Automated Feedback Improves Teachers' Questioning Quality in Brick-and-Mortar Classrooms: Opportunities for Further Enhancement *

Dorottya Demszky†
Jing Liu ‡
Heather C. Hill§
Shyamoli Sanghi¶
Ariel Chung‖

ABSTRACT: AI-powered professional learning tools that provide teachers with individualized feedback on their instruction have proven effective at improving instruction and student engagement in virtual learning contexts. Despite the need for consistent, personalized professional learning in K-12 settings, the effectiveness of automated feedback tools in traditional classrooms remains unexplored. We present results from 224 Utah mathematics and science teachers who engaged in a pre-registered randomized controlled trial, conducted in partnership with TeachFX, to assess the impact of automated feedback in K-12 classrooms. This feedback targeted "focusing questions" — questions that probe students' thinking by pressing for explanations and reflection. We find that teachers opened emails containing the automated feedback about 53-65% of the time, and the feedback increased their use of focusing questions by 20% ($p < 0.01$) compared to the control group. The feedback did not impact other teaching practices. Qualitative interviews with 13 teachers revealed mixed perceptions of the automated feedback. Some teachers appreciated the reflective insights, while others faced barriers such as skepticism about accuracy, data privacy concerns, and time constraints. Our findings highlight the promises and areas of improvement for implementing effective and teacher-friendly automated professional learning tools in brick-and-mortar classrooms.

KEYWORDS: computer-assisted instruction; brick-and-mortar classroom; natural language processing; automated teacher feedback; randomized controlled trial; focusing questions; mixed-methods study; K-12 instruction

---

*Demszky and Liu contributed equally to this research. The authors declare no conflict of interest. This paper has been previously submitted to a pre print platform (EdWorkingPapers) but has not been published.

†Dorottya (Dora) Demszky (ddemszky@stanford.edu; corresponding author), is an Assistant Professor in Educational Data Science at Stanford University, Stanford, CA, United States.

‡Jing Liu (jliu28@umd.edu) is an Assistant Professor in Education Policy at the University of Maryland, College Park, MD, United States.

§Heather Hill (heather_hill@gse.harvard.edu) is a Hazen-Nicoli Professor in Education at the Harvard Graduate School of Education, Cambridge, MA, United States.

¶Shyamoli Sanghi (shyamoli@cs.stanford.edu) was a masters student in Computer Science at Stanford University at the time of the study, and currently a machine learning engineer at TeachFX.

‖Ariel Chung (tchung13@umd.edu) is a Ph.D. student in Education Policy at the University of Maryland, College Park.

1 Introduction

Formative feedback grounded in teachers' practices can enhance instruction and improve student outcomes (Kraft et al., 2018; Shute, 2008; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). Instructional coaches or mentor teachers often provide such feedback by observing classrooms, guiding teacher reflections and offering improvement suggestions (Kraft et al., 2018). However, expert coaching is expensive and time-consuming, limiting most teachers' access to consistent, high-quality feedback. In the United States, only ∼40% of schools provide teachers access to an instructional coach (Taie & Goldring, 2017) and in many schools, teachers primarily receive feedback from their principals, who often lack the time and knowledge to support teachers' thorough analysis and synthesis of evaluation data (Firestone & Donaldson, 2019; Rigby et al., 2017).

Technology has emerged as a promising way to fill the gaps in teacher professional learning by providing teachers with data-driven opportunities to facilitate instructional improvement. Computerized tools can help teachers refresh their pedagogical content knowledge and rehearse in simulated environments (Copur-Gencturk et al., 2024; Markel et al., 2023), receive support for responding to students' written explanations in between class sessions (Bywater et al., 2019, 2023), or reflect on recordings of their instruction through video (Chen, 2020) or text analytics (Jacobs et al., 2022; Demszky et al., 2023). Different from tools that provide computer-scaffolded instruction directly to students (e.g., Cognitive Tutor; Anderson et al. (1995)), these teacher-facing tools represent the potential of computer technology to influence human-to-human instructional quality broadly through numerous mechanisms.

Using natural language processing (NLP), tools that provide formative feedback to teachers based on their instruction have shown promise to complement human observation and coaching, improving instructional practice (Jacobs et al., 2024) and even student outcomes in online settings (?Demszky & Liu, 2023). Such automated feedback tools take a recording of a teacher's lesson as input, transcribe and analyze the recording to identify high-leverage instructional practices, and deliver insights to the teacher to facilitate reflection and instructional improvement. Since such feedback is cost-efficient, scalable, and can be delivered privately, quickly, and frequently, researchers and technology providers (e.g. TeachFX, EdThena) are seeking to understand how they can be best put to teachers' service.

Despite the need for scalable K-12 classroom observation and feedback tools a well as encouraging studies with such tools in online teaching settings, to our knowledge, there exists no rigorous experimental evaluation of whether automated feedback might work in K-12 in-person learning contexts. To this end, we present results from an experiment in which we provided teachers with feedback related to focusing questions — a high-leverage teaching practice involving asking questions that probe student thinking and encourage students to reflect on their thoughts and those of their classmates (Alic et al., 2022; Herbel-Eisenmann & Breyfogle, 2005; National Council of Teachers of Mathematics, 2014; Wood, 1998). Our experimental design allows us to causally estimate whether providing teachers feedback about focusing questions increases the number of such questions and whether it yields related improvements in instruction, such as increasing the amount of student talk and student reasoning.

We also add to the literature on human-computer interaction by seeking to understand how teachers engage with and perceive the utility of our automated feedback. Prior research on technology integration in classrooms indicates that teachers' perceived utility of the technology has a strong influence on their technology adoption (Backfisch et al., 2021; Ertmer et al., 2012; Fütterer et al., 2023; Kale, 2018; Scherer et al., 2019; Q. Wang & Zhao, 2023). This suggests that teachers need to see the value of receiving automated feedback on their instruction in order to effectively use the tool. Jacobs et al. (2022, 2024) have found that many teachers see automated feedback as a valuable vehicle for self-reflection, but that perceptions of accuracy can impact their engagement with such feedback. We seek to deepen this knowledge about factors that may impact teachers' perception of and engagement with automated feedback by conducting qualitative interviews with a subset of teachers in the experimental study.

Thus this mixed-method study is the first that combines a pre-registered randomized controlled trial and qualitative interviews to *experimentally* test the impact of and *describe* teacher engagement with automated feedback on instruction in K-12 in-person classrooms. In doing so, we address the following research questions:

1. To what extent do K-12 teachers engage with the automated feedback on focusing questions?

2. Does the automated feedback on focusing questions impact instruction, including teachers' use of focusing questions, student talk time, and student reasoning? [1]

We augmented these questions with a third question, which we seek to answer with our qualitative interviews in this mixed-methods study.

3. How do teachers perceive the automated feedback on both focusing questions and other teaching practices? What are the barriers for them to engage with the feedback?

To answer these questions, we partnered with TeachFX, a company that delivers feedback to teachers based on classroom recordings via a phone application. We leveraged TeachFX's newly established partnership with the state of Utah to facilitate professional learning for mathematics and science teachers. We randomly assigned teachers to a treatment or control condition based on whether they received automated feedback on focusing questions. We collected recordings of their instruction, post-study surveys, and interview data to understand the impact of the treatment as well as teachers' engagement with and perceptions of automated feedback.

In the following sections, we begin with an overview of related work on technology-based professional learning for teachers and teachers' technology integration. Subsequently, we

[1] The first two research questions are pre-registered (`https://www.socialscienceregistry.org/trials/11258`). The pre-registration also included an additional research question about how the feedback changes teachers' perception of their own instruction, and it also included heterogeneity analyses (e.g. how the impact of the intervention varies by teacher characteristics). However, due to a low survey response rate (n=95), 18% of the original sample), we had to exclude these research questions from our study.

provide a background for our current study (Section 3), with details on the technology platform and participants. In Section 5, we describe the experimental design, including the randomized setup, study procedures and the interview protocol. In Section 6, we provide an overview of the approach we took to answer each research question. In Section 7, we provide the results of our research questions. We conclude by discussing the implications of these results for both research and practice related to using computerized tools in teacher professional learning.

2 Related Work

2.1 Productive Teacher Talk

A large body of education research has shown that teacher talk that encourages students to verbalize, share and co-construct knowledge improves student learning, agency and sense of belonging (Alexander, 2020; Asterhan et al., 2015; Chapin et al., 2009; Howe et al., 2019). Conceptualized under several related frameworks (dialogic instruction, accountable talk, academically productive talk (Michaels et al., 2008; Resnick et al., 2010)), such teacher talk includes moves like pressing students for reasoning, challenging their ideas, and inviting them to engage with each others' ideas. These moves can facilitate individual reflection and social cognition processes that enrich learning across subjects (Adey et al., 2002; Chapin et al., 2009; Topping & Trickey, 2007; Webb et al., 2014).

One such set of moves is focusing questions, which are the topic of this study. Focusing questions are a core building block of dialogic instruction, seeking to evoke student reasoning and allow teachers to take up student contributions (Herbel-Eisenmann & Breyfogle, 2005). As such, they are a primary focus of teacher professional development (Oliveira, 2010; Pehmer et al., 2015). Unlike funneling questions that "lead students to a desired procedure or conclusion, while giving limited attention to student responses that veer from the desired path," focusing questions "attend to what the students are thinking, pressing them to communicate their thoughts clearly, and expecting them to reflect on their thoughts and those of their classmates" (National Council of Teachers of Mathematics, 2014). For example, "What is the slope of this line?" is a funneling question, while "What do you mean by the angle of the line?" or "What do you think Kara means?" are examples of focusing questions (Herbel-Eisenmann & Breyfogle, 2005). Teachers' use of focusing questioning patterns has been linked to better student learning outcomes and confidence in mathematics (Franke & Kazemi, 2001; Hagenah et al., 2018).

Although studies confirm the academic and developmental advantages of dialogic talk moves like focusing questions, establishing student-centered discussion environments represents a major pedagogical shift that many teachers struggle with in practice (Chen, 2020). Teacher-led recitation still dominates most classrooms (Demszky & Hill, 2023; Kane & Staiger, 2012; O'Connor & Snow, 2017; Resnick, Asterhan, Clarke, & Schantz, 2018) likely due to inadequate training and school-level support for teachers (Resnick, Asterhan, & Clarke, 2018; Resnick, Asterhan, Clarke, & Schantz, 2018). Thus, professional learning aimed at enhancing teachers' knowledge, beliefs, and self-efficacy around dialogic talk moves can potentially improve learning environments for students (Chen, 2020; Jacobs et al., 2024).

2.2 Technology-Based Professional Learning Tools for Facilitating Teacher Talk

Recent efforts have sought to complement expert-led professional learning (e.g. instructional coaching) by leveraging technology to facilitate teachers' self-guided improvement at scale, allowing them to practice in simulated environments (Copur-Gencturk et al., 2024; Markel et al., 2023) and to revisit (Chen, 2020; Sherin & Dyer, 2017) and receive automated feedback on their recorded lessons (e.g., Jacobs et al., 2022). Automated feedback is generated by automatically transcribing classroom recordings, computationally analyzing the transcripts, and surfacing insights from these analyses to the teacher. In addition to measuring the *quantity* of talk (e.g. teacher talk time) (Z. Wang et al., 2013), researchers have developed several NLP measures that can analyze the *quality* of teacher and student talk in classroom transcripts. Such NLP measures tend to focus on detecting teacher dialogic talk moves such as pressing for reasoning, connecting students' ideas, building on students' ideas (Demszky et al., 2021; Donnelly et al., 2017; Jensen et al., 2020; Kelly et al., 2018; Samei et al., 2014), as well as growth-mindset and autonomy-supportive talk (Hunkins et al., 2022). Moving beyond measurement to teacher feedback, Suresh et al. (2021) introduces the TalkMoves application that provides teachers with information on the extent to which they use dialogic talk moves, including pressing for accuracy and revoicing student ideas. Similarly, **?** introduces the M-Powering Teachers application that provides feedback to teachers on their talk time and uptake of student ideas.

While new methods and tools for automated teacher feedback are emerging, the field still lacks data and rigorous evidence about whether such tools indeed improve teaching and student outcomes. For the limited number of tools studied by scholars, the results appear promising. Jacobs et al. (2022, 2024) found that the TalkMoves application was perceived positively by K-12 mathematics teachers. Using a pre/post design with 21 participants, Jacobs et al. (2024) found that teachers increased their use of accountable talk moves, especially revoicing, across the two years of the study.

To date, two randomized controlled trials in online environments have demonstrated the success of the M-Powering Teachers tool in improving instruction. In an online computer science course, researchers randomly assigned half of the instructors to receive an email reminder to check automated feedback on their uptake of student ideas. They found that instructors in the treatment group improved their frequency of taking up student ideas by 13 percent compared to the control group (**?**). Similarly, in an online, one-on-one tutoring program that aimed to improve high school students' research skills, tutors who were offered automated feedback improved their uptake of student contributions by 10 percent compared to tutors who did not have access to the feedback (Demszky & Liu, 2023). Findings from these studies also suggest that students taught by instructors who received feedback had more favorable perceptions of their learning experience compared with instructors who did not receive such feedback and were more likely to complete assignments. The current study fills the gap by being — to our knowledge — the first randomized controlled trial to experimentally test the impact of automated discourse-based teacher feedback in in-person K-12 instruction settings.

2.3 Factors that Impact Teachers' Technology Integration

While randomized experiments may provide valuable evidence about the impact of technology on teachers' behavior, understanding teachers' perceptions of technology and the factors that influence their use of technology is critical to the development of effective and user-friendly tools that support teachers. A long line of research describes factors influencing teachers' technology adoption. Many studies use the Theory of Planned Behavior (TPB) proposed by Ajzen (1991) as a framework to explain how teachers' attitudes toward technology, perceived social norms, and perceived degree of control shape their intention and actual behavior (e.g., An et al., 2022; Eksail & Afari, 2020; Seufert et al., 2021; Stinken-Rösner et al., 2023; Teo, 2011; Watson & Rockinson-Szapkiw, 2021). For example, Watson & Rockinson-Szapkiw (2021) found that key TPB constructs, including attitudes and beliefs towards technology, perceived social norms, and perceived behavior control (i.e. ease of use) predict pre-service teachers' intention to use technology-enabled learning. Scherer et al. (2019) analyzed findings from 114 survey studies (N=34,577 teachers), finding that self-efficacy and perceived usefulness largely predicted teachers' intention to use and frequency of using technology. The key role of perceived utility in teachers' technology integration is corroborated by several other studies (Backfisch et al., 2021; Ertmer et al., 2012; Fütterer et al., 2023; Kale, 2018; Scherer & Teo, 2019).

Several factors may inhibit teachers' perceived utility of and thereby adoption of technology, including technostress (Q. Wang & Zhao, 2023), reliability (Butler & Sellbom, 2002), time constraints (Bauer & Kenton, 2005; Francom, 2020) and privacy concerns (Dinc, 2019). Technostress refers to the negative impacts of technology usage on an individual's attitude, psychology and behaviour (Jena, 2015; Tarafdar et al., 2019), and is affected by factors such as techno-complexity (i.e. having to learn technical knowledge continuously), techno-overload (i.e. having to change work habits and increase work speed), techno-invasion (i.e. invasion of work into private time via technology) and techno-insecurity (i.e. worries about one's job being replaced due to technology) (Tarafdar et al., 2019). Q. Wang & Zhao (2023) found techno-complexity and techno-insecurity to harm teachers' intended technology adoption as they increase teacher effort. Techno-overload and techno-invasion were found to have a positive effect on technology integration by some studies (Farhoomand & Drury, 2002; Q. Wang & Zhao, 2023) and a negative one by others (Hung et al., 2015; Li & Wang, 2021), suggesting that these stressors can be moderated by other factors such as teachers' perceived control.

These stressors and challenges can be mitigated by providing better support to teachers, through improved pre-service teacher training, positive school culture, and continuous professional learning opportunities (Spiteri & Chang Rundgren, 2020). Lachner et al. (2021) found that providing teachers with training on how to best advance learning goals through the technology (also referred to as technological pedagogical content knowledge or TPACK) improved their TPACK as well as self-efficacy and their perceived support for technology integration. As for school-level support, technical help, positive culture, and community and encouragement to use technology are all factors that facilitate teachers' technology adoption (Spiteri & Chang Rundgren, 2020; Tondeur et al., 2016). Our qualitative interviews build on this line of work to investigate factors that influence teachers' adoption of technology-based feedback tools.

3 Feedback via The TeachFX Platform & Email

In this section, we describe TeachFX (Section 3.1), the platform we partnered with to deliver feedback as part of the study. We then provide an overview of how we delivered feedback on focusing questions via the platform (Section 3.2) and an email (Section 3.3).

3.1 The TeachFX Platform

We conducted the study in partnership with TeachFX[2], an education technology platform that provides teachers with automated feedback to improve their instruction. Teachers use a mobile application to record their instruction. TeachFX then automatically transcribes and analyzes the recording using NLP tools. Within a day of the recording, teachers receive an email to view their class report on the TeachFX platform.[3] The class report includes the full transcript of the recording as well as insights related to student and teacher talk in the class, including teacher talk time, wait time, the incidence of longer student contributions, and a word cloud representing the frequency of terms used by the students and the teacher. Appendix A contains a screenshot of the class report and a list of insights available to teachers at the time of the study. All teachers in the study had access to the TeachFX class report, but only a random subgroup of teachers had access to feedback on focusing questions via the platform and an email, as described below. Section 5 explains the experimental setup in greater detail.

3.2 In-Platform Feedback on Focusing Questions

Figure 1 shows a screenshot of the focusing question insight we deployed within the TeachFX platform. The insight displayed a transcript heatmap with markers indicating where focusing questions occurred. Teachers could click on these markers to listen to relevant recording

[2] `https://teachfx.com/`

[3] Given that emails and the TeachFX platform are the only methods used by TeachFX to deliver automated feedback, we also use these methods to deliver feedback on focusing questions in our study.

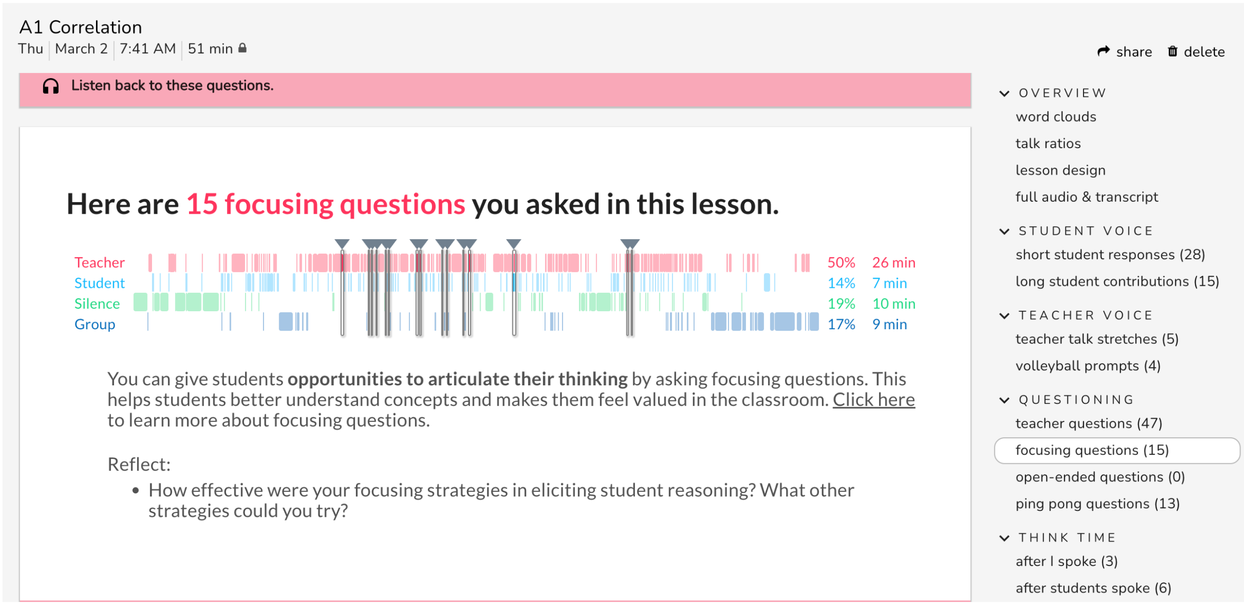


Figure 1: Screenshot of the Focusing Questions insight within the TeachFX app.

segments. The insight also included a brief summary of focusing questions, and a link to a more detailed blog post[4] explaining what focusing questions are and how to ask more focusing questions. Finally, we included a prompt "How effective were your focusing strategies in eliciting student reasoning? What other strategies could you try?", to encourage teachers to reflect on their use of focusing questions and to set goals for their next lesson.

Detecting focusing questions. Focusing questions were identified in classroom transcripts by our machine learning model. Given a transcript of a class recording, we extracted teacher utterances and provided them as inputs to a binary classification model, which told us whether each utterance was a focusing question. We obtained this model by fine-tuning BERT-base (Devlin et al., 2018)[5], a pre-trained language model, on labeled data from the NCTE elementary mathematics classroom dataset (Alic et al., 2022; Demszky & Hill, 2023), which we augmented with 694 annotated examples from TeachFX transcripts to facilitate adaptation to the target domain. To obtain labels for TeachFX data, we recruited two experienced mathematics instructional coaches to annotate teacher utterances for the pres-

[4] `https://medium.com/dorademszky/resources-for-focusing-questions-47bc6cdd9953`

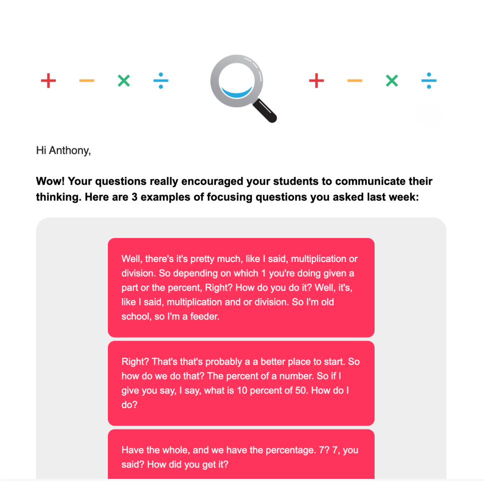
[5] We experimented with RoBERTa-base (Liu et al., 2019) as well but found that BERT performed better.

ence of focusing questions using the annotation guide described in Alic et al. (2022). We then fine-tuned the BERT model on the combined dataset from NCTE and TeachFX, with 200 randomly sampled examples in the held-out validation set and ∼2500 examples in the training set. The data included 25% positive examples (focusing questions) and 75% negative examples (all other teacher utterances). The full training procedure is described in Appendix B. The fine-tuned model achieves a 90% accuracy on the held-out set (precision: 80%, recall: 55%, F1 score: 64%). We acknowledge that sparse practices such as focusing questions are challenging for classifiers to detect with high recall and there is room for improving our model, for example by incorporating additional context (e.g. wait time after questions, or subsequent questions) or augmenting the training data. In Section 6, we discuss the implications of having measurement errors in focusing questions for our analysis.

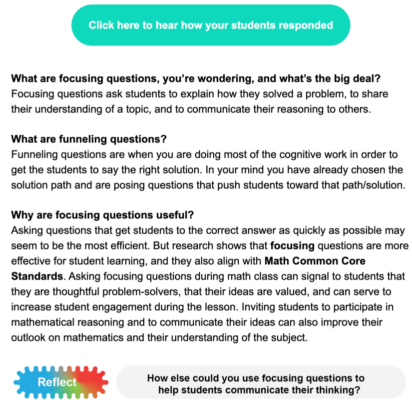
3.3 Email Feedback on Focusing Questions

Teachers in the treatment group also received targeted feedback on their use of focusing questions via a weekly email, illustrated in Figure 2. The email contained both the number of focusing questions the teacher asked in all class recordings in the previous week as well as up to three of their top focusing questions. The top questions (Figure 2a) were identified by two expert annotators, mathematics instructional coaches with decades of experience to ensure that we selected the best questions to reinforce and extend this teaching move.[6] The email included an explanation of focusing questions (Figure 2b), a link to the insight page on the TeachFX app, and a blog post elaborating on focusing questions (Figure 1). Further, it included the top three focusing question starters seen in the week across treatment group teachers (Figure 2c), identified by the same expert annotators.

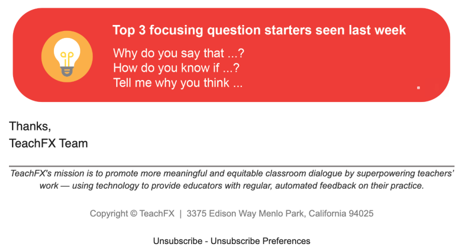
[6]We conducted the selection step manually because we do not yet have an automated way of ranking focusing questions, and we wanted to ensure that the best examples are indeed picked. Although the selection of the top three questions was manual, we conducted a comparison that showed that annotators performed this selection from the automatically identified questions seven times faster than if they had to select examples from all teacher questions. Future work can explore effective approaches to automate the ranking process entirely.



(a) Email Part 1: Focusing question examples.



(b) Email Part 2: Button linking to insight on TeachFX platform (Figure 1), definition of focusing questions, reflection questions.



(c) Email Part 3: Example focusing question starters derived from transcripts of study participants.

Figure 2: Email about focusing questions, sent once a week to treatment group teachers.

4 Participants

TeachFX had recently formed a new research partnership with the state of Utah, as part of which mathematics and science teachers were encouraged to use TeachFX and received professional development opportunities related to automated feedback. Because these new users were not biased by exposure to automated feedback beforehand, they were ideal participants for the study. Furthermore, our target construct – focusing questions – applies to both mathematics and science instruction (National Council of Teachers of Mathematics, 2014; Hagenah et al., 2018; Lemke, 1990; Smart & Marshall, 2013). Our experiment involved all 523 Utah users who made their first recording with TeachFX after the beginning of the study. During post-processing, we excluded participants who did not meet our study criteria, as described below.

Selection criteria for the analytic sample. Our original, unfiltered sample included 523 users from Utah who made their first recording after October 10, 2022. We used this sample to randomize participants to study conditions, and all participants within this sample received the final survey (see survey details in Section 5.3). After the experiment, we excluded participants from the analytic sample who did not meet our pre-registered study criteria. Since our feedback on focusing questions is designed for mathematics/science instruction, we excluded teachers who indicated that they did not teach mathematics/science in the endline survey. If they did not respond to the survey, we observed their recordings and excluded teachers for whom most recordings were in subjects other than mathematics/science. We also focused on teachers in self-contained classrooms and thus excluded participants who indicated in the endline survey that they do not have their own classroom and/or if they indicated via TeachFX that they were recording on behalf of another teacher (e.g. as an instructional coach). These manual filtering steps were performed blind to teachers' assigned conditions in the study. We then excluded lessons made past week 1 that were shorter than 10 minutes, which led to the removal of a few additional teachers from the sample.[7] **Our final analytic sample includes 224** teachers.

Demographic characteristics. Among teachers in the analytic sample who filled out the endline survey (n=60), 87% are female, 10% are male, and 3% preferred not to report their gender. 83% are White or Caucasian, 5% are Hispanic or Latinx, 3% are Asian, and the rest of the teachers identified themselves as multiracial. In terms of teaching experience, the majority of teachers have 8 or more years of experience (61%); 16% have 3-4 years of experience; the rest vary. 37% of teachers teach in elementary grades, 33% of teachers teach

[7] We only applied the duration filter to recordings past week 1 because recordings from the first week only served to provide baseline (pre-feedback) covariates rather than predictors or outcome variables. Further, these teachers still received feedback even if their first recording was short. Thus, to maximize the sample size, we kept these short baseline recordings.

in middle grades, and 28% of teachers teach in high school grades. About 85% of teachers have a regular (non-special education) classroom.

5 Experimental Design & Procedures

We conducted a randomized controlled trial to evaluate the effectiveness of providing feedback to mathematics and science teachers on focusing questions. The study was approved under institutional IRB. Our experiment ran for **five months** between October 10, 2022, and March 10, 2023. We ended the study in March because of the start of the standardized testing season, which interfered with teachers' bandwidth to use the tool. After a teacher completed five weeks of recordings during the study period, we considered their participation in the study complete.

Below we describe the experimental design and procedures, including the randomization (Section 5.1), incentives for recording (Section 5.2), the endline survey (Section 5.3) and semi-structured interviews conducted at the end of the study (Section 5.4).

5.1 Randomization

We randomly assigned participants to treatment and control conditions to test the impact of feedback on focusing questions on their instruction (RQ2). Randomization helps balance participant characteristics across conditions. Teachers were randomly assigned to a condition once they made their first recording during the study period. The random assignment was made via a hashing function within the TeachFX platform, which is similar to a coin flip.

During the study period, all participants (in both conditions) had access to TeachFX's platform, standard feedback that TeachFX offers, and a class report email that nudged teachers to view the platform once they made a recording. Teachers in the treatment group received additional feedback by having access to the focusing insight page (Section 3.2) and an email about focusing questions (Section 3.3). The email was delivered every Tuesday early morning to teachers in the treatment group who recorded in the past week.[8] This experimental setup allowed us to test the effectiveness of sending teachers one additional distinct piece of automated feedback on a high-leverage teaching practice among teachers who already had access to TeachFX's automated feedback. While not a direct test of the effects of automated feedback vs. no such feedback, our approach has the benefit of not denying any participating teacher access to TeachFX services.

5.2 Incentives for Teachers to Record Lessons

Because of declining participation as the school year progressed, TeachFX incentivized teachers in both conditions to record via a raffle. These raffles involved the opportunity to win Amazon gift cards worth $250. TeachFX offered four raffles during the study period to all users in Utah, regardless of study condition. The raffles incentivized teachers to record at least once per week during each of the four different periods (Oct 10-Nov 10, Dec 5-Dec 23, Feb 6-Feb 17, Feb 27-March 10) for a chance to win a $250 gift card during that period. To remind teachers about the raffles, TeachFX sent out emails to all Utah users each month who have made at least one recording that month. This email included a reminder about the raffle periods and encouraged teachers that they were on track to be considered for the raffle prize if they continued to record every week.

5.3 Endline Survey

By default, the only information TeachFX collects about users is their role in the school (Administrator, Instructional Coach, Teacher). To better understand the demographics of our sample, TeachFX administered a survey to participants who completed the study (response rate=18%, n=95 out of 523, see questions in Appendix C). All teachers were offered a $10 Starbucks gift card for completing the survey. Teachers' assigned condition

[8]We decided to send the email early mornings to maximize the chances of teachers reading it before their workday; we sent it on Tuesday rather than Monday as we expected teachers to be the busiest on Mondays.

did not have an impact on survey completion (see Section 6.2). The survey asked teachers to self-identify their gender, race/ethnicity, subjects taught, grade levels taught, years of teaching experience, and whether they teach their own classrooms or instead offer special types of instruction (e.g. small groups, tutoring). The low response rate heavily limits the representativeness of the survey information.

5.4 Semi-structured Interviews

To better understand teachers' perception of the automated feedback, barriers to access, and suggestions for improvement (RQ3), we conducted qualitative interviews that asked teachers about their experiences engaging with the TeachFX platform and feedback and, for teachers in our treatment group, engagement and perceptions of the feedback on focusing questions. We conducted these interviews with thirteen teachers shortly after the experiment finished. We recruited interviewees by working with TeachFX to email all participants who completed the post-treatment survey. Among the fifteen teachers who signed up to be interviewed, we reached out to a diverse group of study participants based on their reported demographics, years of teaching experience, and treatment status. We interviewed teachers in two waves between March and May of 2023. We interviewed six during the first wave and seven during the second wave. Among the thirteen interviewees, seven were from the treatment group, and the other six were from the control group. Eight were working in elementary schools, four in secondary schools, and one in both. All teachers have more than eight years of experience in teaching. In terms of teaching roles, eight interviewees are general education teachers, three are special education teachers, and another two are former teachers who now serve as instructional coaches in their district.

Interviews were conducted virtually over Zoom. Each interview lasted 1-1.5 hours. The interviews asked teachers about their teaching background, previous experience incorporating technology into their teaching, experiences setting up the environment for recording, experiences with the feedback tools, and any feedback they might have regarding the feedback tool. Appendix D includes the full interview protocol. All interviews were recorded and transcribed using Zoom. Two graduate assistants and a Ph.D. student checked and corrected the auto-generated transcript for accuracy. Participants were compensated $50 per hour for the interview.

6 Analytic Approach

This section outlines the analytic approach we took to answer each of the three research questions, leveraging quantitative data from TeachFX and qualitative data collected from the interviews.

6.1 Engagement with Automated Feedback on Focusing Questions (RQ1)

To understand how often teachers engage with the automated feedback, we quantify their interactions with the email and the feedback page. We quantify engagement with the feedback on focusing questions by tracking whether a teacher opened the focusing questions email and whether they viewed the focusing question insight on the TeachFX platform.

6.2 Impact of Feedback on Focusing Questions on Teachers' Instruction (RQ2)

To understand the impact of our intervention on instructional practice, we quantify several discourse features within recordings collected during the study. We conducted regression analyses to compare the use of these discourse features by study condition.

Collected recordings. The analytic sample consists of 1,509 recorded lessons. This sample was obtained after our pre-registered filtering processing (see Section 4) and removed recordings that teachers made past five weeks. Teachers in the sample on average made 1.85 recordings per week over 3.64 weeks (SD=1.25, range=2-5), for a total of 6.74 unique recordings per teacher (SD=5.02, range=2-37). The average duration of recordings is 34 minutes (SD=20, range=0.1-120). TeachFX computes the percentage of student talk transcribed, which can be an indication of recording quality (e.g., noisiness), which affects downstream transcription, especially for student speech that tends to generally cause performance issues for transcription systems. In the sample, 41% of student talk is transcribed on average (SD=19%, range=0-100%). Given the variation in the data in terms of recording quality, we control for this variable in our analyses.

We conducted analyses at the teacher-week level by first combining transcripts within the same week and computing the rate of discourse features on the combined transcripts. This process yields 533 combined recordings. We combined recordings within a week in order to avoid short recordings having a similar weight in the analyses as longer, and thus likely more representative, recordings.

Discourse features. Our primary discourse-related outcome is the rate of focusing questions per hour. In addition, we quantify the percentage of student talk time, the rate of teachers' uptake of student ideas (Demszky et al., 2021), and the rate of student reasoning per hour (Demszky & Hill, 2023) because these are expected downstream instructional outcomes from teachers' use of focusing questions. We use hourly rates instead of raw counts to account for differences in recording duration. For uptake and student reasoning, we use off-the-shelf classifiers described in the corresponding papers that had been validated on existing datasets of math classroom recordings and had been implemented by TeachFX to provide feedback to teachers. We conducted manual checks to ensure that these models work appropriately for our current data when student speech is transcribed, but measurement error is still expected. Assuming that the measurement error is randomly distributed based on condition (e.g., as Table 1 indicates, the amount of student talk transcribed is similar across conditions), we can still use these models to estimate differences in discourse patterns by condition.

We use these features not only because we hypothesized that an increase in focusing questions would lead to an improvement along these discourse features, but also because we observed positive correlations between them and student outcomes in prior work (Demszky et al., 2021, 2023; Demszky & Liu, 2023; Demszky & Hill, 2023). In fact, in our pre-intervention transcripts, too, the rate of focusing questions correlates significantly with student talk percentage ($\rho = 0.2$, $p < 0.001$), teachers' uptake of student ideas ($\rho = 0.18$, $p < 0.001$) and student reasoning ($\rho = 0.61$, $p < 0.001$). Thus, we expected that an improvement in focusing questions would increase student talk time, teachers' uptake of student ideas, and student reasoning during the intervention.

Validating randomization. To verify whether our randomization created groups that were balanced on observable variables, we evaluate whether the demographics of instructors and survey response rates in the treatment and control groups differ statistically. We also compare instructors' discourse features measured in their first recorded lesson prior to receiving feedback. We use all participants in the study regardless of the number of weeks of recordings or whether their recordings are valid. As Table 1 shows, other than the duration of the first recording being marginally different, we do not find statistically significant differences between conditions in any of the teacher demographics and discourse features of the first recording. This analysis suggests that any differences we observe later in the study are likely driven by the effects of the intervention.

Attrition analyses. We also conducted an attrition analysis to examine whether the treatment and control conditions suffered from differential attrition. The results are presented in the bottom panel of Table 1. Attrition in our data occurred when teachers recorded for fewer than five weeks (and thus did not complete the study) or made invalid recordings. We

	Control Mean	Treatment Mean	P Value	N
Survey completed	0.17	0.2	0.4	523
Female	0.82	0.82	0.98	95
White	0.8	0.88	0.29	95
Teaches Mathematics	0.76	0.84	0.31	95
Teaches Science	0.27	0.36	0.33	95
Teaches Elementary	0.42	0.46	0.71	95
Teaches Middle School	0.31	0.2	0.22	95
Teaches High School	0.31	0.22	0.32	95
Duration (minutes)	27.03	30.36	0.07	523
Focusing rate	28.15	26.88	0.58	523
Uptake rate	4.81	5.04	0.78	520
Student reasoning rate	3.35	3.32	0.96	520
Student talk percentage	21.91	21.78	0.94	523
Proportion of student talk transcribed	0.47	0.46	0.51	501
Week of first recording	7.53	7.79	0.62	523
Opened TeachFX class report email	0.13	0.12	0.57	523
Attrition				
Number of weeks teacher recorded	2.48	2.62	0.32	523
Number of unique recordings	1.69	1.89	0.26	523
Invalid recording	0.36	0.33	0.55	523

Table 1: Data used here include all participants *before* data filtering. A randomization check shows that the treatment and control group characteristics do not differ significantly. Rates for different discourse moves describe frequencies per hour. The week of the first recording represents the week during the RCT period when the teacher made their first recording. The attrition values show that attrition in the data (due to validity or lack of recording) is not affected by the randomization.

find no differential attrition in the sample, suggesting that the intervention did not have a significant impact on teachers' likelihood of recording valid lessons.

Regression analyses. We conducted a preregistered intent-to-treat (ITT) analysis with an ordinary least squares regression, at the teacher-week level. This analysis compares teachers' discourse features regardless of whether they chose to engage with the automated feedback. We compare participants by condition rather than by compliance with treatment (checking the feedback), since the latter may introduce selection bias, interfering with our causal estimation: teachers who comply may have certain characteristics (e.g. time or motivation to use self-directed professional development) that non-compliant teachers do not. Furthermore, understanding the overall impact of offering feedback on instruction (rather than the impact of taking up feedback) is likely more relevant to schools and districts that are considering adopting this technology.

The models are specified as below:

$$Y_{iw} = \beta_1 T_i + \beta_2 \boldsymbol{X}_i + \beta_3 \boldsymbol{M}_{iw} + \varepsilon_{iw} \tag{1}$$

where Y_{iw} refers to a particular dependent variable for teacher $i's$ and week w; T is a binary variable that indicates the treatment status, with a value of 1 indicating the treatment condition and 0 otherwise; $\boldsymbol{X}$ is a vector of teacher-level covariates, $\boldsymbol{M}$ is a vector of transcript metadata; β_1 is the parameter of interest, which measures the treatment effects of our intervention on teacher outcomes; and ϵ indicates the residuals. We clustered standard errors at the teacher level to account for repeated observations within a teacher.

We fitted this model to estimate the effects of the treatment on each of the following dependent variables: the number of times a teacher asked a focusing question per hour, the number of times they took up student ideas per hour, the number of student reasoning instances per hour, and student talk percentage. Since the outcomes are computational estimates, they might contain measurement error. Unlike the classical measurement error issue in independent variables, having measurement error in dependent variables would not bias our estimates as long as it does not correlate with our independent variable (Wooldridge, 2019; Pischke, 2007). In our context, we believe that the measurement error in outcomes do not correlate with the treatment status due to the random assignment. However, it might increase the standard errors in our estimates, affecting our statistical inferences. We discuss our findings with caution regarding this issue.

We used the following binary variables as teacher covariates $\boldsymbol{X}_i$ across all models, derived from the final survey: female, White, having had at least 5 years of teaching experience, teaching mathematics, and teaching in elementary, middle or high school. Missing survey data was assigned a value of zero and we included a binary indicator for whether the data was imputed. We also included teachers' baseline discourse features in their first week of recording as covariates: the rate of focusing questions, rate of teachers' uptake of student ideas, rate of student reasoning, student talk percentage and percentage of student talk transcribed. We also included two variables as covariates $\boldsymbol{M}_{iw}$: indicators for the week w of recording for teacher i during the RCT and percentage of student talk transcribed for teacher i in week w (as an indication of recording quality).

To verify that our estimates are not significantly influenced by the choice of demographic control variables, especially given the low survey response rates, we also estimated a version of our models without those variables.

6.3 Teachers' Perception of Automated Feedback (RQ3)

We conducted a thematic analysis of the interview data. We adopted a convergent approach (Fetters et al., 2013), analyzing quantitative and qualitative data simultaneously and comparing the results from both analyses to see how the interview data confirms or helps explain quantitative findings. We started with inductive and deductive coding of the interview transcripts using the NVivo software. Based on the literature in Section 2, the research team developed a codebook to capture teachers' barriers to engagement, suggestions to enhance engagement, treatment group experience, and feedback specific to TeachFX. Two Ph.D. students worked on coding all interview transcripts. The coders used one interview from the treatment group and one from the control group for training and to ensure that intercoder agreement reached 100% before coding interviews individually. After the first round of coding, coders reviewed each other's coded interviews to check for confusion or disagreement. All coding differences were discussed until an agreement was reached. To answer RQ3, the research team came together to discuss the coded data and summarize major themes. The codebook and examples are included in Appendix E.

7 Results

7.1 RQ1: To what extent do teachers engage with the automated feedback on focusing questions?

We tracked two key metrics of engagement with the automated feedback: email opens and views of the focusing question insight on the TeachFX platform. Overall, treatment group teachers opened the emails with the feedback at a much higher rate than viewing the insight on the platform. Between 53-65% of teachers opened their emails across weeks (65% for 1st email, 53% for 2nd, 61% for 3rd, 55% for 4th and 56% for 5th), but only 17-23% of them viewed the focusing insight page (23% for 1st week, 21% for 2nd, 21% for 3rd, 17% for 4th, 25% for 5th week). Similarly, while 76% of teachers in the treatment group opened at least one email, only 43% of them viewed the insight page at least once. On average, teachers opened 2.14 weekly emails (SD=1.65) throughout the RCT. These results suggest more teachers accessed the feedback via the email than the platform, likely because accessing the platform required an extra step (teachers clicking the link in the email or visiting the TeachFX platform separately). However, there is room for improving teachers' consistent engagement with both the email and the platform.

7.2 RQ2: What impact does automated feedback on focusing questions have on teachers' instruction?

As shown in Table 2, the treatment significantly increased teachers' use of focusing questions. On average, treatment group teachers asked 4.61 more focusing questions per hour, indicating a 20% increase compared to the control group ($p < 0.01$). This effect size is greater than observed in prior interventions related to automated feedback on teachers' uptake of student ideas (Demszky et al., 2023; Demszky & Liu, 2023). The small standard errors mitigate potential concerns about measurement error in focusing questions. In contrast, the intervention

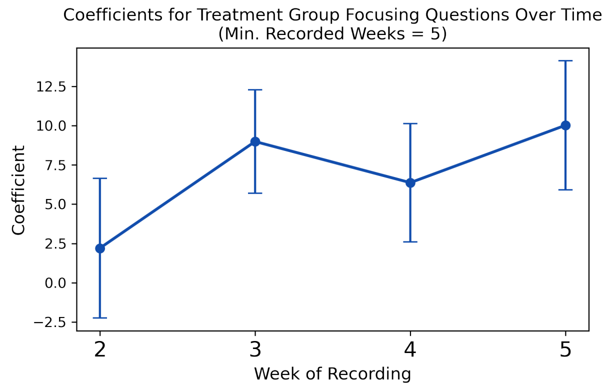
	(1) Focusing rate	(2) Uptake rate	(3) Student reasoning rate	(4) Student talk percentage
Treatment	4.612**	0.274	0.655	0.001
	(1.741)	(0.523)	(0.485)	(0.012)
Control Mean	22.565	3.772	2.896	0.160
R^2	0.346	0.319	0.226	0.244
Observations	533	533	533	533

Table 2: Standard errors are in parentheses. ** p<0.01. These models estimate the effect of the automated feedback on focusing questions (treatment) on teachers' discourse features. Talk move rates are calculated per hour. All models include covariates listed in Section 6.2. We observe a statistically significant impact on focusing questions but not the other discourse features. The control means are the averages of the outcome variables for the control group, which are presented here to contextualize the effect sizes.

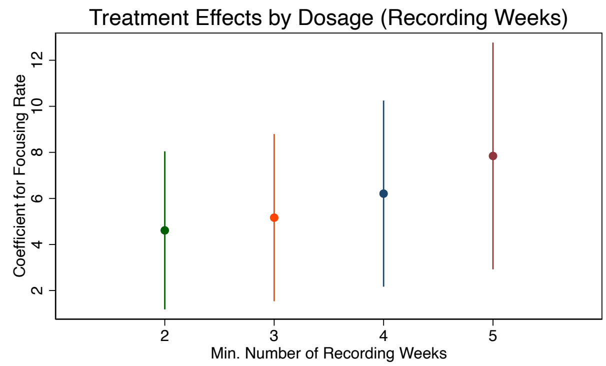
did not have an impact on other discourse features besides focusing questions. Specifically, we do not see any statistically significant effect on student talk time, student reasoning, or teachers' uptake of student ideas, rejecting our hypothesis that improving teachers' use of focusing questions would lead to an improvement in other aspects of instruction.

To examine how effect sizes for focusing questions change over time, we analyzed the sample of teachers who completed the RCT, recording for at least five weeks (n=83). Given that teachers' persistence in recording depends on their motivation and perception of the feedback, our estimates are only suggestive rather than indicating a causal relationship between number of recordings and impact on instruction. As Figure 3a shows, the effect size increases from week 2 to week 3 and then generally stagnates. The greatest effect size hovers around 10 additional focusing questions per hour in week 5, indicating a 59% increase compared to the control group. This pattern aligns with previous studies (Demszky et al., 2023; Demszky & Liu, 2023), indicating that feedback effects typically emerge after 2-3 weeks before stagnating or diminishing. This suggests the potential for varying feedback after the initial weeks.

To shed light on how teachers' engagement with the tool might impact treatment effects, we explored the correlation between effect sizes and the minimum number of weeks a teacher recorded. As Figure 3b shows, when we include teachers who recorded as few as two weeks in



(a) Coefficients for the rate of focusing questions in the treatment group, plotted over week of recording. This sample only includes teachers who recorded for at least five weeks.



(b) Coefficients are from regressions conditional on the total number of recorded weeks. For example, the first coefficient uses a sample where teachers recorded at least 2 and up to 5 weeks. The plot shows that the intervention had a greater impact when teachers used the feedback on a consistent basis for a longer period of time.

Figure 3: Treatment effects displayed over recorded weeks (a) and by dosage (b).

our regression (i.e., our main analytic sample in Table 2), the treatment effect is 4.48 focusing questions per hour. When we gradually increase the minimum of recorded weeks, we see a linear growth pattern of treatment effects. Specifically, when we only include teachers who recorded their lessons for a total of five weeks in our analytic sample, the treatment effect reaches 7.78, which is a 73% increase compared to our reported effect. This finding suggests that teachers who decide to record more consistently experience greater benefits from the feedback. This relationship is correlational rather than causal, as increased treatment effects may be explained by the characteristics of participants who decided to record more (e.g. available time, motivation, perception of the tool).

7.3 RQ3: How do teachers perceive the automated feedback on focusing questions?

In addition to documenting teacher engagement with automated feedback on focusing questions quantitatively, we explored teachers' subjective experiences through interviews. We examined both their perceptions of the feedback on focusing questions and their overall experience with the TeachFX platform. Interviews revealed three themes that help explain factors that impacted teachers' engagement with automated feedback. The first theme, which describes variations in awareness of the feedback on focusing questions, suggests that this feedback did not reach all members of the treatment group. The second theme, general barriers to engagement, references the various challenges teachers face when trying to engage with automated feedback tools on a consistent basis. The third theme describes strategies that mitigated these barriers and facilitated teachers' use of automated feedback.

Varied Awareness & Understanding of Feedback on Focusing Questions. Our interview data suggest that using emails to deliver automated feedback on focusing questions was only successful for a portion of teacher participants. Among the seven treatment group teachers we interviewed, only three were clearly aware of the emails and saw them as distinct from class reports they had already been receiving from TeachFX. This finding helps explain the moderate email open rates described in RQ1. Teachers who did notice the feedback said they found the treatment emails and feedback helpful in improving their use of focusing questions and inviting more student input in their classroom. For example, one teacher shared how using the transcript to locate and learn from her own focusing questions helped her build more student-centered discussions, "I've been looking a lot at the focusing questions and I like to not just see the number [of focusing questions], but kind of click on it, and see the transcript and skip to the next one and the next one and read my focusing questions and see what were the questions that I asked. And how did the students respond to my questions?" This example illustrates how teachers might benefit from the feedback by using it as a tool to revisit their practice.

In contrast, five out of the seven treatment teachers we interviewed were unaware of the emails containing information about focusing questions. For example, when the interviewer shared screenshots from the treatment email, one teacher noted, "I have not seen an email like that." Others only had vague memories of the treatment emails and were unable to differentiate them from the more general TeachFX class reports they received.

Since the treatment emails clarified and reinforced information about focusing questions, teachers who noticed the in-plaform insight without seeing the email had trouble interpreting focusing questions, sometimes confusing them with other types of questions identified on the TeachFX platform (e.g., volleyball questions, ping-pong questions, open-ended questions, etc. that analyze talk time dynamics rather the language of the talk – see documentation in Appendix A). Such confusion may have prevented teachers from utilizing the feedback on focusing questions or have led them to misinterpret it and ask other kinds of questions. As one teacher noted when being asked about focusing questions, "... I never really figured out like what the underlying element was that was tying them [the focusing questions] all together. And so I had a hard time really gaining much valuable information from that...." This suggests that including more details about focusing questions on the platform or suppressing feedback on other question types could have helped draw attention to and clarify focusing questions for users who did not read the email.

In sum, teachers' varied engagement with our treatment emails limited the impact of our automated feedback. While some teachers fully capitalized on the information we provided and actively adjusted their teaching practice, others were not even aware of it. This finding points to the important role of the delivery mechanism in achieving the desired benefits of automated feedback.

General Barriers to Engagement. Interview participants reported other factors that inhibited their use of the TeachFX feedback platform generally. Three teachers expressed hesitation to use TeachFX due to concerns about data privacy and association with teacher evaluation and accountability. While TeachFX emphasized that the feedback is private to the teachers and intended only for their professional development, one teacher noted, "Nobody likes listening to themselves and being observed and things like that, so like finding ways to be able to share things that we're happy about without feeling like... I don't know, like you're going to be criticized." One instructional coach also observed reservations from the teachers he worked with, "Can it be accessed by principals? Or can it be accessed by parents, or whose data it actually is? This has been on [teachers'] mind."

Second, most (12 out of 13) teachers identified transcription inaccuracy to be a barrier to their engagement with the feedback. They reported issues with configuring the application to recognize their voice and to obtain accurate transcriptions in noisy classroom environments. For example, one teacher shared her experience configuring the recording by saying "I had a hard time with the system reading my voice at the beginning. I had to do a lot more [recording], and it took me a long, long time for the system to work on my phone. . . .. There was an update that came out, I think, around November or October, and then, when I did that one, then [it] started working. But before that it wasn't working on my phone." Some teachers also noted that their automated transcripts contained errors, especially for student speech, rendering the feedback less precise than what teachers would have preferred. One teacher shared how both she and her colleagues were troubled by the imprecise transcripts, "It has some really obvious flaws in the recording. And so a lot of us are like, 'Oh, I did not say that.' . . . I know that that's a hang-up for a lot of teachers." Another teacher noted, "And there still are parts were, like, it would say "student voice detected" [instead of transcribing what they said]... You know what they said If I went in and listened to the recording most of the time. Because I know my students most of the time, I could figure out what they were saying." These findings corroborate those of Jacobs et al. (2022), showing that teachers are less likely to engage with the feedback they deem inaccurate.

Lastly, all 13 teachers we interviewed thought that the platform was straightforward and user-friendly, especially for the data visualization and reflection questions embedded in the platform. However, time constraints appeared to be a crucial factor that prevented some teachers from taking up the feedback. Nine out of 13 teachers identified having limited time as a major reason for not using TeachFX as often as they would like. Some teachers reported that they were too busy to read through all the feedback and information provided in the summary email or on the platform. One teacher shared, "I think for me the hard [thing] is like I didn't have time to sit and read it when it would come in, and then I would forget about it." Another teacher further elaborated on the timing issue, saying, "Sometimes it [the TeachFX class report email] wouldn't come to me until the afternoon, and by then I was done and doing my planning for the day."

Strategies for Supporting Engagement with Feedback. Teachers also shared multiple strategies that mitigated aforementioned barriers and motivated them to engage with TeachFX's feedback. Most (12 out of 13) interviewees noted that having a human component (e.g. a coach or a peer-learning group) during the TeachFX onboarding process was very useful for them to better trust, understand, and utilize the feedback. As one teacher explained, "Well, the math coach came in, and she demonstrated it. She taught a lesson, and she had it all set up on a little iPad that I get to use for a year... she set it up and showed me how to use it, and then showed me how to pull it up, and then she made sure that I had it on the iPad, and then I used it and talked to her about it the first time, so I had some help. It was nice."

Seven teachers identified material support and incentives to be an important motivating factor. Some schools distributed tablets to their teachers to use for recording, which teachers appreciated. Teachers also noted the effectiveness of the raffle incentive. As one of them said, "TeachFX did like a raffle where it's like once a week in the month of January they will enter you into this raffle, and I was like, 'hey, I'm all about raffles.' You know, so yes, I'll admit there is some external motivation going on there. And that is, I'll admit a lot of us are externally motivated. So... I'm gonna get a $250 gift card? Heck yes, I'll record my class".

Finally, nine interviewees mentioned that setting personal goals while using the tool was a helpful source of intrinsic motivation for utilizing the feedback. Those who set personal goals based on automated feedback generally found the feedback more valuable. One teacher noted, "This piece just kind of helped me like, did a mental check. Okay, you're, you're heading towards your goal. You're you're kind of on track what you know. And then asking myself what I can do differently. That, I found useful."

8 Discussion

8.1 Implications of Study Findings

Computerized tools are emerging as a scalable complement to human-based solutions for teacher professional learning. In particular, NLP-powered formative feedback grounded in teaching practices has been proven to be effective in a few online learning contexts (Demszky et al., 2023; Demszky & Liu, 2023). The present study was among the first to investigate the impact of automated feedback in brick-and-mortar classrooms using a randomized controlled trial targeting a high-leverage practice: teachers' use of focusing questions. By targeting focusing questions, the study sought to test a common element in many instructional improvement efforts — inducing teachers to ask more open-ended probing questions. Qualitative data complemented experimental evidence, revealing how teachers engaged with automated feedback on focusing questions and identifying broader barriers to its use. Our study provides three key findings, with implications for educational theory, schools and districts that are interested incorporating automated feedback, and designers of instructional technology.

First, offering feedback to teachers on focusing questions led them to ask an average of 4.6 additional focusing questions per hour, an increase of 20% over the control group ($p < 0.01$). The effect follows a clear dose-dependent pattern, with teachers who recorded their lessons more consistently reaping greater benefits. This result corroborates Jacobs et al. (2024)'s pre-post study, indicating that automated feedback can be a promising tool to improve teaching in in-person K-12 classrooms. They also augment mixed empirical results from prior work on the effectiveness of light-touch, message-based interventions that seek to motivate teachers towards professional development (Azzolini et al., 2023; Hanno, 2023) and the use of technology (Banerjee et al., 2023). For schools and districts considering the use of automated feedback in their classrooms, the fact that a simple email intervention, despite a 50% open rate, can significantly improve a targeted teaching practice within five weeks is remarkable and suggests the promise of this low-cost intervention. If the findings from focusing questions generalize to other aspects of instruction, districts and schools can expect modest but meaningful improvements in targeted instructional practices. For technology developers, the dosage-dependence combined with a lower-than-ideal email open rate points out the importance of finding engaging ways to draw teachers' attention.

Second, our experiment did not yield observable improvements in student talk time, student reasoning, or teacher uptake of student ideas, all valued outcomes in mathematics and science classrooms. While we presented teachers with questioning strategies that sought to help them elicit more student talk and build on such talk, a more elaborate intervention may be necessary to bring about change in these areas. Such an intervention may need to help educators connect focusing questions to these downstream elements of classroom instruction (Putnam & Borko, 2000). As one teacher in our interview sample shared, for instance, teachers may need to notice the ways focusing questions result in student talk – e.g., talk with substantive reasoning or with only shorter speech and procedural recitation.

Third, our qualitative interviews reveal a range of factors that prevented teachers from fully engaging and actively using the feedback, including different levels of awareness of the feedback delivered in emails, concerns about transcript precision and data privacy, and motivation and time commitment to use the feedback. Jacobs et al. (2024) has identified similar barriers to teachers' engagement with automated feedback. These findings highlight the need for technology developers to improve core features and user experience, and for districts to facilitate seamless integration of such tools into teachers' professional learning. Below we outline specific directions for future work to address these concerns.

8.2 Future Directions for Designing & Implementing Technology-Based Teacher Feedback Tools

Our findings suggest several avenues for improving the design and implementation of technology-based teacher feedback tools. While some of these directions may only apply to tools that analyze classroom discourse, others may be relevant to a broader set of technologies that facilitate teacher professional learning.

First, it is critical to improve the quality of automated transcription in noisy classrooms, which currently serves as a bottleneck to accurate feedback (Bokhove & Downey, 2018; Jensen et al., 2020) and thereby teachers' perceived utility of the feedback. Recent work indicates that training or fine-tuning speech recognition models on noisy classroom data can improve the performance of these models in this domain compared to off-the-shelf systems (Attia et al., 2023, 2024; Southwell et al., 2024). Enhanced transcription could facilitate the development and expansion of feedback tools to a broader range of teaching practices, including ones that rely heavily on accurate transcriptions of student talk.

Second, since teachers have busy schedules that require them to constantly multi-task, it is crucial to develop innovative strategies to fit technology seamlessly and engagingly into their current routines (Francom, 2020; Ruggiero & Mong, 2015). Enhancing the content of the feedback (e.g., diversifying its focus, adding more visuals, embedding more interactivity) as well as its delivery (e.g., using texts, gamified and customized nudges) could improve visibility and engagement, and thereby the impact of the feedback. Our qualitative findings and prior evidence on coach-enhanced technology integration (Grierson et al., 2024; Zimmer & Matthews, 2022; Liao et al., 2021) suggest the promise of integrating automated feedback into instructional coaching, while safeguarding teachers' control of their data. For example, instructional coaches can provide scaffolding for automated feedback by reviewing the feedback that the teacher chooses to share and complementing it with their questions and actionable suggestions (Jacobs et al., 2022). Such integration into existing coaching cycles can thus remove the perception of automated feedback as "one more thing" on teachers' plates. Substantively, such scaffolding can provide teachers with skills to interpret the feedback in a way that can lead to comprehensive and long-term instructional improvement — for example, the noticing of cause-and-effect relationships in teaching, by observing the quality of student answers in response to different questions. Instructional coaches could even help complement inaccuracies in the transcript or the feedback by listening to pertinent segments of the recording and providing their expert interpretation.

Last but not least, our qualitative findings as well as prior work on the value of purposeful technology integration (Backfisch et al., 2021; Stinken-Rösner et al., 2023) suggest that facilitating teachers' goal-setting might enhance the effectiveness of automated feedback. Including video clips modeling high-quality instruction and concrete next steps based on the feedback that align with teachers' personal teaching goals could support such goal-setting (Jacobs et al., 2024; Mumtaz, 2000; Okumuş et al., 2016). Furthermore, in addition to providing teachers with static reflection questions and question starters, an adaptive and interactive reflection assistant (human or AI-based) that follows high-quality reflection practices (Korthagen & Nuijten, 2022) could also enhance actionable goal-setting based on automated feedback.

8.3 Limitations

While this study makes several contributions to the literature on technology-based professional learning, one principal limitation is the absence of robust metadata, including information about students' and teachers' backgrounds, learning outcome data, and data on students' beliefs and motivations. Obtaining such data is a major priority for future work, as it would help investigate the impact of the automated feedback on downstream student outcomes, the potential heterogeneity by teacher and student characteristics, as well as the underlying mechanisms of such an intervention.

Second, since our intervention was an add-on piece of feedback on top of the rich automated information already provided by TeachFX, we could not test the impact of automated feedback itself. A promising area of future work would be to design a control condition that does not receive automated feedback or a similar comparison that allows us to disentangle the impact of providing teachers with automated feedback.

9 Conclusion

This study provides the first experimental evidence of the impact of automated feedback on focusing questions in K-12 brick-and-mortar classrooms. We find that such feedback significantly increases teachers' use of this high-leverage instructional practice, showing the potential of automated feedback to enhance teacher professional learning. We also highlight critical challenges in effectively engaging teachers with these tools, especially around transcription accuracy, feedback interpretation, and time constraints. The results emphasize the need for providing teachers with targeted support, e.g. via instructional coaches, to better realize the benefits of automated feedback. Future research should focus on refining these tools, integrating them into teacher routines and exploring their broader impact on teaching practices and student outcomes.

References

Adey, P., Robertson, A., & Venville, G. (2002). Effects of a cognitive acceleration programme on Year I pupils. *British Journal of Educational Psychology*, *72*(1), 1–25. doi: 10.1348/000709902158748

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. doi: 10.1016/0749-5978(91)90020-T

Alexander, R. (2020). *A dialogic teaching companion.* New York: Routledge.

Alic, S., Demszky, D., Mancenido, Z., Liu, J., Hill, H., & Jurafsky, D. (2022). Computationally identifying funneling and focusing questions in classroom discourse. In *Proceedings of the 17th workshop on innovative use of NLP for building educational applications (BEA 2022)* (pp. 224–233). Association for Computational Linguistics.

An, F., Yu, J., & Xi, L. (2022). Relationship between perceived teacher support and learning engagement among adolescents: Mediation role of technology acceptance and learning motivation. *Frontiers in Psychology*, *13*, 992464.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*(2), 167-207. doi: 10.1207/s15327809jls0402_2

Asterhan, C., Clarke, S., & Resnick, L. (2015). *Socializing intelligence through academic talk and dialogue.* Washington, DC: The American Educational Research Association.

Attia, A. A., Demszky, D., Ogunremi, T., Liu, J., & Espy-Wilson, C. (2024). Continued pretraining for domain adaptation of wav2vec2. 0 in automatic speech recognition for elementary math classroom settings. *arXiv preprint arXiv:2405.13018*.

Attia, A. A., Liu, J., Ai, W., Demszky, D., & Espy-Wilson, C. (2023). Kid-whisper: Towards bridging the performance gap in automatic speech recognition for children vs. adults. *arXiv preprint arXiv:2309.07927*.

Azzolini, D., Marzadro, S., Rettore, E., Engelhardt, K., Hertz, B., & Wastiau, P. (2023). Raising teacher retention in online courses through personalized support. evidence from a cross-national randomized controlled trial. *Journal of Research on Educational Effectiveness*, *16*(2), 300–325. doi: 10.1080/19345747.2022.2100850

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. doi: 10.1016/j.compedu.2021.104159

Banerjee, S., Jambrina-Canseco, B., Brundu-Gonzalez, B., Gordon, C., & Carr, J. (2023). Nudge or not, university teachers have mixed feelings about online teaching. *Humanities and Social Sciences Communications*, *10*, 232. doi: 10.1057/s41599-023-01691-1

Bauer, J., & Kenton, J. (2005). Toward technology integration in the schools: Why it isn't happening. *Journal of Technology and Teacher Education*, *13*(4), 519–546. Retrieved from https://www.learntechlib.org/primary/p/4728/

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. doi: 10.1177/2059799118790743

Butler, D. L., & Sellbom, M. (2002). Barriers to adopting technology for teaching and learning. *Educause Quarterly*, *2*(1), 22–28.

Bywater, J. P., Chiu, J. L., Hong, J., & Sankaranarayanan, V. (2019). The teacher responding tool: Scaffolding the teacher practice of responding to student ideas in mathematics classrooms. *Computers & Education*, *139*, 16–30. doi: 10.1016/j.compedu.2019.05.004

Bywater, J. P., Lilly, S., & Chiu, J. L. (2023). Examining technology-supported teacher responding and students' written mathematical explanations. *Journal of Mathematics Teacher Education*, *26*(6), 785–807. doi: 10.1007/s10857-022-09546-3

Chapin, S. H., O'Connor, M. C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn, Grades K-6.* Sausalito: Math Solutions.

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*, 103670. doi: 10.1016/j.compedu.2019.103670

Copur-Gencturk, Y., Li, J., Cohen, A. S., & Orrill, C. H. (2024). The impact of an interactive, personalized computer-based teacher professional development program on student performance: A randomized controlled trial. *Computers & Education*, *210*, 104963. doi: 10.1016/j.compedu.2023.104963

Demszky, D., & Hill, H. (2023). The ncte transcripts: A dataset of elementary math classroom transcripts. In *Proceedings of the 18th workshop on innovative use of nlp for building educational applications.*

Demszky, D., & Liu, J. (2023, July). M-powering teachers: Natural language processing powered feedback improves 1:1 instruction and student outcomes. *L@S '23: Proceedings of the Tenth ACM Conference on Learning @ Scale*.

Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2023). Can automated feedback improve teachers' uptake of student ideas? evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis*, 01623737231169270.

Demszky, D., Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., & Hashimoto, T. B. (2021). Measuring conversational uptake: A case study on student-teacher interactions.

In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 1638–1653).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. doi: 10.48550/arXiv.1810.04805

Dinc, E. (2019). Prospective teachers' perceptions of barriers to technology integration in education. *Contemporary Educational Technology*, *10*(4), 381–398. doi: 10.30935/cet.634187

Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017). Words matter: Automatic detection of teacher questions in live classroom discourse using linguistics, acoustics and context. In *LAK '17: Proceedings of the seventh international learning analytics & knowledge conference* (p. 218–227). New York: Association for Computing Machinery. doi: 10.1145/3027385.3027417

Eksail, F. A. A., & Afari, E. (2020). Factors affecting trainee teachers' intention to use technology: A structural equation modeling approach. *Education and Information Technologies*, *25*(4), 2681–2697. doi: 10.1007/s10639-019-10086-2

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, *59*(2), 423–435. doi: 10.1016/j.compedu.2012.02.001

Farhoomand, A. F., & Drury, D. H. (2002). Managerial information overload. *Association for computing machinery. Communications of the ACM*, *45*(10), 127–127.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research*, *48*(6pt2), 2134–2156. doi: 10.1111/1475-6773.12117

Firestone, W. A., & Donaldson, M. L. (2019). Teacher evaluation as data use: What recent research suggests. *Educational Assessment, Evaluation and Accountability*, *31*, 289–314. doi: 10.1007/s11092-019-09300-z

Francom, G. M. (2020). Barriers to technology integration: A time-series survey study. *Journal of Research on Technology in Education*, *52*(1), 1–16. doi: 10.1080/15391523.2019.1679055

Franke, M. L., & Kazemi, E. (2001). Learning to teach mathematics: Focus on student thinking. *Theory into Practice*, *40*(2), 102–109. doi: 10.1207/s15430421tip4002_4

Fütterer, T., Scherer, R., Scheiter, K., Stürmer, K., & Lachner, A. (2023). Will, skills, or conscientiousness: What predicts teachers' intentions to participate in technology-related professional development? *Computers & Education*, *198*, 104756. doi: 10.1016/j.compedu.2023.104756

Grierson, A. L., Gallagher, T. L., & St Hilaire, R. (2024). Forging the role of the digital technology coach: Flexibility, responsiveness, and resourcefulness in supporting teacher professional learning. *Professional Development in Education*, *50*(1), 205–220. doi: 10.1080/19415257.2022.2038657

Hagenah, S., Colley, C., & Thompson, J. (2018). Funneling versus focusing: When talk, tasks, and tools work together to support students' collective sensemaking. *Science Education International*, *29*(4), 261–266.

Hanno, E. C. (2023). Nudging early educators' knowledge, beliefs, and practices: An embedded randomized controlled trial of text message supports. *Early Education and Development*, *34*(2), 530–550. doi: 10.1080/10409289.2022.2053476

Herbel-Eisenmann, B. A., & Breyfogle, M. L. (2005). Questioning our patterns of questioning. *Mathematics teaching in the middle school*, *10*(9), 484–489. doi: 10.5951/MTMS.10.9.0484

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences*, *28*(4-5), 462–512. doi: 10.1080/10508406.2019.1573730

Hung, W.-H., Chen, K., & Lin, C.-P. (2015). Does the proactive personality mitigate the adverse effect of technostress on productivity in the mobile environment? *Telematics and Informatics*, *32*(1), 143–157. doi: 10.1016/j.tele.2014.06.002

Hunkins, N., Kelly, S., & D'Mello, S. (2022). "Beautiful work, you're rock stars!": Teacher analytics to uncover discourse that supports or undermines student motivation, identity, and belonging in classrooms. In *LAK '22: Proceedings of the twelvth international learning analytics & knowledge conference* (pp. 230–238). New York: Association for Computing Machinery. doi: 10.1145/3506860.3506896

Jacobs, J., Scornavacco, K., Clevenger, C., Suresh, A., & Sumner, T. (2024). Automated feedback on discourse moves: teachers' perceived utility of a professional learning tool. *Educational Technology Research and Development*, 1–23. doi: 10.1007/s11423-023-10338-6

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*, *112*, 103631. doi: 10.1016/j.tate.2022.103631

Jena, R. (2015). Technostress in ict enabled collaborative learning environment: An empirical study among indian academician. *Computers in Human Behavior*, *51*, 1116–1123. doi: 10.1016/j.chb.2015.03.020

Jensen, E., Dale, M., Donnelly, P. J., Stone, C., Kelly, S., Godley, A., & D'Mello, S. K. (2020). Toward automated feedback on teacher discourse to enhance teacher learning. In *Chi'20: Proceedings of the 2020 CHI conference on human factors in com-*

puting systems (p. 1–13). New York: Association for Computing Machinery. doi: 10.1145/3313831.3376418

Kale, U. (2018). Technology valued? observation and review activities to enhance future teachers' utility value toward technology integration. *Computers & Education*, *117*, 160–174. doi: 10.1016/j.compedu.2017.10.007

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Bill & Malinda Gates Foundation. Retrieved from `https://eric.ed.gov/?id=ED540960`

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. doi: 10.3102/0013189X18785613

Korthagen, F., & Nuijten, E. (2022). *The power of reflection in teacher education and professional development: Strategies for in-depth teacher learning.* New York: Routledge.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88(4)*, 547–588. doi: 10.3102/0034654318759268

Lachner, A., Fabian, A., Franke, U., Preiß, J., Jacob, L., Führer, C., ... Thomas, P. (2021). Fostering pre-service teachers' technological pedagogical content knowledge (tpack): A quasi-experimental field study. *Computers & Education*, *174*, 104304. doi: 10.1016/j.compedu.2021.104304

Lemke, J. L. (1990). *Talking science: Language, learning, and values.* Ablex Publishing Corporation.

Li, L., & Wang, X. (2021). Technostress inhibitors and creators and their impacts on university teachers' work performance in higher education. *Cognition, Technology & Work*, *23*(2), 315–330. doi: 10.1007/s10111-020-00625-0

Liao, Y.-C., Ottenbreit-Leftwich, A., Glazewski, K., & Karlin, M. (2021). Coaching to support teacher technology integration in elementary classrooms: A multiple case study. *Teaching and Teacher Education*, *104*, 103384. doi: 10.1016/j.tate.2021.103384

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*. doi: 10.48550/arXiv.1907.11692

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). GPTeach: interactive TA training with GPT based students. In *L@s '23: Proceedings of the tenth acm conference on learning @ scale* (pp. 226–236). NY: Association for Computing Machinery. doi: 10.1145/3573051.3593393

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 philosophy and education*, *27*, 283–297. doi: 10.1007/s11217-007-9071-1

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. doi: 10.1080/14759390000200096

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* Reston: Author.

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*, 1227–1249. doi: 10.1007/s11423-016-9455-4

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *47*(4), 422–453.

O'Connor, C., & Snow, C. (2017). Classroom discourse: What do we need to know for research and for practice? In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (2nd. ed., pp. 315–342). New York: Routledge. doi: 10.4324/9781315687384

Pehmer, A.-K., Gröschner, A., & Seidel, T. (2015). Fostering and scaffolding student engagement in productive classroom discourse: Teachers' practice changes and reflections in light of teacher professional development. *Learning, Culture and Social Interaction*, *7*, 12–27.

Pischke, S. (2007). Lecture notes on measurement error. *London School of Economics, London*.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational researcher*, *29*(1), 4–15.

Resnick, L. B., Asterhan, C. S., & Clarke, S. N. (2018). *Accountable talk: Instructional dialogue that builds the mind*. Geneva: The International Academy of Education (IAE) and the International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Resnick, L. B., Asterhan, C. S., Clarke, S. N., & Schantz, F. (2018). Next generation research in dialogic learning. In G. E. Hall, L. F. Quinn, & D. M. Gollnick (Eds.), *The Wiley handbook of teaching and learning* (pp. 323–338). Hoboken: John Wiley & Sons, Inc. doi: 10.1002/9781118955901.ch13

Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well structured) talk builds the mind. In D. Preiss & R. Sternberg (Eds.), *Innovations in educational psychology:*

Perspectives on learning, teaching and human development (pp. 163–194). New York: Springer.

Rigby, J. G., Larbi-Cherif, A., Rosenquist, B. A., Sharpe, C. J., Cobb, P., & Smith, T. (2017). Administrator observation and feedback: Does it lead toward improvement in inquiry-oriented math instruction? *Educational Administration Quarterly*, *53*(3), 475–516. doi: 10.1177/0013161X16687006

Ruggiero, D., & Mong, C. J. (2015). The teacher technology integration experience: Practice and reflection in the classroom. *Journal of Information Technology Education*, *14*, 161–178. doi: doi:10.28945/2227

Samei, B., Olney, A. M., Kelly, S., Nystrand, M., D'Mello, S., Blanchard, N., ... Graesser, A. (2014). Domain independent assessment of dialogic properties of classroom discourse. In *Proceedings of the 7th international conference on educational data mining (EDM)* (pp. 233–236). London: International Educational Data Mining Society.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. doi: 10.1016/j.compedu.2018.09.009

Scherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. *Educational Research Review*, *27*, 90–109. doi: 10.1016/j.edurev.2019.03.001

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre-and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, *115*, 106552. doi: 10.1016/j.chb.2020.106552

Sherin, M. G., & Dyer, E. B. (2017). Mathematics teachers' self-captured video and opportunities for learning. *Journal of Mathematics Teacher Education*, *20*, 477–495. doi: 10.1007/s10857-017-9383-1

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi: 10.3102/0034654307313795

Smart, J. B., & Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *Journal of Science Teacher Education*, *24*(2), 249–267.

Southwell, R., Ward, W., Trinh, V. A., Clevenger, C., Clevenger, C., Watts, E., ... Whitehill, J. (2024). Automatic speech recognition tuned for child speech in the classroom. In *Icassp 2024-2024 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 12291–12295).

Spiteri, M., & Chang Rundgren, S.-N. (2020). Literature review on the factors affecting primary teachers' use of digital technology. *Technology, Knowledge and Learning*, *25*, 115–128. doi: 10.1007/s10758-018-9376-x

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Education Finance and Policy*, *10*(4), 535-572. doi: 10.1162/EDFP_a_00173

Stinken-Rösner, L., Hofer, E., Rodenhauser, A., & Abels, S. (2023). Technology implementation in pre-service science teacher education based on the transformative view of TPACK: Effects on pre-service teachers' TPACK, behavioral orientations and actions in practice. *Education Sciences*, *13*(7), 732. doi: 10.3390/educsci13070732

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J., & Sumner, T. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse:

The TalkMoves application. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Proceedings of the artificial intelligence in education - 22nd international conference, AIED 2021 (volume 12749)* (pp. 344–348). AIED.

Taie, S., & Goldring, R. (2017). *Characteristics of public elementary and secondary school teachers in the united states: Results from the 2015-16 national teacher and principal survey. first look. (NCES Report No. 2017-072).* Washington, DC: National Center for Education Statistics.

Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta-techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal*, *29*(1), 6–42. doi: 10.1111/isj.12169

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628-51. doi: 10.1257/aer.102.7.3628

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education*, *57*(4), 2432–2440. doi: 10.1016/j.compedu.2011.06.008

Tondeur, J., Forkosh-Baruch, A., Prestridge, S., Albion, P., & Edirisinghe, S. (2016). Responding to challenges in teacher professional development for ICT integration in education. *Educational Technology and Society*, *19*(3), 110–120. Retrieved from `https://www.jstor.org/stable/pdf/jeductechsoci.19.3.110.pdf?seq=1`

Topping, K. J., & Trickey, S. (2007). Collaborative philosophical inquiry for schoolchildren: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology*, *77*(4), 787–796. doi: 10.1348/000709907X193032

Wang, Q., & Zhao, G. (2023). Exploring the influence of technostress creators on in-service teachers' attitudes toward ict and ict adoption intentions. *British Journal of Educational Technology*, *54*(6), 1771–1789. doi: 10.1111/bjet.13315

Wang, Z., Miller, K., & Cortina, K. (2013). Using the LENA in teacher training: Promoting student involement through automated feedback. *Unterrichtswissenschaft*, *41*(4), 290–302.

Watson, J. H., & Rockinson-Szapkiw, A. (2021). Predicting preservice teachers' intention to use technology-enabled learning. *Computers & Education*, *168*, 104207. doi: 10.1016/j.compedu.2021.104207

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., & Turrou, A. C. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research*, *63*, 79–93. doi: 10.3389/fpsyg.2022.992464

Wood, T. (1998). Alternative patterns of communication in mathematics classes: Funneling or focusing. In H. Steinbring, M. Bussi, & A. Sierpinska (Eds.), (pp. 167–178). Reston: NCTM.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, *211*(1), 137–150.

Zimmer, W. K., & Matthews, S. D. (2022). A virtual coaching model of professional development to increase teachers' digital learning competencies. *Teaching and Teacher Education*, *109*, 103544. doi: 10.1016/j.tate.2021.103544

Appendix A Other TeachFX Insights

Teachers could see a range of insights about pedagogical moves and classroom observations on the TeachFX app during the study. Figure A1 shows a view of the class report and a full list of insights are listed below:

1. Word Clouds for teacher talk and student talk
2. Talk ratios (percentage of teacher talk, student talk, group talk and silence)
3. Short Student Responses
4. Long Student Contributions (talk stretches where at least one student spoke for at least 7 seconds)
5. Student Questions
6. Teacher Talk Stretches
7. Volleyball Prompts (teacher 'passing the ball' back to students)
8. Teacher Questions
9. Open Ended Questions

   ↪ Although this insight may seem similar to focusing questions, it analyzes talk dynamics rather than the content (language) of the teacher utterance. Specifically, the insight shows every teacher question that is closely followed by a long student talk, while allowing for some silence and short teacher talk in between, in the case the teacher calls on a student. The start time is the start of the teacher question, and the end time is the end of the long student talk. This insight ignores any teacher question that contains "can __ read" with any number of words in the blank.

10. Ping Pong Questions (teacher playing "ping pong" with students)

11. Think Time After Teacher Spoke (Wait Time 1)

12. Think Time after Student Spoke (Wait Time 2)

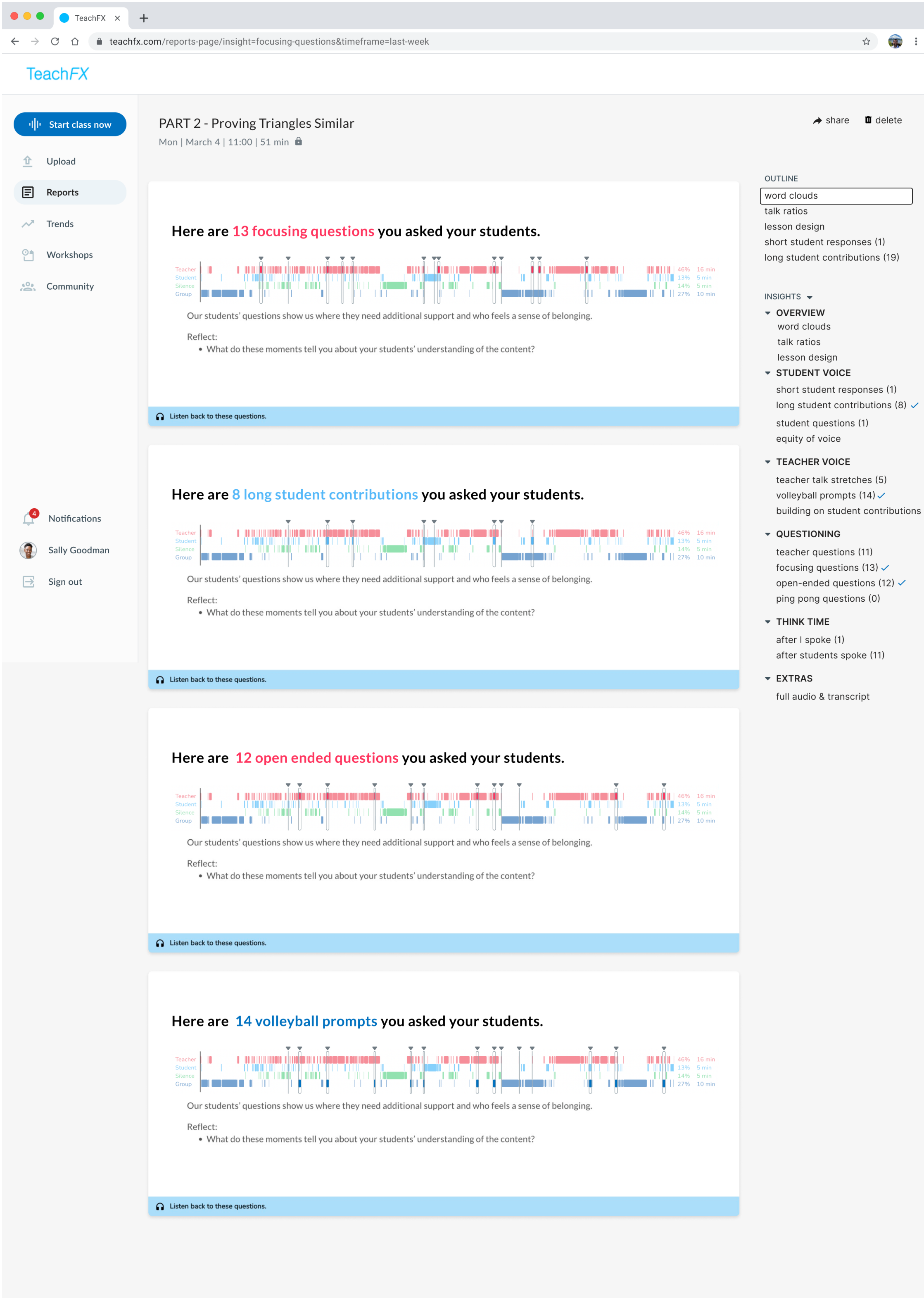


Figure A1: A view of the TeachFX class report at the time of the study.

Appendix B Focusing Questions Model

Appendix B.1 Original Model

We refer to a model that does *not* include data from TeachFX in its training data, only previously labeled data from NCTE (Demszky et al., 2021; Demszky & Hill, 2023), as the Original Model. We used this model for obtaining focusing questions for the first week of the experiment, and for sampling data for re-training purposes (see below). We obtained this model by fine-tuning BERT base (Devlin et al., 2018) for a binary classification task of identifying whether or not a teacher utterance was a focusing question or not on the following labeled data. We used ∼2000 NCTE teacher utterances that which had annotations for whether the teacher utterance was a focusing question, a funneling question or neither (we considered the funneling questions and neither category in the category of non-focusing questions). This data had a 90:10 split of non-focusing vs focusing questions. To handle the class imbalance, we over-sampled the minority class to enable the training set to have an approximately equal number focusing and non focusing questions. The final model hyperperameters we used were: Learning rate: 2e-5; Max embedding length: 256 tokens; Number of epochs: 3; Optimizer: Adam; epsilon for Adam (to avoid divide by zero error): 1e-8.

Appendix B.2 Re-Trained Model

We refer to the model that we used for identifying focusing questions during the second week of the experiment and onwards as the Re-Trained Model. We developed this model in order to address the domain shift that arose from our model being trained on NCTE data but used for inference on TeachFX transcripts. To obtain this model, we fine-tuned BERT base on both TeachFX and the aforementioned NCTE data. We describe our method of re-training below.

To re-train the model, our goal was to add a small sample (few hundred examples) from

TeachFX to the training data that included a large percentage of focusing questions in order to demonstrate to the model how focusing questions looked like in the TeachFX domain. Furthermore, we wanted to carefully select examples of focusing vs non-focusing questions from the TeachFX data that the Original Model would have misclassified, to improve the model's prediction accuracy in this new domain. This process involved multiple steps. First, we collected 22 TeachFX transcripts from the first week of our study – from *both* treatment and control group teachers – and extracted all teacher utterances from these transcripts. We considered one utterance as 3 consecutive sentences of teacher talk, computed using a sliding window. Using our Original Model, we obtained 444 teacher utterances that were predicted as focusing questions and approximately 700 which were not. We asked an expert annotator to label which of the predicted 444 focusing questions were indeed focusing questions. From this, we got 276 labeled focusing questions and 168 labeled non-focusing questions. In order to augment the number of non-focusing questions (so that the fine-tuned model does not see too many positive examples while training), we manually verified and extracted 250 non-focusing questions of the 700 non-focusing question predictions. Thus, we obtained 694 teacher utterances (the aforementioned 444 + 250 teacher utterances) from the TeachFX data. We added these utterances to our original NCTE dataset. From this dataset, we randomly sampled and chose 200 validation utterances for the held out set. We used the remaining utterances for training the BERT model. For this training set, we over-sampled the minority class such that we got approximately a 50:50 focusing question vs non-focusing question split.

Appendix C Teacher Survey

Appendix C.1 Survey Email Text

"Thank you for using TeachFX to reflect on your teaching practice! To learn how to better support teachers and improve our app insights, we are sending you a survey about your background and the teaching practices used in your math/science lessons. This survey will take no more than 5 minutes to complete, and your identity will remain strictly confidential. To show our appreciation for your time, we will send you a $10 Starbucks gift card for completing the survey."

Appendix C.2 Survey Questions

Appendix C.2.1 1. Thinking about your mathematics/science teaching, please indicate your opinion about each of the statements below:

(Scale: "not at all" to "all of the time")

a) My questions elicit students' mathematical/scientific thinking and reasoning.

b) My students talk about their mathematical/scientific ideas.

c) I pose open-ended questions.

d) I engage my class(es) in discussion.

e) I require students to explain their reasoning when giving an answer.

Appendix C.2.2 2. This set of questions seeks your opinion on how K-12 mathematics/science should be taught. Please indicate the extent to which you agree with the following statements:

(1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree)

a) It is valuable for students to see and hear other students' mathematical/scientific explanations.

b) Having students talk about their thinking helps other students make sense of mathematical/scientific ideas.

c) Teachers should ask students to explain how they got an answer.

d) Having a student clarify their thinking can surface mathematical/scientific reasoning in a way that helps other students learn.

e) Teachers should listen to all students' ideas, even if they are unusual or incorrect.

Appendix C.2.3 3. Which subject(s) do you teach?

(Select all that apply)

a) Mathematics

b) Science

c) Other(s): (Please specify)

Appendix C.2.4 4. At which grade level(s) do you teach mathematics/science this year?

(Select all that apply)

a) Kindergarten

b) 1st grade

c) 2nd grade

d) 3rd grade

e) 4th grade

f) 5th grade

g) 6th grade

h) 7th grade

i) 8th grade

j) 9th grade

k) 10th grade

l) 11th grade

m) 12th grade

Appendix C.2.5 5. Please indicate your role in the school:

a) I am a regular classroom teacher and teach a general/mixed population of students.

b) I have my own classroom, but exclusively teach a specific population of students (e.g., special education, emergent bilingual students)

c) I do not have my own classroom, but instead offer small-group instruction, tutoring or other kinds of special assistance to students from other teachers' classrooms.

Appendix C.2.6 6. How many total years of experience do you have teaching mathematics?

(Select one)

a) less than 1 year

b) 1-2 years

c) 3-4 years

d) 5-6 years

e) 6-7 years

f) over 8 years

Appendix C.2.7 7. Which of the following best describes your race-ethnicity?

(Select all that apply)

a) Asian

b) Native Hawaiian or Other Pacific Islander or Pacific Islander

c) Black or African American

d) Hispanic or Latinx

e) Native American, Alaska Native, or Indigenous

f) White or Caucasian

g) Multiracial or Biracial

h) Other race-ethnicity:

Appendix C.2.8 8. With which of the following do you identify?

(Select one)

a) Male

b) Female

c) Nonbinary

d) Prefer to self-identify:

e) Prefer not to say

Appendix D Interview Protocol

Appendix D.1 Treatment Group

Background

a) Please tell us a little bit about yourself: What subject and grade do you teach? What are your years of experience in teaching?

b) Please share with us a little bit about your goal as a teacher towards student learning

c) What drew you to try out TeachFX as a feedback tool?

d) Were you required to use the tool by either the school or the district? What were the requirements?

e) How do you feel about incorporating technology into your teaching practices?

Experience using TeachFX

a) Please walk us through your experience when you first used TeachFX.

b) Where did you click or not to get more information? What did you find helpful? How did you interpret or make use of the feedback?

c) Please tell us a little bit about your experience using TeachFX.

d) What did you like about the tool? What did you not like?

e) How easy was it for you to record and upload your lessons?

f) What are some obstacles you faced?

g) How easy was it for you to navigate on the platform?

h) Which aspect of the tool did you utilize the most?

i) What are some obstacles or difficulties you experienced when using the tool?

j)Do you use TeachFX regularly? Why or why not?

k) How easy was it for you to incorporate the feedback in your teaching?

l) In which aspect do you think TeachFX can best support you in terms of your teaching?

m) Would you recommend this tool to a colleague? Why or why not?

Feedback on focusing questions

a) Do you remember receiving emails like these and seeing these feedback?

b) What do you remember about the emails and the feedback?

c) Did you know about focusing questions before using TeachFX?

d) If not, were the explanation and examples provided by TeachFX helpful for you?

e) What would you say are the differences between focusing and funneling questions?

f) Could you walk us through how you made use of the emails and the feedback tools?

g) What do you think about the weekly email? How helpful was it for you? (Question prompt, blog post and other resources, etc)

What do you think about the language of the feedback?

a) Are you familiar with the terms and language TeachFX used?

b) Was it clear? Are certain terms inaccurate?

c) How helpful are the reflect questions for you?

Suggestions for improving the tool

a) Is there any other feature that you would like to see on TeachFX?

b) Is there any other kind of feedback that you would like to receive?

c) What about the format and delivery of the feedback?

d) Is there anything that you would drop or would want differently?

e) What other advice or support would you like to receive more regarding posing focusing questions in classes?

Final thoughts

a) Do you have any other thoughts to share?

b) Do you have any questions for me?

Appendix D.2 Control Group

Background

a) Please tell us a little bit about yourself: What subject and grade do you teach? What are your years of experience in teaching?

b) Please share with us a little bit about your goal as a teacher towards student learning

c) What drew you to try out TeachFX as a feedback tool?

d) Were you required to use the tool by either the school or the district? What were the requirements?

e) How do you feel about incorporating technology into your teaching practices?

Experience using TeachFX

a) Please walk us through your experience when you first used TeachFX.

b) Where did you click or not to get more information? What did you find helpful? How did you interpret or make use of the feedback?

c) Please tell us a little bit about your experience using TeachFX.

d) What did you like about the tool? What did you not like?

e) How easy was it for you to record and upload your lessons?

f) What are some obstacles you faced?

g) How easy was it for you to navigate on the platform?

h) Which aspect of the tool did you utilize the most?

i) What are some obstacles or difficulties you experienced when using the tool?

j) Do you use TeachFX regularly? Why or why not?

k) How easy was it for you to incorporate the feedback in your teaching?

l) In which aspect do you think TeachFX can best support you in terms of your teaching?

m) Would you recommend this tool to a colleague? Why or why not?

What do you think about the language of the feedback?

a) Are you familiar with the terms and language TeachFX used?

b) Was it clear? Are certain terms inaccurate?

c) How helpful are the reflect questions for you?

Suggestions for improving the tool

a) Is there any other feature that you would like to see on TeachFX?

b) Is there any other kind of feedback that you would like to receive?

c) What about the format and delivery of the feedback?

d) Is there anything that you would drop or would want differently?

Final thoughts

a) Do you have any other thoughts to share?

b) Do you have any questions for us?

Appendix E Interview Codebook & Examples

Code		Description and Examples	
B	Barriers to engagement	When an interviewee talks about barriers that they or their colleagues run into which prevent them from engaging with the feedback	
B1	(Attitude) Lack of trust toward technology	When an interviewee mentions a lack of interest, a sense of distrust, hesitancy, or resistance towards technology as a barrier to utilizing automated feedback tools	"There's always pushback from teachers on like, okay. Is this something new like, you know. I have to do something new." (0411) "...can it be accessed by principals? Or can it be accessed by parents, or whose data it actually is? This has been on Teacher's mind" (0411)
B2	(Interaction) Trouble overcoming technical difficulties	When an interviewee mentions technical difficulties (e.g., issues with recording, handling devices, navigating the uploading process, transcript accuracy, etc.) they run into	"It has some really obvious flaws in the recording. And so a lot of us are like, 'Oh, I did not say that.' … I know that that's a hang-up for a lot of teachers." (0319)
B3	(Interaction) Trouble understanding feedback	When an interviewee expresses confusion about the feedback received (this includes confusing one type of feedback from another, lack of clarity of what feedback is provided, inability to understand feedback due to information overload, etc.)	"The only thing that I felt like I didn't take advantage of as much is really looking at the platform, because I just would look at... like when you when you get on the platform there's kind of like this graph ... and so it's just those visuals because I just don't have a lot of time." (0418)

Code(cont. 1)		Description and Examples (cont. 1)	
B4	(Incorporating feedback) Time constraint	When an interviewee mentions that they do not have enough time to make full use of the feedback tool (this includes not enough time to wait for the feedback to be ready, to understand what is on the platform, and to understand the feedback in a more detailed fashion)	"I think for me the hard [thing] is like didn't have time to sit and read it when it would come in, and then I would forget about it." (0428) "... sometimes I do have time to read through my emails and other times I don't, and if it's a busier time, then if it's an email that's not crucial. I might not take the time to look at it." (0403)
B5	(Incorporating feedback) Challenges to translate feedback in teaching practices	When an interviewee talks about challenges they face when trying to use the feedback in their teaching (e.g., failure to see how the feedback might be helpful for their teaching, inability to see effects, etc.)	"I think that's where the difficulty arises. Taking the data, taking the reflection, and then actually putting that into practice, saying, okay, my talk ratios were 80-20, teacher talk to student talk. How do I change that, or, or, do I want to change it right? Like, what, what, what steps need to be there for me to realize that" (0411)
B6	Other barriers	Other barriers to engagement that the interviewee brings up	

Code(cont. 2)		Description and Examples (cont. 2)	
E	Ways to enhance engagement	When an interviewee talks about the different factors and habits that support their engagement with the feedback	
E1	(Attitude) Comfortability with technology	When an interviewee expresses comfortability and competence with using technology in their classrooms (e.g., technology skills, experience with technology, open-mindedness towards technology, willingness to understand new tools through trial and error, etc.)	"For me [the recording process] was pretty easy, I think the only thing I had to do was clear up a little bit of space on my phone, so it would record a whole period." (0317) "so we use a lot of those tools again, so very familiar with apple products, very familiar with iPads" (0319)
E2	(Interaction) External support to engage with the tool	When an interviewee mentions help from other people or sources (e.g., an instructional coach, an instructor from TeachFX, a colleague, a professional learning group, previous knowledge and skills, etc.) who enhance their understanding of, reflection on, and engagement with the feedback	"Well, the math coach came in, and she demonstrated it. ... she set it up and showed me how to use it, and then showed me how to pull it up, and then she made sure that I had it on the iPad, and then I used it and talked to her about it the first time, so I had some help. It was nice." (0421)

Code(cont. 3)		Description and Examples (cont. 3)	
E3	(Interaction) User-friendly information delivery	When an interviewee describes the medium of communication of feedback (e.g., the language used, the visual presentation, the organization of different types of feedback, etc.) as factors enhancing their engagement with the feedback	"So sometimes, adding some new visual components to TeachFX, so it's like, oh. oh, that wasn't there before! Let me click on that, and see what that is, that would catch my attention if visually the platform changed a little bit like those." (0505) "Sometimes, even when I'm just showing the teachers, they will look at the reflection questions versus me asking them from a coaching standpoint, so they use those to reflect on their own teaching without somebody there to guide them." (0501)
E4	(Incorporating feedback) Frequent and consistent use of feedback tool	When an interviewee talks about the usage frequency and/or consistency they find to be the most useful	"I try to do it at least once a week" (0317) "I teach math 5 days a week, and so I would pick one of the days, and I tried to pick different days" (0513)
E5	(Incorporating feedback) Strategies to translate feedback in personalized teaching practices	When an interviewee describes how they incorporate the feedback into their own teaching contexts (e.g., setting personal goals, establishing accountability, working with peers, incorporating into their teaching routines, etc.) and how that has helped them engage with the feedback	"This piece just kind of helped me like, did a mental check. Okay, you're, you're heading towards your goal. You're you're kind of on track what you know. And then asking myself what I can do differently. That, I found useful." (0428)

Code(cont. 4)		Description and Examples (cont. 4)	
			"... a word cloud is just always fun, very visual, because a lot of, not just kids, but adults, too, are very visual" (0505)
E6	(Content) Useful feedback type	When an interviewee identifies the types of feedback that they utilize the most and find to be the most helpful for their teaching	"we try anything to get participation. So, looking at like student talk, looking at how much they are, you know, interacting, and then also looking at wait time after questions, because sometimes it can feel like I'm giving plenty of time for them to think. But it's really not as much as you think it is, right?" (0506)
E7	Other ways to enhance engagement	Other ways of engagement that the interviewee brings up	"... if we were willing to try out using TeachFX in our classroom on a consistent basis, and then do some writing about it some a few assignments and things they will, they will pay us a stipend. And so that is what got me initially" (0317)

Code(cont. 5)		Description and Examples (cont. 5)	
T	Treatment group experience	When the treatment group interviewee mentions an experience specific to the treatment group	
T1	Awareness of focusing questions	When treatment group interviewee talks about whether they are aware of the existence of feedback on focusing questions on the platform	"I've been looking a lot of the focusing questions and I like to not just see the number, but kind of click on it, and see the transcript and skip to the next one and the next one and read my focusing questions and see what were the questions that I asked. And how did the students respond to my questions?" (0317)
T2	Email for information dissemination	When treatment group interviewee talks about the treatment email they receive	"No, I haven't [seen the email]. But ... since I was unsure about focusing questions, this (the examples of focusing questions I asked)... [is] the information that I need." (0403)

Code(cont. 6)		Description and Examples (cont. 6)	
T3	Incorporating focusing questions in teaching	When treatment group interviewee talks about how they make use of focusing questions on their teaching based on the feedback they received	"Especially I would look at the focusing questions and look at the student responses, and I tried to pay attention to who I was calling on, and also what my follow-up questions were, ... And of course, you know there is going to be some change as it goes, because students are going to say things that you're not expecting, ... so that I could make sure that I was allowing the students to have more of a role in following up." (0513) "I felt like when I would listen to the focusing questions, I never really figured out like what the underlying element was that was tying them all together. And so I had a hard time really gaining much valuable information from that, because I felt like some of the focusing questions that it recorded were really poor questions almost felt like they would fall into a ping pong question category or something like that." (0403)
T4	Knowledge about focusing questions	When treatment group interviewee talks about their content knowledge about focusing questions and how the knowledge can support their instruction	"I learned about this when I was getting my master's degree. But I already knew that kind of vocabulary. I already knew that I care about focusing questions and want them in my classroom, but I'd never had a tool to measure how often I was using them." (0317)

Code(cont. 7)		Description and Examples (cont. 7)	
X	TeachFX-specific experience	When an interviewee talks about experiences and feedback that is specific to TeachFX	
X1	Feedback for the tool	When an interviewee provides feedback that is specific to TeachFX's use scenario	"The system was very straightforward and easy to set up and use" (0316) "I think the only thing that was confusing me today was, how do I get back to the main screen now that I'm in a report, and I've already forgotten how I get back." (0317)
X2	Introduction to the tool	When an interviewee talks about how they were first introduced to TeachFX	"… it was a PD offer online through the district after school for several different sessions where they just kind of introduced us to the program and how we could use it. And then we would … do a recording, and then come back and talk about it." (0506) "It was actually from the district office, and they offered it for training, and they gave us a little bit of a stipend" (0421)
X3	Navigating the platform	When an interviewee provides detailed and TeachFX-specific experience of how they navigate the feedback platform	"The system was very straightforward and easy to set up and use" (0316) "I discovered something now that it was on my phone. When I have the Teach FX on my computer, I don't see the graph." (0318)

Appendix F Table 2 Without Demographic Controls

	(1) Focusing rate	(2) Uptake rate	(3) Student reasoning rate	(4) Student talk percentage
Treatment	4.517*	0.284	0.514	0.004
	(1.805)	(0.563)	(0.491)	(0.013)
Control Mean	22.565	3.772	2.896	0.160
R2	0.316	0.276	0.190	0.169
Observations	533	533	533	533

Table F1: Standard errors are in parentheses. * $p<0.05$. These models estimate the effect of the automated feedback on focusing questions (treatment) on teachers' discourse features. These models exclude demographic control features obtained from the survey. We still observe a significant impact on focusing questions, but not the other discourse features.