Lessons learned from instruction: Results from a study of upper-elementary mathematics classrooms

Heather Hill, David Blazar, David Braslow, Mark Chin, Charalambos Charalambous, Claire Gogolen, Corinne Herlihy, Andrea Humez, Katie Lynch, Dan McGinn

NCTE National Conference

April 24, 2014

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C090023 to President and Fellows of Harvard College. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Overview

- Alternative title: Educational Production Function meets Mathematics Education Research
- Project set out to link:
  - Value-added indicators of teacher quality
  - Measures derived from the "ground up" – from what we think matters
    - Teachers' subject matter knowledge for teaching
    - Teachers' knowledge of their students
    - Teacher efficacy, preparation to teach mathematics
    - Teachers' general pedagogical practices (CLASS)
    - Mathematical Quality of Instruction (MQI)
- Side bonus in this talk: Implementation of Common Core

Overview

- First look at:
  - What mathematics instruction in upper-elementary classrooms looks like
  - What *mathematics-specific* aspects of teaching predict teacher value-added scores
    - Do now: What would you do??
  - What characteristics of *teachers* predict teacher value-added scores

Study Basics

- National Center for Teacher Effectiveness main study
  - Over 300 fourth and fifth grade teachers in 4 districts
  - Value-added scores for teachers
    - *Method*: Typical within-district multi-level model (student prior achievement, demographics, peer & cohort effects)
    - *Test 1*: State standardized test scores for ALL students for up to 4 years
    - *Test 2*: Alternative test scores (fall & spring) for NCTE students for up to 2 years
  - Two years of videotaped lessons (up to 6 lessons per teacher)
    - Coded with the Mathematical Quality of Instruction instrument (MQI) and Classroom Assessment Scoring System (CLASS)
  - Other alternative indicators
    - Teacher knowledge, teacher preparation for teaching, knowledge of students' mastery of mathematical content, teacher efficacy, background variables

WHAT DOES MATHEMATICS INSTRUCTION LOOK LIKE?

What math instruction look*ed* like

"Mathematics teaching in most elementary classrooms emphasizes rules, procedures, memorization, and right answers (Goodlad, 1984; Stodolsky, 1988). Students seldom confront serious mathematical problems and are rarely expected to reason about mathematical ideas. Teachers stand at the board, show students how to do a particular procedure or type of problem, and assign practice exercises. Students then work quietly on these, asking the teacher for help if they get stuck. When students are done, the teacher checks their answers, marks the ones that are wrong, sometimes goes over the steps once again, and then students fix their incorrect answers." (Cohen & Ball, 1990, p. 234)

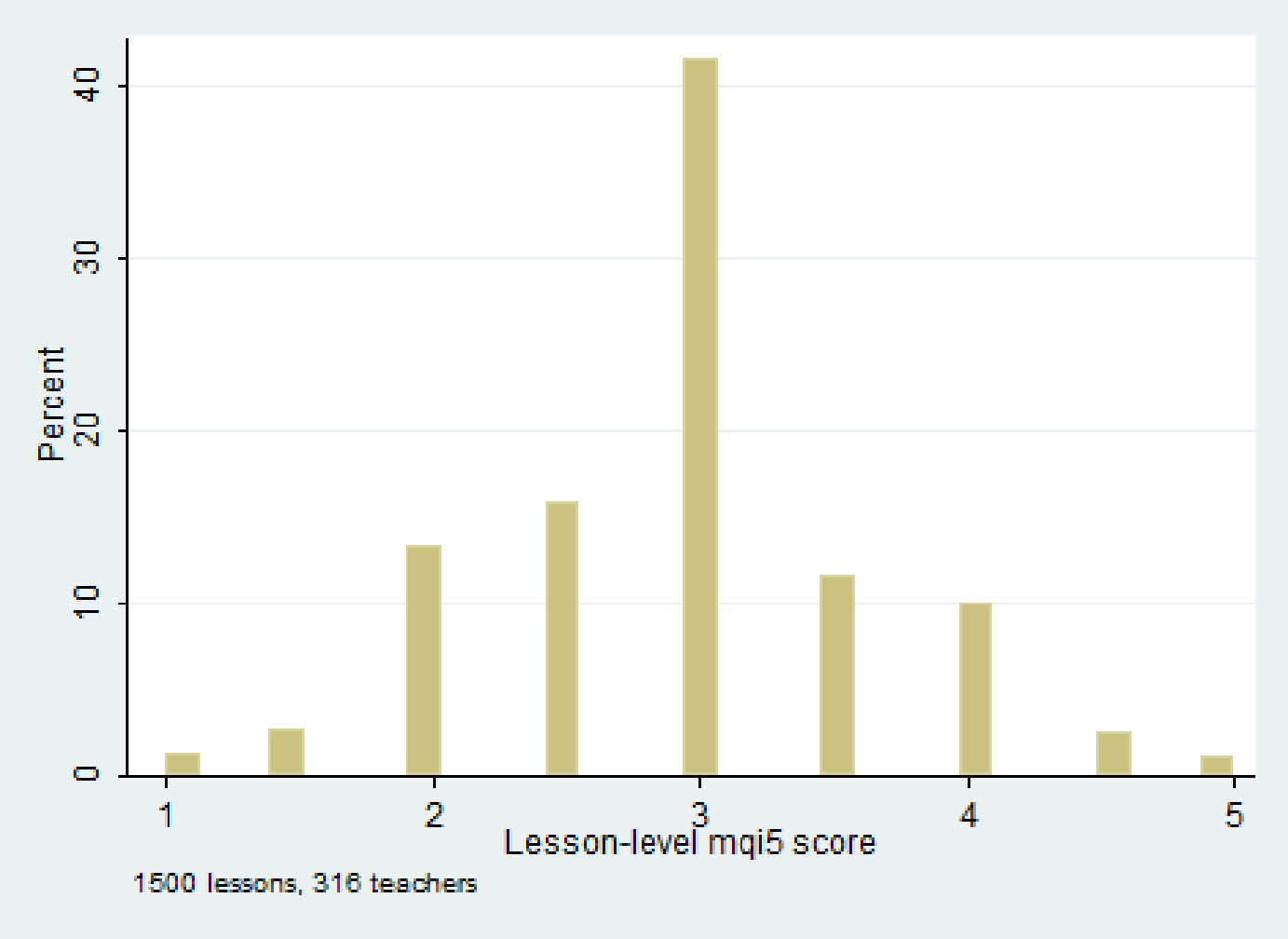
What does instruction look like now?

- 30 years past the initial reforms aimed at more ambitious instruction
- At the cusp of changes driven by CCSSM?

Method: Scoring Instruction

- Mathematical Quality of Instruction "overall lesson score"
  - 1: Systematic and major mathematical errors, serious lack of coherence in instruction, lack of focus on mathematics
  - 3: Either:
    - Nothing terrible, nothing good
    - Slightly positive, slightly negative
  - 5: Strong focus on meaning of mathematics, at least moderately strong student participation in the development of mathematical ideas
- Each lesson scored by two raters

Instruction according to MQI Raters



Christa

- Fifth grade
- Students have been confused about area and perimeter
- So teacher draws Venn diagram and asks them to identify features they share and features that are specific

Christa

Positive characteristics of "Mid" (3) lessons

- Isolated instances of meaning-making
  - Sense-making around word problems ("does an answer of 7/8ths make sense?")
  - Use of multiple representations (but without careful linking between)
  - Multiple methods for solving single problems (but without comparing/connecting those messages)
  - Occasional mathematical explanations, usually for "local" problems (e.g., "I knew Roberto had 28 apples because he had two more than Lauren, who had 26.")
- Some student participation
  - Usually limited to an answer to a single "why" question
  - …Or a turn and talk about a mathematical idea
  - Also a number of student *mathematical* questions ("Why do you divide both sides by 2?")

Not positive aspects of "Mid" (3) lessons

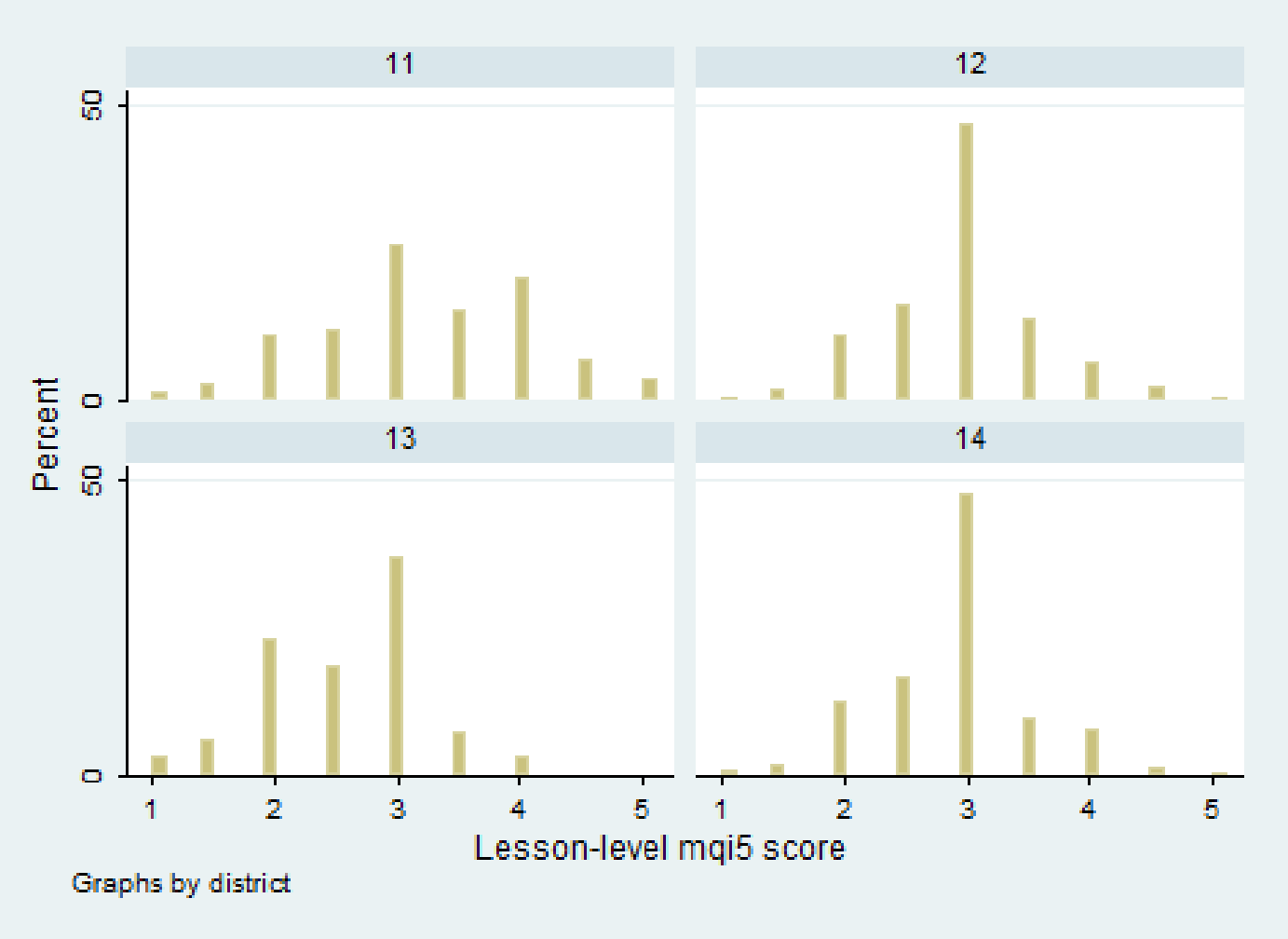
- Meaning-making, mathematical practices, and student participation is usually fleeting, and may be unconnected to central part of the lesson
- Lots of teacher showing/telling
- May be occasional lack of clarity around the mathematics
- May be "meh" sequencing of activities and problems
- Many missed opportunities

Lessons at a 4.5 or 5

- Student reasoning, explanation, and work on cognitively demanding problems
- Teacher able to use student ideas to develop the mathematics of the lesson
- Mathematics much more “present” in the lesson—particularly explanations and connections between representations/methods
- Teacher uses mathematical language and encourages students to do the same

But where are these lessons?

MQI Instruction by District



What do we make of this?

- Strong district differences among our data
  - D11: State test picks up more intellectually ambitious work; adopted text is NCTM-aligned; investments in teacher professional development (focusing on teacher content knowledge and inquiry pedagogy) since 2000; coaches in schools but largely controlled by math coordinator; teacher leaders; professional development for principals.
  - D12: Adopted text is NCTM-aligned; state test picks up more intellectually ambitious work.
  - D13: Teachers told to "make up their own" lessons; little math-specific professional development; no coordination with principals; extreme accountability system; conventional state test.
  - D14: Conventional curriculum materials; basic skills state test; frequent personnel changes in the math offices; math coaches but funded/controlled by schools

Conclusion to this paper

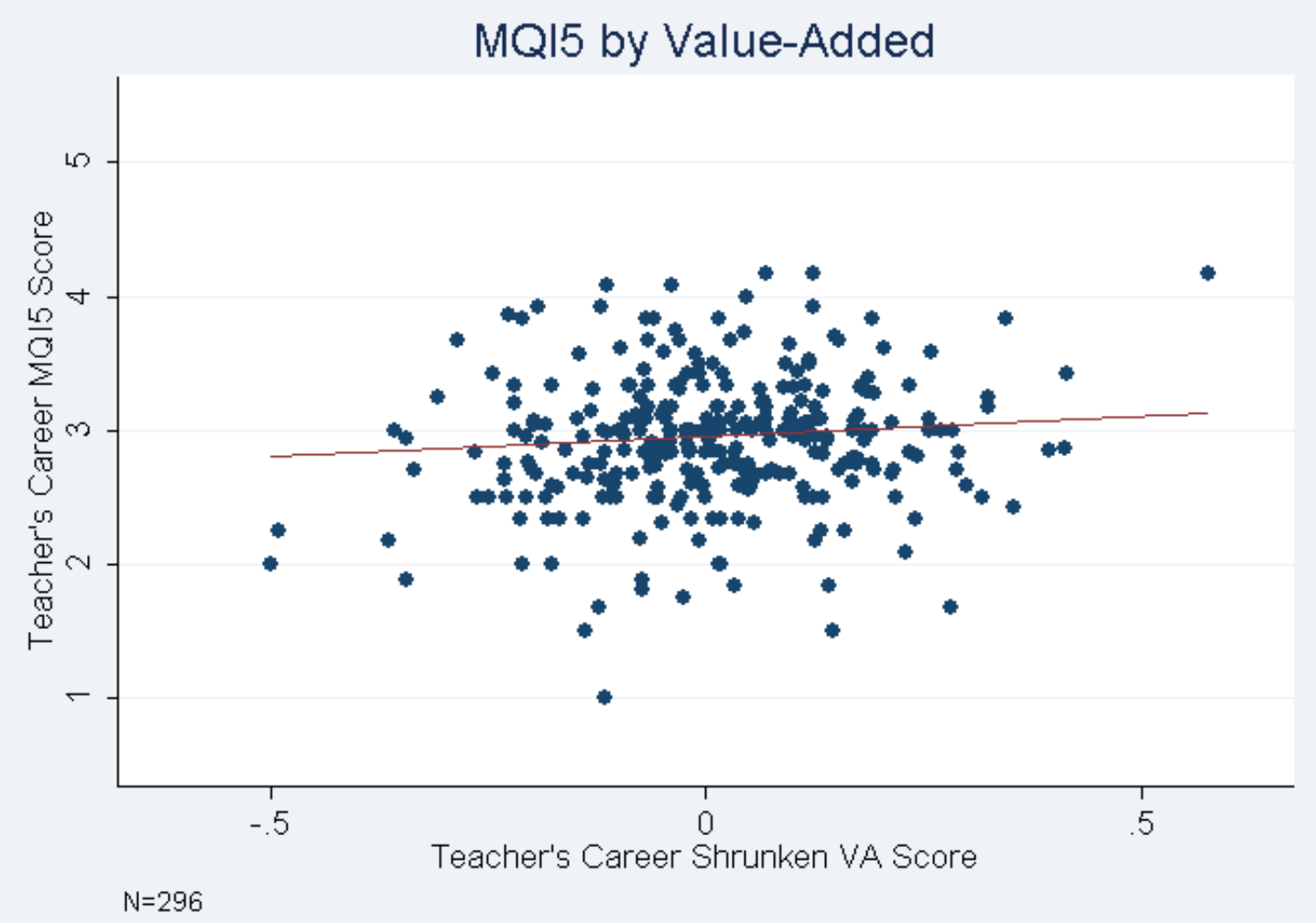
- Common Core Instruction DOES exist, and it's not isolated cases
- Deeply problematic instruction occurs at about the same rate as Common Core instruction
- "New normal" in mathematics classrooms– but new normal is not close to the Common Core ideals
- Serious district effects

WHAT ELEMENTS OF *TEACHING* PREDICT STUDENT OUTCOMES?

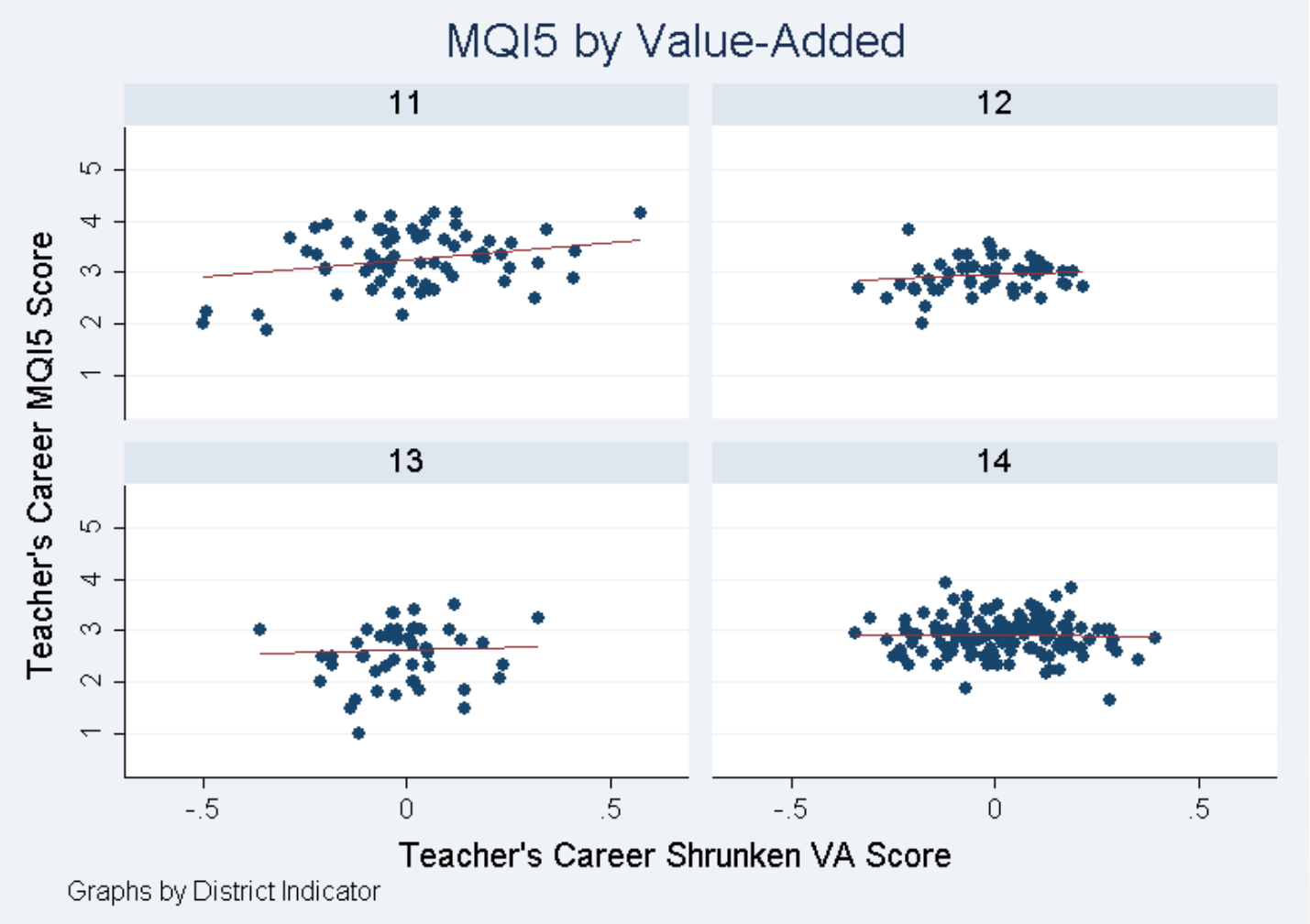
Data for this analysis

- Collected student-level data from state standardized assessments; NCTE test
  - Up to 4 years of teacher-classrooms, all used in the model
- Up to three years of lesson observations
  - MQI: Scored with 2 raters per video
  - CLASS: Scored with 1 rater per video
- Total sample size: 296 teachers
  - D11: 68
  - D12: 53
  - D13: 49
  - D14: 126

MQI and teacher VAM scores



MQI and state teacher VAM by district



MQI and Value-Added Scores

	State			
Variable	Overall	11&12	13	14
CWCM	.0989*	.3139**	-.0233	.0360
Rich	.0849	.2243~	-.0489	-.0930
WWS	.1678*	.3204**	-.1257	-.0316
Errors	-.1137	-.2212	.0092	-.0434
CCSP	.1427*	.2646*	.0240	.0157
MQI5	.1276~	.2746*	.0245	-.0491

CWCM= classroom work connected to mathematics (+)

Rich = Richness of the mathematics

WWS = Working with students' ideas/misconceptions (+)

Errors= Teacher errors

CCSP= Common Core Student practices (+)

MQI 5 = overall lesson quality (+)

MQI and Value-Added Scores

	State				NCTE			
Variable	Overall	11&12	13	14	Overall	11&12	13	14
CWCM	.0989*	.3139**	-.0233	.0360	.0706	.2299*	-.0797	.0713
Rich	.0849	.2243~	-.0489	-.0930	.0730	.1183	.1452	-.0317
WWS	.1678*	.3204**	-.1257	-.0316	.1302*	.2533**	.0446	-.0836
Errors	-.1137	-.2212	.0092	-.0434	-.0893	-.0968	-.2305~	-.0085
CCSP	.1427*	.2646*	.0240	.0157	.1110~	.2455**	-.0040	-.0401
MQI5	.1276~	.2746*	.0245	-.0491	.1139~	.2024*	.1289	-.0358

CWCM= classroom work connected to mathematics (+)

Rich = Richness of the mathematics

WWS = Working with students' ideas/misconceptions (+)

Errors= Teacher errors

CCSP= Common Core Student practices (+)

MQI 5 = overall lesson quality (+)

Just to lay this out

- *It's the district*: Common Core-ish instruction "works" – but only when districts make it happen (literally)
- In districts 11 and 12, effect sizes are large when compared with the average "teacher effect size"
  - Average teacher effect size for state test is 0.17 in our data
  - In 11 & 12, most effects are approximately double that
- But other districts struggled to put the conditions in place for Common Core Instruction. They *knew*, but could not do.
-

The kicker (math ed crowd)

Math Coordinators see teacher evaluation as an impediment to improving instruction

And…

Not huge returns (so far) for new teacher evaluation systems

Or so say folks like Dan Goldhaber

What would you do?

- Do you know what type of district (or state) you are?
  - If not, what's stopping you from finding out?
- If you are District 11, what do you do next?
  - How do you maintain the focus on improving instruction toward Common Core ideals?
- If you are District 14, what do you do next?
  - What are the barriers to improving instruction?
  - What resources can you mobilize to remove those barriers?
  - Will improvements toward the Common Core result in better student test scores?
  - What data do you need to determine whether this is working? How would you monitor the situation?

What I would do

1. Exit about 5-10% of teachers from the mathematics classroom
   a. A high-surveillance group gets yearly, intensive scrutiny
   b. Group membership triggered by poor value-added scores, observation flags, novice teacher status
   c. Exit those we can
2. Leave everyone else alone
   a. Much less frequent high-stakes evaluations
   b. Low stakes
3. Empower the Curriculum and Instruction folks to do their work
   a. Start talking to these folks
   b. Start creating coherence across the district with regard to instruction and instructional initiatives

WHAT *TEACHER* CHARACTERISTICS PREDICT INSTRUCTION?

Measured….

- Conventional predictors from educational production function literature (math methods/content courses taken; years of experience, bachelor's and master's degree)
- More proximal indicators
  - Teachers' mathematical knowledge
  - Teachers' knowledge of their kids' proficiency
- Practices you can't measure via video
  - Curriculum alignment – their enacted curriculum vs. test

Teacher training and experience

	State Test	NCTE Test
Master's Degree	0.031	0.020
Advanced Math Courses	-0.001	0.006
Math Content Courses for Teachers	0.046*	0.044*
Math Major/Minor or Grad. Degree	0.030	-0.004
Bachelors in Education	0.077*	0.039
Elementary Math Certified	0.050	-0.038
4-10 Years Experience	-0.006	-0.000
10+ Years Experience	-0.041	-0.033
Alternative Certification	0.050	0.075
No Certification	0.002	-0.026
Student-level Prior Achievement	x	x
Student-level Background	x	x
Classroom-level Composition	x	x
School-level Composition	x	x
N Students (teachers)	7843 (283)	7843 (283)
Adjusted Pseudo R^2	0.08	0.13

Teacher knowledge variables

	State Test	NCTE Test
More accurate knowledge of student performance	0.031*	0.025~
Knowledge of likely student misunderstandings	-0.017	-0.021
Teacher mathematical knowledge	0.029*	0.039**
Student-level Prior Achievement	x	x
Student-level Background	x	x
Classroom-level Composition	x	x
School-level Composition	x	x
N Students (teachers)	7843 (283)	7843 (283)
Adjusted Pseudo R^2	0.07	0.21

Other survey measures

	State Test	NCTE Test
Test Prep - Has decreased quality of instruction	-0.014	-0.011
Test Prep - Specific Activities	-0.001	-0.015
Efficacy	-0.023	-0.012
Content Coverage - Numbers/operations	-0.004	0.025
Content Coverage - Algebra	0.044**	0.016
Effort & Formative Assessment	0.036*	0.009
Student-level Prior Achievement	x	x
Student-level Background	x	x
Classroom-level Composition	x	x
School-level Composition	x	x
N Students (teachers)	7843 (283)	7843 (283)
Adjusted Pseudo R^2	0.09	0.07

Effort	Formative assessment
Grading mathematics assignments	I evaluate student work on mathematics assessments or assignments using a written rubric.
Gathering and organizing mathematics lesson materials (e.g., locating and copying supplemental material, preparing manipulatives)	I provide detailed written feedback on student mathematical work in addition to a numeric score.
Reviewing the content of specific mathematics lessons (e.g., reading the teacher manual, seeking additional information about the content)	I differentiate mathematics assignments based on students' individualized learning needs.
Preparing for a mathematics lesson by trying out explanations, or working through examples or problems	I change my lesson plans based on what I learn from analyzing student work.
Helping students learn mathematics before or after school hours	I examine student work to understand the process students use to solve mathematics problems.

Conclusions

- Selected variables related to outcomes
  - Teacher capacity seems very important for our test
- Small effect sizes
  - Typical for educational production function literature
  - But suggests that good classrooms are associated with many different factors, not just one overriding factor
- Total amount of variance in teacher value-added scores explained
  - 30% for state test
  - 45% for our test
  - Still a lot unaccounted for … but can we measure it?

PREDICTING INSTRUCTIONAL QUALITY

Predicting Instruction

- Question: What teacher background and knowledge predict high-quality instruction?
- Why ask?
  - Policy debates over preparation route
  - Many TE programs require certain preparation pathways
  - Personal resources often thought to contribute to instructional quality
- Method:
  - A bit of data reduction – “Richness,” “Working with Students” and “CCASP” collapsed based on factor analysis to “Ambitious instruction”
  - Used two CLASS dimensions, Emotional Climate and Classroom Organization
  - Regress these over a variety of teacher-reported characteristics
- Two models:
  - One for background characteristics only
  - One for kitchen sink – background, personal, institutional

Predicting Instruction: Conventional Predictors

Regressions of Domains of Instructional Quality on Conventional Teacher Characteristics

	Classroom Work Connected to Math	Ambitious Instruction	Mathematical Content Errors	Classroom Emotional Support	Classroom Organization
Novice Teacher	0.1419	-0.1033	-0.0818	-0.0192	-0.6767**
	(0.173)	(0.195)	(0.219)	(0.205)	(0.229)
Number of Math Methods/Content Courses	0.0580	0.1505*	-0.0308	0.0471	0.0587
	(0.062)	(0.070)	(0.063)	(0.067)	(0.080)
BA in Education	-0.2148	-0.1431	-0.0595	0.1196	0.1899
	(0.147)	(0.144)	(0.148)	(0.140)	(0.141)
Master's Degree	-0.0018	0.0502	0.0453	-0.1082	-0.2415~
	(0.153)	(0.146)	(0.153)	(0.149)	(0.136)
Certified in Elementary Math	0.0487	-0.1321	-0.0304	0.3676*	-0.0879
	(0.179)	(0.161)	(0.156)	(0.168)	(0.157)
Traditional Certification	-0.2434	-0.0477	0.1107	0.2255	0.1803
	(0.179)	(0.208)	(0.192)	(0.185)	(0.228)
Adjusted R-Squared	0.005	0.009	-0.018	0.014	0.064

Notes: ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses. Each column represents a separate regression. Sample for all regressions includes 273 teachers.

Predicting Instruction: Personal Resources

Regressions of Domains of Instructional Quality on Teachers' Personal Resources

	Classroom Work Connected to Math	Ambitious Instruction	Mathematical Content Errors	Classroom Emotional Support	Classroom Organization
Mathematical Knowledge	0.0722	0.2510***	-0.3562***	0.0577	0.0872
	(0.073)	(0.055)	(0.060)	(0.061)	(0.055)
Knowledge of Student Peformance-Distance	0.0468	0.1060*	-0.0421	0.0425	0.0076
	(0.062)	(0.051)	(0.066)	(0.064)	(0.066)
Preparation for Teaching/Formative Assessment	0.0097	-0.0219	0.1707**	-0.0048	0.0395
	(0.066)	(0.060)	(0.065)	(0.075)	(0.089)
Teacher Efficacy	-0.0454	-0.0461	-0.0659	0.1119~	-0.0911
	(0.063)	(0.049)	(0.058)	(0.068)	(0.063)
Joint Tests for Groups of Variables (*F*-statistic, *p*-value)					
Personal Resources	0.649 0.628	6.890 0.000	12.547 0.000	1.192 0.315	1.344 0.254
Adjusted R-Squared	0.041	0.383	0.225	0.045	0.112

Notes: ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses. Each column represents a separate regression. Sample for all regressions includes 273 teachers.

Predicting Instruction: Institutional Resources

Regressions of Domains of Instructional Quality on School/District Resources

	Classroom Work Connected to Math	Ambitious Instruction	Mathematical Content Errors	Classroom Emotional Support	Classroom Organization
Test Preparation Activities	0.0866	-0.0261	0.0842	0.1044	0.2047**
	(0.062)	(0.055)	(0.059)	(0.073)	(0.074)
Testing Has Changed Instruction	-0.0430	-0.1003~	-0.0271	-0.0856	-0.0382
	(0.063)	(0.055)	(0.062)	(0.064)	(0.055)
School Resources	0.0543	0.0544	0.0527	-0.1342~	0.0708
	(0.074)	(0.064)	(0.066)	(0.079)	(0.077)
District 2	-0.6910***	-1.3153***	-0.1687	-0.0676	-0.1209
	(0.197)	(0.168)	(0.160)	(0.198)	(0.171)
District 3	-0.4903*	-1.2672***	0.3727~	-0.2814	-0.4287~
	(0.191)	(0.206)	(0.216)	(0.214)	(0.245)
District 4	-0.5950***	-1.3005***	0.4912**	0.3791*	0.1072
	(0.168)	(0.170)	(0.163)	(0.190)	(0.164)
Joint Tests for Groups of Variables (*F*-statistic, *p*-value)					
District/School Resources	4.210 0.000	15.208 0.000	5.520 0.000	2.695 0.015	3.358 0.003
Adusjted R-Squared	0.041	0.383	0.225	0.045	0.112

Notes: ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Standard errors in parentheses. Each column represents a separate regression. Sample for all regressions includes 273 teachers.

Other NCTE papers (not mine!)

- Can knowledgeable observers predict teachers' value-added score? (Not well; chance = 0.33; our value = 0.48)
- Are MKT and "pure" mathematical knowledge distinguishable? (No! Ack!)
- Are self-reports and observational indicators of instructional practice correlated? (Yes! 0.2)
  - Which better predicts instruction? (Self-reports! Ack!)
- What best predicts student outcomes, mathematics-specific indicators or generic indicators? (Math-specific!)
- How much teacher-level variability in student outcomes can be accounted for by the kitchen sink model? (0.25 for state, 0.44 for our test)

Lessons Learned

- Don’t try this at home
- There is still a lot of unexplained variance in value-added scores
  - Ever explainable? I don’t know.
- Tremendous variability in relationship between VA and instruction
  - Sample?
  - Tests?
- Other than teachers’ content knowledge, not many explanations for quality of instruction