ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-873

Practice-Based Teacher Education Pedagogies Improve Responsiveness: Evidence from a Lab Experiment

Zid Mancenido
Harvard University

Heather C. Hill
Harvard University

Jeannette Garcia Coppersmith
Harvard University

Hannah Carter
Harvard University

Cynthia Pollard
Stanford University

Chris Monschauer
Harvard University

Practice-based teacher education has increasingly been adopted as an alternative to more traditional, conceptually-focused pedagogies, yet the field lacks causal evidence regarding the relative efficacy of these approaches. To address this issue, we randomly assigned 185 college students to one of three experimental conditions reflective of common conceptually-focused and practice-based teacher preparation pedagogies. We find significant and large positive effects of practice-based pedagogies on participants' skills in eliciting and responding to student thinking as demonstrated through a written assessment and a short teaching episode. Our findings contribute to a developing evidence base that can assist policymakers and teacher educators in designing effective teacher preparation at scale.

VERSION: November 2023

Suggested citation: Mancenido, Zid, Heather C. Hill, Jeannette Garcia Coppersmith, Hannah Carter, Cynthia Pollard, and Chris Monschauer. (2023). Practice-Based Teacher Education Pedagogies Improve Responsiveness: Evidence from a Lab Experiment. (EdWorkingPaper: 23-873). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/mm5s-b648

Practice-Based Teacher Education Pedagogies Improve Responsiveness: Evidence from a Lab Experiment

Zid Mancenido[1], Heather C. Hill[1], Jeannette Garcia Coppersmith[1],

Hannah Carter[1], Cynthia Pollard[2], Chris Monschauer[1]

[1] Harvard Graduate School of Education

[2] Stanford Graduate School of Education

Author Note

Zid Mancenido https://orcid.org/0000-0003-2172-5011

Heather C. Hill https://orcid.org/0000-0001-5181-1573

Jeannette Garcia Coppersmith https://orcid.org/0009-0000-8669-9729

Hannah Carter https://orcid.org/0009-0007-4539-1323

Correspondence concerning this article should be addressed to Zid Mancenido, 445 Gutman Library, Harvard Graduate School of Education, 6 Appian Way, Cambridge, MA 02138. Phone: 646-467-1027. Email: mancenido@g.harvard.edu. This research was generously supported by the National Science Foundation (ECR-1920616). We are grateful to participants at the Research Designs and Measurement in Teacher Education conference and the Association for Education Finance and Policy conference who provided feedback and comments on an earlier version of this paper. All errors and omissions are our own.

Abstract

Practice-based teacher education has increasingly been adopted as an alternative to more traditional, conceptually-focused pedagogies, yet the field lacks causal evidence regarding the relative efficacy of these approaches. To address this issue, we randomly assigned 185 college students to one of three experimental conditions reflective of common conceptually-focused and practice-based teacher preparation pedagogies. We find significant and large positive effects of practice-based pedagogies on participants' skills in eliciting and responding to student thinking as demonstrated through a written assessment and a short teaching episode. Our findings contribute to a developing evidence base that can assist policymakers and teacher educators in designing effective teacher preparation at scale.

Keywords: teacher preparation, practice-based teacher education, causal evaluation

Practice-Based Teacher Education Pedagogies Improve Responsiveness: Evidence from a Lab Experiment

Over the past two decades, teacher education has increasingly adopted practice-based pedagogies focused on developing pre-service teachers' skills for enacting complex teaching practices, and away from pedagogies focused on building pre-service teachers' knowledge and conceptual frameworks for teaching (Ball & Cohen, 1999; Grossman, 2018; Lampert, 2010; McDonald et al., 2013). This shift responds to growing concerns that teacher education programs have overvalued knowledge and frameworks for teaching at the expense of supporting the development of complex teaching practices such as leading discussions and modeling content, thus underpreparing novices for the competencies needed day-to-day in classrooms (Grossman et al., 2018). Shifts towards more practice-based approaches, however, have not been accompanied by rigorous studies evaluating their relative effectiveness, studies that are sorely needed given the strong theoretical bases underpinning both conceptually-focused and practice-based approaches to educating teachers (Cochran-Smith et al., 2016; Mancenido, 2023; see also National Academies of Sciences, Engineering, and Medicine, 2022: p. 105-6).

To address this gap, this paper reports findings from a lab experiment aimed at generating some of the first causal evidence on the relative effectiveness of conceptually-focused versus practice-based teacher education pedagogies. We randomly assigned 185 college undergraduates to experience one of three one-hour interventions, all aimed at developing their skills in eliciting and responding to student thinking within the context of elementary math instruction. These three interventions reflected common forms of conceptually-focused and practice-based pedagogies currently used in the field. The conceptually-focused intervention involved reading and discussing research on teaching, followed by further reflection and discussion. The practice-based intervention involved decomposing a video of expert teaching, and then approximating practice with peers and a teacher educator. The mixed-condition intervention involved reading and discussing research on teaching, followed by decomposing a video of expert teaching.

We find significant and large positive effects of the practice-based and mixed conditions on eliciting and responding to student thinking as demonstrated through a written assessment and a short teaching episode. Notably, we find that only the practice-based condition improved participants' responsiveness on a teaching task that was distinct from the one featured in the intervention. These findings provide causal evidence of the relative effectiveness of practice-based teacher education pedagogies in developing complex teaching practices. Together, they contribute to developing a more rigorous evidence base that can help policymakers and teacher educators design effective teacher preparation at scale.

Literature Review

Over the past two decades, teacher effectiveness scholars and some U.S. policymakers have identified high-leverage teaching practices that pre-service teachers should master prior to entering their own classrooms — for example, designing lessons and lesson sequences, explaining and modeling content, eliciting and responding to student thinking, and building teacher-student relationships (e.g., Council of Chief State School Officers, 2013; Darling-Hammond & Bransford, 2005; Grossman, 2018; Michigan Department of Education, 2018). To ensure pre-service teachers (PSTs) achieve these learning goals during teacher education programs, teacher educators must select among different sets of pedagogies.

One such set of pedagogies – hereafter, *conceptually-focused* teacher preparation – privileges the development of conceptual tools for planning, analyzing, and reflecting on teaching (Grossman, Hammerness et al., 2009). Drawing on decades of research showing that teachers' conceptual tools precede and therefore determine their classroom practices (Fang, 1996; Pajares, 1992; Maggioni & Parkinson, 2008; Richardson et al., 1991), advocates for conceptually-focused pedagogies argue that pre-service teacher preparation is most effective and efficient when it provides PSTs with the appropriate mental models, knowledge, beliefs, and/or dispositions to make high-quality decisions regardless of situation or context (Clark, 1988; Pajares, 1993; Nespor, 1987; Wideen et al., 1998). Indeed, some researchers argue that when teacher educators do not focus on conceptual tools but instead focus on specific instructional skills that are effective in certain situations and contexts – as in the 'micro-teaching' experiments of the 1960s – teachers can become technicians implementing "decontextualized moves" rather than professionals using their informed judgment to engage in complex decision-making and practice (Wideen et al., 1998; Zeichner, 2012).

Conceptually-focused teacher education pedagogies have long been the modal approach in teacher preparation program coursework (Grossman, Hammerness et al., 2009; Wideen et al., 1998). They often take several forms: PSTs might read about the intended practice, discuss what they have read with others, reflect on how what they are learning relates to their pre-existing beliefs about teaching, and synthesize their learning via reflection or planning. During these discussions and reflections, teacher educators sometimes model the specific practices being discussed, providing concrete examples for PSTs. Whatever the form, a common focus is interrogating and revising (incorrect) mental models PSTs have developed through the partial view of teaching formed through their apprenticeship of observation as a PreK-12 student (Lortie, 1975).

Another set of teacher education pedagogies – hereafter *practice-based* teacher preparation – privileges the development of skills for complex teaching by specifying and supporting PSTs to acquire “the strategies, routines, or activities that novices need to learn to do and from which they will continue to learn teaching” (Lampert, 2010: p. 26). Advocates of this approach argue that teaching is complex– particularly ‘ambitious’ teaching that prioritizes student thinking and reasoning – but that learning specific teaching practices and how to apply those practices appropriately can enhance novice teachers’ pedagogical reasoning (Kavanagh et al., 2020), judgment (Lampert et al., 2013) and practice (Kavanagh & Rainey, 2017). Scholars who study and use practice-based pedagogies also argue that they more effectively prepare PSTs because they make the complexity of teaching more tractable for novices, reducing cognitive load when learning to teach, and supporting skill-building, automaticity, and the development of professional vision (Grossman, Compton et al., 2009; Grossman, Hammerness et al., 2009). For example, Windschitl et al. (2012) argue that practicing instructional activities in low-stakes simulated environments reduces the cognitive load for new teachers once they get into their classrooms. Similarly, Stroupe & Gotwals (2018) argue that practice-based approaches reduce PSTs’ anxiety and increase their self-efficacy, which makes them more confident and able to engage in the responsive, interactive elements of classroom practice.

Two popular forms of this practice-based approach are *decompositions* and *approximations* (Grossman, Compton et al., 2009). *Decompositions* involve novices observing and analyzing complex teaching practices as enacted by expert teachers, whether in person or on video. Different than simple modeling of practice in the teacher education classroom, as can occur in a conceptually-focused approach, teacher educators support novices’ observations of actual classroom teachers’ practice by drawing attention to and breaking down specific events and decisions. This deepens both novices’ understanding of the specific components of a complex practice and how and why experts may have chosen to use that practice in the moment.

Decompositions thus help PSTs attend to and learn about the essential elements of a practice as enacted by expert teachers, potentially building their skills in enacting it themselves in the future. *Approximations* are opportunities for novices to engage in "deliberate practice" (Ericsson, 2002) through experiences, such as simulations and rehearsals, that are proximal to classroom teaching. Although approximations can vary depending on the "authenticity" with which they approximate day-to-day classroom practice (Schutz et al., 2019), what is common is that teacher educators actively shape the experience to "provide opportunities for students to experiment with new skills, roles, and ways of thinking with more support and feedback than actual practice in the field allows" (Grossman, Compton et al., 2009: p. 2076-7). For instance, in the case of rehearsals or other simulated practice (Kavanagh et al., 2020), this can include interjecting to provide positive feedback in the moment, making suggestions on how to elevate instruction, offering a "rewind" to have another go, or holding discussions about critical events or decisions. Over the past two decades, practice-based approaches like decompositions and approximations have gained increasing popularity within teacher education program coursework (Grossman, 2018; Kavanagh et al., 2020; Lampert et al., 2013; McDonald et al., 2013).

Despite well-developed theory motivating the use of both conceptually-focused and practice-based teacher education pedagogies, no studies have rigorously evaluated their relative effectiveness – or the effectiveness of their combination – in preparing PSTs for complex teaching practice. Several studies have used non-experimental designs, however, and shown positive correlations between practice-based pedagogies and PST outcomes. For example, Kavanagh & Rainey (2017) analyzed video of teacher education coursework and PSTs' classroom instruction, finding instructional practices taught primarily through practice-based approaches were more frequently observed in PSTs' subsequent classroom instruction. Other studies have used experimental designs but not been able to provide theoretically-aligned evidence of the relative effectiveness of either type of pedagogy. For example, Strasser et al. (2021) compared video analysis, deliberate rehearsal, and a more conceptually-focused approach. However, for logistical reasons, the authors varied the content between the practice-based conditions (language stimulation skills for pre-K classrooms) and the conceptual condition (letter knowledge, phonemic awareness, print awareness), preventing a clean comparison between practice-based and conceptually-focused teacher education pedagogies. Another study, Sims et al. (2023), randomly assigned PSTs to either read about effective questioning for retrieval or watch an expert teacher model the practice, finding positive effects for the modelling condition. While the authors note that modelling plays an important role in some practice-based pedagogies, they acknowledge their study generated evidence only of the effects of modelling itself, not on the sorts of practice-based pedagogies more commonly used and studied in the field (Kavanagh et al., 2020; Shutz et al., 2019).

The Present Study

To fill this gap in the literature, we designed a lab experiment comparing the relative effectiveness of conceptually-focused pedagogies, practice-based pedagogies, and a mix of both. We developed interventions to reflect, to our knowledge, common forms of conceptually-focused and practice-based pedagogies currently used in the field. We then tested these approaches in the context of preparing PSTs to elicit and respond to student thinking during math instruction, a relatively complex practice requiring the noticing and interpretation of student thinking and the formulation of a response based on content and instructional goals (Barnhart & van Es, 2015; Boerst et al., 2011; Shaughnessy & Boerst, 2018). Given its relationship to student learning, eliciting and responding to student thinking is generally agreed to be an essential teaching practice for novice teachers to acquire during teacher preparation (Blazar, 2015; Bransford et al., 1999; Kane et al., 2011; Nystrand & Gamoran, 1991). Thus our primary research question guiding study design and analysis was:

> **Research Question #1**: Which teacher education pedagogy is most effective in preparing study participants to elicit and respond to student thinking during math instruction: (1) reading then reflecting on research (*conceptually-focused*); (2) decomposing a video of expert teaching and then approximating practice with peers and a teacher educator (*practice-based*); or (3) reading research and then decomposing a video of expert teaching (*mixed*)?

We answered this question using two outcome measures: vignettes that presented a short snippet of mathematics instruction and then asked participants to write how they would respond to a particular student; and live enactments that required participants to teach mathematics for up to 10 minutes to two trained actors. Given theory that suggests conceptually-focused pedagogies are more comprehensive (i.e., preparing PSTs with mental models that guide practice regardless of situation and context) and some scholars' concerns that practice-based pedagogies may be too narrow in their focus (i.e., leading PSTs to learn "decontextualized moves"), we designed the live enactments to measure eliciting and responding to student thinking across two different teaching tasks. One teaching task (teaching students who provide correct answers) aligned closely with participants' training during the interventions; the other task (teaching students who provide incorrect answers) allowed research participants to elicit and respond to students, but was not the focus of the interventions. We assessed participants on these two different tasks to answer the following question:

Research Question #2: Are there differences across treatment conditions in performance on teaching tasks that are more proximal vs more distal to the intervention focus?

Figure 1 summarizes the experimental design and study protocol of the lab experiment. In what follows, we provide a brief overview of study protocol and intervention development and then describe the interventions in detail.

The study protocol and interventions were developed iteratively in a three-stage process. First, we synthesized existing research on practice-based teacher preparation and observed practical examples of decompositions and approximations to identify key mechanisms that distinguished them from one another and from more conceptually-focused approaches to teacher education. We then designed a study protocol and interventions that allowed for the distillation and evaluation of these mechanisms. To ensure we were evaluating only differences in pedagogies, we tried as much as possible to keep the actual content of the three conditions parallel by using similar phrasing, problems and examples across conditions. Second, to enhance internal and external validity, we sought feedback on our protocol and interventions from teacher education researchers and experimental design methods experts. We did this because we wanted our interventions to reflect what both practice-based and conceptually-focused teacher educators were doing in their courses, and we wanted our interventions to be as specified and our design to be as rigorous as possible. Finally, we piloted the study protocol and interventions with graduate and undergraduate students to ensure procedures were easy to understand, and thus that any identified effects could be attributable to the interventions and not due to issues with engaging in the study.

The study began with an intake survey (described below) after which participants scheduled themselves for a two-hour lab session. During each session, facilitators spent the first ten minutes introducing participants to the focus of the study (e.g., “our focus is to improve how you elicit and use student ideas to help students achieve their learning goals”). Facilitators defined the practice, solicited participants’ personal experiences of observing effective eliciting and responding to student thinking, and then reinforced the importance of this practice for student learning, particularly within the context of elementary math teaching. This introduction ensured participants, regardless of eventual treatment condition, were equivalently oriented to the focus of the study and had a shared language for describing the practice they were to develop. Following this introduction, participants were randomly assigned to one of the three hour-long interventions. After finishing the intervention, participants then completed the posttest outcome measures as described below. Participants were compensated $60 for their time and effort once they completed all study procedures, about two and a half hours of effort.

Conceptually-focused pedagogies condition

We designed the conceptually-focused condition to reflect both the theoretical underpinnings of the approach and its most commonly used forms: reading and discussing research on teaching, followed by further reflection and discussion. Facilitators began the intervention by explaining how reading and discussing research on teaching helps novices learn what teaching practices are effective and why, so they can have the conceptual tools to flexibly enact effective practices in future. Participants then spent twelve minutes reading a short chapter that summarized research on eliciting and responding to student thinking, gave examples of teachers doing the practice well and poorly when students provide correct or partially correct responses, and offered key principles for enacting the practice effectively (National Council of

Teachers of Mathematics, 2014). Participants then engaged in a facilitated discussion. The facilitator ensured that each participant shared during the discussion and highlighted key ideas, including: that students sharing and discussing their mathematical ideas deepens their understanding; that different types of questions (gathering information, probing thinking, making the mathematics visible, encouraging reflection and justification) should be purposefully selected and patterned to enable students to engage in meaningful mathematical discussion; and that responding to student thinking involves pushing students to clarify their ideas, even when they provide correct answers, in order to deepen mathematical understanding in line with lesson goals. The discussion concluded with participants identifying key takeaways they would remember for the next time they taught.

At the half hour mark, participants transitioned to a reflection activity where they linked their learning to prior experiences and made commitments for future action. Facilitators began by motivating the activity, noting how consolidating learning makes it more likely that we remember it and put it into practice next time we teach. Participants were then given 20 minutes to individually respond to two prompts: (1) How do the key concepts raised in the previous reading and discussion about eliciting and responding to student thinking align with your own experiences as a student and/or teacher?; and (2) What are two principles about effective teaching that you are taking away and/or re-committing to based on the previous reading and discussion? Facilitators suggested participants write roughly 250 words per question. Facilitators monitored the shared document to ensure participants were engaged. In the remaining 10 minutes, the facilitator had each participant share one principle about eliciting and responding to student thinking they were taking away or recommitting to, encouraging them to make explicit how their thinking had changed over the previous hour. During this discussion the facilitator modeled effective eliciting and responding to student thinking by asking probing questions and making connections between ideas.

Practice-based pedagogies condition

We designed the practice-based condition to reflect commonly enacted forms of decompositions and approximations: analyzing video of expert teaching, and practicing teaching with peers and a teacher educator. Facilitators began the intervention by motivating the value of analyzing expert teaching. Participants then watched a short video of a master teacher helping students solve a single-digit addition problem by eliciting and responding to their thinking. To assist with identifying the component parts of this teaching practice, participants recorded what the teacher said or did in a graphic organizer. Facilitators then led a discussion in which all participants shared what they noticed the teacher saying or doing to elicit and use student thinking. Participants then watched a segment of the clip again to break down the decisions the teacher made, hypothesizing about why they might have chosen to respond to students in particular ways. Facilitators then elicited and responded to participants' share-outs, highlighting key practices the expert teacher used, including: revoicing student ideas; posing a series of factual questions to clarify students' thinking; drawing on and probing students' thinking to highlight mathematical ideas; and comparing students' ideas to illuminate the focal concept of efficient strategies. To conclude the discussion, facilitators prompted participants to identify takeaways from the video and discussion, as well as two strategies that they would remember for the next time they taught.

At the half hour mark, participants transitioned to an approximation activity, where they prepared to teach the same problem taught in the video they just analyzed to at least one other participant (or research assistant [RA] acting as a confederate) and the facilitator. We used the same mathematics problem in the decomposition and approximation activities to reflect sequences of practice-based pedagogies as described in the research literature (McDonald et al., 2013). Participants were given five minutes to prepare, with optional planning questions such as "how would you respond if…?" and "What clarifying or probing questions can you ask students to make their thinking clearer to other students?" After the five minutes, participants took turns playing the teacher or a student, with the facilitator interjecting at appropriate intervals to enact four signature teacher educator strategies during approximations: providing positive feedback in the moment, suggesting what the participant could do to elevate instruction, suggesting a rewind to give the participant another chance, and pausing briefly to hold a quick discussion about what the participant should do next and why (Kavanagh et al., 2020; Shutz et al, 2019). In the last five minutes, facilitators convened a short discussion, prompting participants to share reflections from the experience and identify two strategies they will use in future teaching.

Mixed condition

We designed the mixed condition to reflect – to our knowledge – a commonly enacted form of teacher education that blends conceptually-focused and practice-based pedagogies: reading and discussion followed by video analysis (e.g., van Es et al., 2017). In the first half hour, participants completed the reading and discussion of research on teaching. In the second half hour, participants completed the decomposition of a video of a master teacher helping students solve a single-digit addition problem. Both activities were run as described above.

Method

Sample and Setting

The study was conducted at two highly selective private universities in the northeast United States. As our initial power analyses suggested we needed more participants[1] than the total population of pre-service teachers attending both universities, we decided instead to recruit students from the general undergraduate population. We recruited 185 undergraduates to complete an intake survey and attend one of 45 lab experiment sessions held between February-July 2022. Participants were recruited through departmental and student group mailing lists, flyers posted in classrooms and across campuses, social media networks, and referrals from participants who had already completed the study. Column 1 in Table 1 provides descriptive statistics of demographic and pretest characteristics of the sample overall. On average, participants were approximately 20 years old, approximately half identified as white, a majority identified as female, and nearly two-thirds reported majoring in STEM fields. This was an academically very high achieving sample, with an average percentile rank on the math portion of the SAT or ACT of 96.15.

Although not demographically representative of the average student enrolled in a teacher preparation program in the US (Ingersoll et al., 2018), our recruitment strategies aimed to mitigate concerns about generalizability to the pre-service teacher population in two ways. First, we purposefully sought to recruit participants who were curious about or interested in teaching, and thus who potentially could choose to teach following graduation (neither university we recruited participants from offered an undergraduate teacher preparation pathway). Second, we identified three observable characteristics that distinguish pre-service teachers (*interest in teaching*, *experience with teaching*, and *courses taken in education*), collected participant data on these characteristics, and then checked they were evenly distributed across all three treatment groups. As shown in Table 1, 77% of participants reported an interest in teaching as a career; nearly all participants reported experience leading or teaching children (e.g., as a tutor, teaching assistant or aide); and just over one-third of participants had taken at least one education-related college course. This suggests that although our participants were not pre-service teachers at the time, they were not unfamiliar with teaching – both in terms of interest and experience.[2] Regardless, we test whether findings are sensitive to these characteristics by controlling for them in our conditional model (described below).

Columns 2-4 in Table 1 present the results of regressions designed to assess whether random assignment created, by chance, an imbalance between treatment conditions. Column 2 shows the sample average of pre-treatment covariates for participants in the conceptually-focused condition, and columns 3 and 4 show the results of our statistical tests. Overall, we found participants to be well-balanced on demographic variables, experience and interest in teaching, academic variables, baseline performance, and pretreatment survey outcomes. We found no statistical differences between groups on observable characteristics beyond what might be expected to occur due to chance given the number of variables tested. An F-test of the null hypothesis confirms there were no statistically significant differences between groups across all covariates tested ($F(38, 106) = 1.02$, Prob > F = 0.4507), suggesting groups did not differ on observable measures due to random chance, and any differences we observe in participant outcomes are likely due to the effects of the treatments. Regardless, we present model specifications that include baseline covariates to improve precision and correct for any pre-test imbalances.

Study Administration

Because of the ongoing COVID-19 pandemic and potential for restrictions on in-person research activities across campuses, sessions were conducted online. These sessions were scheduled with a maximum of nine participants and minimum of two. Both the mean and median session had four participants. When sessions involved fewer than six participants, research assistants (RAs) acted as confederates during the interventions to ensure all participants had an equivalent experience with the opportunity to reflect, discuss, and/or practice with peers. These RAs also doubled as the actors.

Each intervention was facilitated by a trained instructor. Instructors were provided with comprehensive and standardized study materials, including (when appropriate) scripts, facilitator notes, and activity guides. To become an instructor, one first had to shadow an instructor through all parts of the study. To guard against concerns that the effects of interventions may be a result of particular instructors, we purposefully assigned instructors to deliver different interventions across sessions. To ensure an equivalent experience was provided across all sessions, the lead author also variously shadowed research team members throughout the study period, and a selection of sessions were recorded and reviewed for fidelity of implementation of study procedures.

Measures

We developed, piloted, and then implemented two measures to assess participants' skills in eliciting and responding to student thinking. The first required participants to view six vignettes of elementary math instruction (three at pre-test, three at post-test), each presented in a comic strip-like format and each featuring situations in which one or two student(s) responded to a teacher question. Vignettes covered diverse content within elementary math, featured a mix of student correct and incorrect answers, and ended with a student response to a teacher question. Participants then recorded what they would say to the class next by typing out their responses (see Figure 2 for an example). Based on the teacher-student interactions, participants could simply affirm a student response ("good job"), probe the student for more information ("can you tell me what has to be true about the five parts?"), or use a student response to further discussion

("I like how Emory said that the figure was divided into *five parts*. What else do people notice about the five parts?"). Vignettes were initially written by a working group of mathematics educators for another project, piloted with math teachers, and then refined. To ensure they would be accessible to our participants (untrained in teaching or mathematics), we piloted and refined them again with a general survey panel of college-educated adults.

The second measure required study participants to spend ten minutes preparing and then ten minutes teaching two math problems to two RAs trained to pose as students (hereafter, *enactments*). One was a two-digit addition problem, and the other was a simple arithmetic word problem. A screenshot of the problems taught is provided in Figure 3. To reduce sources of variability in the enactments, each RA was trained to respond to participants in as consistent a way as possible, including being given standardized answers to the two problems and standardized responses to common participant follow-up questions (e.g., "what was your first step?"; "how did you calculate that?"; "why did you choose that method?").

Scripting RA responses also ensured that participants responded to one correct student answer and one incorrect student answer. This allowed us to measure participants' eliciting and responding to student thinking across one proximal and one distal teaching task: (a) responding to students providing correct responses (the task focused on during the interventions); and (b) addressing student misconceptions (a task not focused on during the interventions, but during which eliciting and using student thinking can occur). We chose these two tasks to capture growth on content taught in all three conditions related to eliciting and responding to students, but to also allow differentiation between the conditions in improving performance with a more generalized task. In addition, the second task may be more difficult; previous research has shown that pre-service teachers find it more challenging to respond to incorrect student answers than correct responses (Shaughnessy & Boerst, 2018; Shaughnessy et al., 2021).

Scoring

We coded both the vignettes and enactment responses using the rubrics shown in Tables 2a and 2b. We iteratively developed these rubrics by first adapting codes from the Mathematical Quality of Instruction observation tool (Hill, 2014) and then refining these codes using data collected during pilot tests of the vignettes and enactment. The coding scheme was then confirmed as fit-for-purpose by ensuring rater agreement amongst all authors of at least 80% using ~10 randomly selected vignette and enactment responses from the lab experiment sample.

Participants' responses to each of the six vignettes and each of the four enactments were double-coded by trained RAs. RAs were randomly assigned to participant responses[3] and were blind to treatment condition. RAs' initial agreement on vignette codes was 73.4%, and enactment codes was 68.3%, both deemed reasonable given the codes required high-inference judgments. Discrepant codes were reconciled by one or two of the authors, with authors discussing a handful of codes and providing feedback to raters during twice-weekly coding meetings.

We then analyzed coded data to arrive at outcome variables. For the vignettes, participant responses were initially coded on two dimensions (*eliciting student thinking; responding to student thinking*), but factor analyses appeared to indicate one-factor models at both pre- and post-tests.[4] As such, we averaged scores across both dimensions for the three items in the pre-test (α=.57) and the three items in the post-test (α=.54). Lower-than-expected reliabilities derive from the small number of items used at each time point, and would have the effect of making any treatment effects in our data harder to detect.

For the enactments, we could not undertake factor analysis because correct and incorrect items were nested within problems, and there were only two items per construct (i.e., two opportunities where participants were tasked to respond to a student's correct response, and two opportunities where participants were tasked to respond to a student's incorrect response). Therefore, based on theoretical expectation, we combined the two correct answers (α=.61; *EnactmentCorrect* in Table 2b) and the two incorrect answers (α=.60; *EnactmentIncorrect* in Table 2b) into single variables for analysis. Again, low reliabilities derive from the limited number of items per construct, and would have the effect of making the detection of treatment impact less likely.

Analytic Models

To examine the average treatment effect of engaging in more practice-based teacher preparation pedagogies, we estimated the following model:

$$Y_i = \beta_0 + \beta_1 PracticeBased_i + \beta_2 Mixed_i + \beta X_i + e_i \quad (1)$$

Here, Y_i represents the outcome for participant i on three measures of responsiveness to students during math instruction: (1) participants' average score across all three posttest vignettes (*Vignettes*); (2) participants' average score across the two enactment opportunities where students responded with correct answers (*EnactmentCorrect*); and (2) participants' average score across the two enactment opportunities where students responded with incorrect answers (*EnactmentIncorrect*). $PracticeBased_i$ is an indicator coded 1 if participant i was randomly assigned to the practice-based condition. $Mixed_i$ is an indicator coded 1 if participant i was randomly assigned to the mixed condition.

To improve the precision of our estimates, our model also includes a series of covariates (X_i) controlling for two sets of baseline characteristics. The first set was demographic characteristics, including participant age, race/ethnicity, whether they attended high school in the US, whether their primary caregiver was ever a teacher, experience and interest in teaching, number of undergraduate courses taken in education and math, field of undergraduate major, and SAT/ACT math percentile rank. The second set of baseline characteristics were pretest measures including math knowledge for teaching[5], math self-concept[6], beliefs about math teaching[7], math teaching self-efficacy[8], and responsiveness to students measured through the vignettes.

In this model, β_1 represents the estimated adjusted average treatment effect of the practice-based condition relative to the conceptually-focused condition, and β_2 represents the estimated adjusted average treatment of the mixed condition relative to conceptually-focused condition. We also used a post-hoc significance test to evaluate any differences between the effects of the practice-based and mixed conditions. Prior to implementing any modeling of outcomes over treatments, we pre-registered our analysis plan in the Registry of Efficacy and Effectiveness Studies (#14160.1v1).

Attrition

Although 185 participants completed the intake survey and logged into one of the scheduled lab experiment sessions, we only collected complete outcome data from 179 participants. Facilitators reported no indications that participant attrition may be due to the treatment or treatment assignment; five participants left due to scheduling or connectivity issues (e.g., one participant left due to a babysitting emergency), and one participant's enactment was accidentally not recorded. We find no statistical evidence of differential attrition by treatment condition[9].

Results

Table 3 summarizes the impact estimates of practice-based condition and the mixed condition on participants' eliciting and responding to student thinking as measured across all three outcome measures. Model 1 is an unconditional model (i.e., equation 1 without the series of covariates (X_i)) with vignettes as the outcome measure. Model 2 is the conditional model (i.e., equation 1 above) with vignettes as the outcome measure. The remaining models follow the same pattern, but with different outcome measures: models 3 and 4 use the enactments where participants were tasked to elicit and respond to students' correct responses (*EnactmentCorrect*); models 5 and 6 use the enactments where participants were tasked to elicit and respond to students' incorrect responses (*EnactmentIncorrect*).

In looking at our preferred estimates that include both demographic and pretest covariates (Models 2, 4, and 6), we see that both the practice-based and mixed conditions produce statistically significant positive effects relative to the conceptually-focused condition.

On the vignettes, participants in the practice-based and mixed conditions performed similarly better than participants in the conceptually-focused condition. Controlling for demographic and pretest variables (Model 2), those in the practice-based condition scored on average 0.29 points higher than those in the conceptually-focused condition (ES = .67 SD; p-value <.001). Participants in the mixed condition scored on average 0.31 points higher than participants in the conceptually-focused condition on the three-point scale (ES = .72 SD; p-value <.001). We detected no statistically significant difference in performance on the vignettes between those in the practice-based and mixed conditions.

On the more proximal measure of teaching practice (i.e., eliciting and responding to students' correct answers, a task that participants were trained on), participants in both the practice-based and mixed condition outperformed participants in the conceptually-focused condition. Controlling for demographic and pretest variables (Model 4), participants in the mixed condition scored on average 0.63 points higher than those in the conceptually-focused condition on the four-point scale (ES = .82 SD; p-value <.001). Those in the practice-based condition scored on average 1.19 points higher than those in the conceptually-focused condition (ES = 1.54 SD; p-value <.001). Further, the difference between the purely practice-based and mixed conditions was statistically significant (ES = .72 SD; p-value <.001), favoring the purely practice-based condition.

On the more distal measure of teaching practice (i.e., eliciting and responding to students' incorrect answers, a similar but more challenging task), only the practice-based condition had statistically significant effects. Controlling for demographic and pretest variables (Model 6), participants in the practice-based group scored on average 0.3 points higher than those in the conceptually-focused condition on the four-point scale (ES = .43 SD; p-value <.05).

We note stable estimates across all six model specifications, suggesting that the treatment effects are robust, that randomization likely succeeded in producing balanced groups, and that any chance imbalance in observed demographic characteristics at baseline are not influencing effects.

Taken together, these results suggest practice-based pedagogies can be more effective at preparing participants to elicit and respond to student thinking during math instruction than conceptually-focused pedagogies alone. They also suggest the combination of analyzing expert teaching and then practicing teaching with peers and a teacher educator can improve eliciting and responding to student thinking on a teaching task that is slightly more distal and difficult than what pre-service teachers have been trained on.

Discussion and Conclusion

This study provides the first causal evidence showing that two forms of practice-based approaches to teacher education – decompositions and approximations – enhance teaching performance as compared to more traditional, conceptually-focused approaches. Unlike prior studies which did not use an experimental design (Kavanagh & Rainey, 2017), did not test interventions as closely aligned to practice-based teacher education theory and practice (Sims et al., 2023), or compared interventions that were different in *both* pedagogy and content (Strasser et al., 2021), we used a lab experiment to isolate the specific impact of practice-based pedagogies on practice as measured through written vignettes and live enactments. We found significant positive effects in favor of practice-based pedagogies.

We also tested the effects of practice-based approaches on both the teaching task participants were trained on, and one that was similar but more challenging. Our goal was to generate evidence on whether practice-based approaches may be too narrow in their preparation, and whether conceptually-focused approaches, with their more general focus, are better able to prepare novices regardless of situation and context. We find the opposite: participants who experienced the practice-based condition outperformed those in the conceptually-focused and mixed condition on the more distal teaching task. These results provide initial empirical evidence that validates some theory underlying practice-based approaches, particularly approximations (Grossman, Compton et al., 2009; Grossman, 2018; McDonald et al., 2013), showing that teachers can be more effectively prepared for complex teaching practices by learning in and from appropriately scaffolded opportunities to practice.

Given the limited time available to teacher educators in teacher preparation programs, decisions about what specific pedagogies to prioritize matter. This study offers evidence that adopting practice-based pedagogies – and in particular those that include approximations of practice – may be the most efficient and effective use of time, at least in the case of developing the complex teaching practice of eliciting and responding to student thinking. While we know of no recent study describing typical teacher education practice across the US (see Grossman, Hammerness et al., 2009; Wideen, Mayer-Smith, & Moon, 1998), our conversations with teacher educators in the field generally suggest that conceptually-focused pedagogies remain the modal approach. This suggests improvements in pre-service teacher preparedness could be possible if more teacher educators make greater use of practice-based approaches.

However, before making a stronger recommendation in favor of practice-based pedagogies, this study requires several types of replication. Literal replications – using the same study protocol, treatments and measures – would help verify our conclusions. Conceptual replications could then help further ensure generalizability to theory. These replications should systematically vary elements of how the practice-based and conceptually-focused treatments are designed, and assess the robustness of findings when focused on different teaching tasks (e.g., lesson planning, leading discussions). Additionally, the outcomes of teacher preparation are much broader than what we measured (e.g., teacher identity; skills in analyzing teaching), necessitating testing whether practice-based pedagogies are more effective in preparing novices for those outcomes. Studies like these that test the robustness of our findings can then be used to help develop stronger theory about when to use either practice-based or conceptually-focused pedagogies.

Future research should also test whether our results replicate within the context of teacher preparation programs with pre-service teachers. This could occur through adapting our study materials to run within a single lesson within a general pedagogy, methods, or introductory course (i.e., randomly assigning students within the lesson to different conditions). Or future researchers could expend the materials across a sequence of lessons within a course focused on a particular teaching skill. These studies could also test whether our findings replicate for pre-service teachers across different pathways of teacher preparation, and in stages of their preparation. While challenging to implement, if these studies replicate our findings, they would also help strengthen the evidence on the validity of lab experiments to inform teacher preparation practice.

Our study also raises the possibility of economizing the teacher preparation curriculum by identifying complementarities between pedagogy and content. We results show that practice-based pedagogies can improve participants' skills in eliciting and responding to students providing correct answers, as well as their skills in a slightly more challenging task: eliciting and responding to students providing incorrect answers. Future research could specifically investigate the extent to which preparing teachers for certain tasks translates to other proximal teaching skills. This could potentially help teacher educators better curate the content of teacher preparation.

Most broadly, this study contributes to efforts to generate a more rigorous evidence base on teacher education pedagogies. The field needs more causal evidence on the relative effectiveness of teacher preparation practices, and our experiences attest to recent scholars' concerns regarding challenges with research design and measurement hampering causal research in the field of teacher education (Hill et al., 2021; Mancenido, 2023). We note three specific areas in which further research could assist future researchers. First, we admit to having to think long and hard about how to specify the conditions we tested. Our aim was to isolate the specific impact of practice-based pedagogies vs conceptually-focused pedagogies separate to the content being taught. This required us to specify the pedagogies tested in a way that best represented the theories underlying both approaches, minimize the effects of even slight variations in content, and reflect, as much as possible in the context of a lab experiment, current teacher educator practice. We needed to give each theory a square chance. To achieve this, we relied on expert review, but our efforts would have been enhanced by better information on typical teacher education practice, and more clear research and guidance about how to design programs of research in teacher education. Second, we, like others before us, note the absence of performance measures related to teaching practices that are generally agreed-upon outcomes of teacher preparation (Grossman & McDonald, 2008; Mancenido, 2023). In practice, this meant that we had to create our own measures for the purpose of the study, with mixed results – specifically, sometimes lower-than-desired score reliabilities. Future causal research in the field would be strengthened by field-developed common performance measures that both return reliable scores and are independent of any particular intervention. Third, we are grateful to the growth in training and resources for running experimental and quasi-experimental studies in the broad field of education. However, we note the absence of specific guidance for causal studies in the context of teacher preparation, including discussions on the practical challenges of conducting studies within and in parallel to teacher preparation programs.

We hope that the results of this and future work allow teacher educators and pre-service teachers to direct their valuable time and energy to practices most likely to strengthen teaching skills that drive improvements in student learning. Likewise, we hope future causal research work can build off our study to support teachers in developing the wide range of ambitious teaching skills necessary to effectively educate the next generation of learners.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond, & G. Skyes (Eds.), *Teaching as a learning profession: Handbook of policy and practice* (pp. 3-32). Jossey-Bass.

Barnhart, T., & van Es, E. (2015). Studying teacher noticing: Examining the relationship among pre-service science teachers' ability to attend, analyze and respond to student thinking. *Teaching and Teacher Education*, *45*, 83-93. https://doi.org/10.1016/j.tate.2014.09.005

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review*, *48*, 16-29. https://doi.org/10.1016/j.econedurev.2015.05.005

Boerst, T., Sleep, L, Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record*, *113*(12), 2844-2877. https://doi.org/10.1177/016146811111301207

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* National Academy Press.

Clark, C. M. (1988). Asking the right questions about teacher preparation: Contributions of research on teacher thinking. *Educational Researcher*, *17*(2), 5-12. https://doi.org/10.3102/0013189X017002005

Cochran-Smith, M., Villegas, A. M., Abrams, L., Chavez Moreno, L., Mills, T., & Stern, R. (2016). Research on teacher preparation: Charting the landscape of a sprawling field. In C. Bell & D. Gitomer (Eds.), *Handbook of research on teaching* (pp. 439-547). American Educational Research Association.

Cohen, J., Krishnamachari, A. & Wong, V. C. (2021). Experimental Evidence on the Robustness of Coaching Supports in Teacher Education. (EdWorkingPaper: 21-468). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/dgf9-ca95

Council of Chief State School Officers (2013). *InTASC Model Core Teaching Standards and Learning Progressions for Teachers 1.0*. https://ccsso.org/resource-library/intasc-model-core-teaching-standards-and-learning-progressions-teachers-10

Darling-Hammond, L. & Bransford, J. (Eds.). (2005). *Preparing Teacher for a Changing World: What teachers should learn and be able to do*. Jossey-Bass.

Ericsson, K. A. (2002). Attaining Excellence Through Deliberate Practice: Insights from the Study of Expert Performance. In C. Desforges & R. Fox (Eds.), *Teaching and Learning: The essential readings* (pp. 4–37). Blackwell Publishers. https://doi.org/10.1002/9780470690048.ch1

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational research*, *38*(1), 47-65. https://doi.org/10.1080/0013188960380104

Grossman, P. L. (2018). *Teaching core practices in teacher education*. Harvard Education Press.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205. https://doi.org/10.3102/0002831207312906

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: theory and practice*, *15*(2), 273-289. https://doi.org/10.1080/13540600902875340

Grossman, P., Kavanagh, S. S., & Dean, C. P. (2018). The turn to practice in teacher education. In P. Grossman (Ed.), *Teaching core practices in teacher education* (pp. 1-14). Harvard Education Press.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*(9), 2055-2100. https://doi.org/10.1177/016146810911100905

Hill, H. C. (2014) Mathematical Quality of Instruction. 4-Point Version. Learning Mathematics for Teaching.

Hill, H. C., Mancenido, Z., & Loeb, S. (2021). *Effectiveness Research for Teacher Education*. EdWorkingPaper: 21-252. Annenberg Institute at Brown University. https://doi.org/10.26300/zhhb-j781

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11-30. https://doi.org/10.1086/428763

Ingersoll, R. M., Merrill, E., Stuckey, D., and Collins, G. (2018). *Seven Trends: The Transformation of the Teaching Force – Updated October 2018*. CPRE Research Reports. https://repository.upenn.edu/cpre_researchreports/108.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*(3), 587-613.

Kavanagh, S. S., & Rainey, E. C. (2017). Learning to support adolescent literacy: Teacher educator pedagogy and novice teacher take up in secondary English language arts teacher preparation. *American Educational Research Journal*, *54*(5), 904-937. https://doi.org/10.3102/0002831217710423

Kavanagh, S. S., Metz, M., Hauser, M., Fogo, B., Taylor, M. W., & Carlson, J. (2020). Practicing responsiveness: Using approximations of teaching to develop teachers' responsiveness to students' ideas. *Journal of Teacher Education*, *71*(1), 94-107. https://doi.org/10.1177/0022487119841884

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. https://doi.org/10.3102/0034654318759268

Lampert, M. (2010). Learning Teaching in, from, and for Practice: What Do We Mean? *Journal of Teacher Education*, *61*(1–2), 21–34. https://doi.org/10.1177/0022487109347321

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping It Complex: Using Rehearsals to Support Novice Teacher Learning of Ambitious Teaching. *Journal of Teacher Education*, *64*(3), 226–243. https://doi.org/10.1177/0022487112473837

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the Research Base That Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260–293. https://doi.org/10.3102/0162373719849044

Maggioni, L., & Parkinson, M. M. (2008). The role of teacher epistemic cognition, epistemic beliefs, and calibration in instruction. *Educational Psychology Review*, *20*, 445-461. https://doi.org/10.1007/s10648-008-9081-8

Mancenido, Z. (2023). Impact Evaluations of Teacher Preparation Practices: Challenges and Opportunities for More Rigorous Research. *Review of Educational Research*. https://doi.org/10.3102/00346543231174413

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education*, *64*(5), 378-386. https://doi.org/10.1177/0022487113493807

Michigan Department of Education (2020). *Standards for the Preparation of Teachers: Professional Knowledge and Skills, Middle Grades (5-9) and High School (7-12)*. Michigan State Board of Education.

Mullis, I.V.S. & Martin, M.O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2015/frameworks.html

National Academies of Sciences, Engineering, and Medicine (2022). *The Future of Education Research at IES: Advancing an Equity-Oriented Science*. The National Academies Press. https://doi.org/10.17226/26428.

National Council of Teachers of Mathematics (2014). *Principles to Actions: Ensuring Mathematical Success for All*.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*(4), 317-328. https://doi.org/10.1080/0022027870190403

Nystrand, M., & Gamoran, A. (1991). Instructional Discourse, Student Engagement, and Literature Achievement. *Research in the Teaching of English*, *25*(3), 261–290.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307-332. https://doi.org/10.3102/00346543062003307

Pajares, F. (1993). Preservice teachers' beliefs: A focus for teacher education. *Action in Teacher Education*, *15*(2), 45-54. https://doi.org/10.1080/01626620.1993.10734409

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The Relationship Between Teachers' Beliefs and Practices in Reading Comprehension Instruction. *American Educational Research Journal*, *28*(3), 559–586. https://doi.org/10.3102/00028312028003559

Schutz, K. M., Danielson, K. A., & Cohen, J. (2019). Approximations in English language arts: Scaffolding a shared teaching practice. *Teaching and Teacher Education*, *81*, 100-111. https://doi.org/10.1016/j.tate.2019.01.004

Shaughnessy, M., & Boerst, T. A. (2018). Uncovering the Skills That Preservice Teachers Bring to Teacher Education: The Practice of Eliciting a Student's Thinking. *Journal of Teacher Education*, *69*(1), 40–55. https://doi.org/10.1177/0022487117702574

Shaughnessy, M., DeFino, R., Pfaff, E., & Blunk, M. (2021). I think I made a mistake: How do prospective teachers elicit the thinking of a student who has made a mistake?. *Journal of Mathematics Teacher Education*, *24*, 335-359. https://doi.org/10.1007/s10857-020-09461-5

Sims, S., Fletcher-Wood, H., Godfrey-Fausset, T., & Meliss, S. (2023). *Modelling evidence-based practice in initial teacher training: causal effects on teachers' skills, knowledge and self-efficacy*. Ambition Institute.

Stein, M. K., Correnti, R., Moore, D., Russell, J. L., & Kelly, K. (2017). Using theory and measurement to sharpen conceptualizations of mathematics teaching in the common core era. *AERA Open*, *3*(1). https://doi.org/10.1177/2332858416680566

Stockero, S. L., Leatham, K. R., Ochieng, M. A., Van Zoest, L. R., & Peterson, B. E. (2020). Teachers' orientations toward using student mathematical thinking as a resource during whole-class discussion. *Journal of Mathematics Teacher Education*, *23*, 237-267. https://doi.org/10.1007/s10857-018-09421-0

Strasser, K., Barra, G., & Mendive, S. (2021). The role of video analysis and deliberate rehearsal in the development of language stimulation skills in pre-service preschool teacher preparation: an experimental study. *Teaching and Teacher Education*, *106*, 103459. https://doi.org/10.1016/j.tate.2021.103459

Stroupe, D., & Gotwals, A. W. (2018). "It's 1000 degrees in here when I teach": Providing preservice teachers with an extended opportunity to approximate ambitious instruction. *Journal of Teacher Education*, *69*(3), 294-306. https://doi.org/10.1177/0022487117709742

van Es, E. A., Cashen, M., Barnhart, T., & Auger, A. (2017). Learning to notice mathematics instruction: Using video to develop preservice teachers' vision of ambitious pedagogy. *Cognition and Instruction*, *35*(3), 165-187. https://doi.org/10.1080/07370008.2017.1317125

Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research*, *68*(2), 130-178. https://doi.org/10.2307/1170752

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903. https://doi.org/10.1002/sce.21027

Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of Teacher Education*, *63*(5), 376-382. https://doi.org/10.1177/0022487112445789

Endnotes

1. The total number of pre-service teachers attending both universities was around 50. Under the assumptions of a significance level at 0.05, power at .8, and a minimum detectable effect size of .4SD on teacher practice, our power analysis suggested a minimum sample size of 162 with 54 in each treatment group. We used a slightly more conservative MDES than what Lynch et al. (2019) find in their meta-analysis of STEM professional development and curriculum materials on teacher practice (they report pooled effect sizes of 0.561SD); and more conservative than what Kraft et al. (2018) find in their meta-analysis of teacher coaching on instruction (pooled effect sizes of 0.49 SD).

2. One reason why sample generalizability may be a concern is that Cohen et al. (2021) found that targeted, directive coaching improved teaching practice of pre-service teachers but not for general undergraduates. The researchers attribute this finding to general undergraduates' "lack of schema or prior knowledge about the skills targeted in coaching" (p. 20). We do not believe this is a problem in our study given our sample's interest and experience in teaching is much higher than Cohen et al.'s (2021) sample where only ~43% reported any interest in teaching as a career, and ~63% reported prior experience working with children.

3. As some raters also doubled as actors for the enactments, we reassigned enactments during the rating process so that no rater would score enactments where they were also the actor.

4. For example, for the pretest items, the two factor model could not be identified (χ^2(df=7)=82.93, p<0.001, CFI = 0.79, RMSEA = 0.25, SRMR = 0.20) and the one-factor model fit reasonably (χ^2(df=6)=11.81, p=0.005, CFI = 0.97, RMSEA = 0.11, SRMR = 0.08). We report robust fit statistics as our data are ordinal and we use the WLSMV estimator in the Lavaan R package.

5. We selected 14 items related to number and operations from the broader pool of Mathematical Knowledge for Teaching (Hill et al., 2004). We selected items that were most likely to be broadly accessible given that our participants were not trained math teachers.

6. We used 4 items measuring math self-concept from the TIMSS 2015 Student Questionnaire (Mullis & Martin, 2013). Scores were averaged to form a scale with reliability of 0.90.

7. We adapted items from Stockero et al. (2020) and Stein et al. (2017) measuring belief in the value of eliciting and responding to student thinking when teaching mathematics. We initially identified 19 items based on our assessment of their accessibility given our non-teaching audience. We piloted these items using a survey panel of college-educated young people, then used factor analysis to identify the 10 highest leverage items across two different scales (5 items each). For this study, we averaged scores on each of the two scales: one measuring beliefs related to using student thinking during math instruction with a reliability of 0.73; and one measuring beliefs related to whether teachers should play a more traditional role in math teaching with a reliability of 0.64. We found these reliability co-efficients acceptable given the small number of items.

8. We adapted 9 items measuring math teacher self-efficacy from the TIMSS 2015 Teacher Questionnaire (Mullis & Martin, 2013). Scores were averaged to form a scale with reliability of 0.88.

9. Three participants dropped from the mixed condition, two participants dropped from the practice-based condition, and one participant dropped from the conceptually-focused condition.

Table 1
Baseline Covariate Balance by Treatment Condition

	(1)	(2)	(3)	(4)	(5)
	All Participants	***Conceptually -Focused Condition***	***Practice-Based Condition***	***Mixed Condition***	
Pre-treatment Covariate	Mean	Mean	Coefficient	Coefficient	n
Demographic Information					
Age	20.10	20.12	0.00	-0.02	185
Attended University #1	0.52	0.48	0.03	0.09	185
Primary caregiver were/are teacher	0.31	0.29	0.01	0.06	180
English as primary language at home	0.64	0.71	-0.11	-0.11	180
Attended HS in the US	0.8	0.83	-0.03	-0.07	
Racial/Ethnic Background					
Asian	0.36	0.28	0.12	0.11	185
Black	0.12	0.10	0.00	0.06	185
Hispanic	0.15	0.17	0.03	-0.09	185
White	0.48	0.50	-0.05	-0.01	185
Other	0.04	0.03	0.03	0.00	185
Multi	0.08	0.07	0.00	0.04	185
Gender					
Male	0.32	0.33	-0.01	-0.02	185
Female	0.62	0.63	-0.02	-0.03	185
Other	0.05	0.03	0.03	0.03	185
Experience and Interest in Teaching					
0 yrs exp. teaching children	0.08	0.07	0.00	0.04	185
0-2 yrs exp. teaching children	0.40	0.35	0.05	0.09	185

>2 yrs exp. teaching children	0.52	0.58	-0.05	-0.14	185
0 yrs Exp. teaching math	0.39	0.35	0.02	0.09	185
0-2 yrs exp. teaching math	0.46	0.45	0.02	0.01	185
>2 yrs exp. teaching math	0.15	0.20	-0.04	-0.10	185
No interest in teaching	0.23	0.18	0.01	0.13†	185
Small interest in teaching	0.42	0.45	-0.05	-0.04	185
Some to serious interest in teaching	0.35	0.37	0.04	-0.10	185
Undergraduate Coursework					
0 courses in education	0.63	0.55	0.05	0.18*	185
1-2 courses in education	0.19	0.23	0.01	-1.72*	185
>3 courses in education	0.18	0.22	-0.06	-0.06	185
0 courses in math	0.22	0.32	-0.16*	-0.13†	185
1-2 courses in math	0.47	0.45	0.03	0.03	185
>3 courses in math	0.31	0.23	0.12	0.10	185
Academic Achievement					
SAT/ACT Percentile	96.15	95.62	1.57	0.12	150
Undergraduate Major					
STEM	0.63	0.65	0.00	-0.05	185
Social Sciences	0.29	0.20	0.12	0.13	185
Humanities	0.26	0.27	-0.06	0.03	185
Education	0.15	0.22	-0.07	-0.14*	185
Other	0.05	0.05	0.01	-0.02	185
Baseline Performance and Survey Outcomes					
Math Knowledge for Teaching	0.75	0.76	-0.02	0.00	185
Responsiveness in Vignettes - Pretest	4.31	4.10	0.28	0.33	182

Math Self-Concept	3.88	3.84	0.09	0.02	185
Math Teaching Self-Efficacy	3.28	3.45	-0.21	-0.30*	185
Belief in using student thinking as resource	4.38	4.49	-0.11	-0.23*	185
Belief in traditional role of math teachers	2.17	2.10	0.13	0.07	185

*Note: Demographic information comes from the intake survey. Each row represents results from a separate regression with the same right-hand side specification but a different baseline covariate as the dependent variable. We also conducted a multivariate regression model for all covariates predicted by treatment status: F(38, 106) = 1.02, Prob > F = 0.4507. †p < .10. *p < .05. **p < .01. ***p <.001*

Table 2a

Coding Scheme for Vignettes

Eliciting Student Thinking – for all *Vignettes*

N/A - No teacher utterance or utterance is unrelated.	**0 - No eliciting of student thinking.** Teacher makes a statement, whether approval or correction. No expectation of student response. Asks a question to the whole of the class / not the initial respondent (e.g., "what do others think? Or "Did anyone else have a different way?")	**1 - Some elicitation of student thinking.** Teacher poses a simple proforma clarifying or probing question that does not draw specifically on what the student has shared (e.g., "how did you get that?"; "can you explain to everyone what you mean?").	**2 - Strong elicitation of student thinking.** Teacher poses a question to clarify or probe a specific idea that the student has shared.

Responding to Student Thinking – for all *Vignettes*

N/A - Teacher utterance is not responsive to prompt.	**0 - No or minimal use of student thinking.** Response is plausible in the given situation (e.g. "Oh!). Teacher responds in a pro forma way (e.g., acknowledges student response is correct or incorrect; thanks or praises the student but does not respond to their mathematical ideas; asks for another student answer to the problem; provides direct instruction about the problem to the class).	**1 - Some use of student thinking.** Teacher response goes beyond pro forma to feature some use of student ideas; focus stays at least for a moment on the student(s), but is does not rise to strong use (e.g., brief restatement of student method; brief restatement followed by teacher direct instruction; asking students to repeat their answers; asks class whether student is correct).	**2 - Strong use of student thinking.** Teacher weaves student ideas into the development of the mathematics (e.g., compares student responses; adds mathematical emphasis to student method or solution; fills in mathematical details that were missing from student response; asks student why question; asks class why question based on a student(s) method; asks "How do you know?")

Table 2b

Coding Scheme for Enactments

Eliciting and Responding to Student Thinking – for *EnactmentCorrect*				
N/A - No teacher utterance or utterance is unrelated.	**0 - No eliciting and use of student thinking.** Teacher responds in a non-specific way, without expectation of student response, or the teacher responds with a simple approval or correction.	**1 - Minimal eliciting of student thinking.** Teacher asks a single question or makes a statement inviting a student response, and then moves quickly to stating approval or correction.	**2 - Some eliciting and use of student thinking.** Teacher goes beyond a single question or statement, seeking further clarifications or justifications of the student's ideas. Focus stays at least for a moment on the student(s), but does not rise to strong use.	**3 - Strong eliciting and use of student thinking.** Teacher poses multiple questions that attend specifically to student ideas. Teacher uses questioning to get students to state the key mathematical concept(s) being developed.

Eliciting and Responding to Student Thinking – for *EnactmentIncorrect*				
N/A : Teacher utterance does not acknowledge misconception.	**0 - Teacher does not probe student thinking and/or just provides a correction.** Teacher does not ask questions to identify the misconception and just tells the student what the problem is actually asking them to do / what the problem is not asking them to do.	**1 - Minimal elicitation of student thinking and no attention to it during remediation.** Teacher asks a single simple proforma question to elicit student thinking (e.g., "what did you do?"; "how did you do that?") but then moves straight to correction and does not engage with the student's response.	**2 - Teacher elicits student thinking and uptakes some of that thinking to remediate.** Teacher elicits some information from the student beyond an initial question to briefly "stick with student" and establish what student is doing. Teacher may then use a series of involved funneling questions to lead student to correct answer.	**3 – Teacher engages with student's thinking during remediation.** Must include one/both of: (1) guiding student through the mistake to get to the root cause (not just a simple "why did you do that?" / "how did you get that?"); (2) probing student responses to get them to explain why the method is not sound.

Table 3

Treatment Effects on Participants' Performance Outcomes

	Vignettes		*EnactmentCorrect*		*EnactmentIncorrect*	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Practice-Based Condition	.297*** (.075)	.288*** (.083)	1.200*** (.110)	1.188*** (.118)	.322* (.125)	.296* (.133)
Mixed Condition	.270*** (.075)	.310*** (.086)	.650*** (.110)	.630*** (.124)	.110 (.126)	.108 (.139)
Demographic And Pretest Covariates		X		X		X

*†p < .10. *p < .05. **p < .01. ***p <.001. For interpretability, estimates shown are on the three-point scale for Vignettes (Model 1 and 2) and four-point scale for EnactmentCorrect (Model 3 and 4) and EnactmentIncorrect (Model 5 and 6). For reference, the average scores for the ReadingReflecting group on: (a) the Vignettes was .866 scale points (scale = 0-2); (b) EnactmentCorrect was .983 (scale = 0-3); and (c) EnactmentIncorrect was 1.186 (scale = 0-3). The reference category is the conceptually-focused condition, such that all estimates are the average impact of experiencing the relevant condition relative to the conceptually-focused condition.*

Figure 1

Summary of Experimental Design and Study Protocol

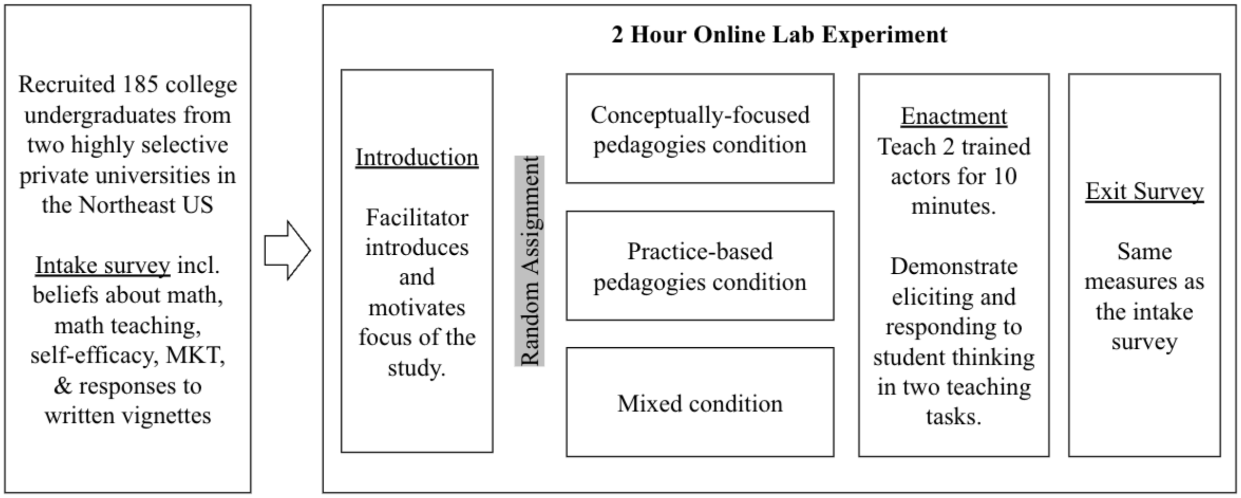
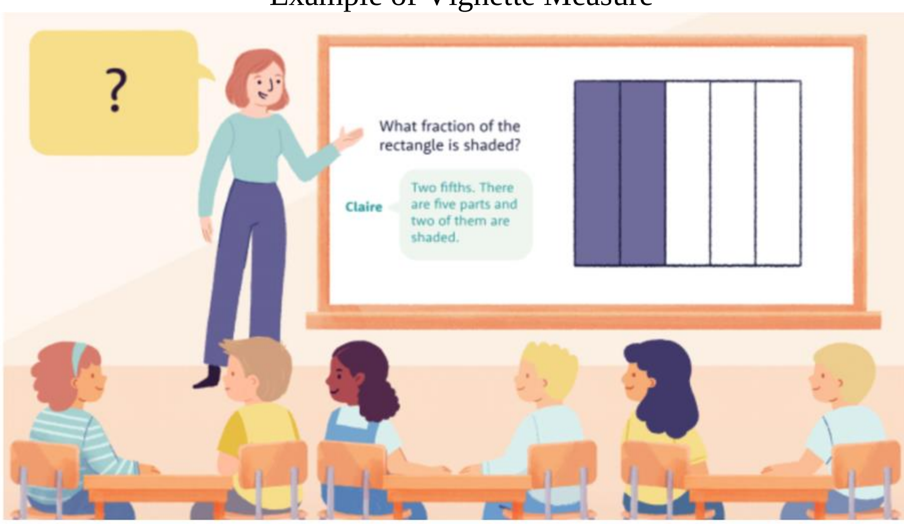


Figure 2

Example of Vignette Measure



Now we ask you to *imagine that you are the teacher* in the animation and to make a decision about what to say to the class next. **Please type what you would say to the class next into the response box below.**

Figure 3

Screenshot of the Online Lab Experiment