Automated Feedback on Discourse Moves:

Teachers' Perceived Utility of a Big Data Tool

Karla Scornavacco, Jennifer Jacobs & Charis Harty

University of Colorado Boulder

Author Notes:

This study was supported by the National Science Foundation (NSF award #1837986). Any opinions, findings and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Paper to be presented as part of the symposium "All In: Student, Teachers and Principals Focus on Achievement Using Transformative Data Practices" at the annual conference of the American Educational Research Association, April 2022, San Diego, CA.

Abstract

Using new technology to provide automated feedback on classroom discourse offers a unique opportunity for educators to engage in self-reflection on their teaching, in particular to ensure that the instructional environment is equitable and productive for all students. More information is needed about how teachers experience automated data tools, including what they perceive as relevant and helpful for their everyday teaching. This mixed-methods study explored the perceptions and engagement of 21 math teachers over two years with a big data tool that analyzes classroom recordings and generates information about their discourse practices in near real-time. Findings revealed that teachers perceived the tool as having utility, yet the specific feedback that teachers perceived as most useful changed over time. In addition, teachers who used the tool throughout both years increased their use of talk moves over time, suggesting that they were making changes due to their review of the personalized feedback. These findings speak to promising directions for the development of AI-based, big data tools that help shape teacher learning and instruction, particularly tools that have strong perceived utility.

Automated Feedback on Discourse Moves:

Teachers' Perceived Utility of a Big Data Tool

Introduction

Classroom Recordings as a Form of Big Data to Support Teacher Learning

Providing teachers with detailed feedback on their classroom discourse addresses a big data challenge of critical importance within an applied domain: teacher professional learning. Video recordings, in particular, are an increasingly popular vehicle to support teachers' learning in and from their own practice (Gaudin & Chaliès, 2015; Major & Watson, 2018). Numerous studies have shown that classroom video holds substantial promise in enhancing teacher cognition and, in some cases, impacting student learning outcomes (Koellner et al., in press).

The rise in teachers' reliance on video for professional learning parallels the increasing use of video in everyday routines as an effective means for capturing, sharing, and generating knowledge. Rapid advances in recording technologies enable teachers to self-record their classrooms, with examples such as "classroom robotics" that are designed specifically for educators to easily film their instructional environments (Franklin et al., 2018; Prince, 2016). In addition, expanded storage capacities and newly developed online repositories aid in the process of managing, sharing and analyzing unstructured big data in the form of video and audio recordings.

A major barrier in the use of recorded lessons as a big data resource for automated analyses and feedback is the capability to generate sufficiently high quality audio. Much

of the critical information from recordings of teacher and student interactions is captured in the speech and language components. Yet, school classrooms are well-known to be noisy environments that are difficult to record with high fidelity without posing a major disruption to teachers and students. At the same time, recent studies have demonstrated that audio from classroom lessons collected with minimal imposition can be successfully processed using automated speech recognition software (Bokhove & Downey, 2018; Donnelly et al., 2016). This automated conversion from speech to text in a sufficiently reliable and robust fashion is key to developing linguistically motivated computational models from large datasets of classroom recordings.

Building on advances in automatic speech recognition, natural language processing, and deep learning, recent research has shown that the development and training of computer models to automate and scale discourse analyses is feasible (Song et al., 2020). Working from recordings of speech from K-12 classroom environments, researchers have developed a variety of AI models to reliably detect discursive features such as instructional talk, authentic teacher questions, elaborated evaluation, and uptake (Demszky et al., 2021; Jensen et al., 2020; Kelly et al., 2018). However, few of these efforts have translated into targeted professional learning opportunities for teachers.

Teachers' Perceptions of the Utility Value of Big Data Tools

Despite a rise in educators' interest in using automated tools to provide discourse-related information (e.g., Donnelly et al., 2017), there remains a dearth of research on how teachers may experience such tools, including what data they perceive

as relevant and helpful to improve their instruction. Of critical importance is demonstrating the utility value of teacher-facing tools (Kale & Akcauglu, 2018), to ensure they will find any given tool meaningful, relevant and useful to their practice. Drawing on seminal research in this area (Davis, 1989; Suramanian, 1994), this paper conceptualizes perceived utility as consisting of the following three dimensions related to teachers' use of an automated professional learning tool: 1) teachers' purposes in using the tool, 2) how relevant the tool is to their everyday teaching, and 3) the degree to which the tool informs their classroom practice.

In the educational literature, perceived utility of a new resource or technological tool is typically discussed in terms of teachers' acceptance and use of that tool in the classroom with their students. For example, Okumus et al. (2016) examined teachers' perceptions of the usefulness and ease of use for two software tools they incorporated in their mathematics classes. Researchers found that the alignment between teachers' instructional goals and the capabilities of the software was one of the most critical factors influencing their ratings of utility as well as their actual use of the tools. Others have found that teachers are more likely to engage in *sustained use* of instructional tools or innovations that have "perceived utility," meaning that they are construed as useful in the context of their everyday practice (Fishman et al., 2004; Fishman et al., 2011). Overall, teachers tend to be most drawn to tools that are straightforward and simple to use, offer information that they find interesting, and are aligned with their goals for improvement (Mumtaz, 2000; Okumus et al., 2016).

At the same time, teachers' beliefs and motivations regarding the utility value of new technologies have been shown to vary greatly, due to numerous factors including the local context, individual differences, and the nature of the tool (Backfisch et al., 2021). These factors play a role in how receptive teachers are to new technologies and resources that are intended to support their learning from and about classroom practice. In general, research on teachers' perceptions of and willingness to use new technologies is relatively sparse but likely to become increasingly critical as these technologies expand into the educational space, especially when their aim is to support instructional improvements.

Personalized Feedback to Improve Instruction

A recent report by the National Academies of Science, Engineering and Medicine (2020) argues that "an array of classroom studies provides evidence that many practicing teachers are not prepared to teach in ways that align with new expectations or that are responsive to a more diverse student population" (pg. 131). At issue is not a lack of opportunity for most teachers to participate in professional learning opportunities; in fact, the latest nationally representative survey of US mathematics, science and computer science teachers found that the majority took part in discipline-focused professional development in the past three years (Banilower et al., 2018). Critically important is that the professional development offered to teachers meets their needs and interests, improves their job commitment and satisfaction, and is perceived to be part of their professional journey as educators (Bautista & Ortega-Ruíz, 2015).

Traditional professional development approaches, in which teachers are passive recipients of “one-shot” workshops have been widely panned as “woefully inadequate” and unlikely to result in instructional change (Borko, 2004). In contrast, practice-based professional development – in which teacher learning is situated in the context of their everyday work (Ball & Cohen, 1999) – appears to be much more effective (Darling-Hammond et al., 2017). Video is a commonly used tool in practice-based professional development, offering teachers the opportunity to engage in reflective noticing that provides guidance for self-improvement (Sherin & Dyer, 2017).

New to the field are AI tools that process data from classroom recordings to provide personalized feedback related to teachers’ everyday practice. Without such tools, receiving personalized feedback relies on the observation of classroom lessons by trained experts, such as instructional coaches, which is a time-consuming and expensive endeavor. As a result, current human-centered approaches to providing personalized feedback are difficult to scale to large numbers of teachers. Automated approaches have enormous potential to support instructional improvement efforts at scale, particularly in the form of tools that use natural language processing to distill relevant linguistic information from classroom recordings.

It is important to recognize that providing feedback to teachers is not a trivial undertaking. Teachers’ mindsets, beliefs, and identity all play an important role as to how receptive teachers are to critical feedback and the possibility of making intentional changes in their practice (Jacobs et al., 2018). Key considerations include the degree to which teachers are interested in receiving feedback, whether the nature of the feedback

matches their professional goals, and how to best deliver feedback so that it is useful to the practitioner (Aguilar, 2013; Drago-Severson & Blum-Steffano, 2017; Gibbons & Cobb, 2017).

Classroom Discourse Patterns and Accountable Talk

There is widespread agreement that students' understanding should be constructed through the process of interacting within a learning community, and that discussions should be a prominent and normative feature within K-12 classrooms (Brenner, 1994; Franke et al., 2015; Sherin, 2002). Over the past decade, a robust literature on accountable talk, and in particular the talk moves that promote rich classroom discussions, has emerged (O'Connor, Michaels & Chapin, 2015; Resnick, Asterhan & Clarke, 2018). By using talk moves, teachers place the "intellectual heavy lifting" and balance of talk toward students and help ensure that classroom conversations will be purposeful, coherent, and productive (Michaels et al., 2010). Talk moves can be used by both teachers and learners to construct conversations in which students share their thinking, actively consider others' ideas, and engage in sustained reasoning (Candela et al., 2020; O'Connor & Michaels, 2019). Prior research has documented that using accountable talk moves in classrooms is associated with increased student achievement (e.g., Chapin & O'Connor, 2012; Webb et al., 2019). In addition, using talk moves can be understood as an equity-focused endeavor by ensuring all students have equal access to participation, subject matter content, and developing appropriate habits of mind (Michaels et al., 2010; O'Connor & Michaels, 2019).

A number of professional development models have emerged to support preservice

and inservice teachers in their efforts to learn about effective discourse patterns and work towards improving their own practice in alignment with accountable talk theory (e.g. Heyd-Metzuyanim et al., 2019; Hofmann & Ruthven, 2018; Kershner et al., 2020). Several of these efforts involved the development of sophisticated technological tools that rely on classroom recordings. For example, TeachFX is a commercially available application that uses natural language processing to provide teachers with automated feedback about the discourse patterns in recorded lessons, focused largely on the degree to which students are talking relative to their teacher. Ford and Welling-Riley (2021) report that teachers who piloted TeachFX saw a 45% increase in student talk.

As another example, Chen and colleagues' (Chen, 2020; Chen et al., 2020; Chen et al., 2015) developed the Classroom Discourse Analyzer, a web-based platform informed by accountable talk theory that enables teachers to visualize their use of specific discourse moves such as turns, speakers, amount of talk within a turn, and classification of the talk within a turn. Although some of the discourse information is automatically extracted, the higher-inference classifications must be carried out manually by trained human coders. In a randomized control trial, treatment group teachers experienced significant increases in their use of certain talk moves that were sustained over time (Chen, 2020). Moreover, students of the treatment teachers significantly increased their mathematics achievement relative to the control group (Chen et al., 2020).

On the whole, studies suggest that teachers are willing and able to use personalized feedback from data driven tools to make discernible changes to their classroom

discourse practices. What remains unclear is teachers' perceived utility of these types of tools, including why teachers are interested in using them and the extent to which the teachers find the tools relevant for their everyday teaching.

Research Questions

This paper reports on the pilot study of a big data tool designed to provide teachers with detailed and personalized information about the conversations taking place in their classrooms. The overall question guiding the research was: What is the utility value of a tool that automatically processes and analyzes classroom recordings to generate feedback on classroom discourse patterns? In addition, drawing from prior literature on teachers' perceived utility of a new technological tool, the study also focused on the following subquestions: (1) What were teachers' goals for using the tool, (2) To what extent did teachers find the tool to be instructionally relevant, and (3) How did teachers' classroom discourse change over time?

Talk Moves Application

Overview of the Application

The TalkMoves application is a deployed system that provides teachers with automated, immediate, and actionable feedback about their mathematics instruction. The application uses automated speech recognition and deep learning models to detect the presence of teacher and student talk moves, drawing on accountable talk theory (Jacobs et al., 2022; Suresh et al., 2021). The application consists of three interrelated components: (1) a cloud-based big data infrastructure to manage and process recordings of mathematics lessons, (2) automated speech recognition and deep

learning models to classify talk moves, and (3) a personalized dashboard to visually display each teacher's feedback analytics for their individual lessons and their lessons over time.

To use the TalkMoves application, teachers first self-record their classroom lessons (or portions of lessons). The application then processes and analyzes their recordings, in a fully automated fashion. Specifically, the application converts speech from the audio files to written text in sentence form and assigns speaker labels. Next deep learning models classify the sentences as containing a talk move or not. The models were trained on a set of ten talk moves (six teacher talk moves and four student talk moves), selected due to their relatively high frequency in the training data corpus, the ability of human coders to establish high inter-rater reliability, and based on suggestions from experts in accountable talk. See Table 1 for the set of talk moves, along with definitions and examples.

Table 1. Teacher and student talk moves included in the TalkMoves application

Category	Talk move	Description	Example
TEACHER TALK MOVES			
Learning Community	Keeping everyone together	Prompting students to be active listeners and orienting students to each other	"What did Eliza just say her equation was?"
Learning Community	Getting students to relate to another's ideas	Prompting students to react to what a classmate said	"Do you agree with Juan that the answer is 7/10?"
Learning Community	Restating	Repeating all or part of what a student said word for word	"Add two here."
Content Knowledge	Pressing for accuracy	Prompting students to make a mathematical contribution or use mathematical language	"Can you give an example of an ordered pair?"

Rigorous Thinking	Revoicing	Repeating what a student said but adding on or changing the wording	“Julia told us she would add two here.”
Rigorous Thinking	Pressing for reasoning	Prompting students to explain, provide evidence, share their thinking behind a decision, or connect ideas or representations	“Why could I argue that the slope should be increasing?”
STUDENT TALK MOVES			
Learning Community	Relating to another student	Using, commenting on, or asking questions about a classmate’s ideas	“I didn’t get the same answer as her.”
Learning Community	Asking for more info	Student requests more info says they are confused or need help	“I don’t understand number four.”
Content Knowledge	Making a claim	Student makes a math claim, factual statement, or lists a step in their answer	“X is the number of cars.”
Rigorous Thinking	Providing evidence or reasoning	The student explains their thinking, provides evidence or talks about their reasoning	“You can’t subtract 7 because then you would only get 28 and you need 29.”

The system generates feedback for each uploaded lesson recording based on output from the deep learning models as well as other discourse-related analyses (such as which math terms were used and how much talk came from the teacher relative to the students). This feedback is displayed on a personalized dashboard that shows the information for a specific lesson, that lesson compared to all of the teachers’ lessons, and that lesson compared to the lessons from all other users of the application. For each lesson the dashboard displays the corresponding video recording as well as a variety of data analytics, primarily using graphics and visual representations (see Figure 1).

Figure 1. Dashboard interface displaying information about the teacher talk moves used during a selected lesson

Ideal Use of the Application

In its current early development stage, the TalkMoves application is intended to be used by individual teachers, independently accessing and viewing only their own feedback. Due to privacy restrictions, the recorded data and corresponding feedback cannot be shared with other individuals, meaning that collaborative review by the participating teachers is not supported by the project team at the present time. For the study discussed in this paper, ideal usage of the TalkMove application is an independent, cyclical process that includes the following steps after recording and uploading video of a classroom lesson, as depicted in Figure 2: (1) make sense of the feedback provided by the application, (2) reflect on appropriate instructional changes based on the feedback, and (3) strive to make instructional changes. Each new

uploaded lesson represents an opportunity for teachers to review additional data and consider the extent to which they have made progress towards their instructional improvement goals.

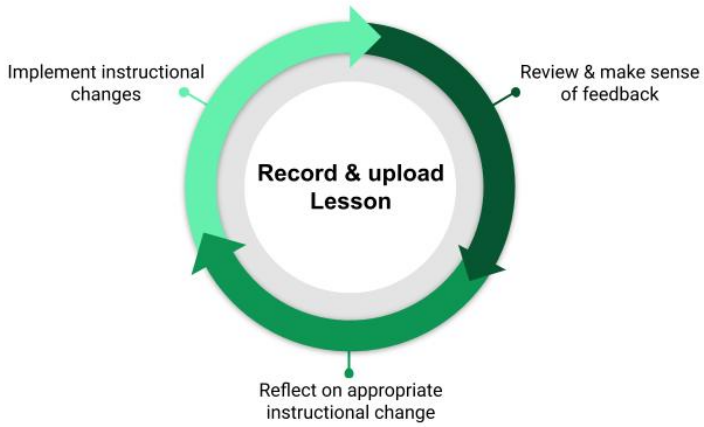


Figure 2. Ideal use of the TalkMoves application in the current study

Method

Participants

Twenty-one teachers from two school districts in the western United States voluntarily consented to participate in a pilot study of the TalkMoves application beginning in the 2019-2020 school year. The teachers spanned grades 4-12 with the majority teaching upper elementary school (71%). The participants were a relatively experienced group of teachers, with a range of 4 to 32 years classroom teaching experience (M=15). Twelve of the teachers continued their participation for a second school year (2020-21). Similar to the full group, the majority of these continuing teachers taught elementary school (n=8) and their average teaching experience was 16 years.

An important backdrop to this study was the COVID-19 pandemic, with face-to-face classroom instruction ceasing in the participating school districts from spring 2020 through at least fall 2020. During this time, the teachers engaged in various instructional models that included in-person, online and hybrid classes. In the first school year of the study, the teachers recorded only in-person lessons. However, during the second school year, most of the teachers recorded both online and in-person (or hybrid) lessons. The research team tracked data on teachers' mode of instruction, though for this study all lessons are included in the analyses to maintain a focus on teachers' perceptions of the TalkMoves application.

Data Sources and Analysis

This mixed-methods study takes an interpretive approach to the teachers' experiences using the data tool, relying on the "the participants' views of the situation being studied" (Creswell, 2013, p. 8) as well as the social construction of meaning by researchers and participants. Qualitative data included responses from five surveys administered throughout the two year period as well as two interviews. The surveys and interviews primarily focused on teachers' use and understanding of the application, as well as its perceived utility value. Quantitative data come from the automated feedback on teachers' recorded mathematics lessons. Examples of this quantitative data include the frequency of each talk move in a given lesson, the total number of talk moves, and the percentage of talk from students.

Qualitative analysis was both deductive and inductive, since some codes naturally connected to the question topic while other codes emerged from the data. For

open-ended questions in the surveys, an open-coding approach (Given, 2008) was used to generate a set of response categories. Two researchers classified each teacher's response(s), tracked inter-rater agreement, and then discussed and reconciled any differences. The researchers also created analytic memos for each teacher that incorporated highlights and trends based on their unique qualitative and quantitative data. To create the memos, researchers engaged in a process of investigator triangulation (Mathison, 1988), with one person drafting an initial memo and the other(s) reviewing and commenting on the draft, with discussion and revision taking place until agreement was reached. The team checked biases, looked for non-examples, and continually returned to the data in an effort to accurately capture teachers' perceptions of purpose and relevance, as well as the actual classroom discourse the data tool captured (and what teachers wanted it to capture).

Results

Teachers' Use of the Tool

Teachers' use of the TalkMoves application is multifaceted and can be analyzed in a variety of ways. Two components of teachers' use are (1) how often they recorded and uploaded lessons and (2) self-reports on how frequently they reviewed their personalized feedback. As shown in Table 2, in Year 1 teachers recorded 10 lessons on average, with a wide range across teachers (recording between 3-21 lessons). Year 2 data are quite similar, with teachers recording 14 lessons on average and ranging from 1-31 recordings per teacher.

Table 2. Teachers' Use of the TalkMoves Application: Recording Lessons

	Year 1 (2019-2020)	**Year 2 (2020-21)**
Number of participating teachers	21	12
Total lessons	210	163
Average # lessons	10	14
Range for each teacher	3-21	1-31
Number of teachers who consistently reviewed their feedback	13	10

As noted, only 12 of the 21 teachers elected to continue recording lessons in the second year of the study. While there are likely numerous reasons for this dropout rate (including the impact of COVID), it may be partly explained by the consistency with which teachers independently viewed their feedback. Of the teachers who participated in the first year of the study, 13 (62%) reported that they logged into the application and looked over their feedback for most or all of the lessons that they recorded. Five of the eight teachers who were infrequent viewers of their data in year 1 elected to not continue in the second year, and three elected to continue participating. During year 2, 10 of the 12 participating teachers (83%) reported that they consistently reviewed their feedback.

Interestingly, the consistency with which teachers reviewed their feedback does not appear to be connected to how often they recorded lessons. Based on the data from year 1, Table 3 shows the number of teachers who were consistent or inconsistent reviewers of their feedback and whether they recorded above or below the average number of lessons. The fact that the number of consistent and inconsistent reviewers

was relatively stable regardless of lesson recordings suggests that these two activities were understood to be largely independent of one another by the participants. It is possible that some teachers may have filmed because they were asked (and paid) to do so by the research team, but they did not feel obligated to review and reflect on their feedback on a regular basis.

Table 3. Number of teachers who recorded above and below the average number of lessons and reviewed their feedback consistently (most/all lessons) or inconsistently

YEAR 1 DATA ONLY

	Consistent Reviewers	Inconsistent Reviewers
Recorded < group mean (10 lessons)	6	4
Recorded ≥ group mean (10 lessons)	7	4

Purpose and Goals for Using the Data Tool

Teachers' stated goals for their own instructional improvement underscore an interest in learning more about talk moves and their own use of these strategies, as well as a desire to shift the discourse in their mathematics lessons toward eliciting more student thinking and peer collaboration. Table 4 shows the breakdown in teachers' goals based on the initial survey and follow-up interviews throughout the two years of the study. When they were first beginning to use the data tool teachers' individual goals varied slightly, but on the whole reflected a generalized aim of improving classroom instruction in alignment with accountable talk theory. Teachers typically commented that they wanted to learn more about their own instruction and hoped that using the

application would help them to gain knowledge and better support their students. One shared, “I am going to study it and start using talk moves. I may be doing some in my classroom but don't know I am” (MC, 12/19). Another teacher hoped the feedback would help her to reflect on: “Are my questioning strategies effective to help students develop critical thinking?” (SF, 11/19).

Table 4. Teacher initial goals and motivation for continued use of TalkMoves

Year 1. Goal Categories	**# Teachers**
Gaining Knowledge Related to Talk Moves	6
Better Supporting Student Thinking and Engagement	5
Increase Student Talk and Peer Collaboration	4
Getting Feedback on Their Instructional Practices	3
Other	3
Year 2. Motivation for Continued Use	
Curiosity – especially the percentage of teacher talk	4
Curiosity – especially the type of talk moves used	3
Teacher focus (goal setting)	2
Teacher focus (general self-check)	2
Student focus	1

Teachers who continued during the second year of the study shared a variety of motivations, which overall were more specific than in the prior year. Most teachers

expressed curiosity about specific data generated by the tool, especially the amount of teacher talk and the nature of the talk moves. They made comments such as, “I like seeing the different kinds of talk moves that both my students and I use. I really enjoy looking at data, so I like digging into all the info.” (KH, 5/21). Two teachers reported using the data in a functional or linear fashion, attending primarily to one talk move category at a time in order to increase the frequency of that discourse strategy. Other teachers reported periodically setting targeted goals, though they appeared to be more motivated by a general “pulse check” than working toward a specific number of talk moves. For example, one teacher said he wanted “to make sure that I hit all of my talking points, and also to make sure that there was a large amount of student engagement” (CC, 5/21).

Perceived Relevance of the Data Tool

On the whole, teachers reported finding the TalkMoves application useful and relevant in both years of the study. In year 1, all but 3 teachers (86%) responded that the data tool was at least “somewhat” useful, and all but 2 (90%) reported that the feedback prompted them to change their instruction to some degree. In year 2 all but one teacher (92%) felt that the tool was relevant to their everyday teaching, and all of the teachers expressed that the feedback either led directly to instructional changes or encouraged their reflection on targeted discursive practices.

Probing more deeply shows that what teachers initially found most relevant was personalized feedback on the amount of student talk relative to teacher talk in their recorded lessons. At first, most teachers did not appear particularly attentive to the

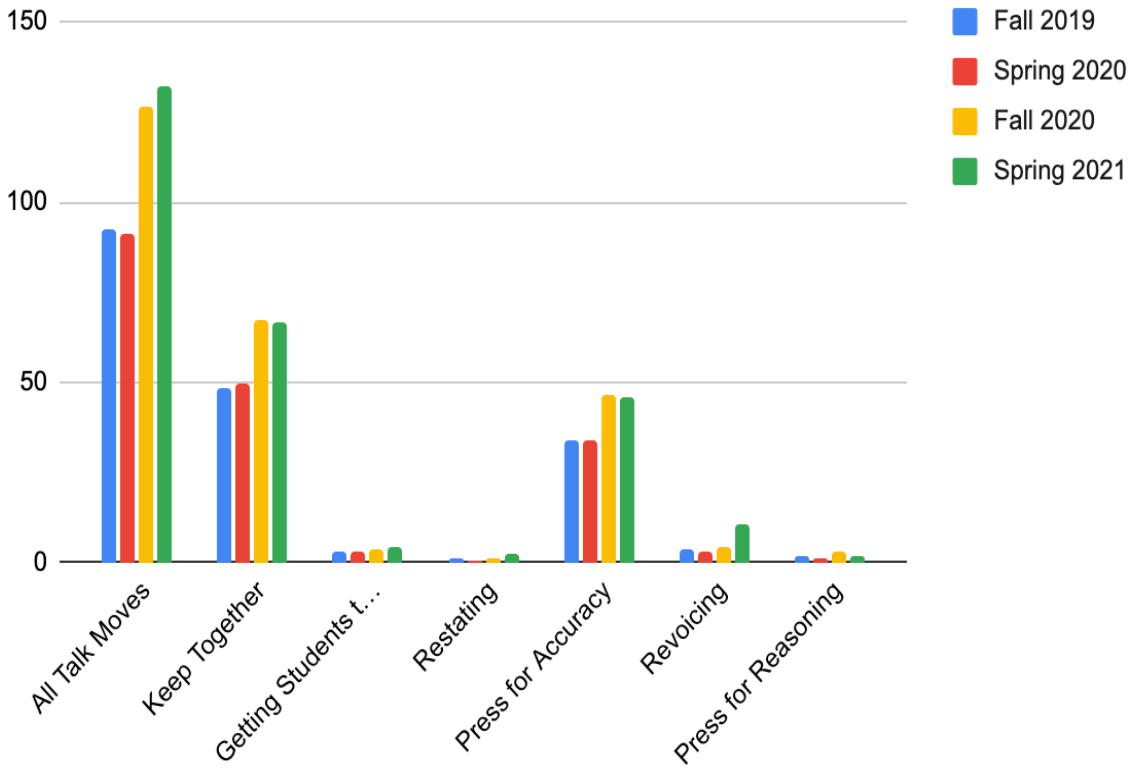
application's unique function and main purpose: providing information about the frequency and types of talk moves that occurred during a math lesson. The first year interview, for instance, revealed that 72% of the teachers initially looked at and were most interested in the relative percentages of teacher and student talk.

Yet, in the second year, as teachers reported that they were increasingly interested in and focused on the use of talk moves, only 25% said that they focused primarily on how much talk came from the teacher relative to the students. This trend toward attending less to the *quantity of talk* and more to the *quality of talk* (use of talk moves) aligns with the developers' intentions, though it also shows that it may take time for teachers to start to become interested in that larger purpose of the tool.

Changes in Classroom Discourse Over Time

Along with analyses of self-report data on teachers' perceptions of the TalkMoves application, this study explored observable changes in classroom instruction based on the lesson recordings that teachers uploaded to the system. Analyses of these recordings suggest that the tool did have a positive influence on classroom instruction over time, specifically in the form of increased use of talk moves. Based on data from the twelve teachers who participated in both years of the project, Figures 3 and 4 provide information as to how much the use of teacher and student talk moves changed during this time period. The figures show the average frequency of teacher and student talk moves by each semester (fall, spring) of the two school years (2019-2020, 2020-21). With a few exceptions, the use of talk moves was not notably different *within* a school year; however observable changes occurred *across* years.

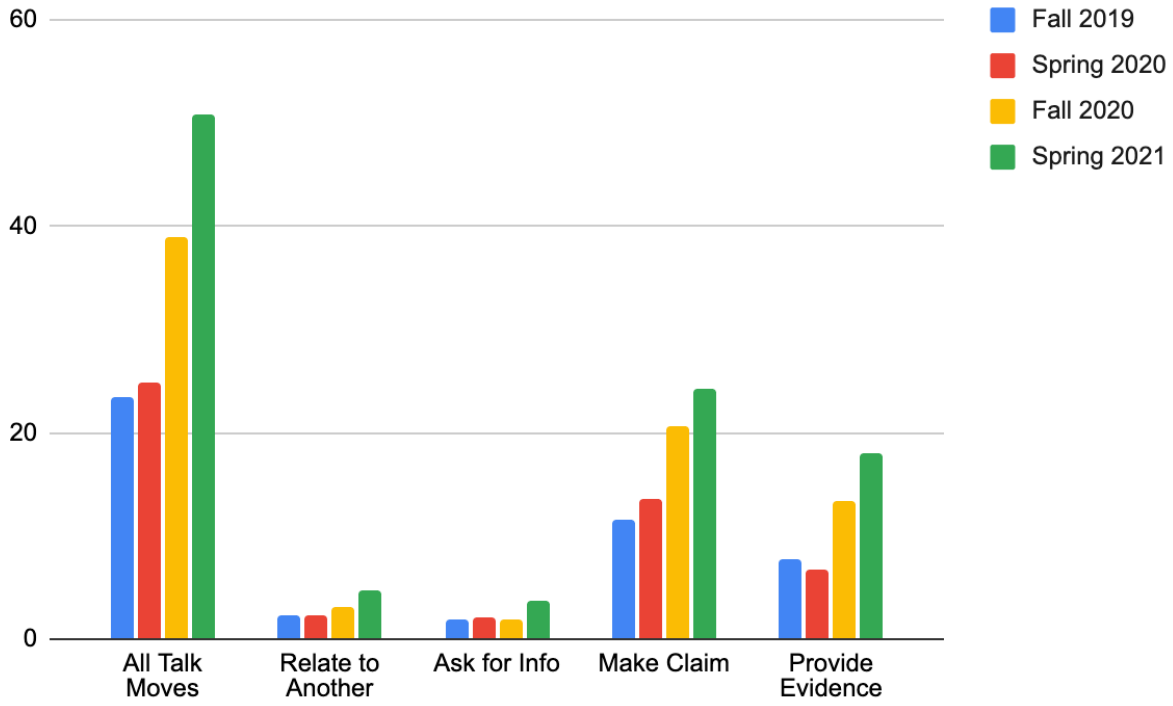
Figure 3. Average frequency of teacher talk moves per lesson during the 2019-20 and 2020-21 school years



As a group, the teachers used an average of 37 more talk moves in their later lessons relative to their earlier lessons (F=5.753, p<.05) and students used an average of 17 more talk moves (F=5.025, p<.0.5); both of these increases in the total number of talk moves per lesson were statistically significant. Moreover there was an increase in the average frequency of each individual talk move per lesson over time, with the percent increase ranging from 15% - 122%. For three of the ten talk moves, these increases from Year 1 to Year 2 were statistically significant: keeping students together

(teacher talk move, F=4.524, p<.05), revoicing (teacher talk move, F=13.354, p<.001), and providing evidence (student talk move, F=6.847, p<.05).

Figure 4. Average frequency of student talk moves per lesson during the 2019-20 and 2020-21 school years



Overall, as evidenced by the amount of talk moves used in a lesson, the results indicate a trend toward an improvement in the *quality of* talk in the mathematics classrooms of teachers who used the data tool across two school years. Due to the small sample size and the lack of a comparison group, these analyses should be considered as exploratory only. Yet, they offer a promising next step in the exploration and development of data tools that can make an impact on teacher learning and instruction.

Conclusion

This study examined the perceived utility value of a big data tool - the TalkMoves application - that provides automated feedback to teachers on their discursive practices in recorded mathematics lessons. During their first year using the tool, the participating teachers reported goals closely aligned with the aim and capacity of the tool to generate feedback on accountable talk moves that support student participation and learning. Moreover, the vast majority of teachers found the feedback relevant and likely to impact their instruction. However, shortly after teachers began using the tool, COVID-19 disrupted face-to-face teaching, creating a serious confound for the project. Only about half of the teachers continued using the tool during the second year of the study, which may be partially due to the challenges related to remote learning, a lack of perceived utility for a subset of teachers, or some combination of the two.

Those teachers who did continue using the TalkMoves application for a second school year reported an ongoing curiosity about their personalized feedback, felt the feedback was relevant to their instructional improvement efforts, and showed a significant increase in the amount of teacher and student talk moves relative to their lessons from the first year. Results from this combination of indicators suggests the tool did have a high perceived utility value, particularly for those teachers who elected to use it throughout two school years.

New technological innovations create new opportunities as well as new choices, including the choice to interact purposefully over a sustained period of time with the technology (Zuboff, 2001). Will the user's investment of time and effort in the new

technology offer sufficient pay off? In the case of the TalkMoves application, which generates information previously unavailable to teachers regarding their classroom discourse patterns, users must not only record and upload their mathematics lessons, but then they must devote time and attention to reviewing their feedback, deriving meaning from the data, and considering whether and how to make purposeful changes to their teaching. Like most novel technological tools, the TalkMoves application was not perceived or used in the same way by all of the pilot teachers in this study. However, on the whole they appeared motivated to utilize their personalized data as a catalyst for self-reflection, leading to noticeable changes in their everyday practice.

The TalkMoves application does not advise teachers on whether or how to make instructional changes. Currently designed as a platform that primarily supports individual reflection, teachers must make sense of their own data, consider whether changes are warranted, and if so, implement specific instructional strategies aligned with their personal goals. While this usage scenario affords teachers full autonomy in their decision making, maintaining active engagement with the application requires a great deal of internal motivation, out of school time and individual effort.

Significance for Developers of Big Data Tools

This study is a fruitful case for centering the importance of teachers' perceptions of and engagement with automated data tools when cultivating equitable education systems. Although tools built on big data infrastructure hold potential for improving instruction, the findings from this study suggest that designing professional learning opportunities based on these tools would also benefit from a clear understanding as to

how teachers feel about and use such tools. The results of this study speak to the importance of adaptability, and especially to providing varied entry points for teachers to engage with these tools no matter how generalized or aspirational the designers' intentions may be. The TalkMoves application that is the focus of this study offered multiple starting points for teachers, such as a focus on the *quantity* of teacher talk and the *quality* of classroom discourse (i.e., the talk moves used by teachers and students during a lesson). When investing money and time into building out more robust big data tools, developers must not only ensure that teachers initially perceive these tools as helpful, but also consider how teachers are valuing and using the tools over time. Professional development efforts that incorporate the TalkMove application, for example, might *begin* with teachers sharing data on the amount of teacher talk relative to their students before shifting to specifically addressing the types of talk moves used - by whom, when, and how often. This approach would align with this study's findings on teachers' shifting attention to the quality of talk (i.e., use of talk moves) after they have spent some time with the tool.

Similarly, it behooves developers of big data tools that capture classroom discourse to ensure that their tools are perceived by teachers as relatively straightforward to use and that the tools help them to learn more about *their own* instruction. The teachers who consistently reviewed their feedback during year 1 were the most likely to continue using the tool a second year, even with challenges of teaching during a pandemic. Certainly the tool is more useful when teachers review their feedback, however it is possible that an *entrypoint* for teachers' use is simply getting in the habit of recording

and uploading lessons. Teachers could then, in their own time or with encouragement and prompting of a colleague, coach, or scheduled professional development session, review their feedback in more detail.

Study Limitations and Next Steps

The paper describes a pilot study, with a small sample of willing volunteers who were interested in receiving personalized feedback about the discourse patterns in their mathematics lessons. The participants were encouraged to both record the classrooms and reflect on the data provided after each lesson was uploaded. However, teachers had full autonomy with respect to how often they filmed, whether and for how long they looked at their feedback, and the degree to which they acted upon their goals for instructional improvement. In fact, teachers' engagement in all of these activities were considered important components of the research team's empirical efforts, and helped to determine the degree to which they perceived the TalkMoves application as having a high utility value.

The fact that COVID significantly interrupted schooling between the first and second years of this study make it challenging to interpret the longitudinal findings. Numerous teachers conveyed that they were overwhelmed by the demands of remote teaching. As one teacher commented in a year 2 survey, "I am strictly remote this year, which is making the priorities look a little different than a normal classroom!" Some of the teachers who did not record lessons in year 2 mentioned that they were interested in continuing their use of the tool during the following school year, once the impacts of COVID were diminished. It is unknown how many of the teachers would have

maintained their motivations for utilizing this type of tool, particularly in a sustained manner over time, if the school year had not been disrupted.

Another avenue worthy of further exploration is the difference in users' experience and perceptions of the tool when used for individual reflection or as part of a collaborative professional development endeavor. Currently, the TalkMoves application is designed for individual teachers to access and use, largely independently. However, professional learning often takes place in a more collaborative environment, such as through professional learning communities and with instructional coaches. An important next step in this work is to design an online feedback environment that is more conducive to such uses, while still ensuring the privacy of individual teachers and their students. Future research, then, might focus on the tool's usefulness as a vehicle for shared inquiry, for instance, building upon teachers' collective curiosities and interests in improving math discourse. Regardless of the specific learning context in which such tools are designed for, it is important to have realistic expectations as to when and how teachers will make shifts in focal areas of equitable classroom discourse. Moreover, information regarding instructional shifts is well complemented by data on other aspects of perceived utility, such as teachers' goals and perceptions of relevance, particularly when considering whether and why a tool is effective.

References

Aguilar, E. (2013). *The art of coaching: Effective strategies for school transformation*. San Francisco, CA: Jossey-Bass.

Backfisch, I., Lachner, A., Stürmer, K., & Scheiter, K. (2021). Variability of teachers' technology integration in the classroom: A matter of utility! *Computers & Education*, *166*, 104159.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond and G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3–31). San Francisco, CA: Jossey-Bass.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+.* Chapel Hill, NC: Horizon Research, Inc.

Bautista, A., & Ortega-Ruíz, R. (2015). Teacher professional development: International perspectives and approaches. *Psychology, Society and Education*, *7*(3), 240–251.

Bokhove, C., & Downey, C. (2018). Automated generation of 'good enough'transcripts as a first step to transcription of audio-recorded data. *Methodological Innovations*, *11*(2), 2059799118790743.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, *33*(8), 3-15.

Brenner, M. E. (1994). A communication framework for mathematics: Exemplary instruction for culturally and linguistically diverse students. *Language and Learning: Educating Linguistically Diverse Students*, 233-268.

Candela, A. G., Boston, M. D., & Dixon, J. K. (2020). Discourse actions to promote student access. *Mathematics Teacher: Learning and Teaching PK-12*, *113*(4), 266-277.

Chapin, S., & O'Connor, C. (2012). Project Challenge: Using challenging curriculum and mathematical discourse to help all students learn. In C. Dudley-Marling and S. Michaels (Eds.), *High-Expectation Curricula: Helping All Students Succeed with Powerful Learning* (pp. 113-127). New York: Teachers College.

Chen, G. (2020). A visual learning analytics (VLA) approach to video-based teacher professional development: Impact on teachers' beliefs, self-efficacy, and classroom talk practice. *Computers & Education*, *144*, 1-15.

Chen, G., Chan, C. K., Chan, K. K., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences*, *29(4-5)*, 642-680.

Chen, G., Clarke, S. N., & Resnick, L. B. (2015). Classroom Discourse Analyzer (CDA): A discourse analytic tool for teachers. *Technology, Instruction, Cognition & Learning*, *10*(2), 85-105.

Creswell, J. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340.

Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2021). *Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial In a large-scale online course*. (EdWorkingPaper: 21-483). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/thn9-wh86

Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017, March). Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 218-227).

Donnelly, P. J., Blanchard, N., Samei, B., Olney, A. M., Sun, X., Ward, B., ... & D'Mello, S. K. (2016, July). Automatic teacher modeling from live classroom audio. In *Proceedings of the 2016 conference on user modeling adaptation and personalization* (pp. 45-53).

Drago-Severson, E., & Blum-DeStefano, J. (2017). *Tell me so I can hear you: A developmental approach to feedback for educators*. Harvard Education Press.

Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences*, *13*(1), 43-76.

Fishman, B., Penuel, W. R., Hegedus, S., & Roschelle, J. (2011). What happens when the research ends? Factors related to the sustainability of a technology-infused mathematics curriculum. *Journal of Computers in Mathematics and Science Teaching*, *30*(4), 329-353.

Ford, K., & Welling-Riley, K. (2021). Student talk in science class. *The Learning Professional*, *42*(3), 58-61.

Franke, M. L., Turrou, A. C., Webb, N. M., Ing, M., Wong, J., Shin, N., & Fernandez, C. (2015). Student engagement with others' mathematical ideas: The role of teacher invitation and support moves. *The Elementary School Journal*, *116*(1), 126-148.

Franklin, R. K., Mitchell, J. O., Walters, K. S., Livingston, B., Lineberger, M. B., Putman, C., Yarborough, R., & Karges-Bone, L. (2018). Using Swivl robotic technology in teacher education preparation: A pilot study. *TechTrends, 62*(2), 184–189.

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41-67.

Gibbons, L. K., & Cobb, P. (2017). Focusing on teacher learning opportunities to identify potentially productive coaching activities. *Journal of Teacher Education*, *68*(4), 411-425.

Given, L. M. (Ed.). (2008). *The Sage encyclopedia of qualitative research methods*. London: Sage Publications.

Heyd-Metzuyanim, E., Smith, M., Bill, V., & Resnick, L. B. (2019). From ritual to explorative participation in discourse-rich instructional practices: a case study of teacher learning through professional development. *Educational Studies in Mathematics*, *101*(2), 273-289.

Hofmann, R., & Ruthven, K. (2018). Operational, interpersonal, discussional and ideational dimensions of classroom norms for dialogic practice in school mathematics. *British Educational Research Journal*, *44*(3), 496-514.

Jacobs, J., Boardman, A., Potvin, A., & Wang, C. (2018). Understanding teacher resistance to instructional coaching. *Professional Development in Education,* 44(5), 690-703.

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 Education, 122*.

Jensen, E., Dale, M., Donnelly, P. J., Stone, C., Kelly, S., Godley, A., & D'Mello, S. K. (2020, April). Toward automated feedback on teacher discourse to enhance teacher learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Kale, U., & Akcaoglu, M. (2018). The role of relevance in future teachers' utility value and interest toward technology. *Educational Technology Research and Development*, *66*(2), 283-311.

Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, *47*(7), 451-464.

Kershner, R., Hennessy, S., Dowdall, K., Owen, H. & Calcagni, M.E. (2020). Teachers as 'natural experimenters': Using T-SEDA to develop classroom dialogue. In Rolls, L. & Hargreaves, E. (Eds.) *Reimagining Professional Development in Schools* (Chapter 8). Routledge.

Koellner, K., Jacobs, J., Borko, H. & Seago, N. (in press). Current trends and tensions in research on teacher professional learning. *International Encyclopedia of Education*, 4th Edition.

Major, L., & Watson, S. (2018). Using video to support in-service teacher professional development: the state of the field, limitations and possibilities. *Technology, Pedagogy and Education*, *27*(1), 49-68.

Mathison, S. (1988). Why triangulate? *Educational Researcher, 17*(2), 13-17.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2010). *Accountable talk sourcebook: For classroom conversation that works*. Pittsburgh, PA: University of Pittsburgh Institute for Learning.

Mumtaz, S. (2000). Factors affecting teachers' use of information and communications technology: A review of the literature. *Journal of Information Technology for Teacher Education*, *9*(3), 319-342.

National Academies of Sciences, Engineering, and Medicine. (2020). *Changing Expectations for the K-12 Teacher Workforce: Policies, Preservice Education, Professional Development, and the Workplace.* Washington DC: The National Academies Press.

O'Connor, C., & Michaels, S. (2019). Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. *International Journal of Educational Research*, *97*, 166-175.

O'Connor, C., Michaels, S., & Chapin, S. (2015). Scaling down to explore the role of talk in learning: From district intervention to controlled classroom study. In L.B. Resnick,

C.S.C. Asterhan, & S.N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 111-126). Washington DC: AERA.

Okumuş, S., Lewis, L., Wiebe, E., & Hollebrands, K. (2016). Utility and usability as factors influencing teacher decisions about software integration. *Educational Technology Research and Development*, *64*(6), 1227-1249.

Prince, T. (2016, March). “Panopto” for Lecture Capture–A First Time User’s Perspective. In *Society for Information Technology & Teacher Education International Conference*(pp. 2550-2565). Association for the Advancement of Computing in Education (AACE).

Resnick, L. B., Asterhan, C. S. C., Clarke, S. N., & Schantz, F. (2018). Next generation research in dialogic learning. In G. E. Hall, L. F. Quinn, & D. M. Gollnick (Eds.). *Wiley handbook of teaching and learning* (pp. 323–338). Hoboken, NJ: Wiley-Blackwell.

Sherin, M. G. (2002). A balancing act: Developing a discourse community in a mathematics classroom. *Journal of Mathematics Teacher Education*, *5*(3), 205-233.

Sherin, M. G., & Dyer, E. B. (2017). Mathematics teachers’ self-captured video and opportunities for learning. *Journal of Mathematics Teacher Education*, *20*(5), 477-495.

Song, Y., Lei, S., Hao, T., Lan, Z., & Ding, Y. (2020). Automatic classification of semantic content of classroom dialogue. *Journal of Educational Computing Research*, *59*(3), 496-521.

Subramanian, G. H. (1994). A replication of perceived usefulness and perceived ease of use measurement. *Decision Sciences*, *25*(5-6), 863-874.

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J.H. & Sumner, T. (2021). *Using transformers to provide teachers with personalized feedback on their classroom discourse: The TalkMoves application*. Paper presented to the 2021 AAAI Conference on Artificial Intelligence in K-12 Education.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273–315.

Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., & Zimmerman, J. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *International Journal of Educational Research*, *97*, 176-186.

Zuboff, S. (1985). Automate/informate: The two faces of intelligent technology. *Organizational dynamics*, *14*(2), 5-18.