

Online-Only Technical Appendix

This appendix provides additional technical information on our use of large language models (LLMs) to analyze supervisor feedback and PST reflections. Here, we describe our coding framework, detail the prompt structure and classification procedure used to extract quality indicators and areas for improvement, and present illustrative examples of coded text. We also report results from multiple validation exercises, including reliability checks, human-LLM agreement rates, and sensitivity and specificity analyses using an independently coded subsample. Together, these materials aim to support the transparency and reproducibility of our text analysis approach. Additionally, we provide a detailed description of how we construct the Best Linear Unbiased Predictors (BLUPs).

Access to Data and Code

All analysis code and replication materials for this study are publicly available at REDACTED FOR ANONYMOUS REVIEW. The data used in this study are available through the openICPSR repository at REDACTED FOR ANONYMOUS REVIEW. Together, these materials enable full replication of the results reported in the main text and this appendix.

Description of Coding Process

As described in the main text, we employed a large language model (LLM) to systematically code the quality of feedback and reflections, as well as to identify the teaching skills highlighted for improvement. Our coding process proceeded in two phases. First, we developed a classification scheme grounded in prior literature and refined it through an iterative process. The four measures of feedback and reflection quality—(1) identification of strengths, (2) use of specific examples, (3) discussion of areas for improvement, and (4) inclusion of actionable next steps—were informed by research from Ellis and Loughland (2017), Hattie and Timperley (2007), Hunter and Springer (2022), and Scheeler et al. (2004). To identify salient teaching skills, we reviewed random samples of feedback and reflections and applied topic modeling techniques to surface commonly discussed pedagogical themes.

Once the quality indicators and areas for improvement were finalized, we iteratively refined a prompt to ensure that LLM outputs aligned with our manual coding. We then used Python to loop over each of the 20,000+ documents in our dataset, embedding each document within a structured prompt that was submitted to the ChatGPT-4o-mini API. The prompt employed single-shot classification—providing a single labeled example—to instruct the model to return binary indicators for the presence of each quality feature and the primary teaching skill targeted for improvement. As these tasks involve relatively low-inference judgments that can plausibly be defined as “objective” by domain experts, we believe single-shot prompting is appropriate for our use case (Liu et al., 2024, 2025).

The LLM returned a JSON object for each document, where each field corresponded to one of the predefined quality indicators or areas for improvement. Each field was coded as a binary variable, equal to one if the feature was present and zero otherwise. Examples of the prompts provided to the LLM for analyzing supervisor feedback and PST reflections are shown in Figure B1 and Figure B2, respectively.

Potential Bias in LLM Coding

An important limitation of our approach is the potential for the LLM to exhibit systematic bias in how it classifies text. Such bias could arise if the model is differentially sensitive to

linguistic patterns, writing styles, or discourse structures. If supervisors from different demographic backgrounds, experience levels, or geographic regions systematically use different language, for example, more formal academic phrasing, colloquial expressions, or regional dialects, the LLM may vary in its accuracy even when the underlying substantive content is similar. In addition, because the LLM’s training data consist primarily of internet and published text, they may not fully represent the specialized discourse used in teacher education contexts. Writing styles that are underrepresented in the training corpus may therefore be classified less reliably.

We took several steps to mitigate these concerns. Our iterative prompt development process involved reviewing outputs across randomly selected documents to identify and correct systematic errors. The high agreement between the LLM and human coders (Table B4), as well as the strong correspondence between LLM classifications and independently conducted qualitative coding (Table B5), suggests that the model captures meaningful pedagogical content rather than surface-level stylistic features. Nevertheless, we cannot rule out systematic bias because our data lack information on supervisor characteristics such as demographics, educational background, or years of experience, preventing direct tests of differential accuracy. Future research that combines LLM-based text analysis with detailed supervisor-level data would allow for more rigorous bias assessment and correction. Researchers using similar methods should carefully validate LLM outputs against human coding and acknowledge the potential for bias when key author characteristics are unobserved.

```
'You are a researcher that analyzes reflections written by pre-service teachers (PSTs) after classroom observations.

Text to analyze:
{text}

Analysis Rules for quality indicators:
1. quality_indicators should identify whether the reflection contains specific measures that are related to reflection
   quality (i.e., specific examples, next steps, strengths, and areas for growth)
- quality_indicators.specific_examples should flag whether the reflection contains specific examples of things that
   occurred during the lesson
- quality_indicators.next_steps should flag whether the reflection contains clear, actionable next steps for the PST
- quality_indicators.strengths should flag whether the reflection contains comments on areas or things that the PST did
   well
- quality_indicators.areas_for_growth should flag whether the reflection contains specific growth areas for the PST to
   improve upon.
- CRITICALLY IMPORTANT: areas_for_growth is a function of area_for_improvement and should only be 1 if
   area_for_improvement != "none"

Examples for quality indicators:
- "One thing that I think I do well is giving examples that the students can relate to in order to keep them engaged. I
   think that I have good discipline and my students
   really respect me because of it. I think that I also had good energy. It is so much easier to keep students engaged
   when you are excited about what you are teaching.
I need to get better about blatantly stating the objective. I can seem to use the right phrasing."
- quality_indicators.specific_examples: false
- quality_indicators.next_steps: false
- quality_indicators.strengths: true
- quality_indicators.areas_for_growth: true

- "This day wasn't as great of a "lesson" exactly. They way timing worked out with schedules and 9 - week exams just
   forced us in to a corner on this one.
   Nevertheless, I still found a way to bring in some teaching points. The students did test corrections, and taught me
   how to analyze the data to go over the most missed questions,
   and evaluate what may have caused the mistakes. I took those most missed questions and showed the kids what they did
   wrong with the use of the document camera and the smart board.
It was somewhat dry at times, but I noticed I got a lot of faces lighting up and noises when we went back through the
   questions together.
It was nice to see my students learning from their mistakes, and because of this process their grades improve on every
   test for the most part."
- quality_indicators.specific_examples: true
- quality_indicators.next_steps: false
- quality_indicators.strengths: true
- quality_indicators.areas_for_growth: false

Analysis Rules for Area for Improvement:
1. area_for_improvement should identify the teaching skill that is most prominently discussed as needing improvement in
   the reflection
- Choose "none" if no areas are identified as needing improvement
- Choose "multiple" if 2+ areas are equally emphasized as needing improvement
- Choose "other" if the main area for improvement doesn't match the listed skills
- Otherwise, choose the single skill most emphasized as needing improvement in the list of target_skills provided
- CRITICALLY IMPORTANT area_for_improvement should be a function of areas_mentioned. For example, if there are multiple
   areas_mentioned flagged as true, this should be "multiple"
- CRITICALLY IMPORTANT area_for_improvement can ONLY be {skills_list}, "other", "none", or "multiple". This variable MUST
   NOT take on any other values.

2. areas_mentioned should ONLY be true (1) when the reflection explicitly indicates that a skill needs improvement
- Set to false (0) if the skill is praised or mentioned positively
- Set to false (0) if the skill is just mentioned descriptively without suggesting improvement
- Set to true (1) ONLY if the reflection suggests this specific skill needs improvement

Example interpretations:
- "My classroom management (using hand signals, certain phrases) is getting better with time. The math subject has always
   made me nervous to teach, but as I practice teaching
   it more I am getting more confident in teaching math and math strategies. What I would have done differently (and what
   I need to improve on) is time management and keeping station time /
   teacher table around 15 - 20 minutes. My goal is to keep better track of students' activities in stations while I am at
   the teacher table (and redirecting off - task behavior in these stations)."
- area_for_improvement: "Classroom Management"
- areas_mentioned.classroom_management: true
- areas_mentioned.lesson_planning: false
- "My classroom management was excellent" * classroom_management: false
- "I need to work on my classroom management" * classroom_management: true
- "I took attendance and managed transitions" * classroom_management: false
- "While my lesson planning was strong, my classroom management needs work" *
- area_for_improvement: "Classroom Management"
- areas_mentioned.classroom_management: true
- areas_mentioned.lesson_planning: false

Analyze the text and respond with ONLY a JSON object (no other text) using the following structure:'
```

Figure B1
ChatGPT Prompt: Supervisor Feedback

Note: Figure displays shows an example of the prompt asked to ChatGPT 4o-mini to generate the quality indicators and area for improvement of the supervisor feedback text.

```
'You are a researcher that analyzes reflections written by pre-service teachers (PSTs) after classroom observations.

Text to analyze:
{text}

Analysis Rules for quality indicators:
1. quality_indicators should identify whether the reflection contains specific measures that are related to reflection
   quality (i.e., specific examples, next steps, strengths, and areas for growth)
- quality_indicators.specific_examples should flag whether the reflection contains specific examples of things that
   occurred during the lesson
- quality_indicators.next_steps should flag whether the reflection contains clear, actionable next steps for the PST
- quality_indicators.strengths should flag whether the reflection contains comments on areas or things that the PST did
   well
- quality_indicators.areas_for_growth should flag whether the reflection contains specific growth areas for the PST to
   improve upon.
- CRITICALLY IMPORTANT: areas_for_growth is a function of area_for_improvement and should only be 1 if
   area_for_improvement != "none"

Examples for quality indicators:
- "One thing that I think I do well is giving examples that the students can relate to in order to keep them engaged. I
   think that I have good discipline and my students
   really respect me because of it. I think that I also had good energy. It is so much easier to keep students engaged
   when you are excited about what you are teaching.
I need to get better about blatantly stating the objective. I can seem to use the right phrasing."
- quality_indicators.specific_examples: false
- quality_indicators.next_steps: false
- quality_indicators.strengths: true
- quality_indicators.areas_for_growth: true

- "This day wasn't as great of a "lesson" exactly. They way timing worked out with schedules and 9 - week exams just
   forced us in to a corner on this one.
   Nevertheless, I still found a way to bring in some teaching points. The students did test corrections, and taught me
   how to analyze the data to go over the most missed questions,
   and evaluate what may have caused the mistakes. I took those most missed questions and showed the kids what they did
   wrong with the use of the document camera and the smart board.
It was somewhat dry at times, but I noticed I got a lot of faces lighting up and noises when we went back through the
   questions together.
It was nice to see my students learning from their mistakes, and because of this process their grades improve on every
   test for the most part."
- quality_indicators.specific_examples: true
- quality_indicators.next_steps: false
- quality_indicators.strengths: true
- quality_indicators.areas_for_growth: false

Analysis Rules for Area for Improvement:
1. area_for_improvement should identify the teaching skill that is most prominently discussed as needing improvement in
   the reflection
- Choose "none" if no areas are identified as needing improvement
- Choose "multiple" if 2+ areas are equally emphasized as needing improvement
- Choose "other" if the main area for improvement doesn't match the listed skills
- Otherwise, choose the single skill most emphasized as needing improvement in the list of target_skills provided
- CRITICALLY IMPORTANT area_for_improvement should be a function of areas_mentioned. For example, if there are multiple
   areas_mentioned flagged as true, this should be "multiple"
- CRITICALLY IMPORTANT area_for_improvement can ONLY be {skills_list}, "other", "none", or "multiple". This variable MUST
   NOT take on any other values.

2. areas_mentioned should ONLY be true (1) when the reflection explicitly indicates that a skill needs improvement
- Set to false (0) if the skill is praised or mentioned positively
- Set to false (0) if the skill is just mentioned descriptively without suggesting improvement
- Set to true (1) ONLY if the reflection suggests this specific skill needs improvement

Example interpretations:
- "My classroom management (using hand signals, certain phrases) is getting better with time. The math subject has always
   made me nervous to teach, but as I practice teaching
   it more I am getting more confident in teaching math and math strategies. What I would have done differently (and what
   I need to improve on) is time management and keeping station time /
   teacher table around 15 - 20 minutes. My goal is to keep better track of students' activities in stations while I am at
   the teacher table (and redirecting off - task behavior in these stations)."
- area_for_improvement: "Classroom Management"
- areas_mentioned.classroom_management: true
- areas_mentioned.lesson_planning: false
- "My classroom management was excellent" * classroom_management: false
- "I need to work on my classroom management" * classroom_management: true
- "I took attendance and managed transitions" * classroom_management: false
- "While my lesson planning was strong, my classroom management needs work" *
- area_for_improvement: "Classroom Management"
- areas_mentioned.classroom_management: true
- areas_mentioned.lesson_planning: false

Analyze the text and respond with ONLY a JSON object (no other text) using the following structure:'
```

Figure B2
ChatGPT Prompt: PST Reflections

Note: Figure displays shows an example of the prompt asked to ChatGPT 4o-mini to generate the quality indicators and area for improvement of the PST reflection text.

Table B1
Quality Indicator Coding Examples

Quality Indicator	Example
Strengths: A general comment highlighting what the preservice teacher is doing well, such as preparation, demeanor, or instructional strategies.	<p>1 Strength:This student teacher has the personality of a life-long learner. Target for Growth:Continue to focus on management skills. There was side talking during your rotations.</p> <p>2 Maintained appropriate student attention throughout less Appropriate materials available for students Involved all student in class activities</p> <p>3 - 18, 2014 - 8:15 - 8:50 Strengths:Effective questioning strategies during content discussion, elaborates after students'responses, receptive to suggestions, highly organized and always prepared for lessons, positive communication with students, eager to learn, resourceful with age appropriate activities, and demonstrates dedication when she stays late with the without being asked. Recommendations:Create opportunities for students to participate in discussions, continue expecting students to follow your behavior management by focusing their eyes on you while giving directions, secure students'attention during transitions, and know when to move on with the lesson activity. 3 knows the importance of including all lesson cycle components as she provides instructional activities for each lesson objective. She has established a positive classroom environment where she listens to the student 's responses then gives appropriate praise. Recommend:Call on more non - volunteers, enforce expectations for behavior stopping the lesson activity when students do n't comply, and require respect from students while teaching.</p> <p>4 is well respected by her students and they are eager to participate in her classroom. She plans lessons that are creative, engaging, and productive. She constantly monitors and questions her students so she can assess their learning throughout the lesson. Her cooperating teacher is very pleased with her classroom management skills, her creative lesson planning, and her work ethic.</p>

Continued on next page

Continued from previous page

Quality Indicator	Example
<p>Specific Examples: A concrete description of an observed behavior, instructional move, or classroom moment that illustrates the teacher's practice.</p>	<p>5 is developing into excellent teacher. Her calm approachable demeanor serves her very well with the middle school population that she serves. The group of students she is currently serving did not pass their state assessment; clearly understands the math "phobias" she is dealing with now. She goes the extra mile to make them feel successful for attempting a task they often feel is unobtainable.</p> <p>1 is an outstanding student teacher! She seems to spend a great deal of time preparing lessons and can implement them as well as a teacher with experience. She includes as much technology as possible and selects materials, such as the commercials today, to keep the students interested. She handles the class as well as teachers with experience.</p> <p>2 presented a strong lesson that took the form of a discussion similar to what one would see in a college classroom. This was very appropriate for the astronomy class she was teaching.</p> <p>3 demonstrated strong discussion skills that focused on found in the book " " and " ". Many questions including HLT questions were asked to motivate the students as she modeled each activity that supported the lesson objectives. Recommendations: Continue developing the Cycle.</p> <p>4 did outstanding teaching today during the introductory lesson on the eighth grade research project and the development of the thesis statement. Although this is 's first observation on the thirteenth day of student teaching, she appeared polished and prepared and made "on your feet" decisions during instruction. She has an excellent rapport with the class and appeared relaxed during the lesson.</p> <p>5 developed her teacher voice and instructional delivery methods throughout the semester. She became confident in delivery instruction and more confident in praising students for their efforts in paying attention during class. She will continue to practice this positive method of securing student cooperation and attention during the last few days of student teaching. At the summative conference, the topics of career fair, job applications, last evaluation and next steps for job seeking were discussed.</p>

Area for Improvement:

Continued on next page

Continued from previous page

Quality Indicator	Example
A statement identifying something the preservice teacher needs to work on, typically tied to a specific teaching skill.	<p>1 -Good classroom management / states expectations several times -Good simulation of roles of king - knight - peasant - serf> feudal system; use of today 's rules related to classroom -Work on mini - transitions in timely manner when notes / lecture to be long -Overall:positive, good interaction, be aware of student body language when lecturing; be aware of maybe " hurrying " during notetaking.</p> <p>2 -Transitions / pacing good / knowledgeable; even though, planning is brief -Procedures explained for activities / be sure to be specific on behavioral expectations -"So"- stopped counting @19 in first 20'-distractor could be a problem -Good rapport; relates to students well -Power point - covering poetry - discussion / back 95 % of time to sts by window / some off task antics took place away from power point table for discussion; so you can see whole class</p> <p>3 continues to plan engaging, meaningful activities in which all students want to participate and are successful. She excelled in moving throughout the room with purpose and providing students with opportunities to self - monitor their work. successfully connected learning to prior knowledge and reinforced concepts as students worked in small groups. It is recommended that consistently implement the behavior management plan she has established and set behavioral expectations at the beginning of each activity.</p> <p>4 I enjoyed watching you work with your kids! You look like you are having a blast! Kids like working with you as they interact easily with you and seek you out for help; Your co-operating teacher notes that you work well with individuals and use strategies to help kids who need to hear it differently; for your next visit, work on waiting for compliance after you give directions. .. reinforce your requests so that you can establish a quiet room where kids can think and work. .. be direct about your expectations. .use that serious teacher voice when making it clear what is OK and what is not. Intervene quickly when you do nt get what you want</p> <p>5 Strength:The teacher walks the floor to control the class by her presence and to reteach when it is necessary. Target for Growth:Continue to focus on management skills.</p>

Specific Next Steps:

Continued on next page

Continued from previous page

Quality Indicator	Example
<p>A clear, actionable suggestion or strategy the preservice teacher should try to improve their teaching.</p>	<p>1 organizes the content of the lesson so that it is meaningful to the student. She presented a lesson today over fantasy and read the book the, she pointed out several examples of how this book was a fantasy and what penguins would really do in real life. She will continue to model her learning expectations in detail in order for students to master their understanding of the.</p> <p>2 -Good, clear introduction to food chains with a simple board diagram to enhance comprehension -Great idea having students read the 'I Can' statements (objectives) for the day - Be more specific in directions in what you want students to do prior to transitioning to a hands - on activity -Persistent in expectations of behavior; stopped minor issues before they escalate into bigger issues</p> <p>3 SEE COMMENT OF SECTION 6 FOR A SUMMARY OF THE 3 - CONFERENCE. This is'FOURTH formal obser-vation. This social studies lesson dealt with people who work at. Primary areas of strength noted during this lesson - motivating introduction, calm demeanor while settling a student dispute, good variety of activities and grouping for-mats, use of music and visuals to present content, high stu-dent interest in lesson, thorough closure and positive class-room climate. Areas for Refinement - continue to increase the number of higher level questions, reinforce students dis-playing good manners (boy picking up another group 's textbook left on the floor.)</p> <p>4 did a good job of creating a fun lesson for the students and did really well at stating her academic and behavioral expectations for the lesson. She went over the expecta-tions and secured the students'attention before beginning the lesson. She made connections to prior learning and also connected the learning to the real world. will continue to develop include the following:Waiting until students are quiet before giving any instructions, practicing procedures as needed, adding higher level questions (especially when checking the math worksheet), making sure students have something to do at all times (especially when waiting to be called back to the tables or to the floor), and giving more positive praises and feedback.</p>

Continued on next page

Continued from previous page

Quality Indicator	Example
5	-Preparation and organization of lessons well done with challenges demonstrates courtesy with students -Behavior management overwhelming challenge; work on consequences regardless if student serves time; you did your job

Table B2
Area for Improvement Coding Examples

Area for Improvement	Example
Classroom Management	<p>The teacher needs to improve routines, transitions, behavioral expectations, or ability to maintain student focus.</p> <p>1 -Challenging warmup; creating interest leading into the new of the day -No behavior expectations stated; reactive to off task behaviors -Good transitions to keep students engaged</p> <p>2 Continues to display confidence. Very good lesson plan. Good use of student names in questioning. Be more assertive in your classroom management. Work on closures.</p> <p>3 Strength:Teacher is aware of herself and student achievement. She is willing to do anything to support the learning of her students. Target for Growth:Continue to monitor time and classroom management. Conference:Teacher is ready for her own classroom. She enjoys her students, and they love her.</p> <p>4 I'm please with the overall teaching performance of for the fall of 2016 at in. has come a long way from the beginning of her assignment. I have seen growth in the areas of, of and and. The area which I feel needs improvement is, specifically keep students on tasks and praising students.</p> <p>5 Strengths:Teacher is flexible. Target for Growth:will continue to be a focus; remind students at each period of what is expected and hold the students accountable. Conference:Teacher is ready to have her own classroom of students. She is doing a great job.</p>
Lesson Planning	<p>The teacher needs to better sequence content, articulate objectives, or integrate lesson components coherently.</p> <p>1 did a solid job of teaching the lesson. She should have brought them back together before either letting them start on their homework or before leaving at the end of the class. At that point in time she could have reviewed the three main points of the lesson. Good teacher voice.</p> <p>2 -Did not have lesson plans available -Reflections good; but few demeanor, clear, strong voice -Begin instructions when you have the focus of whole class -Good activities connecting / learning letter by picking appropriate number to indicate what color to use</p> <p>3 -Stated expectations for research paper, deadlines and final form expectations -Need to state behavior expectations before having students go to independent work -Great improvement in interaction with students</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
<p>Student Engagement The teacher needs to enhance monitoring, participation, motivation, or inclusion of all students during instruction.</p>	<p>4 understand the importance of. Today 's class seemed to be a pulling together of loose ends. Only a small percentage of time was devoted to instruction about mitosis and it 's importance.</p> <p>5 Strength:Teacher used specific praise with students. She told students why they must follow her directions – why they must sit on all four legs of the chair. Target for Growth:Remember your closure.</p> <p>1 , quick, brief but positive approach to handling minor off task behavior -Great challenge to students finding perimeter with odd shapes, x factor and word problem -Bell ringer even with timer bit long; students became off task talking and socializing</p> <p>2 Goal:To increase informal assessment of learning by asking students rigorous questions to check for understanding. I would suggest that the pacing of the warm - up be set at 5 minutes to increase students on - task focus. Also, transitions could be smoother by implementing a new procedure for independent work with textbooks. Moning students during warm - up, notes and independent work time is essential to make sure all students are on - task and using their class time wisely.</p> <p>3 -Knowledgeable of content -Allowed spontaneous answers; few responders; need to call on students to include more participation -Be aware of whole class when presenting using technology -Student efforts and right responses acknowledged</p> <p>4 Strength:Teacher is using new materials to support her lesson. Target for Growth:Focus on management skills. Be sure all students participate.</p> <p>5 -Persistent with student focus during lesson - versed in content; presentation of information clear and consistent - Careful of one group dominating / interrupting and another group quiet / non - participating -Direct, assertive with off task behavior -During instruction may want to move about the room during question / answer to ensure all on task</p>
<p>Communication</p>	

Continued on next page

Continued from previous page

Area for Improvement	Example
<p>The teacher needs to improve clarity, voice projection, presence, or responsiveness in instructional delivery. May also include non-verbal communication skills.</p>	<p>1 Strength:Teacher is always prepared and planned. Target for Growth:You are very trained; have confidence in yourself. Be excited and sell;yourself.</p> <p>2 You are becoming more comfortable as a teacher as I saw growth from last observation. Your lesson had planned engaging activities for the students which was good. However, it did not move as quickly as planned which then caused concerns with behavior. You still need to work on using a firm teacher voice level and being more assertive as the teacher.</p> <p>3 Strength:Teacher is building relationships with the students. Target for Growth:Find your " teacher voice ". Be excited; project your voice as you teach. Be confident in yourself.</p> <p>4 () continues to improve her teaching skills by motivating the students and securing their attention throughout the lesson. Behavior expectations were in place and enforced to keep the students on task. Recommendations:students understanding of the lesson objective, increase praise and positive reinforcement to improve student behavior, and avoid / stop using, and OK. CONFERENCE - 27, 2013 10:30 - 11:00 Strengths:Takes the initiative to do what 's needed without being asked, demonstrates a pleasant attitude, prepares effective materials for student activities that meet their learning styles, strong use of technology with the students, strong organizational skills, and plans a variety of activities for most lessons. Areas to Improve:Continue establishing behavioral expectations for all transitions and activities in the classroom, increase the use of praise and positive reinforcement, follow through with early warnings for inappropriate behavior by issuing consequences, slow down speech and think about what students are being asked to do or answer, and stop saying overused words - " alright " and " so ".</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
	<p>5 -Good demonstrative voice; somewhat rapid in speech may be a concern done introduction - stating objective and use of power point / new topic -Presentation demonstrates preparation, knowledge of topic management techniques good - direct, brief, good eye contact, voice means business -Well organized and handled distractors well</p>
Assessment and Feedback	<p>The teacher needs to strengthen their use of formative assessment or provide clearer, more effective feedback to students.</p> <p>1 I always smile when I see that you are on my schedule cause I enjoy watching you work! You ve done such good work in establishing rapport with your kids and you prepare interesting lessons, so its no wonder the kids work so well with you. What an engaging idea. .. what would you take with you for survival? .. they really got into this lesson. I also noted what a caring job you were doing with a boy that struggles with on task behavior and not blurting out. You sat with him, talked him thru some answers and then let him do his work. .all great ideas for helping to diffuse a brewing situation. Moving forward continue your pursuit for time management techniques and also think of ways to make the target and assessment be more than statements. .. how to make them into meaningful activities. .. being more deliberate in their presentation. .. involve kids in reading them and talking about what is expected as the lesson moves forward. Ask how their prior learning could be helpful in the lesson today? How will you know that they understood the lesson? Make them do this work</p> <p>2 I have observed considerable growth in instruction. A very good lesson that was well organized that connected to a product that the students can learn from. His scaffolding has improved along wit his assessment strategies, divergent questioning, procedures and routines. Needs to monitor all students through out all activities and praise students more often during instruction.</p> <p>3 , you did an exceptional job of teaching a challenging lesson today. The students seemed highly motivated to find all the text features and remained on task to complete the scavenger hunt. I wonder if listening to the students' responses to assess progress or increasing student participation throughout the lesson would change the students'understanding of the material.</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
	<p>4 is making progress in student teacher. The positive relationship she has with the students makes the classroom climate positive and also encourages appropriate behavior. During the conference today, we talked about the need to establish the purpose of learning history and the need to apply what they are learning to real life examples. should continue using independent work to aide in assessing student learning and to let the students apply their knowledge to the topic.</p> <p>5 -Great job of using students to read the objectives and what to they will be doing for the day 's lesson -Good use of to begin class; used the quiz as a warm up up did take an extraordinary amount of time; use a timer sooner</p>
<p>Student Comprehension</p> <p>The teacher should focus on ensuring students are understanding the material, including checking for understanding and adjusting instruction as needed.</p>	<p>1 Very professional demeanor. Good movement around the classroom while teaching the lesson. Descriptions and motions to help explain seemed to help students understand better. You may want to elaborate on student answers for more thorough reinforcement of concepts.</p> <p>2 -Very good teacher voice > clear, distinct, good inflection -Difficult concept - anticipate in planning the lesson - seek a variety of activities to help in reasoning and comprehension -Voice helpful in management of class / firm, direct -Lesson coverage detailed / do not let better students dominate answers / when half the class takes longer to stand with answer done on a sample problem / alternatives may be needed to reach understanding of concept / lesson</p> <p>3 Strengths:Suggestions are immediately applied. Target for Growth:Be sure to check for understanding and include this in your pace. Three - way Conference:Tremendous growth this semester with different populations. is a natural teacher. She will be an asset to any campus.</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
	4 This is 's FIRST formal observation. This was an algebra lesson on multiplying polynomials. Primary areas of strength noted during this lesson - clear explanations, consistent review of major concepts and rules, detailed / complete lesson plans, voice projection and relating today 's lesson to previously learned material. Areas needing refinement - soliciting student participation (volunteers and non - volunteers), allowing time for students to copy notes from board before proceeding, slowing down speed of presentation and informal assessment during the lesson to see that the students are understanding the material being presented.

Continued on next page

Continued from previous page

Area for Improvement	Example
	<p>5 Strengths:and Teaching:1. Effective motivation[148]bell work directed students to draw a chart how fuel produces electricity; after giving students time to draw this diagram, you modeled on the smart board how it happens 2. The bell work identified prior knowledge. 3. The objective was displayed on the smart board, but was not stated 4. The purpose was implied by citing advantages and disadvantages of the different energy resources 5. Students felt very free to ask questions during the lecture after raising their hand 6. While lecturing and answering students'questions, you walked about the classroom moning their work 7. For closure, you asked what are some of the benefits of using each type of energy source and what produces lots of carbon dioxide which individual students answered; you stated that tomorrow we would talk about solar power plant (future learning). 8. Your presentation was highly organized and well presented management / environment:1. Had materials ready when needed 2. was effective 3. Established a climate of courtesy and positive rapport with students 4. You were enthusiastic and sincere. 5. Students were attentive and asked numerous questions for clarification and understanding. 6. Students appeared to know what was expected of them behaviorally. Suggestions for Improvement:and Teaching:1. As part of closure restate the objective or have students state what they learned; how could you have involved all students in the closure? 2. Students asked higher level thinking questions; how could you have incorporated higher level thinking questions? 3. How could you have checked for students'understanding during the presentation? Classroom management / environment:1. At the start of class and anytime you want to give instructions, make certain you have students'attention before giving your instructions. Do not talk over students talking; rather use wait time to obtain compliance. 2. Praise students for their hard work and appropriate behavior when applicable</p>
Differentiation	<p>The teacher should work on adjusting instruction to meet diverse learner needs through varied methods or supports.</p> <p>1 Reinforcement:, transitions, and routines were appropriate and helped keep students engaged. Refinement:Explore systems to ensure equitable response opportunities for students.</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
	<p>2 You are a really talented professional. There are some lucky kids in your future. At our post - conference, it was very difficult to find something you could have improved upon on this particular lesson. It was awesome the was able to share moving your struggling learners to a small group sooner would have been helpful.</p> <p>3 You exhibit a growing confidence, especially now during the full responsibility. You are consistent, fair and calm in dealing with students'questions, behavior and learning needs. You have found the balance in what to prepare for lessons focusing on less, but prepared for anything. You understand that gifted students are truly diverse, intense and need relevant engagement and purpose, too! Challenge and extend their thinking more, they can handle it!</p> <p>4 is a little farther in her teaching skills than I would expect at this time. She needs to work on walking among the students consistently, wait time for questions and figuring out several ways to explain concepts when students do n't understand the first way, using students'names and offering sincere praise. This will come for her and i am excited to visit a second time.</p> <p>5 Overall you showed patience with the concept of elapsed time as it is a difficult concept. Working with the smaller group who needed more instructional time, you might want seat at least some at tables to keep more focused. The strategy was a new one for me, but it really seemed to show the students how to look at figuring it out.</p>
None	<p>Absence of any other area for improvement.</p> <p>1 - is growing and learning well as a clinical teacher in this active second grade classroom. She is gaining confidence and did a fine job on the math lesson today. She exhibited patience, knowledge and confidence as she modeled subtraction using a number line with her students. She wrote and followed a detailed lesson plan. She consistently checked all students understanding and spent a great deal of time helping individual students.</p> <p>2 's cooperating teacher reported that is doing an outstanding job. She is knowledgeable about her subject areas and works well with the 4th grade team. has built strong relationships with her students. is organized and has met all deadlines.</p>

Continued on next page

Continued from previous page

Area for Improvement	Example
	<p>3 There is clear evidence that has made strong strides in her confidence and abilities as a teacher.</p> <p>4 Students appeared comfortable with you and you have built rapport with them. was a lesson on political economic and social characteristics of mission life in. Small group instruction / small group work and independent work was observed.</p> <p>5 has been a steadily evolving young professional all semester. She worked very diligently with her cooperating teacher to become skilled in curriculum delivery as well as classroom management. Her humor with the students has been a great asset as she is able to use it to encourage them gently while she guides them to being confident learners.</p>

Table B3
ChatGPT Agreement Rate, by Task

Task	Feature	Supervisor Feedback (1)	PST Reflections (2)
Quality Indicator	PST Strengths	99.3	99.8
Quality Indicator	Area for Improvement	97.1	99.8
Quality Indicator	Next Steps	97.1	98.5
Quality Indicator	Specific Examples	96.2	99.5
Area for Improvement	Other	99.3	99.7
Area for Improvement	Differentiation	99.1	98.1
Area for Improvement	Assessment & Feedback	98.8	97.9
Area for Improvement	Student Comprehension	98.8	97.8
Area for Improvement	Communication	97.4	97.9
Area for Improvement	Student Engagement	96.9	95.6
Area for Improvement	Lesson Planning	96.8	94.5
Area for Improvement	Classroom Management	96.5	96.7

Note: Data generated by running two iterations of the prompts in [Figure B1](#) and [Figure B2](#). Agreement rate calculated as the percentage of identical codes between the iterations for each task.

Table B4
Validity Check Against Human Coder

Task	Feature	Supervisor Feedback			PST Reflections		
		Prevalence LLM (%) (1)	Prevalence Human (%) (2)	Agreement Rate (%) (3)	Prevalence LLM (%) (4)	Prevalence Human (%) (5)	Agreement Rate (%) (6)
Quality Indicator	PST Strengths	98	98	100	100	99	99
Quality Indicator	Next Steps	39	37	90	96	87	89
Quality Indicator	Area for Improvement	47	44	89	97	92	95
Quality Indicator	Specific Examples	80	70	88	96	89	93
Area for Improvement	Assessment & Feedback	0	1	99	4	1	97
Area for Improvement	Differentiation	0	1	99	6	11	93
Area for Improvement	Other	3	5	92	0	7	93
Area for Improvement	Classroom Management	16	19	91	26	29	89
Area for Improvement	Communication	6	3	91	4	11	87
Area for Improvement	Student Engagement	18	17	87	20	41	69
Area for Improvement	Student Comprehension	1	19	82	8	36	70
Area for Improvement	Lesson Planning	7	20	81	50	54	76

Note: Data are based on a manually coded random sample of 100 supervisor feedback and PST reflection documents. Columns (1) to (3) report results for supervisor feedback; columns (4) to (6) for PST reflections. “Prevalence” columns show the number of times a quality indicator was identified by either the human coder or the LLM, as specified in the column header. Agreement rate reflects the percentage of identical codes between the human and LLM for each quality indicator.

Table B5*Predicting Area for Improvement in Supervisor Feedback with Qualitative Coding from Authors (2025a)*

	Supervisor Feedback Area for Improvement								
	Classroom Management (1)	Lesson Planning (2)	Student Engagement (3)	En- gagement (4)	Communica- tion (5)	Assessment and Feedback (6)	Student Comprehension (7)	Differentiation (8)	No Area for Improvement (9)
Monitoring Student Behavior									
Praise	1.110 (0.097)	1.000 (0.000)	0.762 (0.125)	1.242 (0.184)	2.552*** (0.712)	0.479 (0.266)	1.155 (0.440)	0.237 (0.284)	2.052** (0.521)
Transitions	1.069 (0.100)	0.965 (0.210)	1.084 (0.215)	1.416* (0.235)	0.411 (0.277)	1.258 (0.713)	0.898 (0.553)	4.225 (4.345)	1.687 (0.603)
Attention	1.412*** (0.092)	1.112 (0.108)	1.088 (0.157)	1.336 (0.202)	0.683 (0.249)	0.851 (0.428)	0.702 (0.306)	0.000*** (0.000)	0.305* (0.173)
Non-verbal Techniques	1.576*** (0.126)	0.687 (0.223)	0.718 (0.171)	0.426* (0.167)	0.869 (0.499)	0.301 (0.296)	1.204 (0.572)	2.398 (2.597)	0.000*** (0.000)
Corrections	1.687*** (0.152)	0.734 (0.144)	0.817 (0.152)	1.263 (0.173)	0.478 (0.282)	0.441 (0.314)	0.179 (0.181)	0.000*** (0.000)	0.410 (0.238)
Instructional Development									
Lesson Cycle	0.837 (0.082)	1.925*** (0.235)	1.029 (0.132)	1.168 (0.153)	2.134** (0.592)	0.883 (0.320)	1.128 (0.349)	0.149* (0.140)	0.470 (0.199)
Lesson Connections	1.239** (0.092)	2.251*** (0.286)	0.857 (0.161)	0.974 (0.176)	1.217 (0.400)	1.791 (0.837)	1.570 (0.606)	17.356*** (13.368)	0.141 (0.141)
Student Comprehension	0.878 (0.066)	1.013 (0.034)	1.482*** (0.155)	1.384** (0.171)	3.131*** (0.807)	6.281*** (2.087)	2.398*** (0.624)	3.718 (2.798)	0.636 (0.180)
Lesson Delivery	1.010 (0.088)	1.363** (0.145)	1.158 (0.121)	3.114*** (0.423)	1.101 (0.278)	1.307 (0.378)	1.556 (0.418)	1.716 (0.997)	0.380** (0.122)
No criticism	0.310*** (0.029)	0.234*** (0.036)	0.041*** (0.012)	0.085*** (0.029)	0.056*** (0.043)	0.222** (0.113)	0.156*** (0.072)	0.870 (0.792)	6.813*** (1.174)
N	2176	2176	2176	2176	2176	2176	2176	2176	2176

Note: Risk ratios, with heteroskedasticity-robust standard errors in parenthesis. Each column shows results from separate models estimated via generalized linear where the dependent variable is a binary indicator for whether supervisor feedback contains the area for improvement in the column header.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Estimating Best Linear Unbiased Predictors (BLUPs)

Several of our analyses rely on Best Linear Unbiased Predictors (BLUPs), which are estimated from the variance decomposition model below. This model predicts evaluation scores (standardized within year) while accounting for random intercepts for each PST, placement school, and supervisor:

$$\text{StandardizedEvaluationScore}_{isp} = \theta_i + \theta_s + \theta_p + \varepsilon_{isp}$$

In this equation, θ_i represents the proportion of the variance in the standardized evaluation score attributed to PST i , θ_s represents the proportion of the variance explained by supervisor s , and θ_p represents the proportion of the variance explained by placement school p .

Conceptually, these BLUPs capture the “contribution” of an individual supervisor, school, or PST to a PST’s observation score, akin to the interpretation of teacher random effects in a value-added model as a teacher’s contribution to student test scores. After estimation, the BLUPs are rescaled at each level to be expressed in standard deviation units of the supervisor, school, and PST, respectively. This allows us to interpret a Supervisor BLUP of one as a supervisor that gives evaluation scores that are one standard deviation higher than the average supervisor (i.e., is more lenient).