



USING CLASSROOM OBSERVATION RESEARCH TO INFORM DEBATES ABOUT TEACHING EFFECTIVENESS

Candace Walkington

**Wisconsin Center for Educational Research
University of Wisconsin – Madison**

Matthew Valerius

University of Minnesota

UTOP PROJECT

Development and piloting of a classroom observation instrument (**UTeach Observation Protocol**) to measure characteristics of effective teaching in mathematics and science classrooms

Collaborators: Michael Marder, Mary Walker, Larry Abraham, Denise Ekberg, Gail Dickinson, & Kelli Allen, UTeach Natural Sciences, University of Texas at Austin
GRAs: Prerna Arora, Shasta Ihorn, & Jessica Gordon



OUTLINE

- Context for Work
- Background of UTOP and UTeach
- Pilot study of UTeach graduates
- Large-scale study of 994 classrooms in partnership with MET project



MEASURING EFFECTIVE TEACHING

- What does effective teaching look like when it happens?

“Documenting particular features of teaching that are consistently effective for students’ learning has proven to be one of the greatest research challenges in education” (Hiebert & Grouws, 2007)



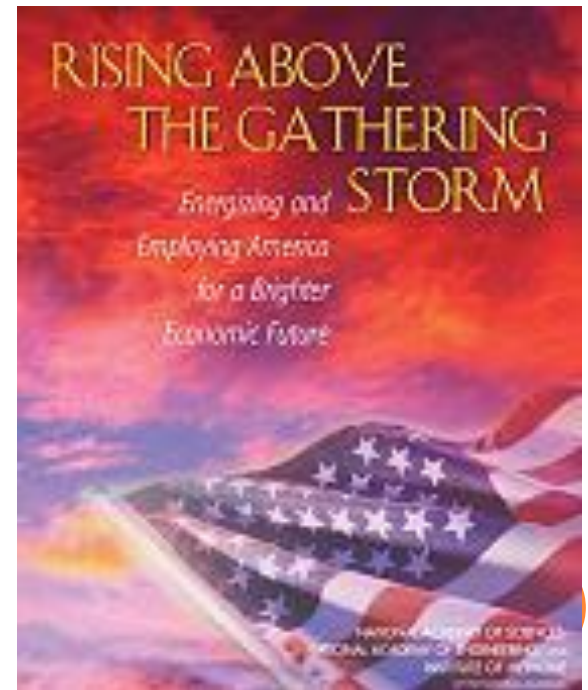
MEASURING EFFECTIVE TEACHING

- What does effective teaching look like when it happens?
- Can classroom observers be trained to make key distinctions in effective teaching practices?
- Can the skills involved with being an effective teacher be successfully trained through a teacher preparation program?



CONTEXT OF WORK

- Measurement of teacher quality lies at the heart of current debates about educational reform
 - Challenges particularly severe in secondary mathematics and science – critical to debate on US competitiveness
 - Teacher preparation under scrutiny



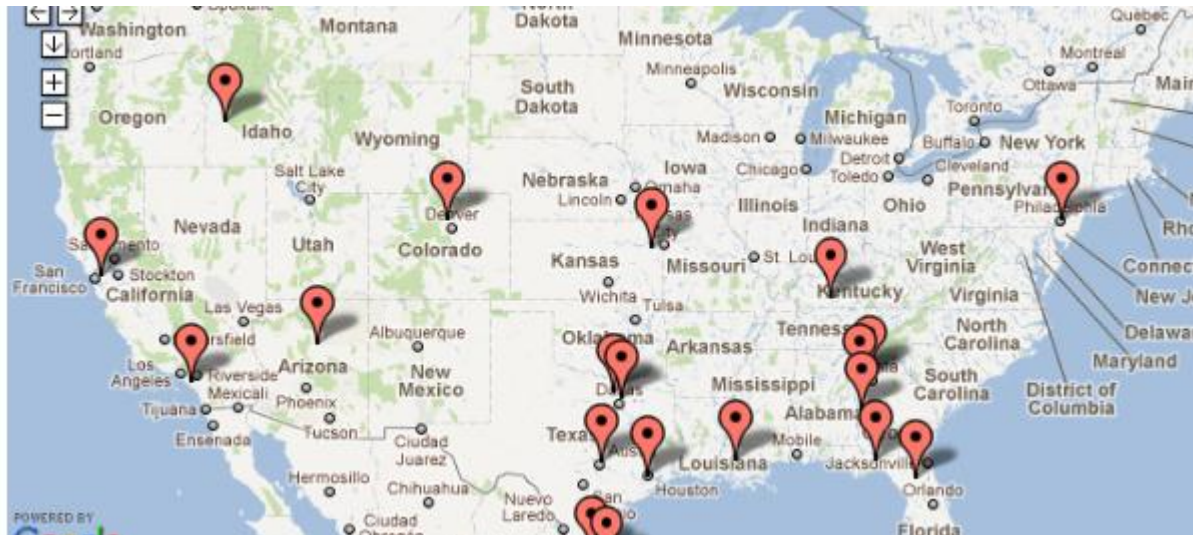
CONTEXT OF WORK

- Initiatives like *Race to the Top* emphasize measuring teacher quality through student standardized test score gains
 - Unclear if tests measure all outcomes of education that we care about
 - Questions about reliability (Baker et al., 2010)
 - Not all subjects assessed
 - Value-added gains “black box” that does little to help us understand good teaching



THE UTEACH PROGRAM

- Steady increase in number of students with strong STEM backgrounds going into teaching
- Replicated at **28** universities in **13** states
- **92%** of graduates go into teaching, **82%** remain 5 years later (compared to **65%** nationally)



SOME KEY FEATURES OF UTEACH PHILOSOPHY

- Organized, well-managed, on-task classroom
- Attention to issues of diversity and access
- Incorporating inquiry/investigative learning
- Using technology for teaching and learning
- Fluid and accurate communication of content
- Fostering student-student collaboration
- Formative assessment of student progress
- Applications and inter-disciplinary connections
- Critical practices of self-reflection
- Facilitating classroom discussion and “student talk”

Research in
Education;
NSES, NRC,
NCTM
Standards



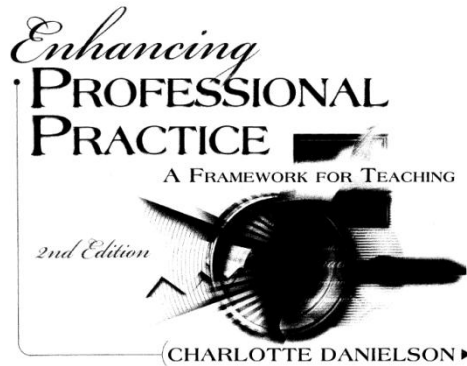
BACKGROUND OF PROJECT

- Persistent requests to evaluate **UTeach Graduates**
- UTeach boosts recruitment and retention, but are UTeach graduates effective teachers?
- Look towards classroom observation



CLASSROOM OBSERVATION PROTOCOLS

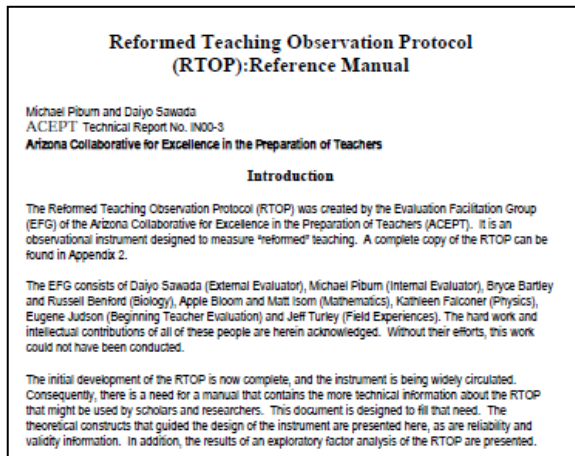
Charlotte Danielson's Framework



CLASS

Shulman (1986): What about content knowledge, and PCK?

RTOP



COP/LSC Protocol

A Study of K-12 Mathematics and Science Education in the United States



DESCRIPTION OF UTOP

- Modified Horizon Research Inc.'s COP (*Inside the Classroom Study*)
 - Tailored to UTeach vision
 - No published indicator or synthesis-level reliability
 - No scoring rubrics



DESCRIPTION OF UTOP

- Full version has 32 indicators (teaching behaviors) in 4 sections
 - Classroom Environment
 - Lesson Structure
 - Implementation
 - Mathematics/Science Content
- 1-5 scale, DK/NA options
- Section Synthesis Ratings

III. RATING SCALES

1 Classroom Environment

Rating Scale: 1= Not observed at all, 2= Observed rarely/ Demonstrated poorly, 3= Observed an adequate amount/ Demonstrated adequately, 4= Observed often/ Demonstrated well, 5= Observed to a great extent/ Demonstrated to a great extent

Rating	Indicator	Evidence
	1.1 The classroom environment encouraged students to generate ideas, questions, conjectures, and/or propositions that reflected engagement or exploration with important mathematics and science concepts.	1.2 Interactions reflected collegial working relationships among students. (e.g. students worked together productively and talked with each other about the lesson). *It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.
	1.3 Based on conversations, interactions with the teacher, and/or work samples, students were intellectually engaged with important ideas relevant to the focus of the lesson.	1.4 The majority of students were on task throughout the class.
	1.4 The teacher's classroom management strategies enhanced the classroom environment.	1.5 The classroom is organized appropriately such that students can work in groups easily, get to lab materials as needed, teacher can move to each student of student group, etc.
	1.6 The classroom environment established by the teacher reflected attention to issues of access, equity, and diversity for students (e.g. cooperative learning, language-appropriate strategies and materials).	1.7 The classroom environment established by the teacher reflected attention to issues of access, equity, and diversity for students (e.g. cooperative learning, language-appropriate strategies and materials).

Classroom culture is non-interactive or non-productive	Classroom culture is interactive only	Classroom culture is adequately productive and interactive	Classroom culture is often productive and interactive, with some collegial interactions	Classroom culture is consistently collegial, interactive, and productive
1	2	3	4	5

UTOP AND ONLINE MANUAL

Rating	Indicator
	<p>1.1 The classroom environment encouraged students to generate ideas, questions, conjectures, and/or propositions that reflected engagement or exploration with important mathematics and science concepts.</p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.2 Interactions reflected collegial working relationships among students. (e.g. students worked together productively and talked with each other about the lesson).</p> <p><i>*It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.</i></p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.3 Based on conversations, interactions with the teacher, and/or work samples, students were intellectually engaged with important ideas relevant to the focus of the lesson.</p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.4 The majority of students were on task throughout the class.</p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.5 The teacher's classroom management strategies enhanced the classroom environment.</p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.6 The classroom is organized appropriately such that students can work in groups easily, get to lab materials as needed, teacher can move to each student of student group, etc.</p> <p>Description Rubric Specific Rating Examples</p>
Evidence:	
	<p>1.7 The classroom environment established by the teacher reflected attention to issues of access, equity, and diversity for students (e.g. cooperative learning, language-appropriate strategies and materials, attentiveness to student needs).</p> <p>Description Rubric Specific Rating Examples</p>

UTOP AND ONLINE MANUAL

Rating	Indicator
	1.1 The classroom environment encouraged propositions that reflected engagement or exploration. Description Rubric Specific Rating Examples
Evidence:	
	1.2 Interactions reflected collegial working relationships among students. (e.g. students worked together productively and talked with each other about the lesson). <i>*It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.</i> Description Rubric Specific Rating Examples
Evidence:	
	1.3 The classroom environment reflected a variety of instructional strategies that were used to engage students in learning. Description Rubric Specific Rating Examples
Evidence:	
	1.4 The classroom environment reflected a variety of instructional strategies that were used to engage students in learning. Description Rubric Specific Rating Examples
Evidence:	
	1.5 The classroom environment reflected a variety of instructional strategies that were used to engage students in learning. Description Rubric Specific Rating Examples
Evidence:	
	1.6 The classroom is organized appropriately such that students can work in groups easily, get to lab materials as needed, teacher can move to each student of student group, etc. Description Rubric Specific Rating Examples
Evidence:	
	1.7 The classroom environment established by the teacher reflected attention to issues of access, equity, and diversity for students (e.g. cooperative learning, language-appropriate strategies and materials, attentiveness to student needs). Description Rubric Specific Rating Examples

This indicator assesses the degree to which students have learned to be collegial, respectful, cooperative, and interactive when working in groups. Evidence of collegial working relationships among students includes collaborative discussions about topics relevant to the lesson and successful distributing of roles and responsibilities within each group...

This indicator should be rated a **1** if there is group work during the lesson, but the group work is highly unproductive. This could include behavior where the majority of the groups are socializing, off-task, arguing, or ignoring each other, as well as regular instances of students copying and/or certain group members doing all of the work.

This indicator should be rated a **2** if ...

Rating of 3 Example: The students were put into debate groups for this class period - one group would debate another group, while the rest of the student groups were in the audience. The groups worked together smoothly - the students were able to pick who was doing what part of the debate, coordinate their arguments, and split the time slots when necessary. The audience also would occasionally compare their notes during breaks...

PILOT STUDY

- Test UTOP on some of our graduate's classrooms
- Conducted 83 observations of:
 - UTeach Graduates (N=21)
 - Non-UTeach Graduates (N=15)
- Novice teachers (most 0-3 years exp)
- Math, science, and computer science classes



PILOT STUDY

- After starting out at similar levels, UTeachers grow more in UTOP scores over time
 - Teaching experience significant predictor of UTOP scores for **UTeach** group ($p < .05$)
- **Noyce Scholars** rated significantly higher on UTOP than other groups, ($p < .01$)
- **Key Question:** Is the UTOP a valid and reliable instrument that measures important components of effective teaching?



NMSI/MET STUDY

- UTOP study conducted in partnership with the Gates Foundation's *Measures of Effective Teaching* project, and the National Math and Science Initiative
- Examine reliability, consistency, factor structure
- Connect teaching behaviors on UTOP to teacher value-added gains

MEASURES *of*
EFFECTIVE TEACHING



THE MET PROJECT

- 3000 teachers from 7 school districts, 7 states
- Various subjects (mathematics, English, science) and grade levels
- Multiple measures of effectiveness (observations, value-added, student surveys, teacher exams)
- Multiple video lessons of each teacher
- Multiple classroom observation instruments
 - Charlotte Danielson's FFT
 - CLASS protocol
 - MQI Rubric
 - UTOP



NMSI/MET STUDY

- 99 raters (math and science master teachers with LTF), scored 994 video lessons of 250 teachers using UTOP
- All lessons grades 4-8 mathematics
- One third of videos double-scored



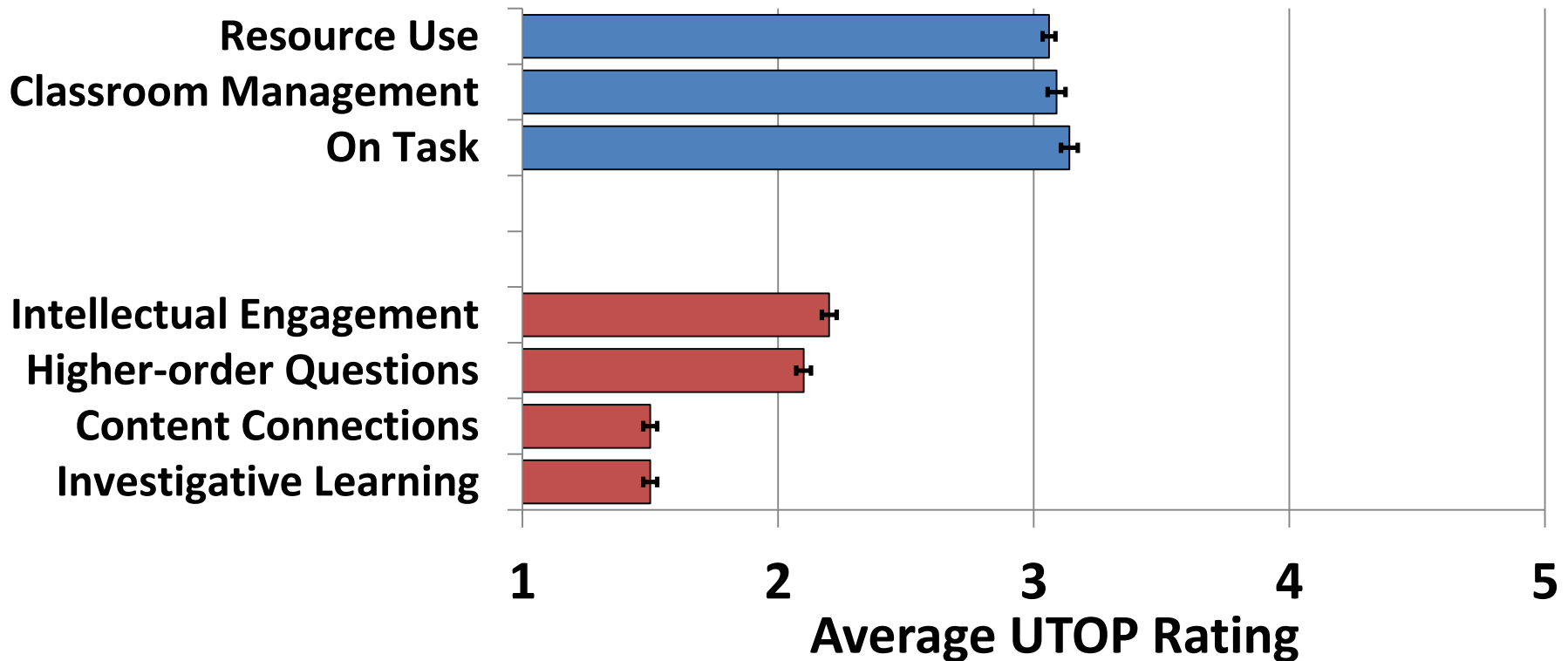
RESULTS

- Most of the 4-8 math video lessons from this national sample did not score highly on the UTOP
- Many middle school math teachers teaching problematic content; many formulaic/key word type approaches.
 - Raters identified problematic content issues in around one half of all lessons



RESULTS

- Surface-level engagement often seen, but little emphasis on conceptual understanding
- “Orderly but unambitious”



PSYCHOMETRIC ANALYSIS OF UTOP

- Factorial Structure
- Inter-rater Reliability
- Connection to Value-Added



FACTOR ANALYSIS OF UTOP

- What **macro-constructs** relating to effective teaching behaviors are being measured by the indicators on the UTOP?



FACTOR ANALYSIS OF UTOP

Cluster 1: Fostering Surface Engagement

- On task & involved
- Class management
 - Group work
- Lesson organization

Cluster 2: Fostering Deep Conceptual Understanding

- Inquiry/investigation
- Higher-order questioning
- Intellectual engagement

Cluster 3: Content Accuracy and Fluidity

- Verbal & written accuracy/fluidity
- Effective use of abstraction

Cluster 4: Making Content Connections

- To real world (authentic)
 - To “big picture”
- To history/current events

FACTOR 1: FOSTERING SURFACE LEVEL ENGAGEMENT

- Classroom management
- Majority “on task”
- Group-work dynamic
- Time management
- Lesson Organization
- Appropriate Resources
- Issues of equity & access
- Teacher critical of lesson



FACTOR 2: FOSTERING DEEP, CONCEPTUAL UNDERSTANDING

- Students generate ideas/conjectures
- Students intellectually engaged
- Students explore content
- Use of higher-order questions
- Use of inquiry/investigation



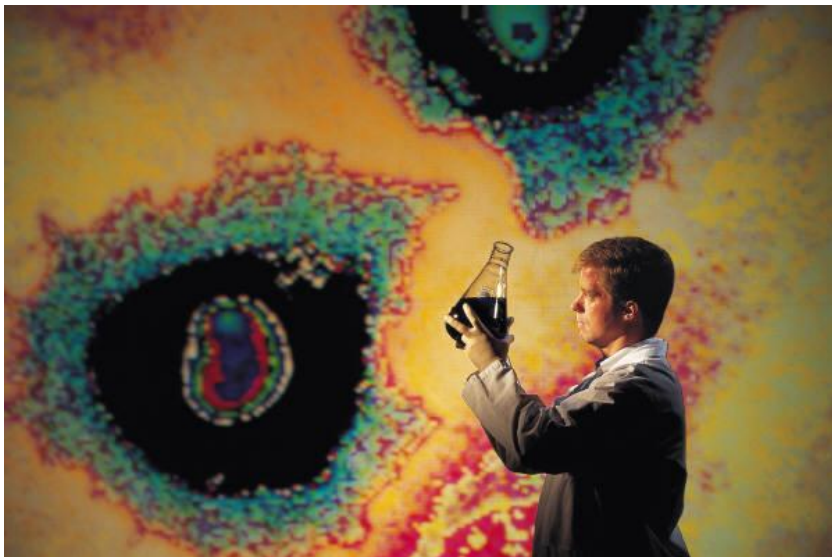
FACTOR 3: CONTENT ACCURACY & FLUIDITY

- Accurate written content information
- Accurate & fluid verbal communication of content
- Appropriate use of abstraction



FACTOR 4: CONTENT CONNECTIONS

- Connect content to “real world” and other disciplines
- Connect content to history & current events
- Connect content to the “big picture” of the discipline



INSTRUMENT RELIABILITY

- Can we **consistently** measure **teaching effectiveness** on the UTOP, beyond the biases of individual raters, or the characteristics of particular lessons?
- Goal: **60-80%** of the variance in teacher scores on the instrument attributable to the **stable** characteristics of the **individual teacher**



INSTRUMENT RELIABILITY

Schedule: 1 observation/year, 1 rater

	% total variance explained
Lesson	13.3%
Teacher	32.77%
Rater	0%
Residual (Rater x Lesson)	53.9%



INSTRUMENT RELIABILITY

Schedule: 4 observations/year, 4 different raters

	Reliability Coeff
Classroom Environment	67%
Lesson Structure	62%
Implementation	64%
Mathematics Content	40%
Overall (Avg Syn)	66%



VALUE-ADDED CORRELATIONS

- Are the teaching behaviors measured on the UTOP associated with **higher student learning gains**, on standardized assessments and tests of conceptual understanding?



VALUE-ADDED CORRELATIONS

	Corr with Prior Year's VA	Corr with Underlying VA	Diff bt/ top and bottom quartiles (sd)
CLASS	0.18	0.25	0.08***
FFT	0.13	0.18	0.06***
MQI	0.09	0.12	0.05*
UTOP	0.27	0.34	0.11***

0.25 standard deviations = 1 school year

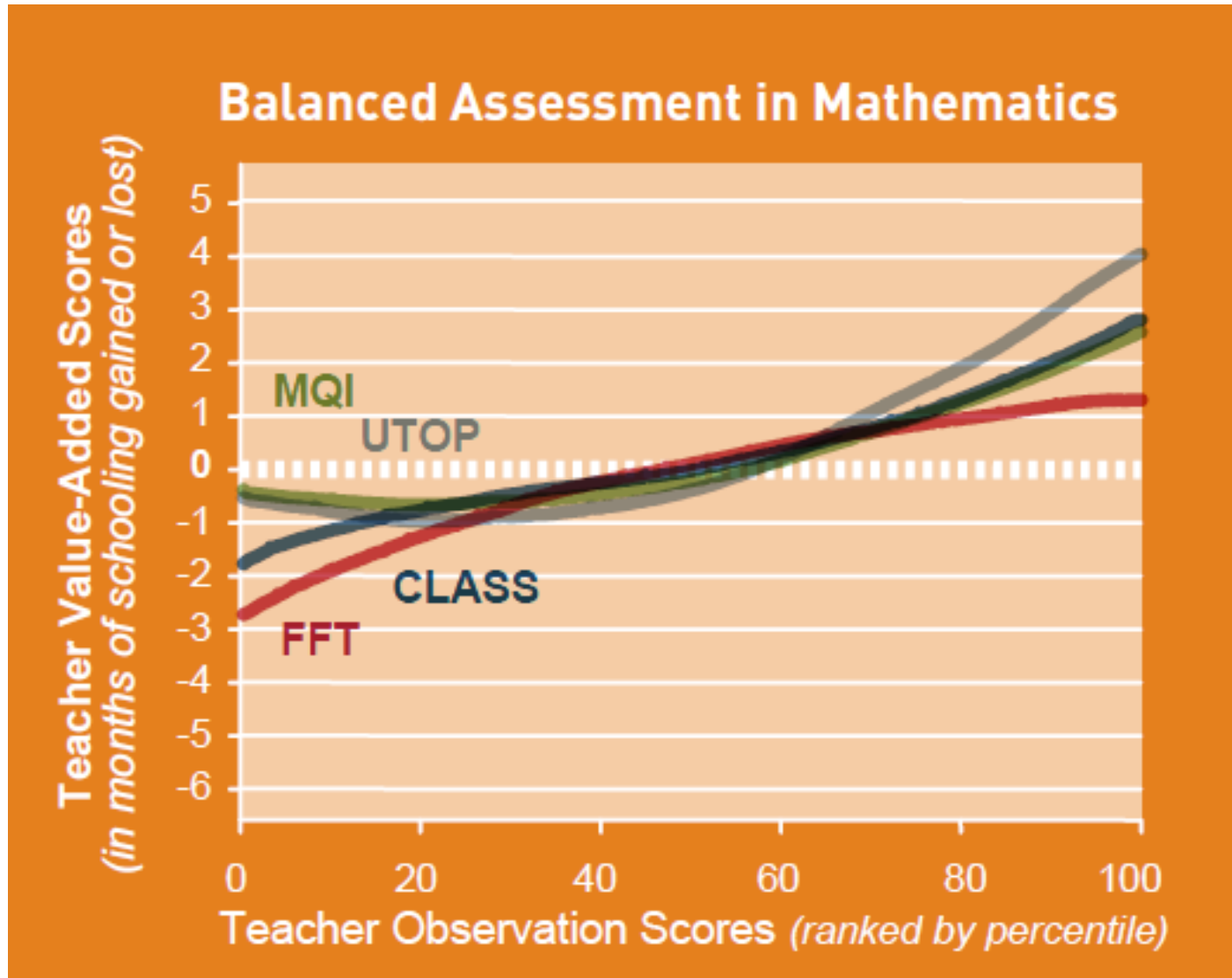
This table is copied from the released 2012 MET Report

VALUE-ADDED CORRELATIONS

	Difference between top and bottom quartiles in:		
	BAM Gain	Student Effort	Positive Emotional Attachment
CLASS	0.05*	0.07**	0.16***
FFT	0.08***	0.11***	0.16***
MQI	0.08***	0.04***	0.03***
UTOP	0.11***	0.13***	0.20***

This table is copied from the released 2012 MET Report

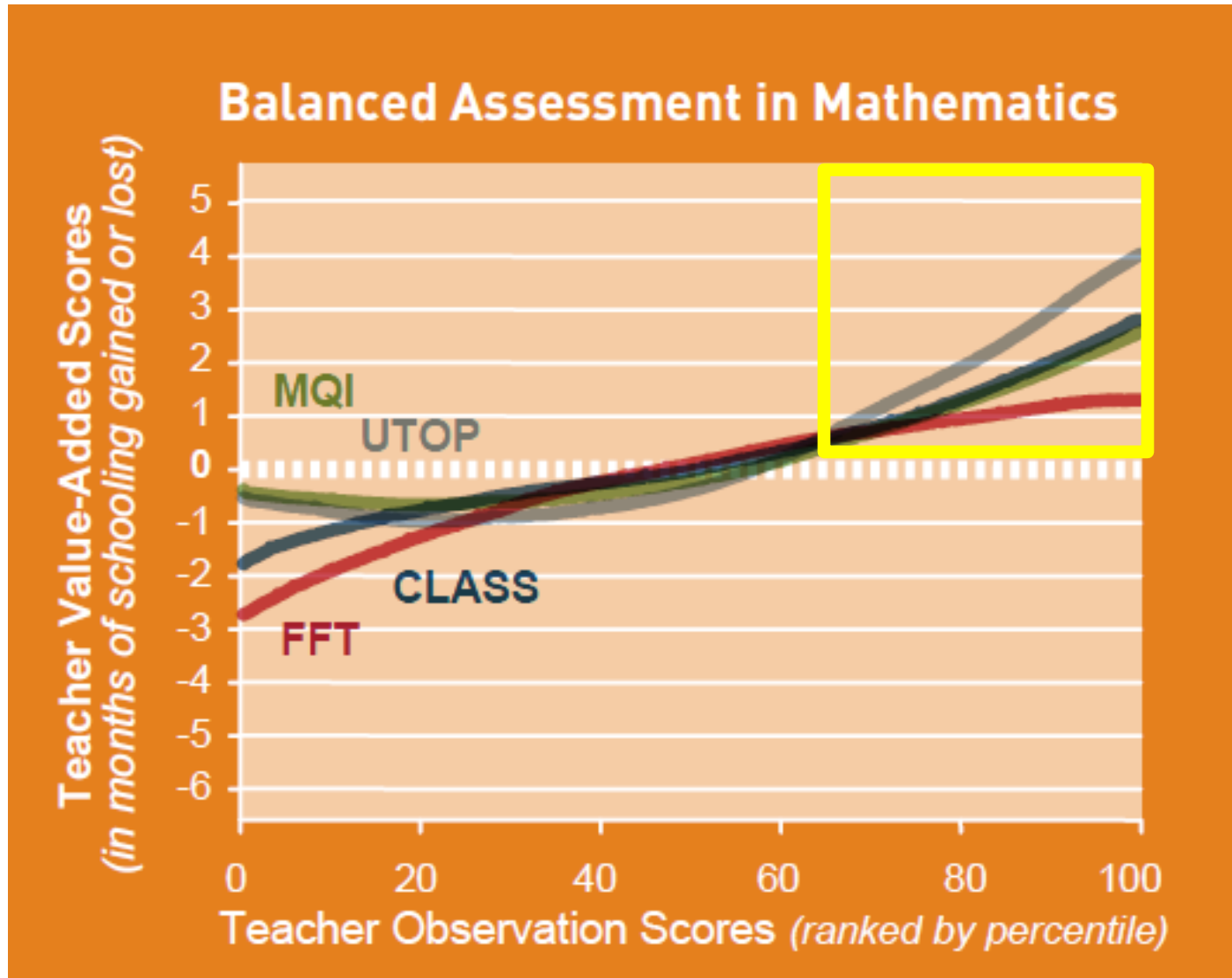
VALUE-ADDED CORRELATIONS



This graph is copied from the released 2012 MET Report



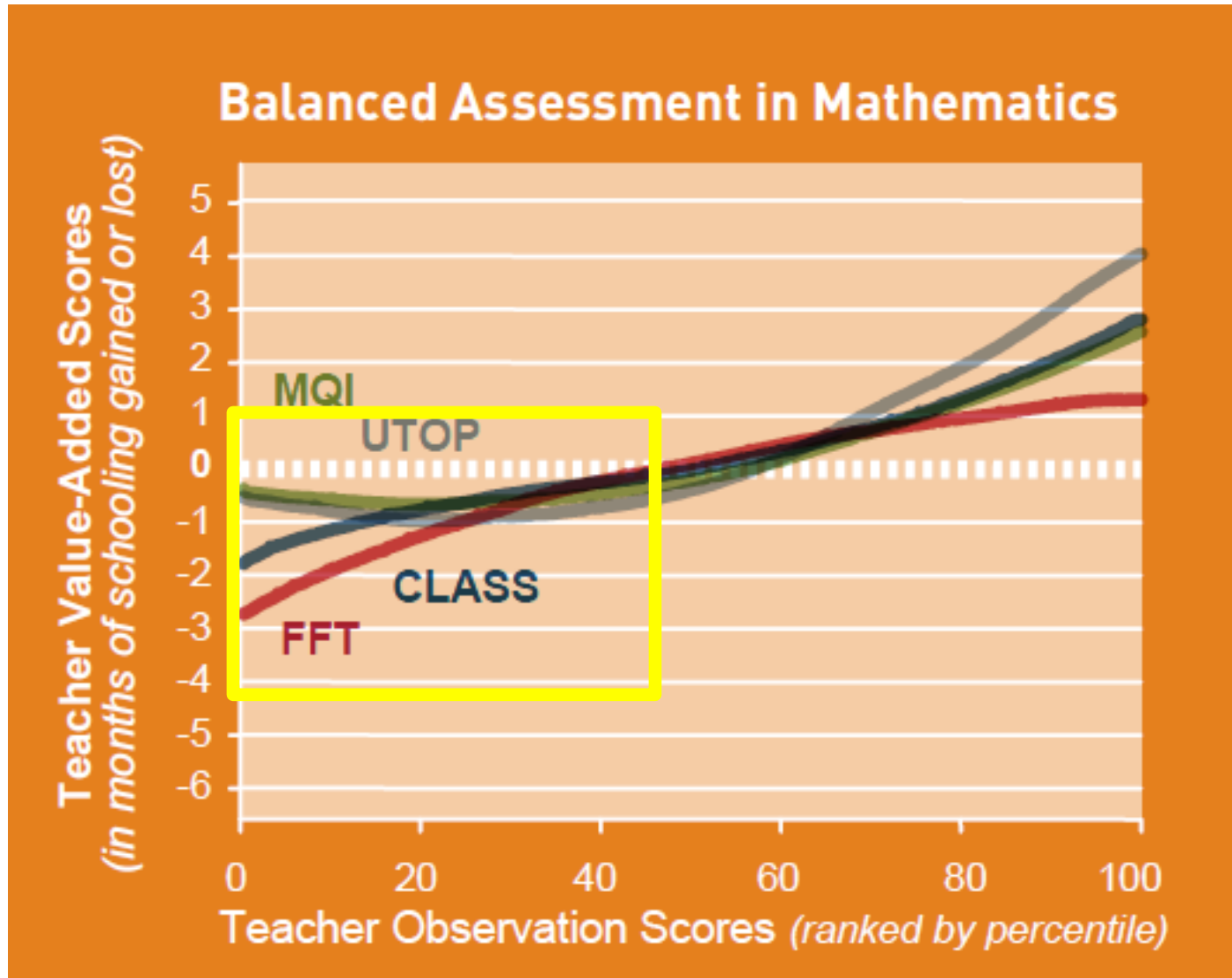
VALUE-ADDED CORRELATIONS



This graph is copied from the released 2012 MET Report



VALUE-ADDED CORRELATIONS



This graph is copied from the released 2012 MET Report



SUMMARY & IMPLICATIONS

- UTOP measures 4 factors of effective teaching
- UTOP has reasonable correlations with value-added – may better detect strong teachers
- Need multiple observations, multiple raters to conduct classroom observation
- Multiple measures of teaching effectiveness (value-added, observations, student surveys, teacher exams, etc.)



FUTURE DIRECTIONS

- Connect specific teaching behaviors to teacher value-added – what really matters?
- Investigate why the UTOP might be more effective at identifying excellent teaching
- Use of UTOP to compare classrooms at project-based school (with UTeach graduates) to those at traditional school, same low income school district



**QUESTIONS,
COMMENTS,
SUGGESTIONS?**

