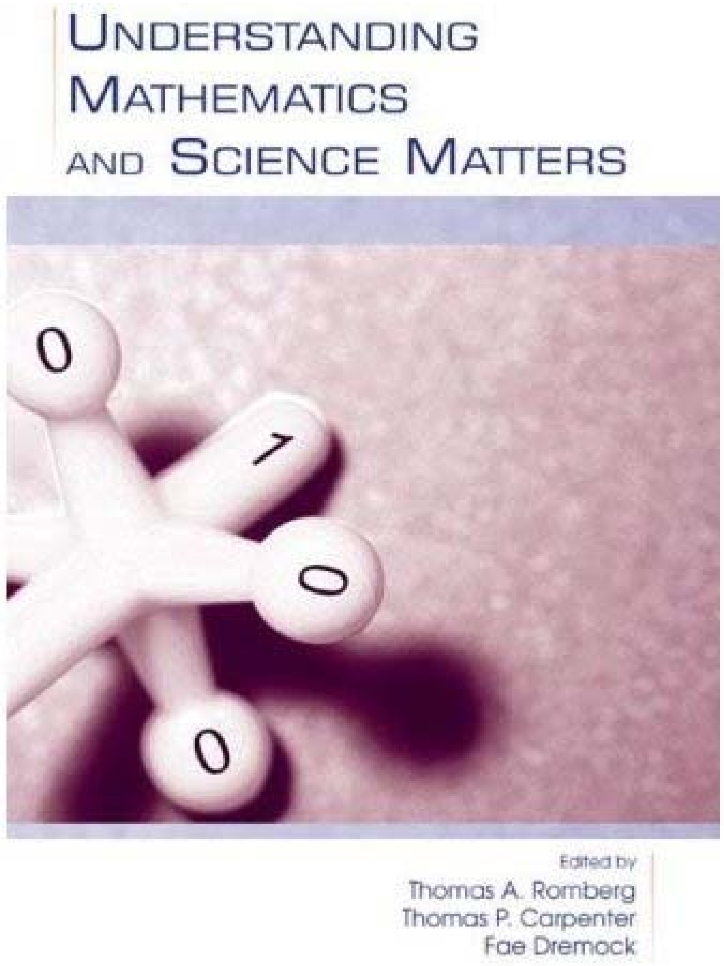
Curriculum Change and Causal Inference: Experimental Effects on Elementary Science Achievement

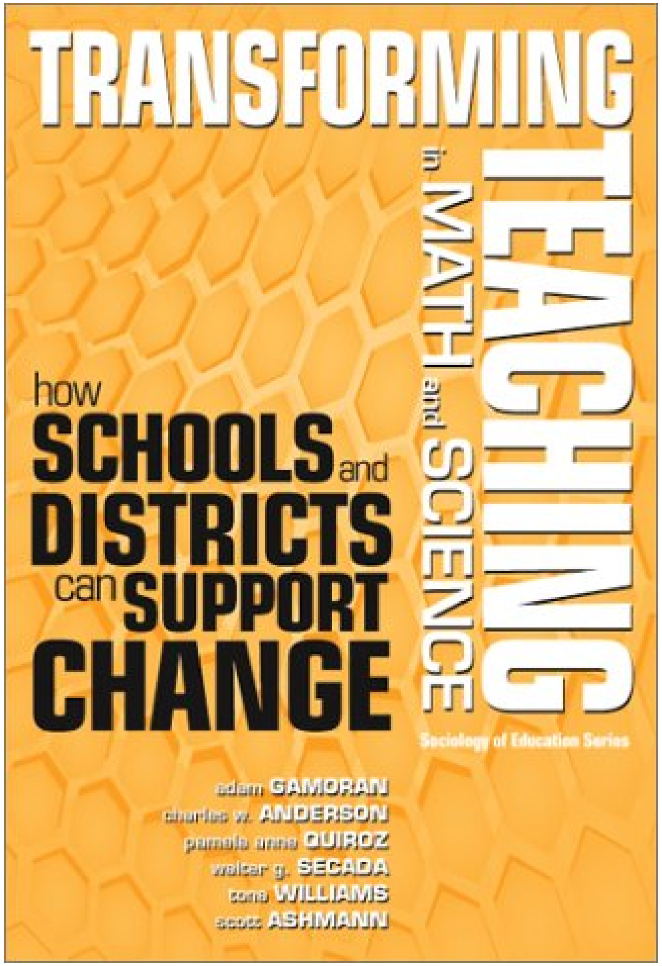
Adam Gamoran, Geoffrey D. Borman, and Jill Bowdon

University of Wisconsin-Madison

Teaching Science for Understanding

- Much interest in “teaching for understanding”
  - A focus on student thinking…
  - …about powerful mathematical and scientific ideas…
  - …in equitable classroom communities.

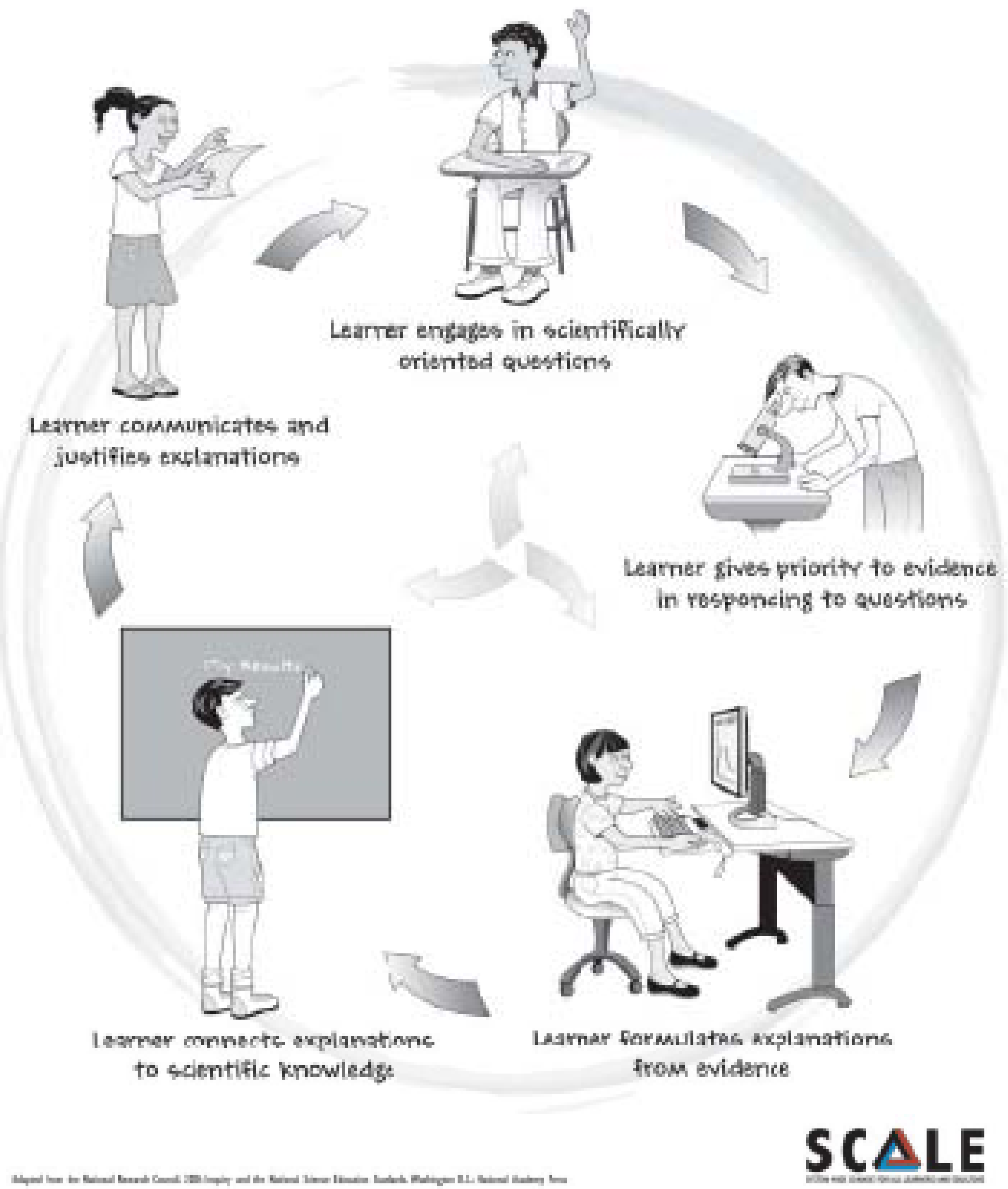

UNDERSTANDING
MATHEMATICS
AND SCIENCE MATTERS
0
1
0
0
Edited by
Thomas A. Romberg
Thomas P. Carpenter
Fae Dremock


TRANSFORMING
TEACHING
in MATH and SCIENCE
how
SCHOOLS and
DISTRICTS
can SUPPORT
CHANGE
Sociology of Education Series
adam GAMORAN
charles w. ANDERSON
pamela anne QUIROZ
walter g. SECADA
tona WILLIAMS
scott ASHMANN

Teaching Science for Understanding

- One approach to teaching for understanding in science: “Immersion” teaching
  - Students and teachers become “immersed” in the practice of science
  - A full cycle of inquiry, from question to hypothesis to data to interpretation

SCIENCE INQUIRY MAP



Curriculum Change and Causal Inference

- Little rigorous evidence on the impact of inquiry-oriented teaching
  - NRC report *Evaluating Curricular Effectiveness*
    - No rigorous evaluations of NSF mathematics curricula
  - What Works Clearinghouse
    - No evaluations of interventions in science
    - One inquiry-approach in mathematics shows “potentially positive effects”: *Everyday Mathematics*
      - Real-life problem solving
      - Student communication of mathematical thinking
- Scientific evidence is in short supply

Scaling Up Curriculum Change

- Even when curriculum effects show promise in rigorous evaluations, they fall short when taken to scale
- Example: national study of instructional technology
  - Many small-scale studies have shown benefits of technology-based instruction
  - Federally-sponsored, large-scale study conducted by Mathematica showed zero impact
  - Implementation was limited
    - E.g. in middle school math: 15 minutes per week

Scaling Up Curriculum Change

- Our study assesses the impact of curriculum change at scale
  - Large sample of schools in an even larger district
  - System-wide change in elementary science through teacher development
- Implementation takes time
  - Effects often decline in the first few years
  - "Implementation dip"
  - May take 5 years to see the full effect – if the reform lasts that long

The Study of Science Immersion

- System-Wide Change for All Learners and Educators
  - NSF Comprehensive Math-Science Partnership
  - School districts of Los Angeles, Denver, Providence, and Madison
  - Immersion units in Los Angeles
    - Extended, inquiry-based curriculum units for grades 3-8
    - Accompanied by intensive professional development institutes, followed by mentoring
      - Emphasis on scientific content

The Study of Science Immersion

- Earlier analyses of immersion
  - Increases in self-reported teacher knowledge
    - Pre-post, no comparison group
  - Increases in student learning, particularly among African American students
    - Comparison group with prior achievement controls
    - Small sample, inattention to multilevel structure of data
- Results are promising but vulnerable to selection bias
  - Better teachers could choose immersion teaching, better students could choose immersion teachers

The Context of Los Angeles

- NAEP 2005 Science
  - California is second-lowest only to Mississippi in grade 4 science achievement
  - Los Angeles is the lowest district in California on the trial urban assessment in grade 4 science
- A disadvantaged population
  - 75% Latino, 9% African American
  - 42% are English Language Learners
  - 81% qualify for free or reduced-price lunch

Design of the Study

- Grades 4-5 Science
  - All teachers have access to the immersion curriculum
  - Random selection of schools to send teachers to professional development institutes
- Our study tests the effects of teacher professional development for immersion teaching on student achievement
- Today: Two-year effects on grades 4 and 5

- First-year results appeared in Borman, Gamoran, Bowdon, *JREE* 2008
- Second-year results are hot off the press!
- AERA symposium Tuesday 10:35am

Design of the Study

- Professional development institutes are labor-intensive
  - Not all 420 elementary schools can participate at the same time
  - Had to be phased in
  - We convinced the district to select schools *randomly* for the professional development opportunities
  - Permits an *unbiased* assessment of impact
    - Not biased by unobserved selection factors

Design of the Study

- Los Angeles Unified School District (LAUSD) divided into 8 local districts
  - Each local district nominated approximately 20 schools
  - A total of 191 schools were nominated
    - We selected randomly 10 schools in each local district
    - Then, we randomly assigned 5 to treatment and 5 to control
  - Yields a sample of 40 treatment and 40 control schools

Design of the Study

- No differences between treatment and control samples, and no differences between study sample and nominated sample
- The 191 nominated schools had student populations that were less disadvantaged than the district as a whole
  - Our results are unbiased, but may not generalize to the whole district
  - However, generalization is better in our study than in most, because most studies are confined to volunteers
  - A closer approximation to implementation at scale than most experiments in education

Design of the Study

- Statistical power
  - As designed: .20 minimum detectable effect size
  - As implemented: .16 minimum detectable effect size
  - Powerful covariates (prior achievement) increased the precision of our estimates, allowing us to detect a smaller effect
- If teacher development for science immersion affects achievement, we should be able to detect it

Design of the Study

- Dependent variables
  - Periodic Assessments
    - G4 Life Science & G5 Earth Science
    - 20 multiple choice items, 1 constructed response
    - Aligned with state science standards
    - Immersion curricular unit aligned with these standards as well
  - State standardized assessment: G5 Science
    - Also aligned with content standards and immersion unit
- Other study aspects not examined today
  - Observations, interviews

Implementation

- Of the 40 schools selected for the intervention
  - 27 sent grade 4 teachers to the professional development institutes in Year 1
  - 3 additional schools sent teachers in Year 2
  - 17 schools sent grade 5 teachers in Year 2
- Participation in the intervention by Year 3
  - 30 schools in grade 4
  - 22 schools in grade 5
  - 36 schools in either grade 4 or grade 5, or both

Analytic Methods

- Testing the effects of the intervention
  - Watch out for sequencing of instruction
    - Three units each year: life, physical, and earth science
    - “Immersion” schools may select immersion topic first
      - Life science in grade 4, earth science in grade 5
    - Students are younger and have covered less science, so may score lower
  - An artifact of curricular sequencing
  - An unintended consequence of implementation
  - Not an issue for state standardized test which occurs in the spring in all schools

Analytic Method

- Two-level hierarchical linear models of students nested in schools, with control for sequence
- Fully specified Model:
  - Level One:
    - $Y_{ij} = b_{0j} + b_{1j}(\text{SEQUENCE})_{ij} + r_{ij}$,
  - Level Two:
    - $b_{0j} = \gamma_{00} + \gamma_{01}(\text{PRETEST})_j + \gamma_{02}(\text{TREATMNT})_j + u_{0j}$,
    - $b_{1j} = \gamma_{10}$,

Analyses and Sample Sizes

- Grade 4 Life Science, Year 1
  - 6,385 students (73%) in 74 schools
- Grade 4 Life Science, Year 2 (same schools)
  - 7,205 students, 74 schools
- Grade 5 Life Science, Year 2 (same students)
  - 6,391 students, 74 schools
- Grade 5 Earth Science, Year 2
  - 6,352 students 74 schools
- Grade 5 California Standards Test, Year 2
  - 7,166 students, 74 schools

Multilevel Models for Grade 4 Life Science, Year 1

	Empty Model	Model 1	Model 2	Model 3
Student Level Sequence Dummy				
School Level Intercept	0.558*** (0.013)			
Treatment				
Science Pretest Composite				

Multilevel Models for Grade 4 Life Science, Year 1

	Empty Model	Model 1	Model 2	Model 3
Student Level Sequence Dummy				
School Level Intercept	0.558*** (0.013)	0.607*** (0.039)		
Treatment		-0.053* (0.023)		
Science Pretest Composite				

Multilevel Models for Grade 4 Life Science, Year 1

	Empty Model	Model 1	Model 2	Model 3
Student Level Sequence Dummy				
School Level Intercept	0.558*** (0.013)	0.607*** (0.039)	0.220* (0.089)	
Treatment		-0.053* (0.023)	-0.058* (0.020)	
Science Pretest Composite			0.669*** (0.132)	

Multilevel Models for Grade 4 Life Science, Year 1

	Empty Model	Model 1	Model 2	Model 3
Student Level Sequence Dummy				-0.013 (0.033)
School Level Intercept	0.558*** (0.013)	0.607*** (0.039)	0.220* (0.089)	0.220* (0.086)
Treatment		-0.053* (0.023)	-0.058* (0.020)	-0.056* (0.019)
Science Pretest Composite			0.669*** (0.132)	0.678*** (0.132)

Multilevel Models for Grade 4 Life Science, Year 1

	Empty Model	Model 1	Model 2	Model 3
Student Level Sequence Dummy				-0.013 (0.033)
School Level Intercept	0.558*** (0.013)	0.607*** (0.039)	0.220* (0.089)	0.220* (0.086)
Treatment		-0.053* (0.023)	-0.058* (0.020)	-0.056* (0.019)
Science Pretest Composite			0.669*** (0.132)	0.678*** (0.132)

Results: Grade 4 Life Science, Year 1

- Treatment effect is negative
  - Students in immersion schools scored lower than those in comparison schools
- Explanations ruled out
  - Differential attrition
  - Timing of life science assessment
- Most likely explanation
  - Teachers struggle to implement the new curriculum – an “implementation dip”
- What happens in Year 2?

Multilevel Models for Grade 4 Life Science: Years 1 and 2 (Same Schools)

	Year 1	Year 1	Year 2	Year 2
Student Level Sequence Dummy		-0.013 (0.028)		
School Level Intercept	0.607*** (0.039)	0.220* (0.087)		
Treatment	-0.053* (0.023)	-0.056* (0.019)		
Science Pretest Composite		0.678*** (0.132)		

Multilevel Models for Grade 4 Life Science: Years 1 and 2 (Same Schools)

	Year 1	Year 1	Year 2	Year 2
Student Level Sequence Dummy		-0.013 (0.028)		-0.045* (0.021)
School Level Intercept	0.607*** (0.039)	0.220* (0.087)	0.620*** (0.019)	0.632 (0.019)
Treatment	-0.053* (0.023)	-0.056* (0.019)	-.014 (0.017)	0.000 (0.014)
Science Pretest Composite		0.678*** (0.132)		0.505*** (0.071)

Multilevel Models for Grade 4 Life Science: Years 1 and 2 (Same Schools)

	Year 1	Year 1	Year 2	Year 2
Student Level Sequence Dummy		-0.013 (0.028)		-0.045* (0.021)
School Level Intercept	0.607*** (0.039)	0.220* (0.087)	0.620*** (0.019)	0.632 (0.019)
Treatment	-0.053* (0.023)	-0.056* (0.019)	-.014 (0.017)	0.000 (0.014)
Science Pretest Composite		0.678*** (0.132)		0.505*** (0.071)

Multilevel Models for Life Science: Grades 4 and 5 (Same Students)

	Grade 4	Grade 4	Grade 5	Grade 5
Student Level Sequence Dummy		-0.013 (0.028)		
School Level Intercept	0.607*** (0.039)	0.220* (0.087)		
Treatment	-0.053* (0.023)	-0.056* (0.019)		
Science Pretest Composite		0.678*** (0.132)		

Multilevel Models for Life Science: Grades 4 and 5 (Same Students)

	Grade 4	Grade 4	Grade 5	Grade 5
Student Level Sequence Dummy		-0.013 (0.028)		-0.008 (0.020)
School Level Intercept	0.607*** (0.039)	0.220* (0.087)	0.581*** (0.022)	0.584*** (0.023)
Treatment	-0.053* (0.023)	-0.056* (0.019)	-0.023 (0.018)	-0.013 (0.013)
Science Pretest Composite		0.678*** (0.132)		0.730*** (0.091)

Multilevel Models for Life Science: Grades 4 and 5 (Same Students)

	Grade 4	Grade 4	Grade 5	Grade 5
Student Level Sequence Dummy		-0.013 (0.028)		-0.008 (0.020)
School Level Intercept	0.607*** (0.039)	0.220* (0.087)	0.581*** (0.022)	0.584*** (0.023)
Treatment	-0.053* (0.023)	-0.056* (0.019)	-0.023 (0.018)	-0.013 (0.013)
Science Pretest Composite		0.678*** (0.132)		0.730*** (0.091)

Multilevel Models for Grade 5: Earth Science and State Standardized Test

	Earth Sci	Earth Sci		
Student Level Sequence Dummy		-0.053 (0.020)		
School Level Intercept	0.564*** (0.031)	0.598* (0.025)		
Treatment	-0.008 (0.018)	-0.001 (0.015)		
Science Pretest Composite		0.704*** (0.132)		

Multilevel Models for Grade 5: Earth Science and State Standardized Test

	Earth Sci	Earth Sci	CST	CST
Student Level Sequence Dummy		-0.053 (0.020)		
School Level Intercept	0.564*** (0.031)	0.598* (0.025)	344.373*** (5.070)	341.382*** (3.819)
Treatment	-0.008 (0.018)	-0.001 (0.015)	-1.105 (4.639)	0.567 (3.325)
Science Pretest Composite		0.704*** (0.132)		0.997*** (0.057)

Multilevel Models for Grade 5: Earth Science and State Standardized Test

	Earth Sci	Earth Sci	CST	CST
Student Level Sequence Dummy		-0.053 (0.020)		
School Level Intercept	0.564*** (0.031)	0.598* (0.025)	344.373*** (5.070)	341.382*** (3.819)
Treatment	-0.008 (0.018)	-0.001 (0.015)	-1.105 (4.639)	0.567 (3.325)
Science Pretest Composite		0.704*** (0.132)		0.997*** (0.057)

Summary of Treatment Effects: Effect Sizes

Year 1 Grade 4 Life Sci	Year 2 Grade 4 Life Sci	Year 2 Grade 5 Life Sci	Year 1 Grade 5 Earth Sci	Year 2 Grade 5 Earth Sci	Year 2 Grade 5 CST
-0.270*	-0.030	-0.066	0.010	0.005	-0.008

Growing capacity or dissipation?

Growing Capacity or Dissipation?

- Over the 3 years of the study…
- Two changes in superintendent, and many other leadership changes
- Elimination of science lead teachers
- Adoption of new science curriculum
  - Full-Option Science System (FOSS)
  - Also an inquiry focus, but with less emphasis on content and little training for teachers
    - A 1 day orientation as contrasted with a 5-day institute
    - No time for content preparation
  - A change in the counterfactual!

Growing Capacity or Dissipation?

- Teachers in immersion schools have continued to use the immersion curriculum alongside the FOSS curriculum
- Surveys: teachers in immersion schools more likely to use immersion “a lot” and more likely to use immersion as much or more than FOSS, as teachers in comparison schools
- So the immersion curriculum is still being used, inquiry science is still being taught

Conclusions

- Implementation at scale is difficult
- Turbulent environment of large, urban districts is challenging for district-wide reform
- Causal chain may break down at many points
  - Incomplete and inconsistent participation
  - Changes in central district leadership
  - Insufficient school capacity to support reform
  - Shifting counterfactual as curriculum changes

Conclusions

- Future work in this study
  - Examine results from third year of intervention
  - Use data from interviews and observations to illuminate implementation and answer the question of growing capacity versus dissipation
- Future work for other studies
  - Enhance capacity to support reform as part of the reform implementation

Acknowledgments

- Research on this paper was supported by a grant from the National Science Foundation (award ID 0554566) to the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison. Findings and conclusion are those of the authors and do not necessarily reflect the views of the supporting agency.