Bickmore, T.; Schulman, D.; and Yin, L. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24(6):648–666.

Blazar, D. 2015. Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review* 48:16–29.

Copur-Gencturk, Y.; Plowman, D.; and Bai, H. 2019. Mathematics teachers' learning: Identifying key learning opportunities linked to teachers' knowledge growth. *American Educational Research Journal* 56(5):1590–1628.

Feild, J.; Lewkow, N.; Burns, S.; and Gebhardt, K. 2018. A generalized classifier to identify online learning tool disengagement at scale. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 61–70.

Graesser, A. C.; Lippert, A. M.; and Hampton, A. J. 2017. Successes and failures in building learning environments to promote deep learning: The value of conversational agents. In *Informational Environments*. Springer. 273–298.

Heffernan, N. T., and Heffernan, C. L. 2014. The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education* 24(4):470–497.

Lane, H. C.; Hays, M. J.; Core, M. G.; and Auerbach, D. 2013. Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology* 105(4):1026–1035.

Nye, B. D.; Pavlik, P. I.; Windsor, A.; Olney, A. M.; Hajeer, M.; and Hu, X. 2018. SKOPE-IT (shareable knowledge objects as portable intelligent tutors): overlaying natural language tutoring on an adaptive learning system for mathematics. *International journal of STEM education* 5(12):1–20.

Nye, B. D.; Davis, D. M.; Rizvi, S. Z.; Carr, K.; Swartout, W.; Thacker, R.; and Shaw, K. 2020. Feasibility and usability of mentorpal, a framework for rapid development of virtual mentors. *Journal of Research on Technology in Education* 1–23.

Nye, B. D.; Graesser, A. C.; and Hu, X. 2014. Autotutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education* 24(4):427–469.

Pianta, R. C.; Mashburn, A. J.; Downer, J. T.; Hamre, B. K.; and Justice, L. 2008. Effects of web-mediated professional development resources on teacher–child interactions in prekindergarten classrooms. *Early childhood research quarterly* 23(4):431–451.

Scher, L., and O'Reilly, F. 2009. Professional development for k–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness* 2(3):209–249.

Schroeder, N. L.; Adesope, O. O.; and Gilbert, R. B. 2013. How effective are pedagogical agents for learning? a meta-analytic review. *Journal of Educational Computing Research* 49(1):1–39.

Schroeder, N. L.; Yang, F.; Banerjee, T.; Romine, W. L.; and Craig, S. D. 2018. The influence of learners' perceptions of virtual humans on learning transfer. *Computers & Education* 126:170–182.

Sträfling, N.; Fleischer, I.; Polzer, C.; Leutner, D.; and Krämer, N. C. 2010. Teaching learning strategies with a pedagogical agent. *Journal of Media Psychology*.