ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-896

Clinical teaching learning trajectory: Exploring field supervisor written feedback on clinical teacher pedagogy

Andrew Kwok
Texas A&M University

Iliana De La Cruz
Texas A&M University

Michelle Kwok
Texas A&M University

Field supervisors are central to clinical teaching, but little is known about how their feedback informs preservice teachers (PSTs) development. This sequential mixed methods study examines over 3,000 supervisor observation evaluations. We qualitatively code supervisor written feedback, which indicates 2 broad pedagogical categories and 9 separate skills. We then quantize these feedback codes to identify the variation in the presence of these codes across PST characteristics, and then use several modeling techniques to indicate that specific feedback codes are negatively associated with evaluation score. Managing student attention was most detrimental to scores in early observations whereas instructional feedback (e.g., lesson delivery) was prioritized later in clinical teaching. Findings inform teacher preparation policy on understanding PST development and improving supervisory feedback.

VERSION: January 2024

Suggested citation: Kwok, Andrew, Iliana De La Cruz, and Michelle Kwok. (2024). Clinical teaching learning trajectory: Exploring field supervisor written feedback on clinical teacher pedagogy. (EdWorkingPaper: 24-896). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/r9hh-hs29

Clinical teaching learning trajectory: Exploring field supervisor written feedback on clinical teacher pedagogy

Andrew Kwok, Iliana De La Cruz, & Michelle Kwok
Texas A&M University

Corresponding Author:
Andrew Kwok
Texas A&M University
540 Ross Street
College Station, Texas 77840
akwok@tamu.edu
314-651-7598 (cell)

Acknowledgements
We would like to thank Jenna Davis, Katherine Davis, and Paola Guerrero-Rosada for their assistance in writing this manuscript.

Abstract

Field supervisors are central to clinical teaching, but little is known about how their feedback informs preservice teachers (PSTs) development. This sequential mixed methods study examines over 3,000 supervisor observation evaluations. We qualitatively code supervisor written feedback, which indicates 2 broad pedagogical categories and 9 separate skills. We then quantize these feedback codes to identify the variation in the presence of these codes across PST characteristics, and then use several modeling techniques to indicate that specific feedback codes are negatively associated with evaluation score. Managing student attention was most detrimental to scores in early observations whereas instructional feedback (e.g., lesson delivery) was prioritized later in clinical teaching. Findings inform teacher preparation policy on understanding PST development and improving supervisory feedback.

Introduction

Field supervisors hold a consequential role in preservice teacher (PST) development, offering an external evaluation of PST instruction, speaking to objective notions of pedagogy across contexts, and serving as a conduit between the preparation program and the placement (Jacobs et al., 2017). These multifaceted responsibilities correspond to their potential in influencing PST quality, entry into the profession, and where PSTs choose to teach (Bartanen & Kwok, 2021).

However, little is known about the variation in supervisor feedback, particularly in how it can promote PST development across teacher preparation. Feedback is essential for both students (Hattie & Timperley, 2007) and beginning teachers (Hunter & Springer, 2022), but given the contextual difference of clinical teaching—including the unique role of supervisors and the nascent development of PSTs—systematically understanding supervisor feedback is uniquely necessary. Specifically, understanding PSTs' developmental growth through feedback of authentic practice is vital to enhancing teacher preparation and policy. Incremental change in this central structure of clinical teaching could improve preparation across programs.

Our study examines a large-scale set of field supervisor feedback from one teacher education program (TEP). We qualitatively analyze over 3,000 supervisor observational responses of clinical teachers spanning three years to unearth skills targeted for early pedagogical improvement (i.e., feedback codes). Then, we investigate the variation in feedback codes, identifying the extent to which context is associated with supervisor recommendations. Finally, we examine the predictive validity of codes on clinical teaching overall evaluation ratings to understand how specific language influences PST developments. The purpose of these

analyses is to identify a baseline of PST practices and that could inform a framework of development throughout clinical training. The research questions guiding this study are:

1. In what pedagogical areas do clinical teaching field supervisors provide feedback?
2. How do supervisor feedback codes vary by context?
3. What is the relationship between clinical teaching evaluation rating and supervisor feedback codes?
4. How does feedback type affect evaluation rating across observation number?

From supervisor written feedback, two categories of feedback (instructional development and monitoring student behavior) and nine separate skills were identified from qualitative analyses. We examine patterns across the presence of these codes by observation number (i.e., first through fourth observation) and to a lesser extent evaluation score, subject area, and grade level. From this heterogeneity, we test whether there is a significant relationship between these codes and PST clinical teaching evaluation score. We utilize several different modeling techniques to account for the nested nature of the observations within PST as well as PSTs by supervisors. Consistently, three feedback codes were found to be significantly negatively associated with their evaluation score: maintaining student attention, using non-verbal techniques, and lesson delivery.

Further, to understand the longitudinal nature of our data, we explore development through change in feedback codes by observation number. We identify significant changes indicating the early focus on maintaining student attention and non-verbal techniques towards later improvement on lesson delivery, lesson cycle, and behavioral corrections. These results provide an initial framework of clinical teacher development by reiterating foundational skills for authentic pedagogy.

Background

Our examination of field supervision feedback requires an understanding of the role of field supervisors during clinical teaching, alongside the value of instructional feedback within educational settings broadly. Below, we review studies in both fields and synthesize them towards an understanding novice teacher development.

Field Supervision

Throughout clinical teaching, PSTs participate in professional activities under the guidance of a cooperating (or mentor) teacher for at least one full semester (National Council for Accreditation of Teacher Education, 2010). The entirety of this experience has been consistently identified as one of the most influential teacher preparation structures (Anderson & Stillman, 2013; Goldhaber et al., 2022; Ronfeldt, 2021). A recent report from the National Council on Teacher Quality stated that within clinical teaching, an essential component is that a "supervisor from the program observes a candidate at least four times during the semester…providing written feedback with each observation" (Pomerance & Walsh, 2020, p. 4). Field supervision occurs throughout the span of preparation (e.g., early field experiences, Kwok & Bartanen, 2022) with greater resources and time generally dedicated to clinical teaching (Jacobs et al., 2017). Despite the ubiquity of field supervision, it has received limited attention, likely for several reasons.

First, there is a range of individuals who serve or are hired as supervisors, from tenure line professors to adjunct instructors (Jacobs et al., 2017), providing little consistency in role. This variation likely underlies increasing teacher education policies that have reduced the requirements to become a supervisor, despite evidence that effective supervisors can positively influence PST development (Darling-Hammond, 2014; Grossman et al., 2008; Mok & Staub,

2021; Wasburn-Moses & Noltemeyer, 2018). Outright, field supervision has been deemed second class work of clinical field experiences (Slick, 1998; Zeichner, 2021), in which faculty members rarely want to participate and programs often overlook it despite the large time investment and coordination required.

Second, the work of PST supervision is expanding and becoming increasingly sophisticated (Burns et al., 2016). Duties regularly include observation, evaluation, and instructional or social emotional support (Caires et al., 2010). In addition, supervisors are often responsible for dynamics outside of the field classroom, such as building collaborative relationships amongst the PSTs, cooperating teachers, students, and administration (Campbell & Lott, 2010; Nguyen, 2009), alongside facilitating connections between the field classrooms and university courses (Hertzog & O'Rode, 2011; McDonnough & Matkins, 2010; Ward et al., 2011). Supervisors may even be involved in university teacher education efforts of curriculum planning (Turunen & Tuovila, 2012) and research for innovation (Clift & Brady, 2005; Ronfeldt et al., 2013; Sewall, 2009). Altogether, supervisors can play an integral role in teacher development, but the complexity of their role can make it difficult for them to balance their responsibilities (Burns & Badiali, 2016).

Third, given the myriad of responsibilities, supervisors often face internal conflicts in what they can and should do, restricting their effect on PST development. Supervisors deal with a duality between assisting and assessing (Slick, 1997), struggling to maintain relationships with within field experiences, yet hold expectations of the program. This was reiterated by Valencia et al. (2009), who found that supervisors feel constrained in what they can offer as feedback given their outsider status, commitment to preserving harmony, or deference to the cooperating teacher. A critique of the structures of field supervision is that supervisors do not spend enough

time with either the teacher education program, mentor teacher, or PST to build the necessary trust for a strong feedback process and this puts into question to the extent to which they can offer authentic and rigorous analysis of teaching (Richardson-Koehler, 1988; Sandholtz & Shea, 2012). This coincides with disparate visions that PSTs have about the purpose, modes of development, and perceptions of quality instruction (Bartanen & Kwok, 2021; Valencia et al., 2009). Even with thoughtfully supportive supervisors, PSTs struggled to reconcile differences in vision between programs and cooperating teachers (Anderson & Stillman, 2013; Matsko et al., 2023). Overall, this contextual dynamic establishes that supervisors need structural support to improve their practice of giving feedback (Levine, 2011; Valencia et al., 2009).

The role of supervision reflects larger calls for strengthening clinical practice within teacher education, particularly from large-scale, mixed methods data (Goldhaber, 2019; Sleeter, 2014). Despite its importance, models of clinical practice have overall been deemed unsystematic or unintentional in their design (Cochran-Smith & Zeichner, 2005; Dennis et al., 2017). Therefore, understanding supervisors' feedback—and its associations with PST quality—can contribute to work that advocates for greater support systems for supervisors who share a vital role in teacher preparation (Cuenca et al., 2011).

Frameworks for Teacher Feedback

The overarching goal of providing PSTs with feedback is to reduce the gap between their current teaching performance and their ideal teaching (Hattie & Timperley, 2007; Scheeler, 2008; Scheeler et al., 2004). Studies in this area generally fall within two domains: how feedback can be received, and the nature of feedback. Several frameworks exist that focus on the emotional and social aspect of receiving and interpreting feedback (Copland, 2010; Ilgen et al.,

1979; Voerman et al., 2014), which overall highlight PSTs' desire to receive reliable feedback (Glenn, 2006).

A more closely aligned body of research to our study elucidates the nature of feedback. In their meta-analysis, Thurlings et al. (2013) confirmed that effective feedback should be "goal oriented, specific, and neutral" (p. 1). Likewise, Scheeler et al. (2004) emphasized that feedback should be positive, corrective, and immediate to elicit changes in teaching. Offering a framework that is immediate and includes actionable steps toward an identified goal, Hattie and Timperley (2007) explained that effective feedback must answer three major questions asked by a teacher and/or by a student: Where am I going? (What are the goals?); How am I going? (What progress is being made toward the goal?); and Where to next? (What activities need to be undertaken to make better progress?). These questions respectively correspond to notions of feed up, feed back, and feed forward. A follow up study by Ellis and Loughland (2017) found that PSTs tended to receive more "feed back" than the other two types of comments.

Most comparably in sample, Hunter and Springer (2022) examine written feedback throughout new teacher evaluations and find a positive association between improvements in performance and feedback that (a) is aligned with an improvement area, (b) discusses the feedback's evidential basis, (c) sets specific improvement goals, and (d) includes actionable next steps. However, feedback should be distinctly different between beginning in-service and PSTs, as the former is the teacher-of-record whereas the latter is still learning about the profession, which we describe next.

Feedback in Teacher Preparation

Teacher preparation programs rely on observation evaluations—primarily during clinical teaching—to develop instructional practice and serve as a form of assessment (American

Association of Colleges for Teacher Education, 2018; Richmond et al., 2019). These observational evaluations are regarded as a data course that needs to be utilized more throughout teacher preparation to raise PST quality (National Council for Accreditation of Teacher Education, 2010). However, there is little guidance for teacher educators on how to effectively implement and improve clinical teaching observations (Boguslav & Cohen, 2023; Caughlan & Jiang, 2014; Dennis et al., 2017; van de Grift et al., 2014), even though receiving feedback from a supervisor could be equally as important as from a cooperating teacher (Anderson & Radencich, 2001).

Whereas many of the aforementioned studies on feedback inform *how* individuals can provide it, there is less evidence about *what* that feedback should be about. Various studies discuss the content and curriculum of teacher preparation (Hollins & Warner, 2021; Wilson et al., 2001), but few argue the timing of content (Feiman-Nemser, 2001). Rather, evidence on novice teacher skill development indicates areas and order of importance for pedagogy (Headden, 2014; Kraft et al., 2020; Ost, 2014; Watzke, 2007). For instance, Malmberg et al. (2010) examined the change in novice teacher classroom quality over time through observational feedback on the pedagogical areas of emotional support, classroom organization, instructional support, and student engagement. The authors found that only classroom organization linearly improved over time whereas the other skills eventually plateaued in observational rating.

Most recently, Bartanen et al. (2023) investigated administrator evaluations of novice teachers and found that classroom management and presenting content were the two most important novice skills, with the former linked to high rates of attrition. The authors recommend that these two skills are fundamental for novices to master first before implementing more complex skills such as questioning. These findings suggest a framework for learning pedagogical

skills for in-service novice teachers, but not necessarily for clinical teachers. Instead, more needs to be understood about how PSTs learn pedagogy to better support them in their development prior to entering the profession. Our study addresses this gap by exploring supervisor feedback to improve clinical teacher development. Given that supervisors provide a unique external perspective and that feedback is crucial for professional development, this intersection is invaluable for informing teacher preparation policy.

Method

Context

We draw our sample from one of the largest Texas undergraduate TEPs, with PSTs earning credentials in early childhood-6th grade or middle grades math/science or English language arts/social studies. This three-semester program concludes with a semester-long clinical teaching experience, where PSTs often get placed in their preferred district. These districts then place PSTs with a campus and cooperating teacher according to their own individual procedures.

Clinical teaching placements are spread throughout the entire state, and the university hires and assigns supervisors according to respective geographic regions. Under state law, supervisors must hold current teacher certification in the same area as the PST's classroom, have at least three years of teaching experience, may not be employed at the clinical teaching placement school, and have completed training for the state-approved teacher observation instrument. Customarily, supervisors are retired educators remaining involved in the profession and are part-time contractors of the university hired solely for this position.

Throughout clinical teaching, PSTs receive four 45-minute formal observations from one field supervisor approximately every three weeks. These observations are designed for developmental purposes to provide PSTs with formative feedback about their teaching practice.

Supervisors contact PSTs prior to receiving their lesson plan or any other pertinent information, and then again afterwards for any additional notes. Twice a semester, supervisors facilitate meetings with the PST and cooperating teacher to collectively discuss PST development, but there are no required interactions otherwise.

Data

We utilize PST clinical teaching observational evaluations from 2017-2019. The TEP observation rubric to evaluate PSTs coincides with the state's implementation of the Texas Teacher Evaluation and Support System (T-TESS) in the 2016-2017 school year. Our focus is on a mandatory comment box and rating at the beginning of the instrument. Field supervisors complete an overall rating (i.e., Exceeds expectations, Proficient, Growth in progress, and Needs significant improvement) and provide an open feedback response within the prompt: "Overall Comments and Recommendations." There were 3349 total observations, with only nine observations that had no ratings for a nearly 100% response rate.

The remainder of the evaluation includes Likert-scale items separated by the pedagogical areas of planning, instruction, and learning environment, though these items are not required for submission. There are comment boxes available for each of these subsections that are similarly optional. Furthermore, although these domains mimic the state T-TESS rubric, the individual items differ. For each of the four observations, PSTs can receive ratings for each item on a 1 "Improvement Needed" to 4 "Accomplished" scale or NA "Not applicable/observed." In cases where an NA was given, the numeric score is missing. Sample descriptive statistics of the data are shown in Table 1. Collectively, the observation data represent 3349 separate evaluations across 967 PSTs within 85 districts and 406 campuses. Most PSTs are at the elementary level, though, the sample is relatively equal across subject areas at the secondary levels.

[Insert Table 1 here]

Data Analysis

We take a sequential mixed methods approach (Ivankova et al., 2006; Tashakkori & Teddlie, 2021) to address our research questions. Due to the open-ended nature of the observation form, we seek to first analyze the data inductively. That is, the supervisors are not instructed to provide written feedback based on any a priori rubric, so we decided to inductively identify patterns for areas of PST development. These identified codes then inform our quantitative analyses. We detail each process below.

Qualitative analysis. The large-scale nature of the data and robustness of responses required multiple steps for analysis. We take an emergent approach because no prior framework appropriately fit our data. There are aspects of Hunter and Springer (2022) and Hattie and Timperley's (2007) frameworks that peripherally aligned with some of our data, but neither of these frameworks are situated in a clinical teaching, field supervisor, or PST context.

First, we identified a unit of analysis throughout the data. In our initial reviews of supervisor feedback, we recognized that feedback responses were generally composed of three distinct parts: a retelling of what was observed, an offering praise for actions observed, and corrective feedback to improve PST pedagogy. We identified retelling portions as generally neutral observations of what was happening in the lesson, and as such, could subsume a large proportion of the overall feedback. However, these retellings did not provide much actionable insight for the PSTs and thus, we exclude retelling from our analysis. The second component, praise, offered more content about PST development but generally did not contain additional information beyond lauding a singular action, so we also excluded statements of praise from our analyses. Instead, corrective feedback was most generative in description and captured specific

areas for PST growth. This decision to limit our sample to corrective feedback is affirmed by previous meta-analyses that indicate lower effect sizes for praise versus critique of a specific task (Hattie, 1999; Kluger & DeNisi, 1996). We acknowledge that there is value in examining the entirety of supervisor responses, but we find that focusing on corrective feedback manageably reduces the data within response and offers the biggest contribution to understanding areas of PST development during clinical teaching.

The next analytical task was to specify the parameters of our unit of analysis. A unit includes distinct, corrective statements about attributes of the PST, aspects of pedagogy, or actions related to instruction, with units spanning from within sentence to multiple sentences. The number of units per response ranged between 1 to 11, with an average of 4 units. Once units were identified, we conducted open coding (Miles et al., 2018) of a random sample of 10% of the overall data (i.e., 300 responses). We routinely met to discuss our developing findings and consolidate the identified patterns by writing summary statements, informed by our analytic memos written during the open coding process (Charmaz, 2006). These summary statements became our initial codes. Throughout, we engage in constant comparison (Glaser & Strauss, 2017) by iterating between the raw data and our evolving interpretations to ensure our codes accurately represent the patterns within the supervisor responses.

Once codes were identified, we classified the codes into two categories: *instructional development* and *monitoring student behavior*. To ensure the coding scheme represented the sampled data, we tested this scheme on 25% of the overall data (i.e., 750 responses), which denotes theoretical saturation of the data (Trotter, 2012). Because our purpose was to identify distinct areas of PST development within supervisor feedback, we decided that units of analysis would only be single-coded. Therefore, we continued to meet to discuss any discrepancies of

coding particular units and adjusted the coding scheme accordingly. Once the coding scheme was finalized,[1] we applied it to all (100%) of the data. In this final step, we did not find any need to adjust the scheme further, as it captures the entirety of our data. The final coding scheme is shown in Table 2.

[Insert Table 2 here]

Throughout our process of analysis, we conducted several tests of reliability. In each stage of sampling the data (i.e., at sampling 10% and 25% of the overall sample set), we compiled a set of responses across years (2017-2019), supervisors (135 separate supervisors), observations (1-4), and PSTs (967) to represent the variability of the data. We conducted interrater reliability tests four separate times to: 1) identify cursory open codes, 2) establish an idea unit, 3) establish codes, and 4) finalize the coding scheme. We repeated tests until achieving a Cronbach's alpha of 0.85 or better across multiple raters. All coding and reliability tests were performed in Dedoose.

Quantitative analysis. To understand how supervisor feedback changes throughout clinical teaching, we quantify the qualitative feedback codes as binary variables representing the presence (or not) of the content for each response. We merged these data with the original observation evaluation data and then descriptively analyzed the presence of codes as well as the variation across PST characteristics (subject area and grade level) and development (time of observation and observation rating). All quantitative analyses are conducted using STATA 18.

[1] There were several rounds of axial coding to consolidate our codes. We took a conceptual approach where we collapsed similar codes. *Lesson delivery* included circulation. Lesson cycle included lesson planning. Monitoring and authority were consolidated into *non-verbal techniques*. We also utilized code frequencies (quantitative) to ultimately identify whether a code remained; *use of technology* was so infrequent (n=16) instances that we combined it with *lesson delivery*.

We first conducted Chi-Square tests of independence on contingency tables of the presence of qualitative codes to determine whether these qualitative variables are related to PST observation evaluation ratings and their characteristics. Contingency tables tested whether the proportions of present qualitative codes by PST characteristics differs from those without the qualitative codes. Statistically significant differences were identified in the proportions across the codes by PST characteristic. Chi-square tests on contingency tables of receiving combinations of categories of feedback versus others or no feedback reveal statistically significant differences of those frequencies across categorical PSTs characteristics of subject area, grade level, and observation number. That is, the proportion of observations that are assigned a feedback code is compared to proportions of observations receiving no feedback and is significantly different across all PST characteristics of interest. This is our first test of feedback codes relating to contextual characteristics of the PST's observation placement.

To answer our second research question, we next chose to test whether feedback would be predictive of the PST's next observation evaluation. Because of the repeated nature of our data, we control the observation times and treat all data equal to establish that a relationship exists between feedback codes and overall rating. The linear regression model explained any relationship between supervisor feedback and the PST's overall rating. The following model was estimated via restricted maximum likelihood:

$$Y_{sgrt} = b_0 + b_k X + e_{sgrt}$$

where Y is a standardized evaluation score, and X represents our feedback codes as predictors and placement and development characteristics as control variables. Indices s, g, r, and t represent subject area, grade level, supervisor, and observation time, respectively. To check for the validity of feedback codes, we calculate the variance inflation factor to determine the

tolerance for each predictor and find no concern for multicollinearity; qualitative feedback variables are not highly correlated with each other (Cohen et al., 2003).

We next account for contextual factors of the PST placement affecting feedback on average overall rating. We control for supervisor and the observation number, focusing on the nested nature our data, or the non-independence between observations conducted by a same supervisor. Using a hierarchical linear model, we next estimate the explained variance and changes in standardized influence of supervisor feedback to the PST overall rating, considering the non-independence within supervisors. Group variation at the clustering by supervisors and their observations is calculated via intraclass correlation coefficients (ICC) at 0.297, just higher than average compared to teacher and school effects (e.g., Cohen et al., 2003).

Finally, we return results from a repeated measures analysis of variance (RM-ANOVA) to examine the effect of each feedback subcode on a PST's observational evaluation overall rating between observations. However, we note the lack of independent observations in that supervisors evaluate various PSTs and assign feedback to multiple PSTs. The within-subject factor was of most interest in this final component of our overall research design—we wanted to measure the effect a feedback code had on a PST's next observation evaluation rating. Tests are conducted using the Huynh-Feldt adjusted *p*-value because the sphericity assumption, or the equal variances across observations, was not met as expected. Additionally, we conduct simple effects analysis to test differences among observation numbers for each feedback code. Marginal effects of feedback codes on overall rating by observation number are reported for significant codes.

Results

RQ1: In what pedagogical areas do clinical teaching field supervisors provide feedback?

We first identify the types of feedback throughout supervisor written responses. We identify patterns of supervisor feedback into two categories: *instructional development* and *monitoring student behavior*, represented in Figure 1. Below, we document each code and provide exemplar quotes across supervisors.

[Insert Figure 1 here]

Instructional Development

The first category of codes relates to academic pedagogical practices, which we refer to as *instructional development*. Supervisors provide corrective feedback pertaining to the lesson cycle, lesson connections, lesson delivery, and supporting student comprehension.

Lesson cycle. The structural elements of a lesson, including stating an objective, transitions, and closing the lesson, is defined as the *lesson cycle*. To begin a class, supervisors recommended that PSTs needed "to state the objective specifically" (8944[2]) and "appropriately" (9051). Supervisors comment on the timing of activities, offering suggestions for improving transitions between activities (e.g., "work on linking all parts of the lesson so there is a seamless connection" (9081)) or ensuring students did not have too much unstructured time ("Way too much dead time (25 minutes) at the end of class" (1015)). Supervisors make suggestions about time management, namely figuring out how to keep students engaged if finished with the work before their peers, or overall if "the lesson was very short; extensions needed" (3216). Nearing the end of the lesson, supervisors offer to close the lesson aligned to the goals of the lesson, with statements of, "closure needed to be clearer; make sure it matches the given objective at the first of the lesson" (1548) and "continue working on closure so that it becomes as much of a routine in your organization as your excellent opening" (9565).

[2] Numbering is according to the data and not in any particular order. Five-digit observation IDs were provided and we simply reduced them to four for the purposes of reducing space.

Lesson connections. Another area of feedback is the content of lessons, particularly in connecting the information to ideas that are relevant to students or their prior learning. Supervisors underscore the importance of purpose setting and relevance: "students need to understand the purpose for the lesson and how it applies to their daily lives" (9074) by "connecting to world experiences" (9749) and "using more examples coming from student's life rather than your own life" (9374). Relatedly, PSTs need to be aware of students' prior learning to enhance lessons. PSTs should "prepare students for the new learning by quickly reviewing previous concepts taught and having students demonstrate they know and understand" (9074). Explicitly recommended by one supervisor: "I suggest you emphasize what they already know and relate the learning to future learning" (9076). Throughout, PSTs should consider the content relative to their students' past and developing understandings.

Lesson delivery. Feedback pertained to aspects of lesson delivery, including clarity of speech and intentionality of movement. Movement centered on circulation, or how PSTs traverse the room to monitor student progress, with stated recommendations of "increased monitoring needs to happen" (9382) and "walk around and through the student desks" (9478). But it also includes intentional attention: "Be sure you actively walk around and make sure everyone has their correct paper out on their desk" (9769), and "monitoring is present, but look at the students' work and comment on it as you monitor" (9809). Another strategy that supervisors suggest is ridding word fillers, stating the "need to eliminate saying, 'um' and 'okay' for pauses in instruction" (8944) or noting that PSTs should "be careful of your use of 'OK'; I recorded 10, but I know I missed some" (2551), in an effort to make the lesson less verbally distracting for students.

Student comprehension. Supervisors have PSTs focus on student comprehension and understanding through targeted classroom questioning. PSTs are recommended that "questioning needs to be individualized more. How do you know that every student is ready to go to the next part of the lesson if you don't question them?" (8946). One supervisor reiterates: "You need to ask most of your questions to individuals. You had good questioning but you need to get every student involved in the questioning…This will allow you to show more student success" (8944). Consequently, supervisors provide suggestions such as, "Lots of questioning but it would help if you spread it out more. Ask more non-volunteers" (1536), or "My only comment was there needs to be a more structured way of dealing with responses to questions (pulling sticks, turn to your partner and then give a thumbs up when you have an answer) rather than raised hands" (1546). Through calculated questioning, PSTs could elicit responses from students to recognize their level of understanding.

Monitoring Student Behavior

The other category of supervisor feedback is *monitoring student behavior*, which often entails ways of preventing or dealing with misbehavior or suggestions to encourage model behavior in the classroom. Five separate codes addressed different aspects of dealing with student behavior.

Praise. Supervisors encourage PSTs to praise their students by saying they "need to praise more consistently" (8946), "praise student for appropriate behavior and effort" (5821), and "use positive reinforcement to encourage appropriate behavior" (1566). Integrating praise allows PSTs to highlight students who excel in the classroom or comply with instruction. As best stated by one supervisor, "Suggestion: Notice out loud what students are doing when they are doing

what you asked instead of noticing students that do not comply" (8957). In such a way, praise highlights model behavior without disparaging students who have yet to exhibit it.

Transitions. Another technique that supervisors recommend is the management of transitions, a key time when students could get off task during changes in the lesson or activities. Supervisors advise that PSTs explicitly communicate the expectations of how students should move around in the classroom, such as "prior to leaving the carpet and after telling the students they will work in pairs, have a brief discussion how we work in pairs." (9076). As explained by another supervisor: "So, the same way you transition students from PE, set students down in chairs afterwards, lower you voice, set expectations, then transition into lesson" (9270). The importance of managing transitions is to "take less time, with the students learning a more proficient way to obtain their materials and getting to the carpet for instruction" (2072), reducing time wasted and having students "get back on track quicker" (3290) after a change in activity.

Attention. Another area that supervisors advise for reducing misbehavior is by maintaining student attention. Supervisors provide similar statements of: "Suggestion: Practice procedure to get and maintain student attention" (8956); "Be sure ALL students are paying attention before you begin a lesson or give instructions" (9081); and "Remember if they stop being attentive or start talking, just stop and use wait time for compliance" (9076). Such statements reinforce that without student attention, PSTs' efforts for instruction would be futile.

Non-verbal techniques. In dealing with misbehavior, supervisors note several non-verbal techniques. They suggest for PSTs to practice visual awareness, or using their eyes to identify off-task behavior, with statements such as, "She still needs to work on her classroom awareness which will develop with time" (9426) and "you need to keep your eyes up and moving in order to see any unacceptable behavior such as two or three students sprinted from one side of

the room to the other" (9896). Another technique revolves around teacher presence, including exhibiting authority through a firm, consistent, and confident demeanor: "Continue to focus on presenting a confident, assertive demeanor in the classroom" (6154). As exemplified by one supervisor, "Continue to work on presenting a confident, assertive demeanor so that students will see you as 'the teacher.' Establishing a strong 'teacher presence' is your priority. It leads to more effective classroom management" (9105).

Verbal techniques. Supervisors also gave feedback on an array of verbal classroom management techniques. One includes emphasizing a teacher voice, which was defined as having a tone that stressed more confidence. As best described by one supervisor, "Assert yourself with confidence, you clearly know the content and let the students feel your excitement that you want to share the content with them. Work on a strong, (deeper) confident voice, and inflection and emphasis to your delivery" (9480). Others discuss it broadly, as supervisors recommend that PSTs should be "working on her teacher voice and also classroom awareness" (9088) and "classroom management was appropriate and very good for this level of students but remember for upper levels you will need to firmer" (9051). In situations when students are off-task or display inappropriate behavior, supervisors remind PSTs to redirect students: "make sure you correct students who are talking during the prior learning exercise" (9071), particularly "correcting students individually, by name, [which] is more effective than punishing the whole class" (8945). Overall, supervisors favor when PSTs provide precise corrections for students to stay on task.

RQ2: How do supervisor feedback codes vary by context?

We leverage our application of the scheme to the entire data by next identifying the variation in supervisor feedback. Figure 2 illustrates the presence of codes throughout the data,

indicating instructional features that are most prevalent across PSTs. Feedback around student comprehension and lesson delivery are the most common codes—19% and 20% respectively—with praise and maintaining student attention as the most frequent behavioral codes.

[Insert Figure 2]

Then, we examine how feedback codes vary by context. Most relevant, we investigate the presence of codes across observation number to see whether feedback changes throughout clinical teaching; Appendix 4 illustrates the presence of codes across remaining characteristics. Figure 3 indicates that lesson delivery is the most present throughout PSTs at 22.84%, with the next most present at least 6% less—student comprehension (16.34%) and maintaining student praise (14.83%)—for PSTs that scored an overall rating of 2-*growth in progress*. This is different than student comprehension and lesson delivery, which tied for most frequent across responses (22.02%) for PSTs being rated as 3-*exceeds expectations*, with the next most prevalent codes being 9% less for praise (13.1%) and maintaining student attention (11.9%). Chi-square tests for each code, shown in Appendix 1, indicate statistically different correlations with overall ratings.[3]

[Insert Figure 3]

RQ3: What is the relationship between clinical teaching evaluation rating and supervisor feedback codes?

Given these statistically significant differences in codes across PST characteristics, we then examine whether feedback could predict for observation evaluation scores.[4] We model overall rating by codes in Table 3, with Model 1 indicating a base linear regression and Models 2

[3] We examine whether similar patterns exist by overall rating, grade, and subject level, shown in Appendix 2, but do not find strong conceptual or further statistical evidence warranting additional analyses in these areas.
[4] Of the entire sample (N=3,349), 63% of observations contained no critiquing feedback of either kind. Those who received feedback critiquing their clinical teaching experience were likely to have lower scores compared to PSTs who received zero constructive feedback. Therefore, it was determined a regression analysis of our predictors of interest on the PST's observation evaluation overall rating would serve to next identify significant feedback codes.

and 3 incorporating placement and observation time variables, respectively.[5] Model 3 showcases that the change in the relationship between feedback code and the dependent variable is minor yet significant when the model controls for supervisor and observation number in addition to subject area and grade level.

[Insert Table 3]

Model 3 displays the changes with these controls. In statistically significant variables, we see a decrease in the coefficients and thus a smaller negative affect on ratings by supervisor feedback codes. This linear regression model results in an adjusted R^2=0.308, meaning over 30% of the variance is explained when controlling for the supervisor and observation time. Comparatively, we examined whether the supervisor and observation number would have an effect on clinical teachers' overall rating of their observation as level 2 predictors in Model 4.

Contextual factors around the clinical teaching experience are accounted for as level 1 predictors, or at the PST-level, in Model 5 including subject area and grade level. We note little to no differences between the nested models in that all feedback codes remain negative and statistically significant. However, changes to overall rating are minimal. *Non-verbal techniques* and *verbal corrections* continue to return greatest amounts of change on overall rating; when these codes are present in a clinical teacher's feedback, their overall rating decreases on average by more than 0.15 units (γ_4 = -0.162, SE = 0.046, $p < 0.001$; γ_5 = -0.156, SE = 0.046, $p < 0.001$). The expected average overall rating stays practically the same at 2.58 with a notably different but similar standard deviation of 0.097. When supervisors provide feedback on whether PSTs should

[5] Appendix 3 indicates estimates of grade level and subject area by feedback code. Grade level placement (K-9, and 12) and subject classes are dummy coded using pre-K and self-contained placements as the reference groups. Linear regression analysis indicate that grade levels are statistically significant predictors of the overall rating score of a clinical teacher's observation (save for 9th grade). Differences in effect on overall rating between pre-K placement compared to grades K-8 and the 12th grade are positive and increasing. Similarly, results indicate that main subjects are statistically significant predictors of overall rating compared to self-contained classroom placements.

maintain student attention, use non-verbal techniques, or their lesson delivery, they are likely to attribute fractionally lower overall ratings as well.

RQ4: How does feedback type affect evaluation rating across observation number?

The hierarchical linear model considers the nested nature of our data—PST repeated observations grouped by their supervisors. Conceptually, we consider the supervisor as a higher-level predictor of a clinical teacher's overall rating seeing as the single supervisor is who conducts repeated observations. Therefore, we organize our multilevel models in the following ways. First, the unconditional model includes PST observation evaluation ratings as the dependent variable and no level 1 predictors in the model and supervisor and observation number as level 2 parameters. The constant starts at 2.47 (SE = 0.074) and is statistically significant ($p <$ 0.001). The random-effects parameters for the unconditional model included supervisor and the observation time with a within observations variance of 0.097 and a variance between supervisors of 0.305. We add indicators in the final two columns of Table 3.

Model 4 in Table 3 introduces the feedback codes that clinical teachers receive from their supervisor as part of their observation evaluations to a hierarchical linear model that partitions PSTs nested by supervisor within observation numbers. The average overall rating with all predictors held constant at zero, meaning no codes were marked present, from the Model 1 that included no control or clustering for contextual factors or development starts at 3.08 (SD = 0.013) and we see a decrease by a more than half a unit (2.55; SD = 0.074) when Model 4 accounts for the non-independence of observations without control characteristics. All feedback codes are also significantly negative effects on PST overall rating. Little change in effect is shown between the nested models with or without controlling for grade level and subject area.

Results from the RM-ANOVA were significant at various confidence levels for five out of the nine feedback codes all shown in Appendix 5. Results for main effect and interaction effects were calculated using the Huynh-Feldt correction. Margins plots for each feedback category of codes are shown in Appendix 6 and Figure 4 shows the change in linearly predicted overall rating for receiving each individual feedback code across observations.[6] The main effect for *Maintaining student attention* with observation number were significant ($F = 6.34$, $p = 0.01$) and its interaction effect using the Huynh-Feldt correction ($F = 3.07$, $p = 0.036$). Similarly, *Verbal corrections* main effect and interaction effect with time were significant ($F = 7.75$, $p = 0.006$; $F = 7.42$, $p < 0.001$). Further, *Non-verbal techniques* had a significant main effect ($F = 8.30$, $p = 0.004$) but an insignificant interaction effect with time ($F = 0.78$, $p = 0.486$). The instructional feedback codes *Lesson cycle* and *Lesson delivery* both returned significant main effects ($F = 5.05$, $p = 0.025$; $F = 15.1$, $p < 0.001$) and a significant interaction effect between *Lesson cycle* and observation ($F = 8.98$, $p < 0.001$).

[Insert Figure 4]

Discussion

This sequential mixed method study investigates clinical teaching field supervisor feedback. We qualitatively code three years' worth of responses, focusing on the pedagogical areas these supervisors detail. Then, we examine patterns in feedback codes by context to ultimately identify relationships between codes and PST observation evaluation score. Finally,

[6] Margins plots for each significant feedback code are shown in Figures 1a and 1b comparing linear predictions of observation evaluation ratings for PSTs who received each feedback code versus those who did not. Whereas Figure 1c presents margins plots for the remaining non-statistically significant feedback codes compared to a single graph of linear predictions for those who did not receive the feedback score because all graphs for the non-statistically significant feedback codes were alike.

we examine how feedback changes throughout clinical teaching. Through these analyses, we approximate PST development throughout clinical teaching, offering several important findings.

Supervisors focus heavily on instructional and behavioral components of pedagogy. This provides a glimpse into the range of content that supervisors focus on and offers evidence of PST developmental needs, reiterating evidence about the early importance of lesson planning (Jacobs et al., 2017) and classroom management (Kwok, 2020). However, our results contribute wider, systematic substantiation and in specific areas that PSTs should centralize their formative learning to succeed within the classroom. That is, the supervisor feedback names pedagogical skills—such as lesson delivery, praise, and non-verbal techniques—to specify actions that should be explicitly trained throughout preparation.

There is foundational clinical teaching pedagogy. Feedback differs by context, specifically observation number, evaluation rating, subject area, and grade level. The type of feedback received is also consequential, as it is associated with lower observation scores. Three feedback codes (maintaining student attention, using non-verbal techniques, and lesson delivery) were consistently significant across regression models, suggesting that these PST actions offer particular importance. This is vital in pinpointing actions that are most crucial—in the eyes of supervisors—in evaluating PST success. In essence, supervisors seem to deem certain skills as foundational for success in early clinical field observations. These results in turn suggest foundational pedagogy for PST development in terms of what initial teacher learning should comprise. PSTs need to master these central pieces of student engagement before learning other aspects of pedagogy, otherwise, observational evaluations could hinder their success moving forward (Bartanen & Kwok, 2021).

Further, foundational pedagogy could be ordered for a clinical teaching learning trajectory. Variation in feedback by observation and RM-ANOVA results posit that feedback on skills systematically changes throughout PST development. Namely, supervisors tend to focus on maintaining student attention and non-verbal techniques during early clinical teaching, versus later improvement on lesson delivery, lesson cycle, and behavioral corrections. These results align with Bartanen et al. (2023), where classroom management and presenting content were fundamental to novice teacher evaluations and even eventual retention. Importantly, we identify a similar priority in pedagogical skills for PSTs and even unearth nuance to these skills, such as classroom management being into five respective codes.

Limitations

This study should be held within the context of its design. First, supervisor feedback was within one teacher education program and one certification area. Future research could examine whether findings are replicable across other contexts. We believe the breadth and depth of data draw to the strength of the study, but we cannot rule out programmatic nuances that guide supervisor feedback.

Second, we do not know the root of the supervisor feedback. That is, we cannot dissociate between what supervisors focus on versus how the PSTs act in the classroom. This could be examined in future studies through observations or video to identify what PSTs are enacting compared to what supervisors write. Similarly, we do not account for whether PSTs utilize supervisor feedback. Studies could observe PST actions after supervisor visits or examine PST responses to supervisor feedback.

Third, we do not know if there were any additional interactions between the supervisor and the PST. Informal, undocumented interactions could have mediated the trajectory of PST

development. Interviews and observations of field supervisor visits could provide necessary programmatic context.

Implications

Despite some empirical restrictions, our findings offer implications for practice, policy, and research. Most relevantly, implications pertain to practice and having teacher preparation programs consider aspects of curricular redesign. Particularly throughout the clinical teaching experiences, programs need to structure PST development around these areas of supervisory focus. Programs should embed supports or coursework around each area of feedback, and allow for extended learning on maintaining student attention, using non-verbal techniques, and lesson delivery. Further, they should train PSTs on how to maintain student attention and use non-verbal techniques early in the program so that they can successfully implement it immediately in clinical teaching. Programs should then transition PST learning to be about lesson delivery, lesson cycle, and behavioral corrections after PSTs have shown mastery with the previous techniques. This would allow for a PST learning progression that affords gradual development of skills.

For policy, findings suggest minimum standards for teacher preparation requirements and a potential framework for clinical teaching experiences. Supervisor feedback emphasizes areas of growth that policymakers may want to consider as foundational to PST development. There could be variation in how programs train on these topics, but they should ensure that programs explicitly address these important pedagogical skills throughout the program, specifically within or reiterated in clinical training.

For research, understanding *how* clinical teaching remains the most influential structure remains pertinent. While there is a significant amount of evidence for cooperating teachers

(Ronfeldt, 2021), more is needed about other structures within this experience. Large-scale and systematic data is vital to unveil PST development (Goldhaber, 2019; Sleeter, 2014), and our work in illuminating supervisor feedback can contribute to preparing a better teaching workforce.

References

American Association of Colleges for Teacher Education. (2018). A pivot toward clinical practice, its lexicon, and the renewal of educator preparation. *A Report of the AACTE Clinical Practice Commission*.

Anderson, L. M., & Stillman, J. A. (2013). Student teaching's contribution to preservice teacher development: A review of research focused on the preparation of teachers for urban and high-needs contexts. *Review of Educational Research*, *83*(1), 3–69. https://doi.org/10.3102/0034654312468619

Anderson, N. A., & Radencich, M. C. (2001). The value of feedback in an early field experience: Peer, teacher, and supervisor coaching. *Action in Teacher Education*, *23*(3), 66–74. https://doi.org/10.1080/01626620.2001.10463076

Bartanen, B., Bell, C., James, J., Taylor, E., & Wyckoff, J. (2023). *"Refining" Our Understanding of Early Career Teacher Skill Development: Evidence From Classroom Observations*. https://doi.org/10.26300/WYKJ-E663

Bartanen, B., & Kwok, A. (2021). Examining clinical teaching observation scores as a measure of preservice teacher quality. *American Educational Research Journal*, *58*(5), 887–920. https://doi.org/10.3102/0002831221990359

Boguslav, A., & Cohen, J. (2023). Different Methods for Assessing Preservice Teachers' Instruction: Why Measures Matter. *Journal of Teacher Education*, 00224871231200279. https://doi.org/10.1177/00224871231200279

Burns, R. W., & Badiali, B. (2016). Unearthing the complexities of clinical pedagogy in supervision: Identifying the pedagogical skills of supervisors. *Action in Teacher Education, 38*(2), 156-174.

Burns, R., Jacobs, J., & Yendol-Hoppey, D. (2016). Preservice teacher supervision within field experiences in a decade of reform. *Teacher Education and Practice, 29*(1), 46–75.

Caires, S., Almeida, L. S., & Martins, C. (2010). The socioemotional experiences of student teachers during practicum: A case of reality shock? *The Journal of Educational Research*, 103, 17–27.

Campbell, T., & Lott, K. (2010). Triad dynamics: Investigating social forces, roles, and storylines. *Teaching Education, 21*(4), 349-366.

Caughlan, S., & Jiang, H. (2014). Observation and teacher quality: Critical analysis of observational instruments in preservice teacher performance assessment. *Journal of Teacher Education, 65*(5), 375–388. https://doi.org/10.1177/0022487114541546

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Clift, R. T., & Brady, P. (2005). Research on methods courses and field experiences. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*, (pp. 309-424*)*. Routledge.

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, (3rd ed.). Routledge.

Copland, F. (2010). Causes of tension in post-observation feedback in pre-service teacher training: An alternative view. *Teaching and Teacher Education, 26*(3), 466-472. http://dx.doi.org/10.1016/j.tate.2009.06.001

Cuenca, A., Schmeichel, M., Butler, B. M., Dinkelman, T., & Nichols, J. R., Jr. (2011). Creating a "third space" in student teaching: Implications for the university supervisor's status as outsider. *Teaching & Teacher Education*, 27(7), 1068–1077. doi:10.1016/j.tate.2011.05.003

Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education*, *89*(4), 547–561.

Dennis, D. V., Burns, R. W., Tricarico, K., van Ingen, S., Jacobs, J., & Davis, J. (2017). Problematizing clinical education. The power of clinical preparation in teacher education, 1-20.

Ellis, N. J., & Loughland, T. (2017). 'Where to next?' Examining feedback received by teacher education students. *Issues in Educational Research*, *27*(1), 51-63.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, *103*(6), 1013–1055. https://doi.org/10.1111/0161-4681.00141

Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Glenn, W. J. (2006). Model versus mentor: Defining the necessary qualities of the effective cooperating teacher. *Teacher education quarterly*, *33*(1), 85-95.

Goldhaber, D. (2019). Evidence-Based Teacher Preparation: Policy Context and What We Know. *Journal of Teacher Education*, *70*(2), 90–101. https://doi.org/10.1177/0022487118800712

Goldhaber, D., Ronfeldt, M., Cowan, J., Gratz, T., Bardelli, E., & Truwit, M. (2022). Room for Improvement? Mentor Teachers and the Evolution of Teacher Preservice Clinical Evaluations. *American Educational Research Journal, 59(5), 1011-1048.* https://doi.org/10.3102/00028312211066867

Grossman, P., Hammerness, K. M., McDonald, M., & Ronfeldt, M. (2008). Constructing coherence: Structural predictors of perceptions of coherence in NYC teacher education programs. *Journal of Teacher Education*, *59*(4), 273–287.

Hattie, J. A. (1999, June). Influences on student learning (Inaugural address, University of Auckland, New Zealand). Retrieved from http://www.arts.auckland.ac.nz/staff/index.cfm?P=8650

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Headden, S. (2014). Beginners in the Classroom: What the Changing Demographics of Teaching Mean for Schools, Students, and Society. *Carnegie Foundation for the Advancement of Teaching*.

Hertzog, H. S., & O'Rode, N. (2011). Improving the quality of elementary mathematics student teaching: Using field support materials to develop reflective practice in student teachers. *Teacher Education Quarterly*, 38(3), 89–111.

Hollins, E. R., & Warner, C. K. (2021). *Evaluating the clinical component of teacher preparation programs*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. National Academy of Education.

Hunter, S. B., & Springer, M. G. (2022). Critical feedback characteristics, teacher human capital, and early-career teacher performance: A mixed-methods analysis. *Educational Evaluation and Policy Analysis*, 44(3), 380-403. https://doi.org/10.3102/01623737211062913

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371. https://doi.org/10.1037//0021-9010.64.4.349

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, *18*(1), 3–20. https://doi.org/10.1177/1525822X05282260

Jacobs, J., Hogarty, K., & Burns, R. W. (2017). Elementary preservice teacher field supervision: A survey of teacher education programs. *Action in Teacher Education*, *39*(2), 172–186. https://doi.org/10.1080/01626620.2016.1248300

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Kraft, M. A., Papay, J. P., & Chi, O. L. (2020). Teacher Skill Development: Evidence from Performance Ratings by Principals. *Journal of Policy Analysis and Management*, *39*(2), 315–347. https://doi.org/10.1002/pam.22193

Kwok, A. (2018). Promoting "quality" feedback: First-year teachers' self-reports on their development as classroom managers. *Journal of Classroom Interaction*, *53*(1), 22–36. https://search.ebscohost.com/login.aspx?direct=true&db=ofm&AN=128122439&site=eds-live

Kwok, A., & Bartanen, B. (2022). Examining early field experiences in teacher education. *Teaching and Teacher Education*, *118*, 103820.

Levine, T. H. (2011). Features and strategies of supervisor professional community as a means of improving the supervision of preservice teachers. *Teaching and Teacher Education*, *27*(5), 930–941. https://doi.org/10.1016/j.tate.2011.03.004

Malmberg, L.-E., Hagger, H., Burn, K., Mutton, T., & Colls, H. (2010). Observed classroom quality during teacher education and two years of professional practice. *Journal of Educational Psychology*, *102*(4), 916.

Matsko, K. K., Ko, E. K., Bressman, S., Mozer, L., & Winter, J. S. (2023). Two sides of a coin: Examining mentor and student teacher relationships during student teaching. *Teaching and Teacher Education*, *130*, 104139. https://doi.org/10.1016/j.tate.2023.104139

McDonnough, J. T., & Matkins, J. J. (2010). The role of field experience in elementary preservice teachers' self-efficacy and ability to connect research to practice. *School Science and Mathematics*, *110*(1), 13-23.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage publications.

Mok, S. Y., & Staub, F. C. (2021). Does coaching, mentoring, and supervision matter for pre-service teachers' planning skills and clarity of instruction? A meta-analysis of (quasi-) experimental studies. *Teaching and Teacher Education*, *107*, 103484.

National Council for Accreditation of Teacher Education. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers. Report of the Blue Ribbon Panel on Clinical Preparation and Partnerships for Improved Student Learning. November*.

Nguyen, H. T. (2009). An inquiry-based practicum model: What knowledge, practices, and relationships typify empowering teaching and learning experiences for student teachers, cooperating teachers and college supervisors? *Teaching & Teacher Education*, 25(5), 655–662. doi:10.1016/j.tate.2008.10.001

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, *6*(2), 127–151.

Pomerance, L. & Walsh, K. (2020). *2020 Teacher prep review: Clinical practice and classroom management*. National Council on Teacher Quality. https://www.nctq.org/publications/2020-Teacher-Prep-Review:-Clinical-Practice-and-Classroom-Management

Richardson-Koehler, V. (1988). Barriers to the effective supervision of student teaching: A field study. *Journal of Teacher Education*, *39*(2), 28–34. https://doi.org/10.1177/002248718803900206

Richmond, G., Bartell, T., Carter Andrews, D. J., & Neville, M. L. (2019). Reexamining coherence in teacher education. *Journal of Teacher Education*, *70*(3), 188–191. https://doi.org/10.1177/0022487119838230

Ronfeldt, M. (2021). Links among teacher preparation, retention, and teaching effectiveness. *National Academy of Education*.

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or preparation? Investigating the effects of teacher characteristics and student teaching. *Journal of Teacher Education*, *64*(4), 319–337. https://doi.org/10.1177/0022487113488143

Sandholtz, J. H., & Shea, L. M. (2012). Predicting performance: A comparison of university supervisors' predictions and teacher candidates' scores on a teaching performance assessment. *Journal of Teacher Education*, *63*(1), 39–50. https://doi.org/10.1177/0022487111421175

Scheeler, M. C. (2008). Generalizing effective teaching skills: The missing link in teacher preparation. *Journal of Behavioral Education*, *17*, 145–159.

Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher education and special education*, *27*(4), 396-407.

Sewall, M. (2009). Transforming supervision: Using video elicitation to support preservice teacher-directed reflective conversations. *Issues in Teacher Education*, 18(2), 11–30.

Sleeter, C. (2014). Toward teacher education research that informs policy. *Educational Researcher*, *43*(3), 146-153. https://doi.org/ 10.3102/0013189X14528752

Slick, S. K. (1998). The university supervisor: A disenfranchised outsider. *Teaching and Teacher Education*, *14*(8), 821–834. https://doi.org/10.1016/S0742-051X(98)00028-6

Slick, S. K. (1997). Assessing versus assisting: The supervisor's roles in the complex dynamics of the student teaching triad. *Teaching and Teacher Education*, *13*(7), 713–726. https://doi.org/10.1016/S0742-051X(97)00016-4

Tashakkori, A., & Teddlie, C. (2021). *Sage handbook of mixed methods in social & behavioral research*. SAGE publications.

Thurlings, M., Vermeulen, M., Bastiaens, T. & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, *9*, 1-15. http://dx.doi.org/10.1016/j.edurev.2012.11.004

Trotter, R. T. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine*, *55*(5), 398–400. https://doi.org/10.1016/j.ypmed.2012.07.003

Turunen, T. A., & Tuovila, S. (2012). Mind the gap. Combining theory and practice in a field experience. *Teaching Education*, *23*(2), 115-130.

Valencia, S. W., Martin, S. D., Place, N. A., & Grossman, P. (2009). Complex Interactions in Student Teaching: Lost Opportunities for Learning. *Journal of Teacher Education*, *60*(3), 304–322. https://doi.org/10.1177/0022487109336543

van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: Calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation*, *43*, 150–159. https://doi.org/10.1016/j.stueduc.2014.09.003

Voerman, L., Korthagen, F. A. J., Meijer, P. C., & Simons, R. J. (2014). Feedback revisited: Adding perspectives based on positive psychology: Implications for theory and classroom practice. *Teaching and Teacher Education*, *43*, 91–98. https://doi.org/10.1016/j.tate.2014.06.005

Watzke, J. L. (2007). Longitudinal research on beginning teacher development: Complexity as a challenge to concerns-based stage theory. *Teaching and Teacher Education*, *23*(1), 106–122. https://doi.org/10.1016/j.tate.2006.04.001

Ward, C. J., Nolen, S. B., & Horn, I. S. (2011). Productive friction: How conflict in student teaching creates opportunities for learning at the boundary. *International Journal of Educational Research*, 50(1), 14–20. doi:10.1016/j.ijer.2011.04.004

Wasburn-Moses, L., & Noltemeyer, A. (2018). Effectiveness of Campus Mentors, an alternative school–university partnership. *Preventing School Failure: Alternative Education for Children and Youth*, *62*(3), 190–197.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). Teacher preparation research: Current knowledge, gaps, and recommendations. In *Center for the Study of Teaching and Policy*. https://www.education.uw.edu/ctp/content/teacher-preparation-research-current-knowledge-gaps-and-recommendations

Zeichner, K. (2021). Critical unresolved and understudied issues in clinical teacher education. *Peabody Journal of Education*, *96*(1), 1–7. https://doi.org/10.1080/0161956X.2020.1864241

Table 1

Sample Descriptive Statistics

Variable	N	Mean	SD	Min	Max
School Level	406				
Districts	85				
Observation Level	3349				
Principals	422				
Supervisors	66				
Female Supervisor		0.81			
PK-3rd Grade Placement	1382	0.41			
4th-8th Grade Placement	1694	0.50			
9th-12th Grade Placement	273	0.09			
Self-Contained	1185	0.35			
ELA/ELAR/ELAW	768	0.23			
Math	814	0.24			
Science	717	0.21			
Social Studies	501	0.15			
Individual Student Level	967				
Cooperating Teachers	889				
Self-Reported Overall Rating		2.997	0.67	1	4
Average Surveyed Overall Rating		3.07	.52	1.5	4
Exceeds Expectations	737				
Proficient	1862				
Growth in Progress	746				
Needs Significant Improvement	3				
1st Observation	967				
2nd Observation	824				
3rd Observation	790				
4th Observation	768				

Note: Table shows descriptive statistics for the analytic sample of clinical teachers observed 1-4 times per semester in their clinical teaching semester from 2017-2019.

Table 2

Field Supervisor Feedback Coding Scheme

Code	Descriptions	Exemplar Quotes
	Instructional Development	
Lesson Cycle	Related to the organizational components of the lesson.	"All students need to be involved in closure to let you know whether or not they achieved the objectives." (5327) "States objective as 'we are going to talk about...' Need to state the objective specifically." (8944)
Lesson Connections	Connecting lesson material to prior or future learning, or relevancy to students' real lives.	"Need to connect to world experiences." (9749) "Did you relate today's learning to previous learning?" (5637)
Lesson Delivery	How effectively the PST delivers the planned material, including clarity of speech, pacing, and monitoring.	"Continue to work on reducing your use of 'OK'." (5642) "Delivery of content was rapid which caused the students to struggle with calculator issues." (9870) "Monitoring is present, but look at the students' work and comment on it as you monitor." (9809)
Student Comprehension	Verifying students understand the lesson. May involve questioning or further activities to promote students' learning.	"Check for understanding after each problem." (1086) "While it is difficult, you need to make sure you include higher order thinking skills in the different phases of the groups." (1216)
	Monitoring Student Behavior	
Praise	Addressing on task or desired behaviors verbally.	"Need to praise more." (8947) "Praise individual students, or tables, or the entire class for appropriate behavior and hard work." (9520)
Managing Transitions	Management of student behavior during changes in the lesson.	"I think the brain break before the lesson led to difficulty settling students down for the lesson." (9270)
Maintaining Student Attention	Obtaining and retaining focused listening from all students in the class.	"The only thing I would recommend is to use a variety of "things" to get the student's attention." (9094)
Non-Verbal Techniques	Strategies to reduce student misbehavior, including eye contact and portraying authority throughout the classroom.	"However, you need to keep your eyes up and moving in order to see any unacceptable behavior such as two or three students sprinted from one side of the room to the other." (9896) "Continue to present a calm, confident, assertive demeanor in the classroom. This will help to strengthen your teacher presence. A strong 'teacher presence' leads to more effective classroom management." (6349)
Verbal Corrections	Addressing off-task or inappropriate behavior verbally.	"The voice level rose again, but you took no action; I suggest you warn them if they cannot stay at voice level 1,then we will go to voice level 0." (5326)

Table 3

The Relationship Between Clinical Teaching Observation Scores and Supervisor Feedback Codes (N=3,340)

		Contextual Models		Nested Models	
Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Monitoring Student Behavior Feedback Codes					
Praise	-0.001 (0.049)	-0.004 (0.049)	-0.002 (0.042)	-0.072* (0.034)	-0.076* (0.034)
Managing Transitions	-0.015 (0.074)	-0.016 (0.074)	-0.019 (0.064)	-0.117* (0.048)	-0.115* (0.048)
Maintaining Student Attention	-0.103*** (0.050)	-0.102*** (0.050)	-0.066*** (0.044)	-0.152*** (0.034)	-0.141*** (0.034)
Non-Verbal Techniques	-0.098*** (0.069)	-0.095*** (0.069)	-0.066*** (0.059)	-0.175*** (0.046)	-0.162*** (0.046)
Verbal Corrections	-0.033 (0.061)	-0.031* (0.060)	-0.012 (0.0520)	-0.164*** (0.040)	-0.156*** (0.040)
Instructional Development Feedback Codes					
Lesson Cycle	-0.039* (0.043)	-0.038* (0.043)	-0.005 (0.037)	-0.072* (0.030)	-0.068* (0.030)
Lesson Connections	0.001 (0.062)	0.002 (0.062)	0.009 (0.053)	-0.130** (0.043)	-0.124** (0.043)
Student Comprehension	-0.052*(0.037)	-0.052 (0.038)	-0.020 (0.032)	-0.045 (0.025)	-0.037 (0.025)
Lesson Delivery	-0.117*** (0.035)	-0.116*** (0.035)	-0.079*** (0.034)	-0.105*** (0.023)	-0.095*** (0.023)
Constant (no criticism feedback)	3.08*** (0.013)	2.85*** (0.098)	2.18*** (0.089)	2.55*** (0.074)	2.58*** (0.097)
Adjusted R-squared	0.056	0.065	0.308		
Controls					
Subject Area		X	X		X
Grade Level		X	X		X
Supervisor			X	X	X
Observation Number			X	X	X

Note: Each column presents results from models introducing external factors as control variables showcasing the difference when accounting for the non-independent observations of clinical teaching experiences by a same supervisor. Models 1-3 are estimated via maximum likelihood. Models 4 and 5 are estimated using mixed effects multi-level modeling in which observation times and supervisors are included as a random-effects parameters. The variance within observations between supervisor was 0.090 and the between supervisor variance was 0.305. Using the variance from the unconditional model, the full contextual model explains an additional 5.7% of the variance. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Figure 1

Field Supervisor Feedback Categories and Corresponding Codes

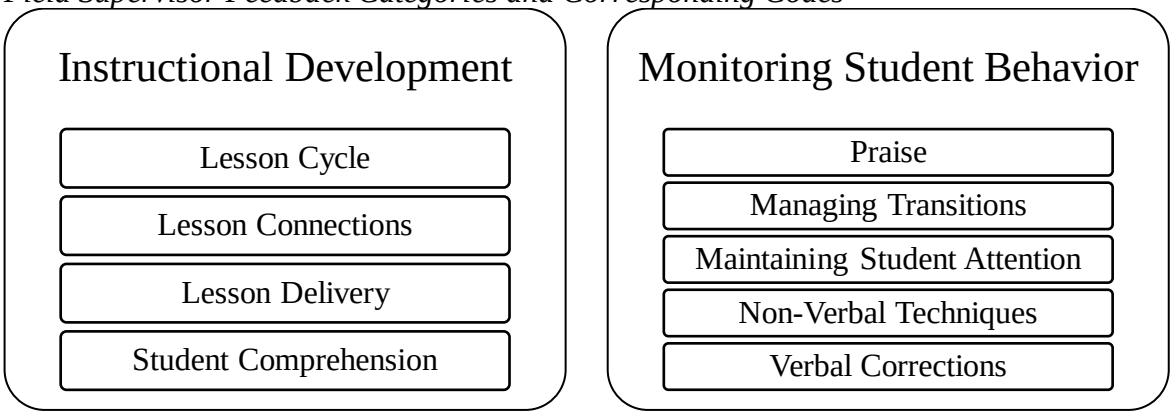


Figure 2

Supervisor Feedback Code Frequency (N=3,349)

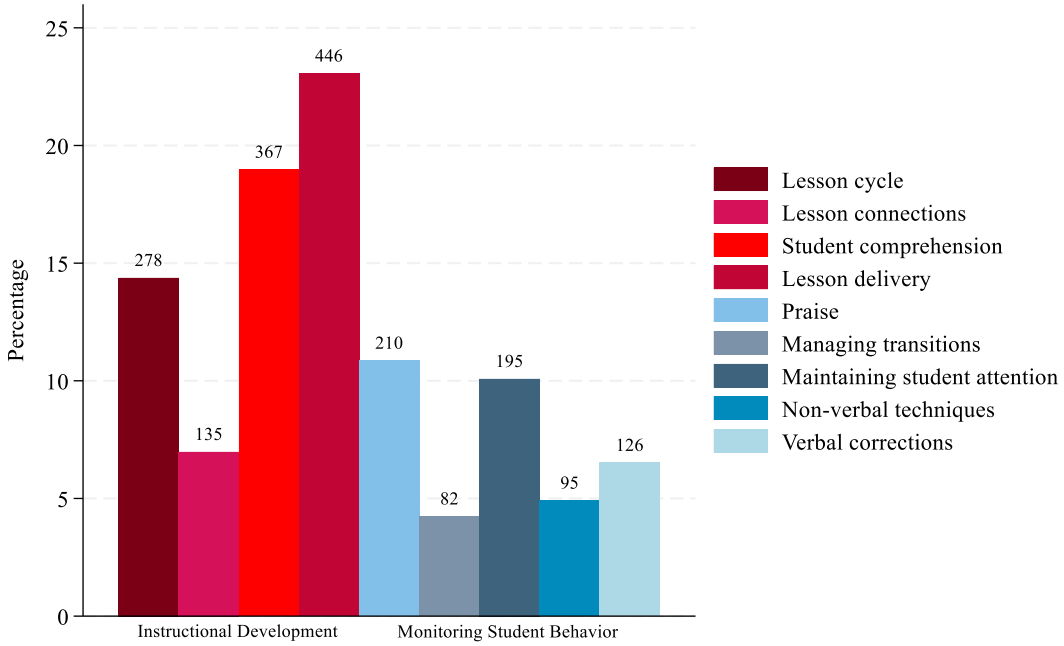


Figure 3

Supervisor Feedback Code Frequency by Observation Evaluation (N=3,349)

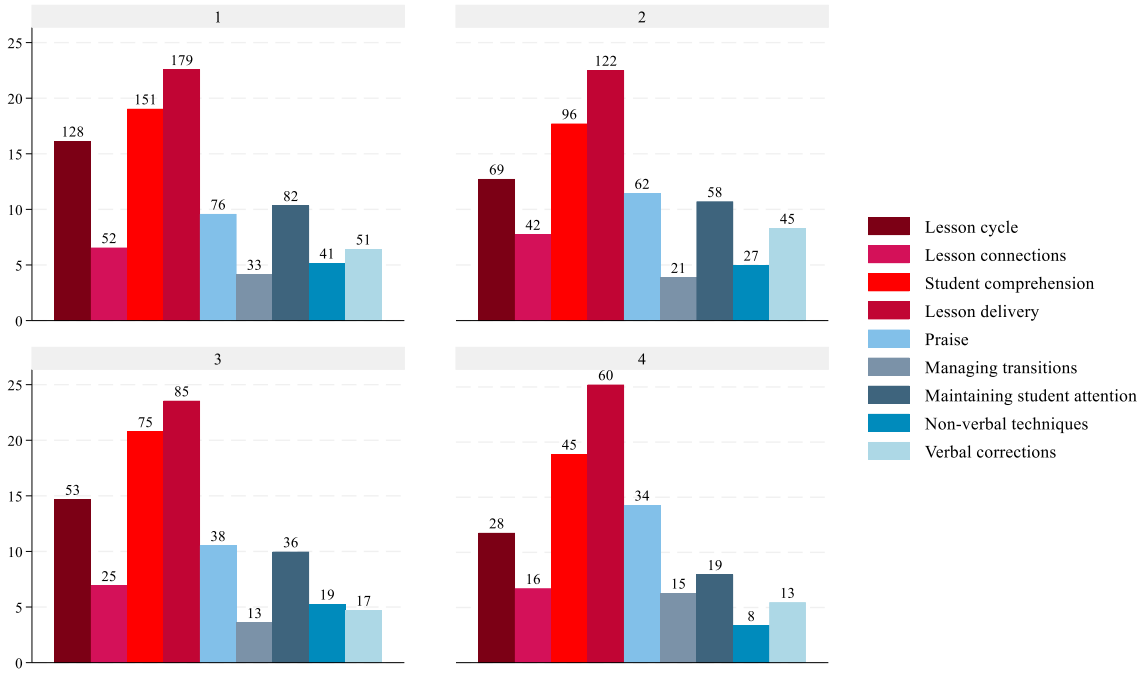
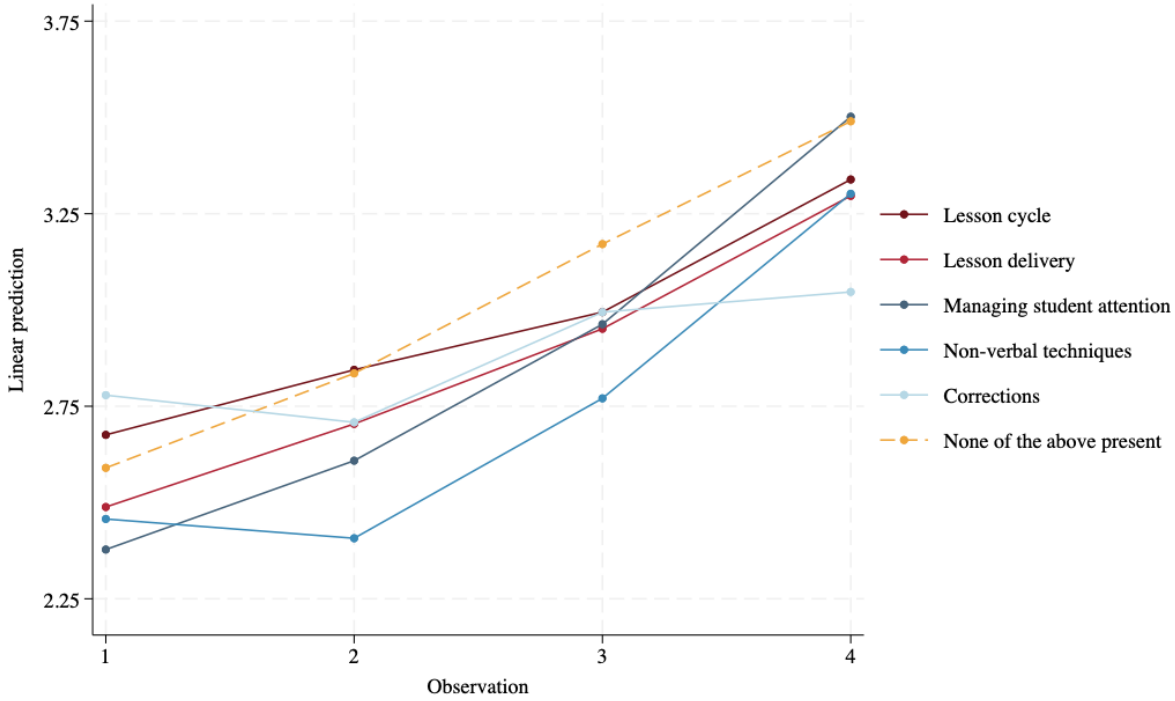


Figure 4

Margins Plots of Present Significant Feedback Codes by Observation Number



Appendix 1

Pairwise Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Observation Number	1.000										
(2) Overall Rating	0.521*	1.000									
	(0.000)										
(3) Praise	-0.060*	-0.063*	1.000								
	(0.000)	(0.000)									
(4) Managing Transitions	-0.040*	-0.051*	0.095*	1.000							
	(0.021)	(0.003)	(0.000)								
(5) Maintaining Student	-0.098*	-0.147*	0.172*	0.093*	1.000						
Attention	(0.000)	(0.000)	(0.000)	(0.000)							
(6) Non-Verbal Techniques	-0.071*	-0.128*	0.082*	0.043*	0.134*	1.000					
	(0.000)	(0.000)	(0.000)	(0.013)	(0.000)						
(7) Verbal Corrections	-0.082*	-0.079*	0.085*	0.131*	0.125*	0.099*	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
(8) Lesson Cycle	-0.126*	-0.085*	0.136*	0.141*	0.092*	0.007	0.123*	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.674)	(0.000)				
(9) Lesson Connections	-0.067*	-0.058*	0.266*	0.085*	0.195*	0.029	0.055*	0.241*	1.000		
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.093)	(0.001)	(0.000)			
(10) Student Comprehension	-0.114*	-0.099*	0.138*	0.068*	0.080*	0.049*	0.101*	0.185*	0.152*	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)		
(11) Lesson Delivery	-0.121*	-0.163*	0.171*	0.080*	0.158*	0.081*	0.098*	0.181*	0.134*	0.197*	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

*$p < 0.1$. **$p < 0.05$. ***$p < 0.01$.

Appendix 2

Table 1a

Chi-Square Test of Feedback Code by Observation Number

	Observation Number				Total
Categories of Feedback	1	2	3	4	
Yes Instructional, Yes Behavioral	171	104	64	47	386
No Instructional, No Behavioral	513	475	537	587	2,112
Yes Instructional, No Behavioral	185	138	129	77	529
No Instructional, Yes Behavioral	98	107	60	57	322
Total	967	824	790	768	3,349

[$X^2(9) = 144.99$, $p < 0.001$]

Table 2a

Chi-Square Test of Feedback Code by Observation Evaluation Overall Rating

	Overall Rating				Total
Categories of Feedback	1	2	3	4	
Yes Instructional, Yes Behavioral	0	154	205	27	386
No Instructional, No Behavioral	1	353	1,151	606	2,111
Yes Instructional, No Behavioral	0	127	325	77	529
No Instructional, Yes Behavioral	2	112	181	27	322
Total	3	746	1,862	737	3,348

[$X^2(9) = 244.90$, $p < 0.001$]

[t] Of the nine observations with no rating only one observation had no comments so remaining comments were included in the dataset with *NA* ratings and thus included in the count here.

Table 3a

Chi-Square Test of Feedback Code by Grade Level

	Grade Level Groupings					Total
Categories of Feedback	PK & K	1-3	4-6	7 & 8	9-12	
Yes Instructional, Yes Behavioral	29	94	147	92	24	386
No Instructional, No Behavioral	260	695	732	268	157	2,112
Yes Instructional, No Behavioral	31	135	194	105	64	529
No Instructional, Yes Behavioral	28	110	120	36	28	322
Total	348	1,034	1,193	501	273	3,349

[$X^2(15) = 92.41$, $p < 0.001$]

Table 4a

Chi-Square Test of Feedback Code by Subject Area

	Main Subject Area					Total
Categories of Feedback	Self-Contained	ELA	Math	Science	Social Studies	
Yes Instructional, Yes Behavioral	104	58	69	56	45	332
No Instructional, No Behavioral	810	277	313	230	121	1,751
Yes Instructional, No Behavioral	147	99	101	83	35	465
No Instructional, Yes Behavioral	120	43	37	38	23	261
Total	1,181	477	520	407	224	2,809

[$X^2(12) = 65.50$, $p < 0.001$]

Appendix 3

Linear Regression Models Including Supervisor Categorical Subject and Grade Levels

		Contextual Models	
Variables	Model 1	Model 2	Model 3
Behavioral Feedback Category Subcodes			
Praise	-0.001 (0.049)	-0.004 (0.049)	-0.002 (0.042)
Transitions	-0.015 (0.074)	-0.016 (0.074)	-0.019 (0.064)
Attention	-0.103*** (0.050)	-0.102*** (0.050)	-0.066*** (0.044)
Non-verbal techniques	-0.098*** (0.069)	-0.095*** (0.069)	-0.066*** (0.059)
Corrections	-0.033 (0.061)	-0.031* (0.060)	-0.012 (0.0520)
Instructional Feedback Category Subcodes			
Lesson cycle	-0.039** (0.043)	-0.038** (0.043)	-0.005 (0.037)
Lesson connections	0.001 (0.062)	0.002 (0.062)	0.009 (0.053)
Student comprehension	-0.052**(0.037)	-0.052 (0.038)	-0.020 (0.032)
Lesson delivery	-0.117*** (0.035)	-0.116*** (0.035)	-0.079*** (0.034)
Subjects (*Self-contained* set as reference group)			
ELA		-0.116*** (0.053)	-0.106*** (0.053)
Math		-0.010*** (0.057)	-0.100*** (0.056)
Science		-0.095*** (0.059)	-0.094*** (0.059)
Social Studies		-0.081*** (0.067)	-0.086*** (0.066)
ELA & Social Studies		-0.064*** (0.057)	-0.062** (0.057)
Math & Science & Social Studies		-0.032* (0.106)	-0.029* (0.106)
Grade Levels (*PK* set as reference group)			
Kindergarten		0.106** (0.105)	0.112** (0.104)
1st		0.126** (0.103)	0.125** (0.102)
2nd		0..150*** (0.111)	0.144*** (0.110)
3rd		0.111*** (0.109)	0.120** (0.108)
4th		0.153*** (0.112)	0.150** (0.111)
5th		0.191*** (0.114)	0.183*** (0.114)
6th		0.203***(0.117)	0.193*** (0.117)
7th		0.221*** (0.116)	0.217*** (0.115)
8th		0.158*** (0.342)	0.140*** (0.340)
9th		0.023 (0.391)	0.020 (0.432)
12th		0.152*** (0.118)	0.157*** (0.118)
Constant (no criticism feedback)	3.08*** (0.013)	2.85*** (0.098)	2.18*** (0.089)
Adjusted R-squared	0.056	0.065	0.308

Note: Model 1 displays the regression model and its standardized coefficients (same as Table 3). Model 2 introduces controls of subject areas and grade levels without controlling for observation number or supervisor. The Model 3 column includes all final control variables.

Appendix 4

Figure 1a

Supervisor Feedback Code by Observation Evaluation Overall Rating (N=3,349)

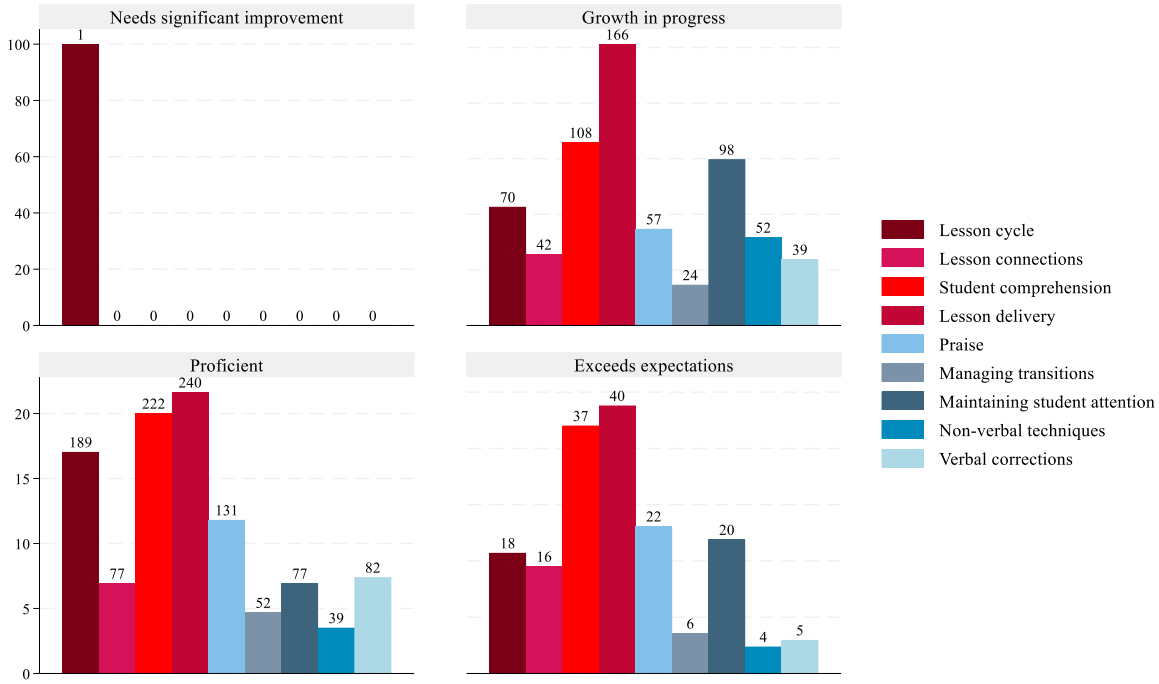


Figure 1b

Supervisor Feedback Code by PST Grade Level (N=3,349)

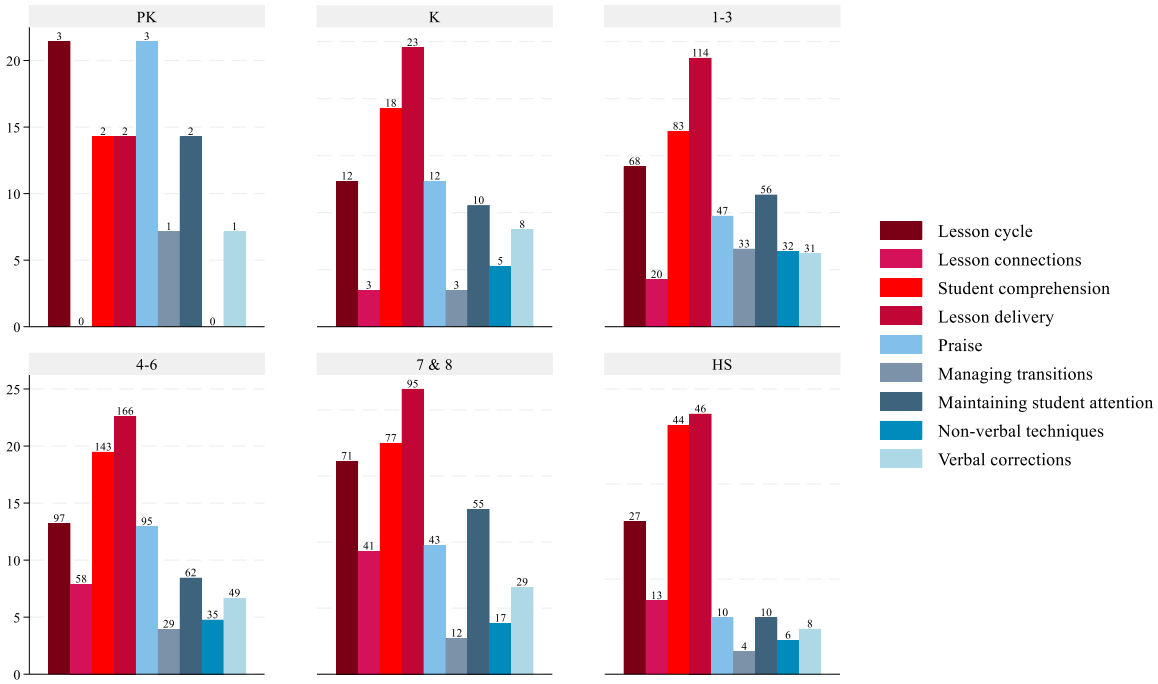
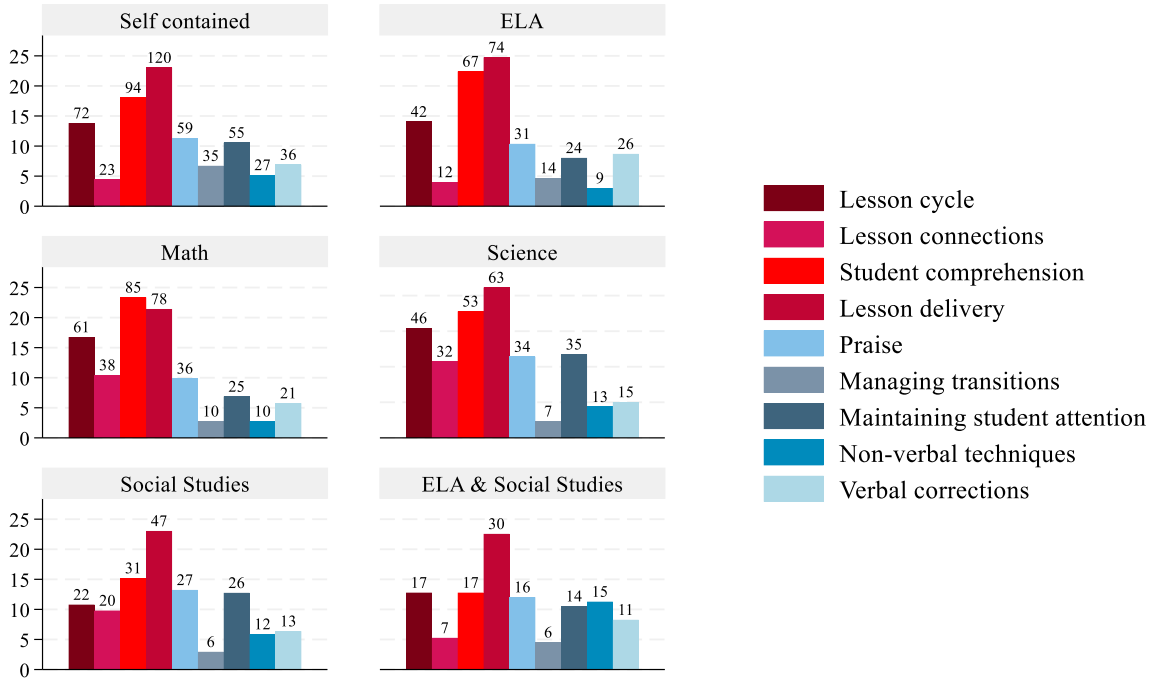


Figure 2c

Supervisor Feedback Code by PST Subject Area (N=3,349)



Appendix 5

Table 1a

Descriptive Statistics and Summary Table for Measures Analysis of Variance of the Effects of Supervisor Feedback Code and Observation Number on Overall Rating Repeated

	Observation				RM-ANOVA *F*		
Feedback Code	1	2	3	4	Feedback code (F)[a]	Observation number (N)[b]	F × N
Praise	2.59 (0.06)	2.76 (0.07)	3.03 (0.09)	3.29 (0.09)	0.37	88.04***	0.83
Maintaining Student Attention	2.35 (0.05)	2.50 (0.08)	2.86 (0.12)	3.47 (0.14)	6.34*	99.68***	3.07*
Managing Transitions	2.64 (0.10)	2.57 0.11)	2.91 (0.08)	3.27 (0.15)	0.72	20.56***	1.42
Non-Verbal Techniques	2.39 (0.08)	2.30 (0.09)	2.74 (0.13)	3.12 (0.30	8.30**	13.24***	0.78
Corrections	2.75 (0.07)	2.58 (0.08)	2.94 (0.10)	2.92 (0.18)	7.75**	8.61***	7.42***
Lesson cycle	2.66 (0.04)	2.78 (0.06)	2.96 (0.07)	3.25 (0.12)	5.05*	47.31***	8.98***
Lesson connections	2.48 (0.07)	2.79 (0.09)	3.20 (0.10)	3.31 (0.18)	0.06	62.23	0.44
Student comprehension	2.58 (0.04)	2.66 (0.06)	3.19 (0.06)	3.24 (0.09)	1.98	96.41***	1.43
Lesson delivery	2.48 (0.04)	2.66 (0.05)	2.92 (0.07)	3.25 (0.07)	15.10***	123.45***	1.86

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. [a] $df = 1$. [b] $df = 3$.

Appendix 6

Figure 1a

Margins Plots of Instructional *Feedback Codes by Observation Number*

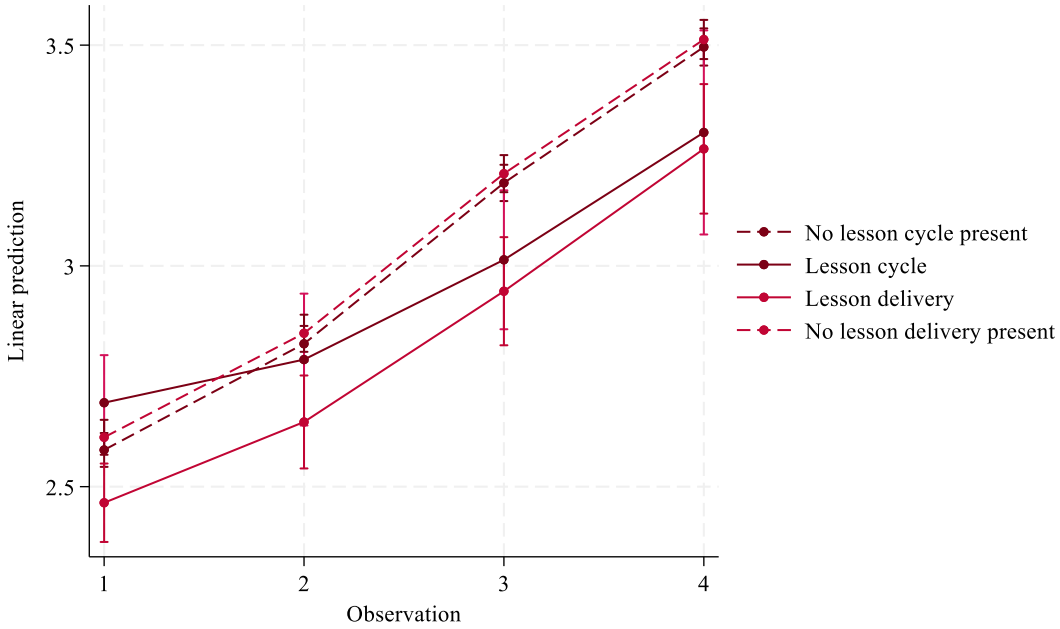


Figure 1b

Margins Plots of Behavioral *Feedback Codes by Observation Number*

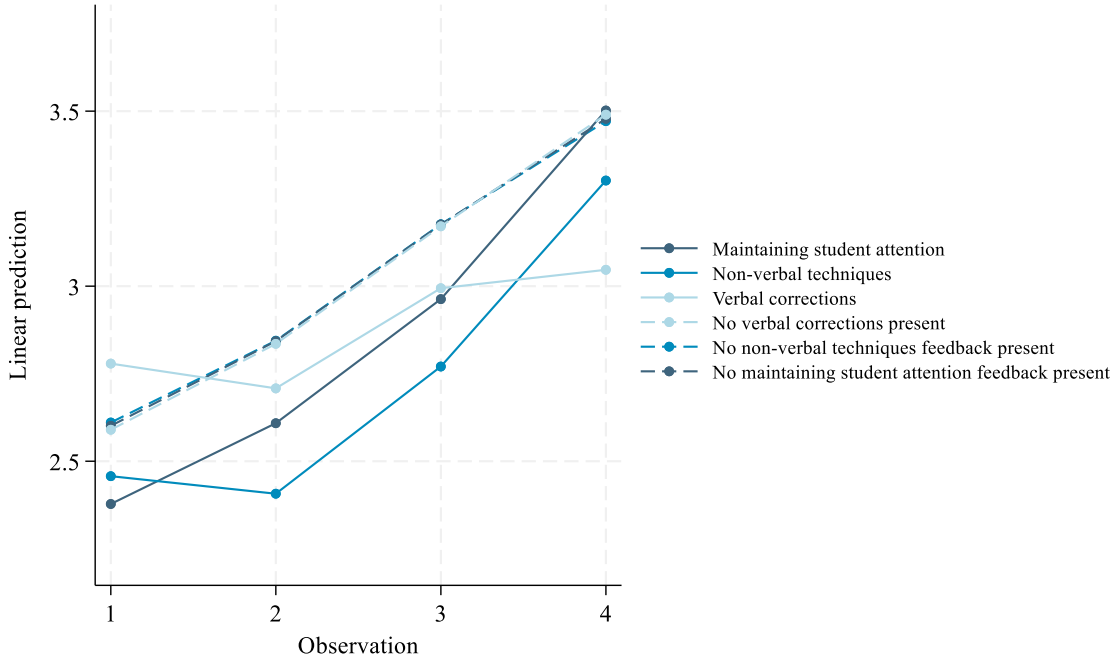


Figure 2

Margins Plots of Remaining Feedback Codes by Observation Number Compared to None Present