

Assessing Classroom Performances of Student Teachers: Articulating Teaching Quality to Strengthen an Educator Preparation Program

Laura Turchi

Margaret Kuczynski

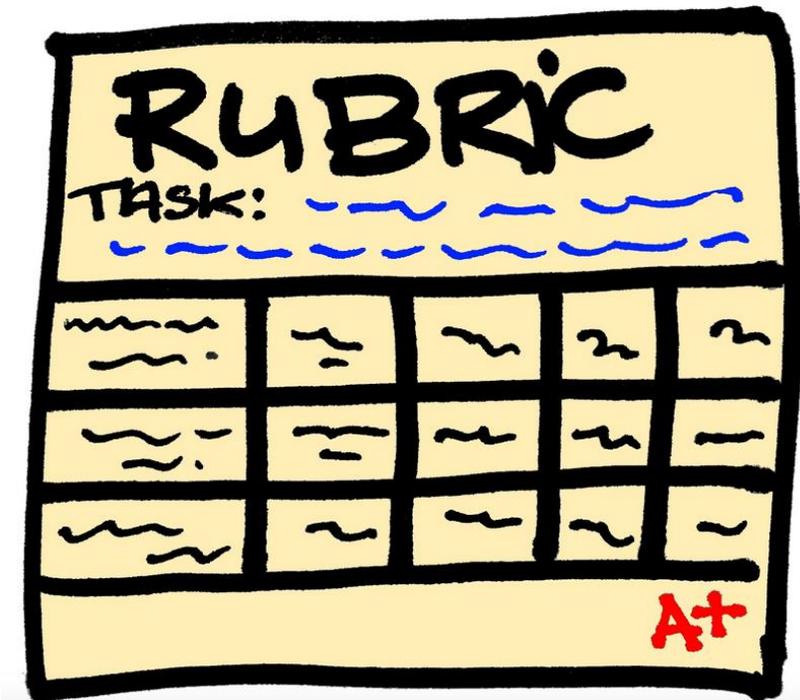
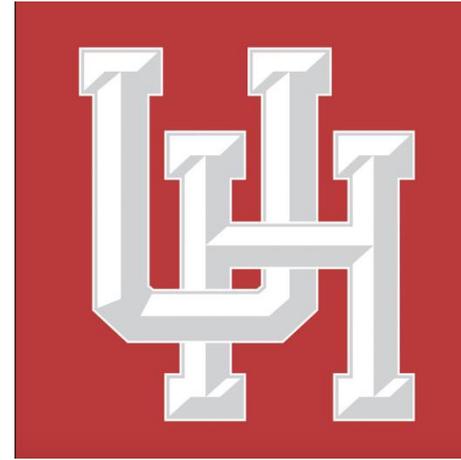
Misty Black

Jane Cooper

Amber Thompson

Overview of speakers/topics

- Turchi – Faculty research on programmatic use of the performance assessment rubric
- Kuczynski – Statistical sanity from Institutional Effectiveness
- Black – Graduate student research
- Thompson – Associate Chair, Teacher Preparation
- Cooper – Site Coordinator



Purpose of our presentation

Peck et al (2016) argue that teacher education is missing rigorous program quality indicators. Without such information, they claim, educator preparation programs are “driving blind.”

If we (teacher preparation programs) are going to be *data* driven, to what extent are student teaching performance assessment scores the data we should be using? Do the scores (and the scoring process) serve our students as well as our program as a whole?

The University of Houston (institutional context)

- Very diverse student body serving very diverse schools
- Student teaching is “year long” (two semesters)
- Founding partner in USPREP (more on that to come – *the role of the site coordinator*)
- New rubrics and new rubric practices (almost) every semester

STUDENT SPOTLIGHT



CUIN doctoral student Saira Siddiqui is working to shine a light on the importance of using Muslim narratives in children's books. She hopes to help curb the rise of bullying and to give Muslim children a sense of belonging.

[Learn more »](#)

DEAN HONORED



Robert McPherson, dean of the University of Houston College of Education, has been selected to help represent his peers at high-research institutions across North America.

[Learn more »](#)

ALUMNI SPOTLIGHT



Teaching & Learning alumna Bethaney Jones received the Rookie Teacher of the Year Award at Dulles Elementary School in December 2016 – just seven months after graduating from the University of Houston College of Education.

[Learn more »](#)

CAMPAIGN LAUNCH



The University of Houston has launched a \$1 billion capital campaign. Chancellor Renu Khator explains: “It’s an ambitious campaign, but when the University of Houston dreams, we always dream big.”

[Learn more »](#)

Relevant Research –where teacher educators could have more voice...

- **The ubiquitous “A”** - Conderman, G., Morin, J., & Stephens, J. T. (2005); Rodgers, A. & Keil, V.L. (2007); Ryan, W. J. (1969).
- **Research protocols are not very helpful** – e.g. Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016).
- **(Practicing) teacher evaluation is not the same thing** - Lash, A., Tran, L., and Huang, M. (2016); Steinberg, M. P., & Garrett, R. (2016).
- **Rubrics are not unproblematic** - Cohen, J. & Grossman, P. (2016); Hill, H. & Grossman, P. (2013); Kennedy, M. M., & American Association of Colleges for Teacher Education (2010).
- Policymakers say: **process should be good for the individual student teacher *and* good for program evaluation** Darling-Hammond (2013); Papay, J. (2012).
- **Buckets of data** Meyer, S. J. (2016); Wineburg, M. S. (2006).

Uses of the TAP Rubric in the Role of the Site Coordinator

Major Themes Addressed

- The role of the Site Coordinator at UH
- How the UH/Modified TAP rubric is used to assess student teachers
- How instruction in class is changed in relation to the rubric
- How the rubric affects communication of best practices

Trends in the Literature

- Strong connection between theory and practice can have positive affect on programs of education (Cochran-Smith, 2009 & Zeichner & Gore, 1990)
- Common assessment can be a bridge between theory and practice (Baumgartner, Koerner & Rust, 2002)
- Rubrics can be come a communication tool to improve community relations → teacher retention (Abdal-Haqq, 1998)

TAP Rubric and Coaching

Changed and structured what we were looking for.

- Pre-Conference
- Post-Conference
 - What questions we want to ask students
 - Explicit connections in response
 - What we felt were best practices before we saw the rubric

TAP Rubric and Coaching

- Something not addressed on the rubric?
- How are site coordinators addressing these?
 - Using a 'close' indicator (are they the same across program)
 - Marking a student lower or higher due to context
 - Using the indicators as a reflection of effort

TAP Rubric and The Student Teaching Class

- Planned curriculum
 - Teaching indicators of the TAP rubric through modules
 - Process of choosing indicators and implications
 - Investigations into the language and the best practices to achieve better scores
 - Use of video to investigate, document and reflect on the indicators

Missing the Point(s)

- Coaching vs. Assessment
- Scale of 1-5, what a 3 means
- Grading

TAP Rubric and The Student Teaching Class

- Frames “just in time teaching”
 - Contextual issues
 - Due to small groupings, allows the rubric to be responsive.
 - Gives language for ‘hard’ conversations when cooperating teacher practice does not align with TAP best practices
- What the students said:
 - 76 % : Analyzing the indicators was effective – extremely effective
 - 26 %: found this task extremely effective
 - Discussions of the TAP Rubric were the second most valued item discussed in ST class, behind issues in your practice.
 - “I found analyzing the TAP rubric was very helpful. As the year went on I became more and more familiar of what was expected of me.”

TAP Rubric and Communication

- All assessments are really a means of communication
 - Communicate within our coaching pairs
 - Within our academic community
 - With our district partners
 - With external bodies

TAP Rubric and Communication

- Communication in Coaching Pairs
 - Aligned conversation directly to success
 - Pre and Post observation conferences
 - Feedback from students survey:
 - 74 % : pre-conferences were of good or excellent quality.
 - 71% : post-conference were of were of good or excellent quality.

TAP Rubric and Communication

- Within our academic community – and program
 - The conversations between site coordinators
 - Conversations about student success initiatives
 - Co-scoring student teaching videos with methods faculty
 - Using data can enhance teacher preparation efforts (Cochran-Smith, 2009)

TAP Rubric and Communication

- With our community partners
 - Governance meetings – with **school and district administration**
 - Have **Cooperating Teachers** participate in the post-conference
 - Framing issues and concerns in the data
 - Having ‘neutral’ conversations

Summary – TAP Rubric and Site Coordinators

- Coaching
- Student teaching class
- Communication
- The TAP rubric and other standards go far beyond shaping what one looks for, but “...systematically guide the organization of different pieces of the field in a way that enable better practice” (Mehta, 2012).

Program Practices	Performance Assessment features
Practice 1 (Spring 2015)	<p>The Performance Assessment Rubric was created by the TEP (aligned with state standards) and included 10 categories for practice, each scored on a 10-point scale (9-10 = A, 8 = B, etc.). Each supervisor evaluation contributed directly to the student teacher’s final grade. Any points “deducted” on an evaluation form were subtracted from a 100-point course total. The rubric uses heading categories of “emerging,” “developing,” and “proficient” with point totals attached.</p>
Practice 2 (Fall 2015)	<p>Supervisors used the same performance assessment instrument (rubric of indicators and descriptors), but instead of using points for each observation, supervisors were asked to give detailed feedback using the rubric and to use all three categories - “emerging,” “developing,” and “proficient” as appropriate to indicate the student teacher’s performance. Student teachers received holistic grades at the mid-term and end of the semester, rather than accumulating a point total. Student teaching supervisors were given opportunities to provide constructive and specific feedback - even when it was negative – without negative consequences to a student teacher’s grade.</p>
Practice 3 (Spring 2016)	<p>[UH joins USPREP] Almost all supervisors receive a three-day training on the TAP Systems rubric; none were required to take the official TAP certification assessment.</p> <p>Nine indicators from the TAP Systems rubric were used; final grades continued to be determined holistically by supervisors. The modified TAP rubric used indicators from Planning and Professional areas as well as instructional dimensions (note that TAP training and assessment inter-rater reliability is only on the instructional dimensions: others are intended to be measured over time within a school setting by multiple evaluators)</p>

Program Practices	Performance Assessment features
Practice 4 (Fall 2016; Spring 2017)	Most site coordinators TAP certified; 6 indicators; rubric scores part of a holistic student teaching grade; 3 Performance Assessment Cycles
Practice 5 (Fall 2017) Where we've landed	All site coordinators TAP trained and certified; 6 indicators; rubric scores translated into specific points for student teaching grades; 2 Performance Assessment Cycles

Next up: Messy Data, Messy Results

How much should the program invest in more efficient data capture – what would we gain as a result?

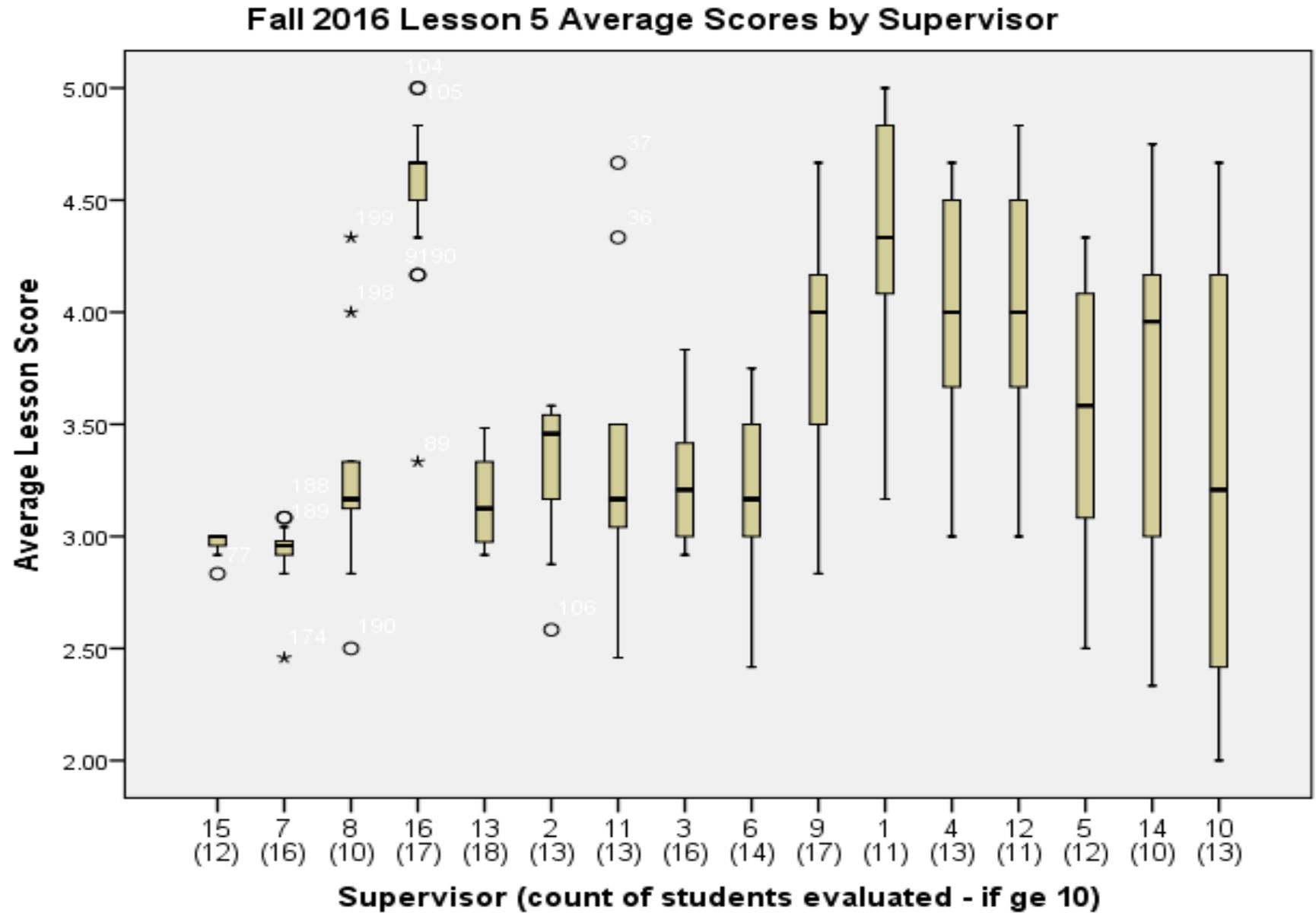
Data Collection and Cleaning

- Data collection
- Data storage
- Data cleaning decisions
- Methods of analyses

Reliability

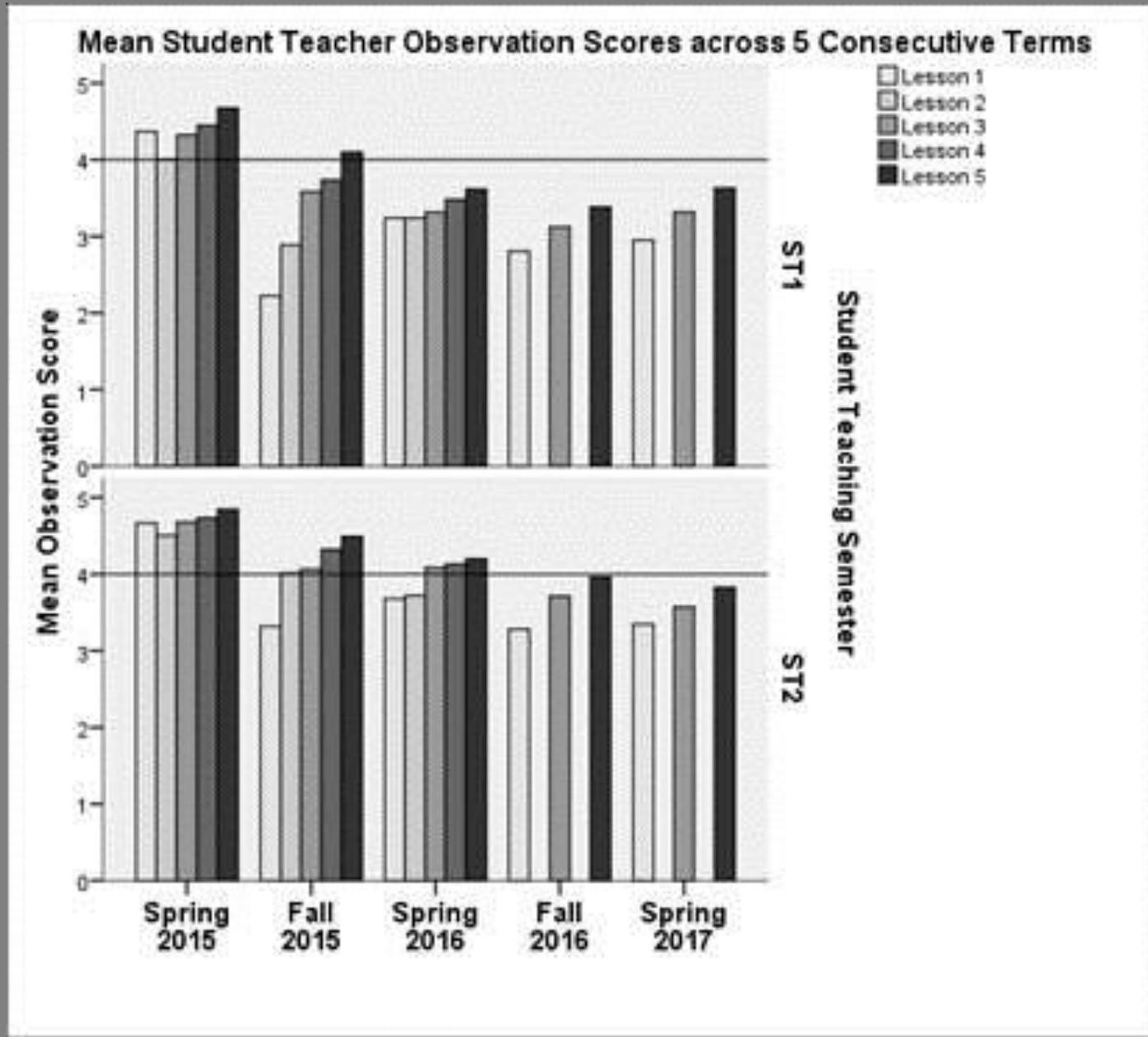
US Prep confirms that the TAP rubric is valid and reliable.
But we must show reliability locally.

To what extent do rubric scores provide evidence of consistency of student teacher expectations?



Inevitable
growth

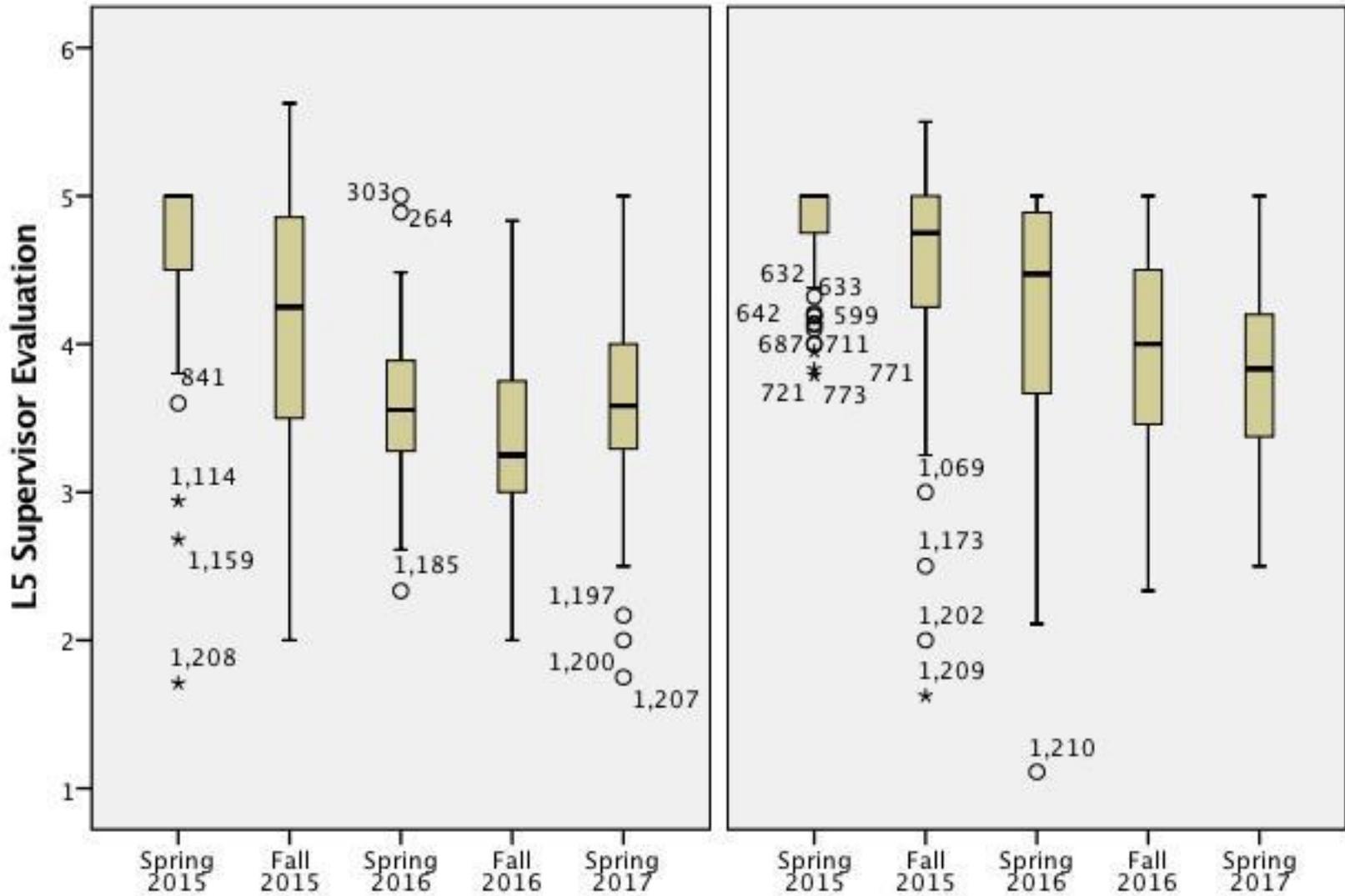
Effects of
changing
rubric?
Effects of
Training?



Student Teaching Semester

ST1

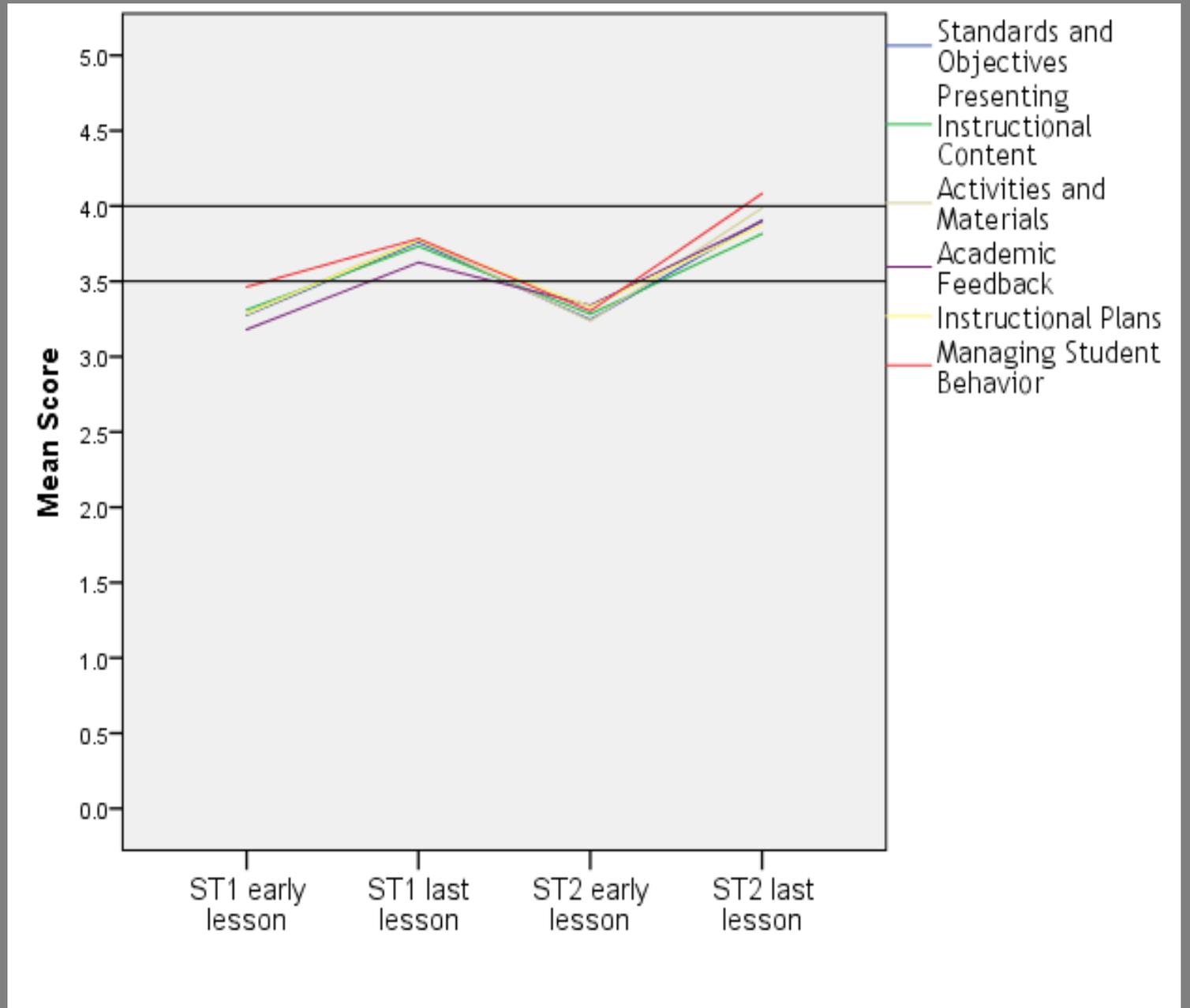
ST2



Once the supervisors were trained to use the rubric differently, did they use it differently? Otherwise said: Do the distributions of scores differ over the practices?

Is there a difference between the first observation's total score in ST1 and the last observation's total score in ST2 for the same individuals under the same instrument?

Is there growth between the first lesson in ST1 and the first lesson in ST2? If not, is the instrument a reflection of where the student's skills lie on a continuum? Similarly, is there growth between the last lesson in ST1 and the last lesson in ST2?



Growth Stories Differ by Certification Area

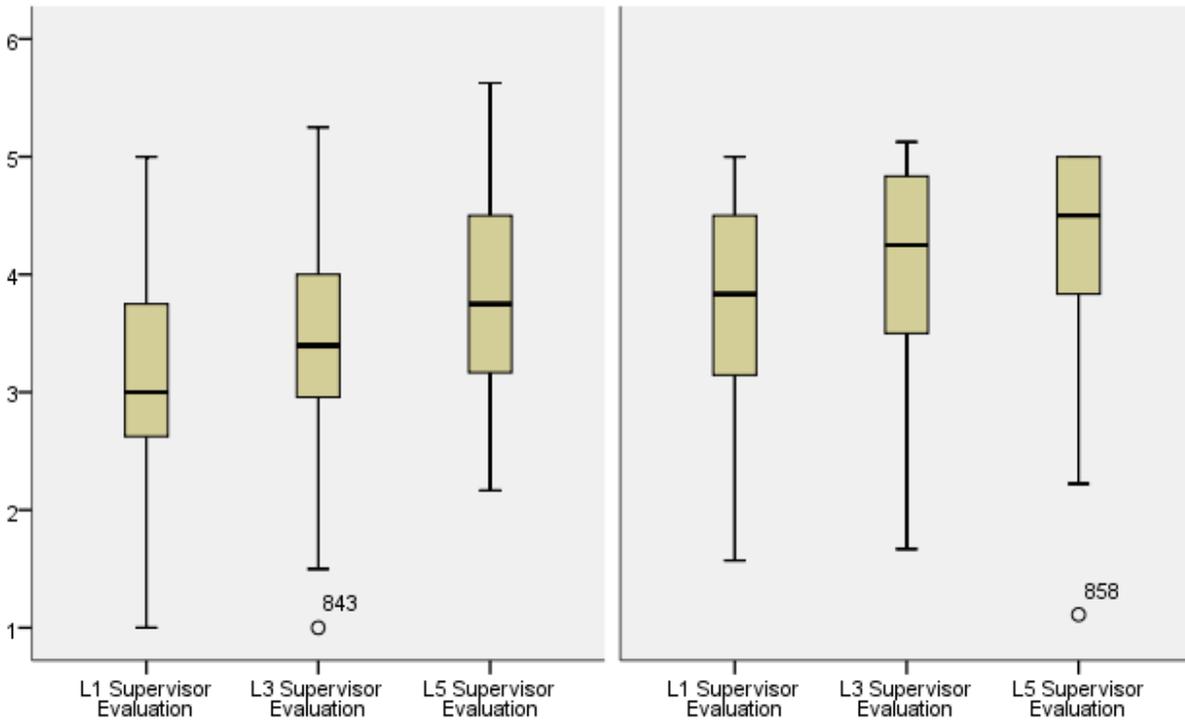
Data over Five Semesters - Spring 2015 through Spring 2017

Certification Area: Generalist EC-6

Student Teaching Semester

ST1

ST2



Student counts ST1 n=307, ST2 n=305 (however 124 evaluations are missing)

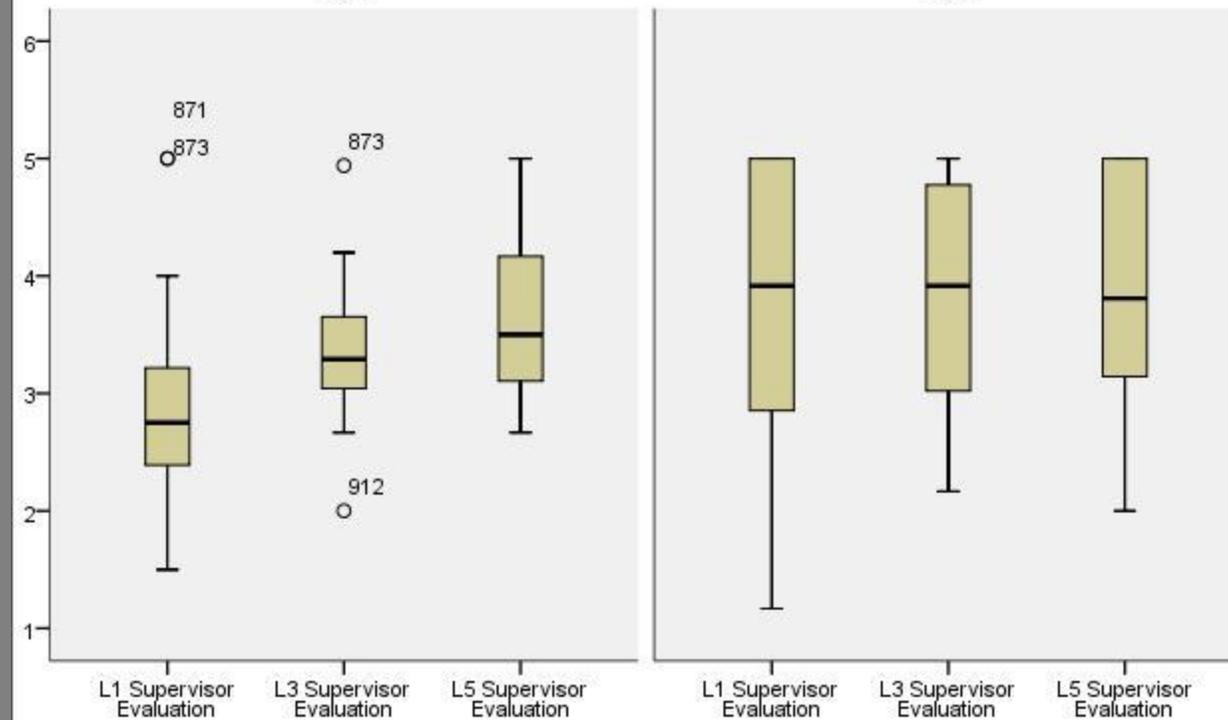
Data over Five Semesters - Spring 2015 through Spring 2017

Certification Area: History 7-12

Student Teaching Semester

ST1

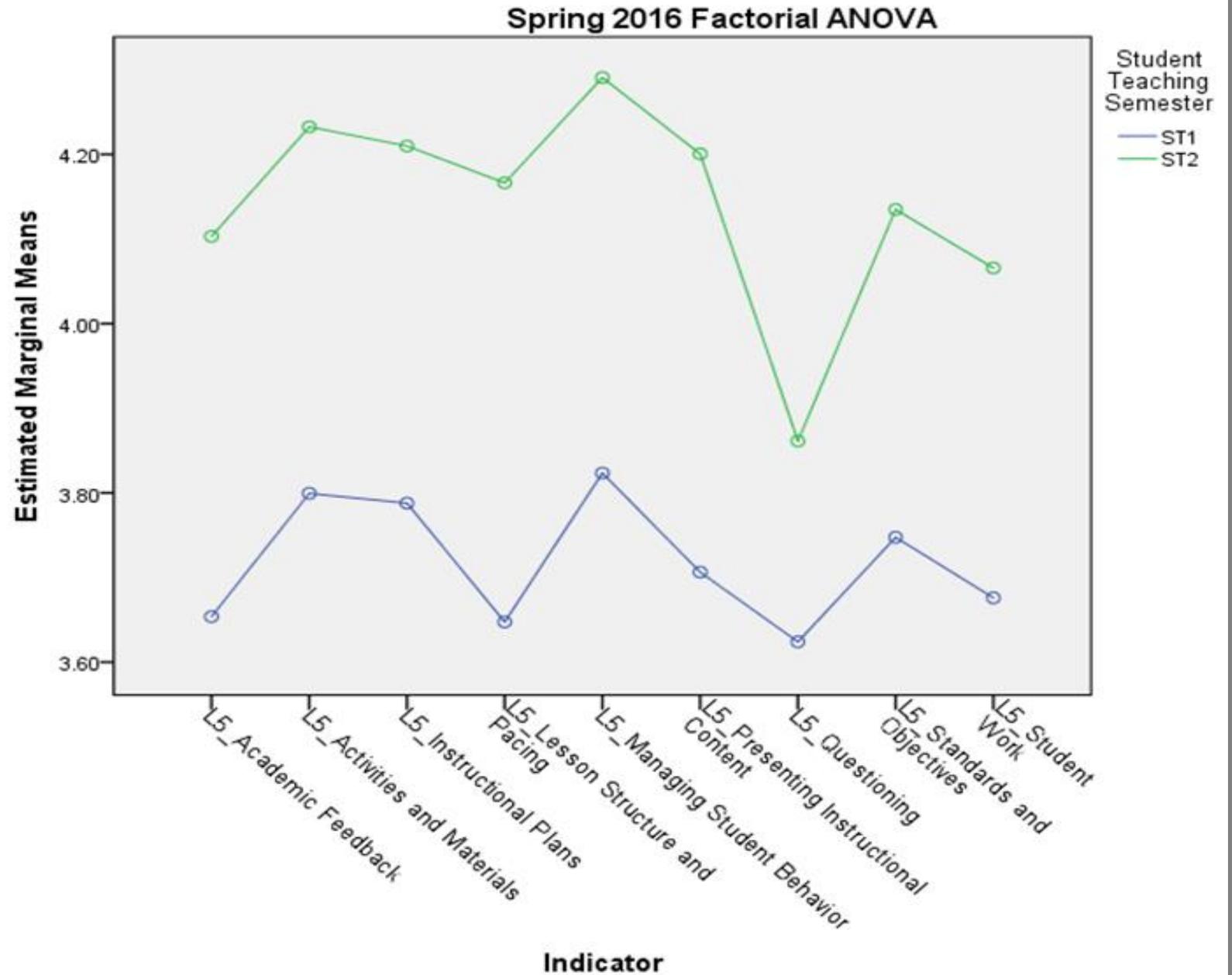
ST2



Student counts ST1 n=26, ST2 n=27 (however 9 evaluations are missing)

Can scores tell us anything about whether program values are promoted in teacher education coursework?

Is there a significant difference between ST1 and ST2?



Supporting Student Teacher growth with Site Coordinators

Comparing Approaches on the first semester of implementation -
Student Scores from Site Coordinators and Student Scores from
Supervisors

Qualitative data across program practices – emergent themes related to scoring and grading

GROWTH Student teachers expect and are expected to “grow” – to be observed, to receive feedback, to be able to improve, and to be praised or at least acknowledged for that growth.

CONSISTENCY Student teachers and supervisors look to the instrument and to the teacher education program to provide consistency of expectations

INTEGRATION Student teachers expect there to be an integration between their program coursework and their field-based learning. Supervisors want this too, and they have opinions about what really matters

GROWTH – where grades and scores may make the most noise

- Because of the changes represented by Practice 2, we expected this group of student teachers would receive more constructive/critical feedback from supervisors under the “cover” of a more holistic grade, and they did.
- Comparing the observation scores from the previous semester (Practice 1) allows us to see that the average score across all formal observations of the Spring 2015 cohort was 9.2 (sd, 1.1); the average score across all formal observations of student teachers in the Fall 2015 cohort averaged 7.3 (sd, 2.8).
- The Practice 1 cohort received consistently high scores each time they were observed by their supervisors, but the Practice 2 cohort found the scores and feedback to be more valuable.

CONSISTENCY – DO student teachers think that our requirements are authentic?

Practice 2 Student Teachers

- Wanted the same expectations/assessments in methods and in student teaching observations (e.g. lesson planning)
- Wanted content-area specific strategies to be reflected in the rubric . Especially student teachers working in early childhood education and those training to work with special populations.

POSSIBLE – As they looked more closely at the rubric, they were less patient with descriptors of “proficient’ classroom practice that did not seem to apply to their context.

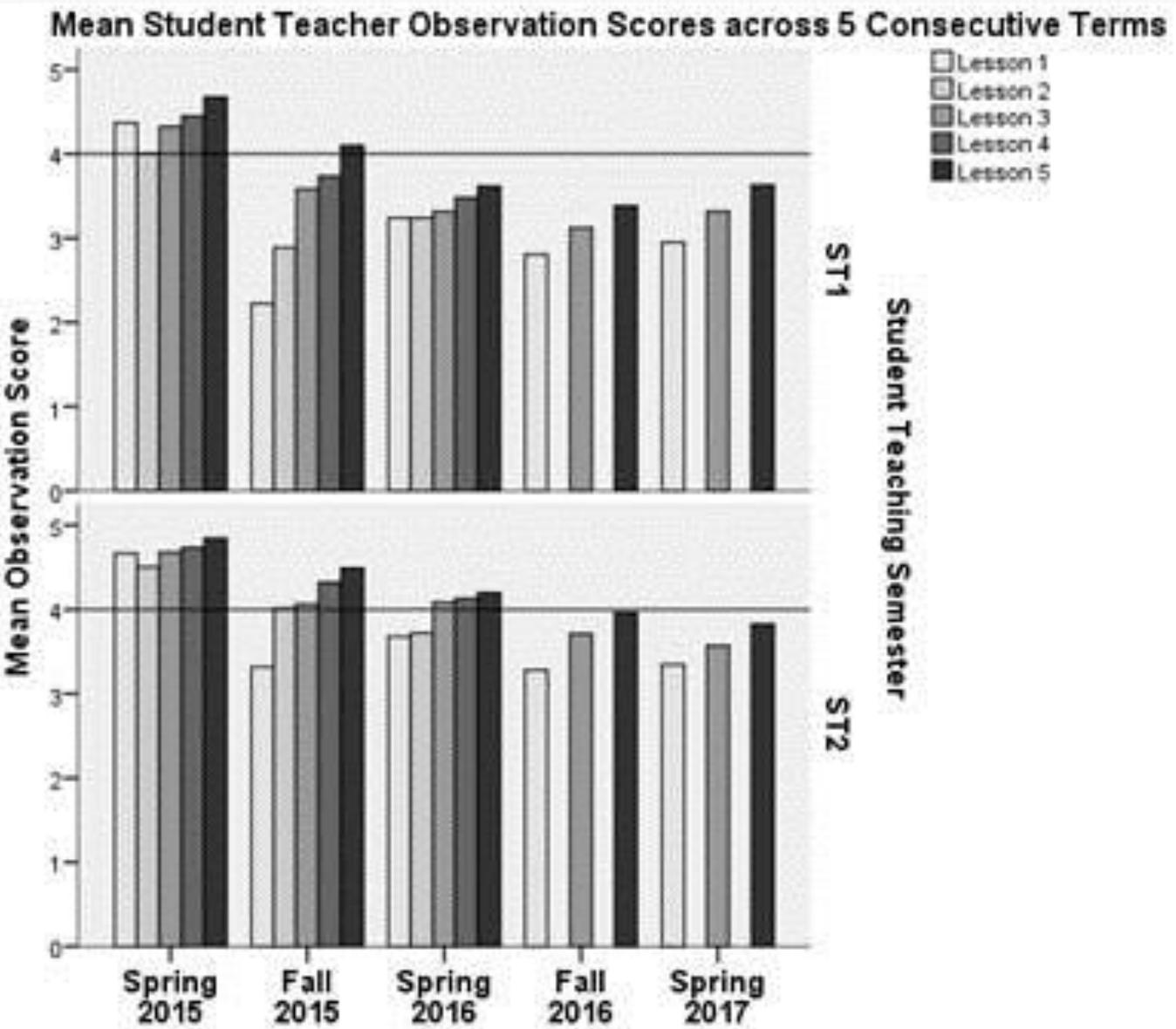
INTEGRATION –

Can TAP provide a common language?

In our survey the supervisors were concerned that the rigorous way they were expected to use the TAP rubric made it a challenge to “account” for the co-teaching of the cooperating teachers, who in sharing instruction might feel themselves to be assessed “harshly” by the supervisor.

On the other side, supervisors said they struggled to use some of the indicators on the performance assessment instrument effectively when the Cooperating Teacher was in effect controlling the classroom. Similar concerns were raised about how to judge a co-created lesson plan (and co-delivered) by the student teacher and the cooperating teacher.

What do student teacher performance assessment scores tell us about our program's quality?



Final Thoughts and Questions