Co-Designing an AI Tool to Support Discourse Based Math Instruction in a High-Dosage Tutoring Context

Chelsea Brown, Jeffrey B. Bush
chelsea.brown-1@colorado.edu, jeffrey.bush@colorado.edu
University of Colorado Boulder

Abstract: TalkMoves is an AI assistive tool that provides automated feedback to educators to support their daily teaching practices. While originally designed for classroom math teachers, this tool can be useful in a broader context. The University of Colorado Boulder and Saga Education formed a co-design team tasked with re-contextualizing TalkMoves for coaches of novice math tutors to use in their ongoing professional development. To effectively adapt an existing technology to a new problem space, the co-design team iteratively exchanged ideas of what exactly TalkMoves could achieve, as well as the specific needs of the coaches. Facilitators used strategies such as communal orientation, expansive dreaming, backcasting, and revoicing to promote productive collaboration. Three main goals emerged: maximize opportunities for user agency, center design around goal setting, and integrate the tool into the existing workflow. Any adaptation of an AI tool would benefit from this approach.

Introduction and Background

AI assistive tools have the potential to make a dramatic impact in education, especially in areas that are chronically under-resourced. Many students have been negatively impacted due COVID-19 and issues of inequity have been compounded; one proven way to address learning loss is through incorporating frequent tutoring (Dorn et al., 2020). Saga Education (Saga) has developed a unique and scalable approach to this challenge by recruiting novice tutors and pairing them with experienced coaches who provide ongoing professional learning (PL).Coach feedback is critical in ensuring that tutors learn high leverage and equitable pedagogy, however, finding time to observe multiple tutors and sessions is both challenging and tedious. TalkMoves is an AI feedback tool that has shown success helping classroom teachers in their daily practice (Jacobs et al., 2022), and could potentially assist coaches by augmenting their ability to provide high quality and evidence-based feedback to tutors about their discourse based tutoring strategies. The current study focuses on the unique integration of TalkMoves into the existing workflow of Saga coaches by exploring three research questions. *RQ1:*What themes arise that explain how the co-design team of practitioners and researchers aligned their goals and priorities to re-contextualize the design of TalkMoves for a high-dosage tutoring context? *RQ2:*What generalizable facilitation strategies were used to enhance productive collaboration in the design process? *RQ3:*What design themes and goals emerged within and across co-design sessions that created synergy between the existing TalkMoves application and the existing Saga workflow?

To address these questions, we drew on many theories. *Co-design*, an established method drawing on sociocultural theory, incorporates expertise from all involved parties and intentionally structures projects to address learning within the context of motivation, identity, belonging, intertwined with cognitive development (Penuel et al., 2007). Collaborative efforts enhance applicability, increase adoption rates, and foster agency through successful partnerships. To adapt an existing tool to growing needs, we borrow the concept of *solutions mapping* from the field of engineering (Lee et al., 2020). Starting with an existing technology (TalkMoves) and finding a novel application (remote tutoring), the technology and the problem continually inform each other resulting in novel adaptations. TalkMoves was first a browser-based tool that took recordings of classroom lessons and used computational modeling to process the dialogue for talk moves (utterance level categorizations of teacher and student speech surrounding math content) and discourse features (e.g. talk ratio: time teacher talks vs. students, wait time: how long until a response, and quantity of math terms used). Talk moves are based on *accountable talk theory* which states that deliberate and intentional forms of communication can lead to more equitable participation in academic conversations (Michaels et al, 2008). A pilot study with classroom math teachers found that highlighting when and where talk moves were present (or not) during a lesson was beneficial and informative for the teachers (Jacobs et al., 2022; Suresh, in press).

Methodology

The co-design team, made up of CU and Saga researchers as well as coaches and coach supervisors, met eight times remotely over four months to discuss the design of the interface for coaches and supervisors, which

produced a set of wireframes (interactive depictions) that captured many of the discussion points. Additional sources include collaborative digital whiteboard spaces, video memos for reflection, and one-on-one interviews of Saga coaches and supervisors. We used multiple strategies to assess the data to ensure triangulation of the emergent features and concepts (Saldaña, 2016). Analytic memos were written for each meeting, capturing the main points and trends. Videos and whiteboards were reviewed to collect all comments that pertained to what participants wanted to incorporate into TalkMoves. Comments were then iteratively reviewed and open coded (Corbin & Strauss, 2015) for similarity and initial groupings. Once consistent categories emerged, comments were inductively coded (Saldaña, 2016). Comparing codes to initial trends in the analytic memos allowed for a comprehensive understanding of the overall themes being captured from a fine grained to broader analysis. A mico-vignette (Barter & Renold, 1999) was created to showcase multiple themes, as well as facilitator strategies. Reviewing these sources of data establishes how researchers and Saga members reached a mutual understanding of how TalkMoves could be applied in a new context and provides insight into the specific features and facilitator moves that brought the main suggestions out.

Findings

RQ1: Creating a shared vision to re-contextualize TalkMoves: Through co-adaptation and reflection about TalkMoves, we see how CU and Saga align priorities and create a shared vision of how the new platform could be mutually beneficial. We know from solution mapping that integrating a solution into a new problem space is an iterative exchange. Our analysis indicates four themes where researchers and Saga converge to align their goals and priorities to re-contextualize the design of TalkMoves in a high dosage tutoring context. First, the co-design team had to identify the user for whom TalkMoves would assist the most. Next, the CU researchers had to comprehend the specific context in which TalkMoves would be used. Simultaneously, Saga coaches had to learn what TalkMoves provided for classroom teachers and reconcile how it would apply in a tutoring setting. Finally, by incorporating the new design of TalkMoves, Saga coaches needed to imagine how their practices would change.

Central to creating a shared vision for the platform was identifying who the platform was being built for and why. First and foremost, CU and Saga researchers had to identify that the Saga coaches were best situated to use the feedback from TalkMoves, instead of tutors as originally thought. Tutors had neither time for co-design or PL about interpreting results, but this tool could be extremely helpful to coaches. Therefore, the next stage focused on understanding both the normal workflow of the coaches as well as the challenges TalkMoves could assist with. Each new layer of understanding informed how the application could best be adapted. The full picture of the coaches workflow was refined across every co-design session. If the co-design team had only met once or twice researchers would not know valuable details such as the challenge coaches faced with taking notes, that coaches were often the only adult in the physical classroom, or that supervisors did not have a standardized way of tracking whether a goal was achieved or not. All these details emerged in later sessions, and all informed the new application. Through trying to understand coaches' process and understanding what AI can achieve, new areas for the app to support were uncovered. Concurrently, the team focused on the tension between different contexts; what is needed to support tutoring versus classroom instruction. Coaches and classroom teachers both need information about what is occurring within a lesson; however, this looks very different across users and settings. In the original form of TalkMoves, teachers would upload recorded sessions themselves and different graphs for teacher talk moves, student talk moves, and other summaries were generated. It was left to the teacher to interpret what was useful. In the re-contextualized app, sessions are automatically uploaded. Coaches can navigate across each of their tutors, set the data presented to be grouped by class or date, and look at trends across specific talk moves.

Saga also adapted their approach to PL based on lessons learned from the co-design process. In addition to exploring TalkMoves and all the information it could provide, Saga also had to reconcile how *accountable talk theory* that guided TalkMoves' development was in line with institutional commitments. This was a delicate balance of both learning as the sessions unfolded and imagining what would be most beneficial in their practices for the future. By incorporating TalkMoves into their daily practice, this also meant that the coaches' practices for supporting tutors could be positively disrupted by this collaboration. An example of this can be seen in how coaches originally had specific details that they look for in the tutor's sessions that pertained to relationship building, rigor and mathematical knowledge. The TalkMoves application provides a new frame of reference to look for these types of details. One of the coaches stated in their final one-on-one interview that they were still looking for all of the correlations between talk moves and Saga frameworks and that they've had to step back and think a lot about how they identify things. So, while the inclusion of the app makes identification faster, it also requires the coaches to learn a new perspective. Situating TalkMoves into a new

context required Saga to understand what TalkMoves could offer as well as its future capabilities, and for CU researchers to understand what the critical issues and challenges Saga coaches experienced in their workflow.

RQ2: Facilitator strategies: Facilitators guided productive collaboration by establishing communal orientation, promoting expansive dreaming and backcasting, and continually revoicing collaborator ideas. While co-design is a mutually enriching experience, getting multiple stakeholders involved is only one piece of the puzzle; well-directed meetings is another vital component (Penuel et al., 2007; Hoadley, 2002). Facilitators ensured productive meetings by providing unique perspectives and implementing critical scaffolds. Two facilitators were boundary spanners (Akkerman & Bakker, 2011); one researcher had been a math teacher, and former employee of a startup that helped design Saga's digital math environment. The Saga facilitator was originally a Saga tutor and coach. Their backgrounds afforded them perspectives from practitioners and developers, increasing the boundary permeability of the design work to create a tool that meets the needs of both. The facilitator strategies provided intentional structure that encouraged new ideas to grow and build on each other. Each session was initiated by being framed around a group goal, creating communal orientation, and ensuring that conversations addressed how to assist coaches in their workflow. This approach promoted the acceptance of TalkMoves; if the focus was limited instead to only 'this is the tool, how would you change your work to use it', this would have stunted the ideas produced and would not have been as well received.

Early sessions incorporated expansive dreaming, while later sessions attempted to backcast those goals onto the technical and learning environment constraints of the tool (Quist & Vergragt, 2006). For example, before viewing TalkMoves, participants watched a five-minute video of classroom instruction and were asked to reflect on the different observations or data that they might expect to use (e.g., what might they expect to see for talk distribution between teacher and students). It was only after everyone had their own expectation set that the data from TalkMoves was presented. Having everyone imagine their typical practice first ensured that they could clearly compare their expectations to that of the app and imagine new uses. Facilitators continually tried to elicit ideas independent from the current app function so that participants would not limit their ideas. Additionally, reflection time, free interaction, and visual aids were used both during and outside of the sessions. Coaches and supervisors created a short reflection recording, had built in time during sessions to explore the app, and visual aids were always used as a jumping off point. Initial TalkMoves' graphs progressed to wireframes that displayed the suggestions produced from the meetings. These detailed examples helped to solidify what and how the coaches and supervisors imagined using the application in their typical workflow. Finally, facilitators were very intentional about how they engaged with collaborators. They frequently revoiced the suggestions or comments, which highlighted the value of everyone's contributions. Revoicing is also a talk move that has been identified as an effective way to promote accountable talk (Michaels et al., 2008), and was a common occurrence which often solidified ideas or uncovered new topics. Multiple elements worked together to nudge everyone to generate refined ideas of what was needed and how exactly TalkMoves could best work.

RQ3: Emergence of critical design themes and overarching goals: By examining all of the comments across sessions, 15 themes and 3 main design goals emerged. Goal one: the interface should enable user agency, and is supported by themes such as Filter (ability to adjust presentation of data through selection, sorting or comparison), Exemplars (specific high-quality examples), Sharing (how information is shared between coach, supervisor, and tutor) and View (specifics of how data is presented). Goal two: the interface should center around goal setting, and is supported by themes such as Trends (displaying data trends), Reflection (ability for tutors to reflect on data, observations and/or feedback), and Goals (relates to goal setting). Goal three: the interface should be integrated into the existing workflow, and is supported by themes such as Perspective (Refers to the different needs across different Saga positions), Bookmarking and note taking (Ability to take notes and mark points within the video/transcript), Classify and Quantify (Additional area identified for data annotation or quantification), Recommender system (Automatic system recommendations), Integration (specific to integrating app into current Saga practice), Alignment (connecting different types of data), Dashboard alert (quick signal on users home page as to status on areas of interest), and Summary (summarization of data or session). These goals were distilled by reviewing the emergent design themes that arose within and across sessions. Design goals are not mutually exclusive, and multiple themes support more than one goal.

Frequency of theme occurrence did not indicate importance. Frequent themes captured important components, however, their continued presence indicated that these areas needed to be elaborated upon or clarified as collaborators' understandings developed. Consistently distributed themes were readily understood concepts that specifically needed refinement in the context of the new tool. Infrequent themes demonstrate how important themes can emerge at different stages. Some themes seen early on do not continue since there was immediate consensus, while others appear later in response to fine-grained decisions. It may be difficult to develop new ideas until a stable picture is established and can only emerge through iterative learning. Themes are interleaved; often a single comment encapsulates multiple themes. For example, one participant stated that

they want the autonomy to send different data (e.g. specific notes, summary, strategy evidence, etc.) to different people since everyone has their own unique way of doing things. This one comment contributes to many themes (Notes, Summary, Sharing). While almost half of the ideas were present from the very first meeting and persisted through subsequent sessions, these ideas solidified as the co-design team built their joint understanding of what was needed and what was possible. This was not a trivial task that was completed in a single session. Instead, differing degrees of learning from each other continued through all of the meetings. Through this exchange we see solution mapping and how the team mutually designed the app interface.

Conclusion

This study provides rich insights into the development and integration of a new AI tool into an existing educational practice. Solution mapping provides an informative frame for how RPPs can co-adapt tools and solutions to new and old problems. Key to re-contextualizing was allowing for time and space to align understanding of both the problem and solution. This process was aided by beginning sessions focusing on dreaming and later sessions having researchers and coaches share complementary expertise to backcast Saga's goals for an ideal future onto the reality of the technology and their workflow. This approach is advantageous for any re-contextualization or implementation of an AI tool.

Additional developments and changes to the tool or practice may only be adoptable after more time has established trust in the current iteration of TalkMoves. TalkMoves may be especially good at automatically identifying many short but critical moments for coaches to observe across multiple sessions. However, coaches maintained that they wanted to be able to follow closely to their normal practice of watching a full twenty minutes of a singular session. Even if the new way of reviewing would save time and give a broader overview, coaches were not ready to solely rely on the app. This is understandable; the coaches need to first trust the app before accepting more dramatic changes to their practice. In a collaboration such as this, designing a new tool can be a slower process in order to establish comfort and trust such that greater change can be accepted later. Trust both in the design as well as in the results is essential, however, it is only through initial use that this latter form of trust can develop. There is a delicate balance between designing a useful tool and designing for change, as these goals are not always in alignment. RPPs must be committed to co-adapting and looking for new ways to improve.

References

Akkerman, S.& Bakker, A.(2011) Boundary crossing and boundary objects. Review of ed. res., 81(2), 132-169

Barter, C., & Renold, E., (1999) The Use of Vignettes in Qualitative Research, *Social Research Update*

Corbin, J., & Strauss, A.(2015) Basics of qualitative research: Techniques and procedures for developing grounded theory. *SAGE*.

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020) COVID-19 and learning loss- disparities grow and students need help. *McKinsey & Company*

Hoadley, C.P. (2002) Creating context: Design-based research in creating and understanding CSCL. *Proceedings of Computer Support for Cooperative Learning (CSCL)*

Jacobs, J., Scornavacco, K., Harty, C., Suresh, A., Lai, V., & Sumner, T., (2022) Promoting rich discussions in mathematics classrooms: Using personalized, automated feedback to support reflection and instructional change. *Teaching and Teacher Ed. (112).* DOI:10.1016/j.tate.2022.103631.

Lee, J.W., Daly, S.R, Huang-Saad, A., Rodriguez, G., & Seifert, C.M., (2020) Cognitive strategies in solution mapping: How engineering designers identify problems for technological solutions. *Design Studies 71* 100967. DOI:10.1016/j.destud.2020.100967.

Michaels, S., O'Connor, C.& Resnick, L.B.(2008) Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life. *Studies in Phil. and Ed., 27,* 283–297. DOI:10.1007/s11217-007-9071-1

Penuel, W.R., Roschelle, J., & Shechtman, N. (2007) Designing formative assessment software with teachers: an analysis of the co-design process. *Research and Practice in Tech. Enhanced Learning. 2(1).*51-74

Quist, J., & Vergragt, P., (2006) Past and future of backcasting: The shift to stakeholder participation and a proposal for a methodological framework, *Futures. 38.* 1027–1045. DOI10.1016/j.futures.2006.02.010.

Saldaña, J. (2016) The coding manual for qualitative researchers (3rd ed.). Thousand Oaks, CA: Sage.

Suresh, A., Penuel, W., Jacobs, J., Martin, M. & Sumner, T. (in press) Using AI tools to provide teachers with real-time feedback on their classroom discourse patterns. In X. Zhai & J. Krajcik (Eds), *Uses of Artificial Intelligence in STEM Education.* Oxford University Press.