EVIDENCE FOR TEACHER EDUCATION

# SeeMeTeach Teacher Education Program Impact Report

Study Purpose: What high-resolution classroom evidence reveals about candidate development after methods coursework and field experience

## Framing the Question

Teacher preparation programs are predicated on the belief that coursework and clinical experiences both contribute meaningfully to a candidate's professional growth, expanding their knowledge and refining their teaching skills over time. However, a critical question remains:

***Do these programs actually improve how future teachers teach?***

To probe this question, we might consider the following:

- What do higher-resolution indicators provide to observers and teachers beyond the standard measures currently used in most teacher education programs?
- What is the impact of teacher education classes and clinical experiences on future teachers' performance in the classroom? Specifically, how has the essential core of teaching been modified to better align with their teaching goals?

To begin answering these questions, this chapter presents:

1. An analysis of teaching behaviors among a cohort of 12 preservice teachers following two semesters of coursework and concurrent clinical experience.
2. A closer examination of individual teacher development, including a comparative analysis of two participants, to determine whether the data yielded sufficient granularity to distinguish meaningful differences in teacher and student interactions.

When using high-resolution indicators to track growth (or lack thereof), the findings revealed surprising variability in skill development. The author's attention was drawn to a subset of candidates whose teaching showed minimal alignment with the intended outcomes of the program's activities. This unexpected result prompts a closer examination of the data that follows.

## Executive Summary

This report examines whether SeeMeTeach can provide teacher educators with convincing evidence about the development of teacher candidates' instructional practice. Twelve future middle and secondary science teachers taught the same 12- to 15-minute inquiry lesson in a controlled mixed reality simulation after completing two science methods courses and a semester-long clinical experience. Researchers video-recorded and analyzed their lessons with SeeMeTeach.

The evidence was sufficiently detailed to show both program-level patterns and meaningful differences among individuals. It revealed areas of clear development, areas of incomplete development, and specific instructional practices that required further attention.

| 12 | 4 / 5 / 3 | 22% vs 66% | < 3.5 sec |
|---|---|---|---|
| teacher candidates | top middle and lowest tiers | student response time A vs B | wait time remained insufficient |

## Central Finding

SeeMeTeach provided actionable evidence that conventional observation summaries would be unlikely to capture. It showed what candidates actually did when questioning and responding to students, how those actions affected student participation, which candidates demonstrated the intended inquiry-oriented interaction pattern, and where program instruction had not yet translated into practice.

## Responsible Interpretation

These results strongly indicate candidate performance after program experiences. A baseline-to-exit design using repeated SeeMeTeach observations would provide stronger evidence of program impact over time.

STUDY DESIGN

# What Was Examined

Research Setting: Teacher candidates taught the lesson using five avatar students from the TeachLive simulated classroom. This controlled setting reduced uncontrolled variables compared with a classroom setting and allowed the program to compare the same instructional skills under similar conditions.

## Program Preparation

The 12 candidates had completed two of three semesters in their teacher-preparation sequence. Their preparation included a three-credit science methods course, a six-credit science methods course, a three-hour daily middle- or high-school clinical experience, and training in using SeeMeTeach to analyze their own teaching.

## Common Teaching Task

Each candidate taught an interactive demonstration about floating and sinking to five avatar students. Candidates were instructed to uncover what students thought and why—not to explain the science content. They were expected to use questioning, responding, and wait time while engaging all five students.

## Evidence Collected

- **Discourse balance:** Teacher talk time compared with student talk time.
- **Question quality: Use of yes-or-no**, recall, speculation, and explanation or justification questions.
- **Response quality: How candidates reacted after students answered, including judgment**, acknowledgment, repetition, clarification, and use of student ideas.
- **Interaction patterns:** Sequences of questions, student responses, and follow-up teacher actions.
- **Student engagement:** Amount, distribution, and quality of participation among the five avatar students.
- **Wait time:** Whether candidates provided enough processing time after questions and student responses.

## What Would Count as Program Success

| Desired Candidate Practice | Expected Evidence |
|---|---|
| Elicit student thinking | More student talk and less teacher talk |
| Ask cognitively demanding questions | Greater use of speculation and explanation questions |
| Probe students' reasoning | Acknowledgment followed by requests to clarify or elaborate |
| Engage all learners | Frequent and evenly distributed student responses |
| Support deeper thinking | Wait time of approximately 3.5 seconds or longer |

GROUP RESULTS

# What the Group Data Revealed

The group averages showed substantial student participation, but the ranges exposed uneven enactment of the practices emphasized by the program.

## Teacher and Student Talk

| Measure | Average | Range |
|---|---|---|
| Teacher talk time | 69.12% | 62–78% |
| Student talk time | 28.06% | 21–35% |
| Teacher reaction to misbehavior | 2.80% | 2–5% |

The 16 percentage-point range in teacher talk directly affected opportunities for students to voice ideas, predictions, and reasoning. Candidates with lower teacher-talk percentages created more room for student thinking to become visible.

## How Teacher Talk Was Used

| Teacher Action | Average | Range |
|---|---|---|
| Presenting information | 17.14% | 7.5–38% |
| Giving directions | 10.84% | 0.29–35% |
| Questions and responses | 41.40% | 29.25–57% |

The 29.25% to 57% range in questioning and responding was especially important. It showed that some candidates spent much more of the lesson interacting with students to elicit ideas, whereas others relied more heavily on presenting information or directing activity.

## Question Types

| Question Type | Share of Questions | Average Count | Range of Lesson Time |
|---|---|---|---|
| Yes or no or multiple choice | 43.0% | 14 | 0–20% |
| Recall | 16.9% | 5.41 | 0.83–8.70% |
| Speculation or prediction | 32.9% | 10.83 | 3–26.2% |
| Explanation or justification | 7.5% | 2.5 | 0–7.14% |

Yes-or-no questions were used most frequently. The wide 3% to 26.2% range in speculation questions strongly differentiated candidates: some regularly invited prediction and deeper thinking, while others relied primarily on closed questions.

INTERACTION QUALITY

# Did Candidates Uncover Student Thinking

SeeMeTeach revealed not only which questions were asked, but whether candidates followed student responses with actions that made reasoning visible.

## Responses After Students Answered

| Response Type | Average Count | Count Range | Average Time Used |
|---|---|---|---|
| Reject response | 0.17 | 0–1 | 0.05% |
| Acknowledge without judging | 9.00 | 0–22 | 3.59% |
| Confirm response | 6.83 | 0–17 | 2.03% |
| Repeat response | 14.92 | 8–40 | 4.27% |
| Clarify for the student | 2.92 | 0–8 | 2.33% |
| Answer student question | 1.58 | 0–4 | 1.05% |
| Ask student to clarify | 7.91 | 0–18 | 3.39% |
| Use a student idea | 0.08 | 0–1 | 0.08% |

## The Most Important Interaction Pattern

For this inquiry task, the most productive sequence was a speculation or explanation question, followed by a student response, a neutral acknowledgment, and then a request for the student to clarify or elaborate. This sequence enabled the teacher to hear the reasoning behind an answer rather than merely determine whether it was correct.

**4 of 12**
demonstrated the full inquiry pattern

**7 of 12**
demonstrated only part of the pattern

**1 of 12**
used neither key action

## Program Interpretation

Candidates generally avoided responses that directly shut down thinking, such as rejection. However, repetition was the most common response, and many candidates failed to follow a student answer with a request for clarification. Seven of the 12 candidates did not consistently complete the interaction sequence needed to uncover students' reasoning.

## Actionable Program Target

Reduce reliance on repeating or confirming answers. Increase the combined use of higher-level questions, neutral acknowledgment, requests for clarification, and use of student ideas.

ENGAGEMENT AND THINKING TIME

# What Students Were Given Opportunities to Do

## Student Engagement

| 28.06% | 21–35% | 34.45 | 19–56 |
|---|---|---|---|
| average student engagement time | range across lessons | average student answers per lesson | range in student answers |

- **Four lessons:** All five avatar students were highly engaged.
- **Seven lessons:** All five students were engaged at a high or medium level.
- **One lesson:** Four students were highly engaged while one student was not engaged.
- **Across lessons:** Engagement was generally distributed throughout the teaching episode.

Engagement was relatively strong. The evidence also showed that quality mattered: many student responses were brief answers to teacher questions rather than extended explanations. More speculation, justification, and clarification would make engagement more cognitively demanding.

## Wait Time

Wait time was the clearest program-wide weakness. Research cited in the source report identifies approximately 3.5 seconds as the threshold at which positive effects on engagement and critical thinking become more likely. Most pauses were so brief that the observer could not trigger the SeeMeTeach wait-time control. Even when wait-time events were captured, there were often too few events to calculate a representative average.

| Wait Time Indicator | Program Finding |
|---|---|
| After teacher questions | Usually too short or absent |
| After student responses | Usually too short to encourage elaboration |
| With higher-level questions | Inconsistently present across candidates |
| Overall target | All candidates needed more deliberate use of wait time |

## Why This Evidence Matters

A conventional observation might describe a lesson as interactive because students answered frequently. SeeMeTeach distinguished participation from deeper thinking by showing whether students received enough time and follow-up prompting to explain their reasoning.

DEVELOPMENTAL PROFILE

# Three Levels of Candidate Development

The evidence sorted candidates into meaningful developmental groups rather than treating the cohort average as the program outcome.

| Developmental Level | Candidates | Observed Teaching Profile |
|---|---|---|
| Top level | 4 | Used more advanced questions, invited speculation, asked students to elaborate, avoided clarifying for students, and generally avoided judging responses. Wait time remained below the desired level. |
| Middle level | 5 | Used more higher-level questions, acknowledgment, and repetition, but did not consistently ask students to expand or clarify their answers. Wait time was usually insufficient. |
| Lowest level | 3 | Relied mainly on yes or no and recall questions, used more judgment, provided fewer acknowledgments, and missed opportunities for elaboration. Wait time was short or absent. |

## What the Distribution Says About Program Impact

The program had not produced a uniform level of performance by the end of the second semester. Four candidates demonstrated the full range of inquiry-oriented questioning and responding. Five showed partial development, and three displayed little evidence of the intended practices. This distribution is more useful than a single cohort average because it shows how many candidates met the program target and how many need additional support.

## Evidence of Program Strength

- **Observable enactment:** One-third of the cohort demonstrated the complete inquiry-oriented interaction pattern.
- **Broad engagement:** Most candidates engaged all five avatar students at a high or medium level.
- **Low use of counterproductive responses:** Rejection and teacher clarification were relatively uncommon.

## Evidence of Needed Program Improvement

- **Closed questioning remained dominant.** Yes-or-no questions represented 43% of questions.
- **Follow-up probing was inconsistent.** Most candidates did not reliably ask students to clarify their reasoning.
- **Wait time was insufficient across levels.** Even the strongest candidates generally remained below the target.

INDIVIDUAL DIFFERENTIATION

# Teacher A and Teacher B

The clearest demonstration of SeeMeTeach's value was its ability to distinguish two candidates' instructional approaches using specific, low-inference evidence.

### High Contrast Indicators

| Indicator | Teacher A | Teacher B | Interpretation |
|---|---|---|---|
| Student response time | 22% | 66% | Students had three times as much response time with Teacher B. |
| Teacher information time | 15% | 8% | Teacher A spent nearly twice as much time presenting information. |
| Directions | 9.3% | 0% | Teacher A devoted substantial time to directing activity. |
| Speculation questions | 0% | 2.7% | Only Teacher B asked questions requiring speculation. |
| Speculation plus justification | 0% | 0.6% | Only Teacher B asked for both a prediction and its reasoning. |
| Confirming student answers | 3.4% | 0.6% | Teacher A confirmed answers more than five times as often. |
| Overall teacher talk | 67% | 30% | Teacher A's lesson was teacher-centered; Teacher B's was student-centered. |
| Overall student talk | 33% | 70% | The discourse balance was nearly reversed. |

### Predominant Discourse Patterns

| Teacher A | Teacher B |
|---|---|
| Short-answer question → student response → teacher repeats or confirms | Yes or no question → student response → another question or neutral response |
| No speculation or justification questions recorded | Some speculation questions followed by acknowledgment |
| Pattern kept the teacher at the center of the exchange | Pattern created substantially more student participation |

### Why This Comparison Is Convincing

Both candidates completed the same program experiences and taught the same lesson in the same controlled simulation. Yet their discourse balance, question types, response patterns, and student participation differed markedly. SeeMeTeach made those differences visible and produced specific improvement targets for each candidate.

PROGRAM USE

# How Teacher Educators Can Use This Evidence

The report demonstrates that SeeMeTeach can support candidate coaching and program evaluation at the same time.

## For Individual Candidate Development

- **Diagnose a teaching pattern.** Identify the questions, responses, wait time, and discourse balance that make up the candidate's teaching fingerprint.
- **Set a precise improvement target.** Replace a general goal such as "be more student-centered" with an observable action such as increasing clarification prompts after student responses.
- **Document growth.** Repeat the same or a comparable observation and compare the candidate's evidence over time.

## For Program Improvement

- **Assess intended outcomes.** Determine how many candidates demonstrate the interaction practices emphasized in methods courses and clinical work.
- **Locate curriculum gaps.** Use cohort patterns to identify practices that require stronger modeling, rehearsal, feedback, or field-based application.
- **Differentiate support.** Provide intensive intervention for candidates in the lowest developmental tier while extending the practice for stronger candidates.
- **Supply accreditation evidence.** Aggregate low-inference indicators across candidates, cohorts, courses, and clinical placements.

## Recommended Evidence Design

| Collection Point | Purpose |
|---|---|
| Program entry baseline | Establish candidates' initial questioning, responding, engagement, and wait-time profile. |
| After first methods course | Assess early acquisition of targeted interaction practices. |
| During clinical experience | Examine transfer from coursework into authentic teaching. |
| Before program completion | Document growth, remaining gaps, and readiness for independent practice. |

## Bottom Line

SeeMeTeach gave the teacher-education program evidence that was specific enough to evaluate the cohort, differentiate individual candidates, and identify the next instructional priorities. Its greatest value is not simply counting classroom events; it is connecting teacher actions to student participation and making candidate development visible over time.