Professor Neil Heffernan
NAE Talk, October 2017

ASSiSTments

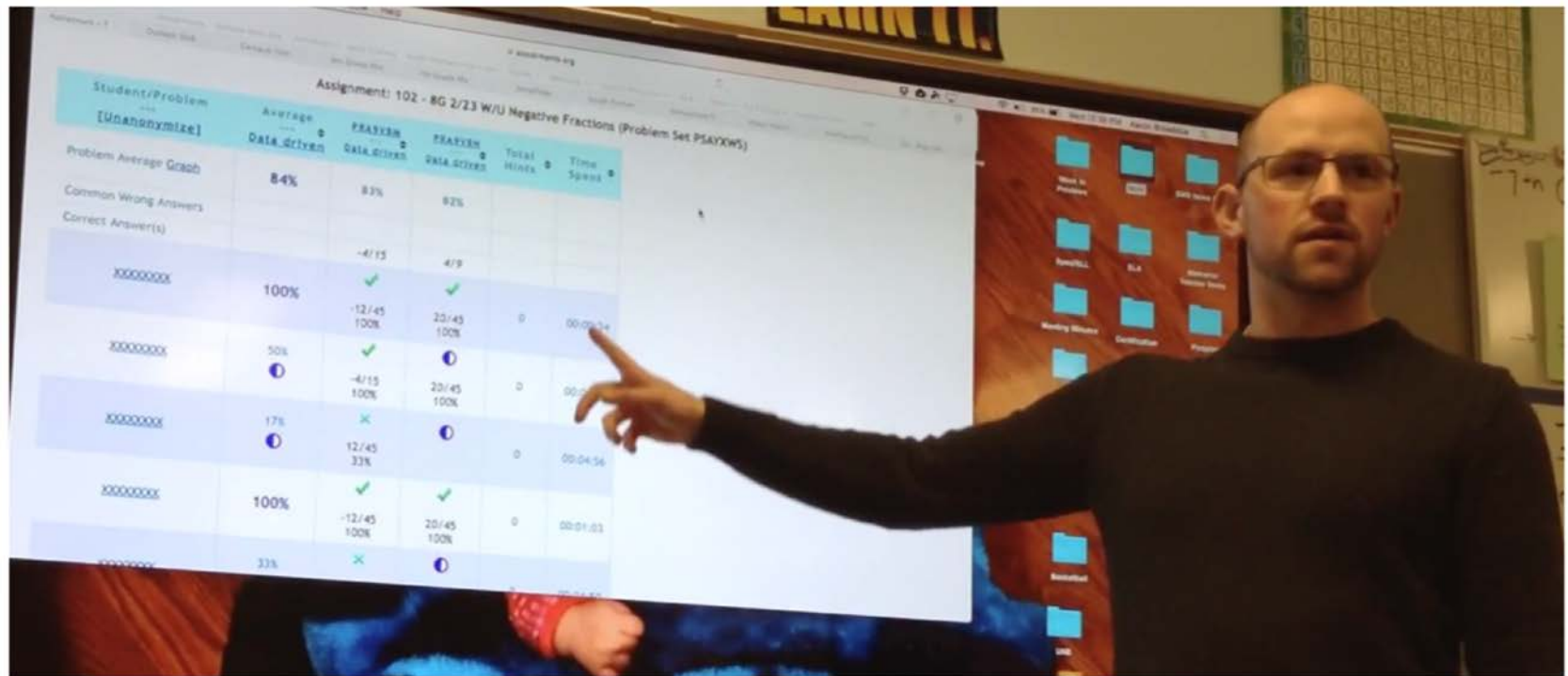# 12 million problems solved by 50,000 students

ASSISTments is Feedback

for students

These students know immediately if they answered correctly or incorrectly

We operate in a space that relies on and supports the teacher

Aaron Broaddus - 7th Grade Math Teacher - South Portland, ME

Student/Problem --- [Unanonymize]	PRAHE5Y ⇕	PRAHE5Z ⇕	
Problem Average Graph	27%	61%	
Common Wrong Answers	1/9^10, 56%	1/5^13, 58% +feedback	
Correct Answer(s)	1/3^10	1/5^3	1/16^2
XXXXXXX	◑	✕ 5^3 0%	✓ 1/16^2 100%
XXXXXXX	✕ 1^-10 0%	✓ 1/5^3 100%	✓ 1/16^2 100%
XXXXXXX	✕ 1/9^10 0%	✓ 1/5^3 100%	✓ 1/16^2 100%

Using the Item Report

During Class

?

Does it work?
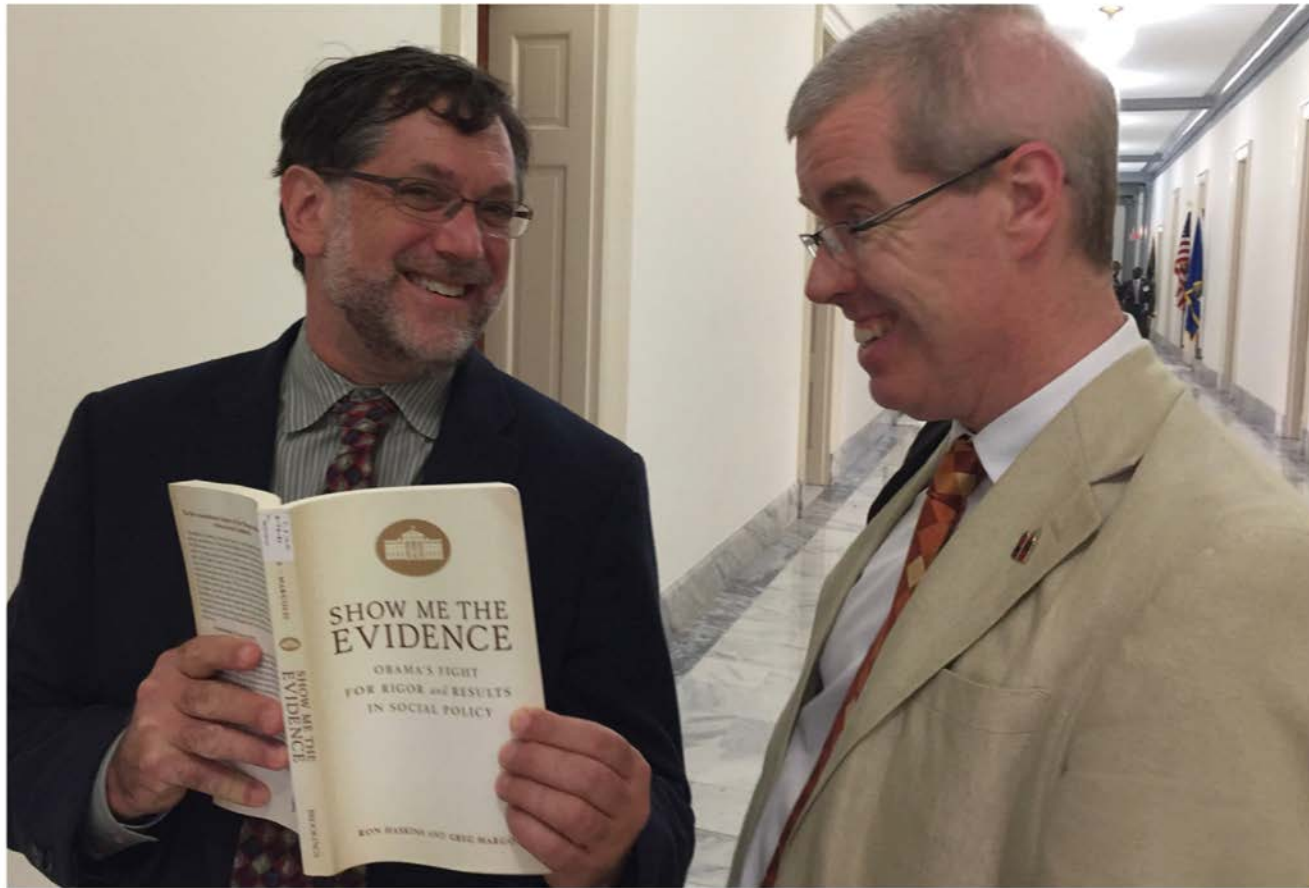
Of the 90 interventions with RCTs funded by the IES since 2002, only 11 have had positive results (Jon Baron)

Report on 26 studies of K-12 intelligent tutoring systems for math

(Steenbergen-Hu and Cooper, 2013)

Findings:

"Effect of ITS [is] small to modest"
and
"might contribute to achievement gap."

(Steenbergen-Hu and Cooper, 2013)

What Works Clearinghouse

For Math

Reliably Positive Interventions = ???

What Works Clearinghouse

For Math

Reliably Positive Interventions = 6

Recruited 44 Schools in Maine

All 7th Grade Teachers from each School

~ 2800 Students
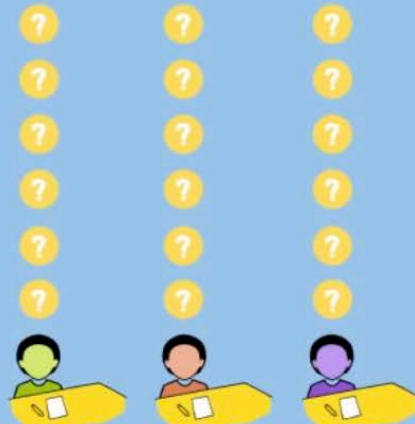
Immediate Feedback on Classwork an...

Students use their te...
worksheet...

2.1 Exercises

Vocabulary and Concept Check

1. **VOCABULARY** Is the quotient of two integers always...
2. **WRITING** Are all terminating and repeating decimals...

Tell whether the number belongs to each of the followin rational numbers, integers, whole numbers.

3. -5 4. -2.16 5. 12

Tell whether the decimal is *terminating* or *repeating.*

7. $-0.4848\ldots$ 8. -0.151 9. 72.72

Practice and Problem Solving

Write the rational number as a decimal.

11. $\frac{7}{8}$ 12. $\frac{1}{11}$ 13. $-\frac{7}{9}$ 14. $-\frac{17}{40}$

15. $1\frac{5}{6}$ 16. $-2\frac{17}{18}$ 17. $-5\frac{7}{12}$ 18. $8\frac{15}{22}$

19. **ERROR ANALYSIS** Describe and correct the error in writing the rational number as a decimal.

$\times \quad -\frac{7}{11} = -0.6\overline{3}$

Write the decimal as a fraction or a mixed number in simplest form.

20. -0.9 21. 0.45 22. -0.258 23. -0.312

24. -2.32 25. -1.64 26. 6.012 27. -12.405

Order the numbers from least to greatest.

28. $-\frac{3}{4}, 0.5, \frac{2}{3}, -\frac{7}{3}, 1.2$ 29. $\frac{9}{5}, -2.5, -1.1, -\frac{4}{5}, 0.8$ 30. $-1.4, -\frac{8}{5}, 0.6, -0.9, \frac{1}{4}$

31. $2.1, -\frac{6}{10}, -\frac{9}{4}, -0.75, \frac{5}{3}$ 32. $-\frac{7}{2}, -2.8, -\frac{5}{4}, \frac{4}{3}, 1.3$ 33. $-\frac{11}{5}, -2.4, 1.6, \frac{15}{10}, -2.25$

34. **COINS** You lose one quarter, two dimes, and two nickels.
   a. Write the amount as a decimal.
   b. Write the amount as a fraction in simplest form.

35. **HIBERNATION** A box turtle hibernates in sand at $-1\frac{5}{8}$ feet. A spotted turtle hibernates at $-1\frac{16}{25}$ feet. Which turtle is deeper?

48   Chapter 2   Rational Numbers

Problem ID: **PRABAEPE**

Comment on this problem

Page 48 #20

Type your answer below (numeric expression):

100% ?

Submit Answer

Show answer

Skill Practice with Skill Builders

Skill Builders are based on one skill – students need to answer 3 questions correctly in a row to complete the assignment

Answer 3 correctly in a row

Convert 31% into... ✗
Convert 36% into... ✗
Convert 42% into... ✓
Convert 60% into... ✓
➡ Convert 84% into... ✓

Assignment: Converting a Percent to a Fraction 6.RP.A.3c

Problem ID: **PRABUU6**   _Comment on this problem_

Convert **84%** into a **fraction**.

Type your answer below (mathematical expression):

84/100   100% ?

Correct!

[Submit Answer] [Next Problem]   [Show hint 1 of 2]

Teachers have access to reports to monitor student progress

Student [Unanonymize]	Status [Show dates]	Days working on assignment	Total time working on assignment	Problems seen
XXXXX	✓	1	00:09:00	3
XXXXX	✗ Exceeded daily limit	1	00:12:45	10
XXXXX	○			
XXXXX	✓	2	00:04:04	3

Jeremy Roschelle
@roschelle63

Can we improve math homework? Presented
@SRI_Education @assistments positive findings
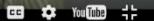at the White House Symposium on Digital
Learning today

SRI's Roschelle Reviews ASSISTments Maine Efficacy Study and its Major Findings
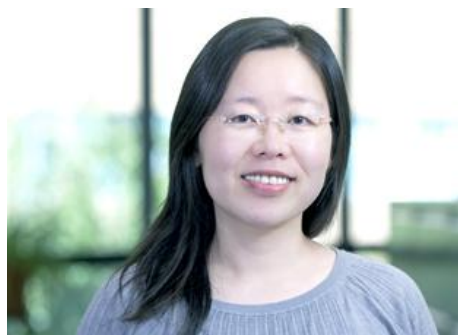
Online Mathematics Homework Increases Student Achievement

Jeremy Roschelle
Mingyu Feng
Robert F. Murphy

SRI International

Craig A. Mason

University of Maine

In a randomized field trial with 2,850 seventh-grade mathematics students, we evaluated whether an educational technology intervention increased mathematics learning. Assigning homework is common yet sometimes controversial. Building on prior research on formative assessment and adaptive teaching, we predicted that combining an online homework tool with teacher training could increase learning. The online tool ASSISTments (a) provides timely feedback and hints to students as they do homework and (b) gives teachers timely, organized information about students' work. To test this prediction, we analyzed data from 43 schools that participated in a random assignment experiment in Maine, a state that provides every seventh-grade student with a laptop to take home. Results showed that the intervention significantly increased student scores on an end-of-the-year standardized mathematics assessment as compared with a control group that continued with existing homework practices. Students with low prior mathematics achievement benefited most. The intervention has potential for wider adoption.

Keywords: *computers and learning, effect size, evaluation, experimental design, hierarchical linear modeling, homework, mathematics education, technology*

Finding 1:
Teachers Reliably Changed Practices

Homework Review
Emphasis on problems that were difficult for students

Finding 2:
Reliable Gain & Meaningful Impact on Student Learning

Finding 3:
Greater Effects for Low Prior Math:
Closed Achievement Gaps!
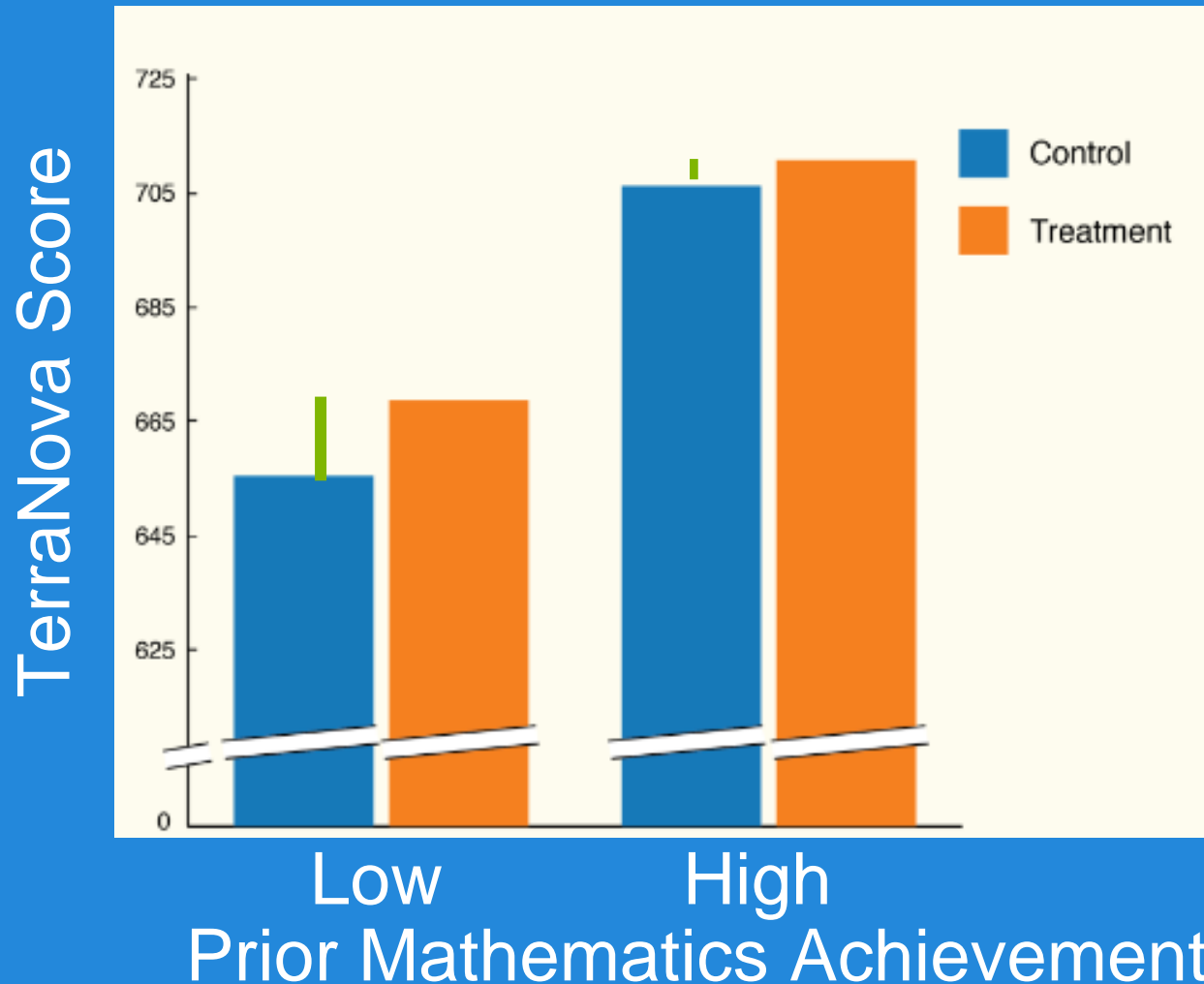
ASSISTments

A Free Public Service of Worcester Polytechnic Institute

What it's not…

McGraw Hill Education | ALEKS®

CARNEGIE LEARNING

IXL

KHANACADEMY

The New York Times Magazine

The Machines Are Taking Over

By ANNIE MURPHY PAUL   SEPT. 14, 2012

Neil Heffernan was listening to his fiancée, Cristina Lindquist, tutor one of her students in mathematics when he had an idea. Heffernan was a graduate student in computer science, and by this point — the summer of 1997 — he had been working for two years with researchers at Carnegie Mellon University on developing computer software to help students improve their skills. But he had come to believe that the programs did little to assist their users. They were built

Illustration by Tim Enthoven

Crowdsourcing Teacher and Student Content and Feedback tiny.cc/NBER

Must Stay Free

Jeremy Roschelle @roschelle63

Can we improve math homework? Presented @SRI_Education @assistments positive findings at the White House Symposium on Digital Learning today

"We only care about effect sizes that are over 0.2."
- White House OSTP, Tom Kalil

Thomas Kalil

Kalil in late 2012

Jeremy Roschelle @roschelle63

Can we improve math homework? Presented @SRI_Education @assistments positive findings at the White House Symposium on Digital Learning today

DIRECTOR OF IES

THOMAS W. BROCK

Commissioner, National Center for Education Research Delegated the Duties of the Director of the Institute of Education Sciences

- IES Director, Tom Brock

Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms

Mark W. Lipsey, Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry,
Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick

:ies NATIONAL CENTER FOR SPECIAL EDUCATION RESEARCH
Institute of Education Sciences

Tran...
Rep...
Education Interventions Into More
Readily Interpretable Forms

And Tom Kalil "got it"

October 24, 2016

Subject: A Better way to think about Effect Sizes

Dear Mr. Kalil,

Nice meeting you Tom, at the OSTP event. I promised to follow up with a pointer to the IES recommendation on how to think about effect sizes. I spoke with several people from the US Dept of Education and asked them, "Are you going to tell Tom that he is using statistics badly and that he needs to read the IES report on Lipsey?" Instead, they encouraged me to reach out to you myself. I apologize if this seems condescending as that is definitely not my intent but please bear with me here. I think I have an interesting and possibly (at least in my mind) better way for you to think about effect sizes.

ies NATIONAL CENTER FOR
SPECIAL EDUCATION RESEARCH
Institute of Education Sciences

Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms

NOVEMBER 2012

Mark W. Lipsey, Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry,
Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick

"In his landmark book on statistical power, Cohen (1977, 1988) … dubbed 0.20 as "small," 0.50 as "medium," and 0.80 as "large." Ever since, these values have been widely cited as standards for assessing the magnitude of the effects found in intervention research despite Cohen's own cautions about their inappropriateness for such general use."
- Lipsey, et al. 2012

"McCartney and Rosenthal (2000) report that for interventions related to heart attacks, the best interventions fall well below the 0.2, which Cohen characterized as "small," but those interventions sometimes correspond to reducing the incidence of heart attacks by about half - an effect of enormous practical significance."
- Lipsey, et al. 2012

Lipsey, et al. indicates we need a better way to think about the practical significance of the size of an effect

"The widespread indiscriminate use of Cohen's generic small, medium, and large effect size values to characterize effect sizes … is thus likewise inappropriate and misleading."
- Lipsey, et al. 2012

Effect size is easy to game

Effect Size = $\dfrac{\text{(Control - Treatment)}}{\text{(Pooled Standard Deviation)}}$

How to game it: Make the standard deviation of the control small

Real measures that folks care about (i.e., standardized achievement tests) are not gameable in this way

Another way to game the effect size:

Use specialized measures (… not evil)

Table 9. Achievement effect sizes from randomized studies broken out by type of test and grade level

Type of Test	Grade Level	N of Effect Sizes	Median	Mean	Standard Deviation
Specialized Topic or Test, Researcher Developed	Elementary	230	.34	.40	.55
Standardized Test, Narrow Scope	Elementary	374	.17	.25	.42
Standardized Test, Broad Scope	Elementary	89	.07	.08	.27

Another way to game the effect size:

Have the control group do nothing, or something obvious (re-read the chapter)

By comparing to the real world (a control group that is also being taught math for a standard test), effects are harder to achieve.

"active comparator"= current standard of care

So what is the right thing to do?

Lipsey's approach allows for better understanding of the practical significance of your effect

Assessing the Practical Significance of Intervention Effects

1. Benchmarking Against Normative Expectations for Academic Growth
2. Benchmarking Against Policy-Relevant Performance Gaps
3. Benchmarking Against the Observed Effect Sizes for Similar Interventions
4. Benchmarking Effects Relative to Cost

Benchmarking Against Normative Expectations for Academic Growth

	Control	Treatment	Difference	Effect Size
Posttest	88	90	2	.2

Benchmarking Against Normative Expectations for Academic Growth

	Control	Treatment	Difference	Effect Size
Pretest	???	???		
Posttest	88	90	2	.2

Benchmarking Against Normative Expectations for Academic Growth

Figure 1. Pre-post change for the three scenarios with the same posttest difference
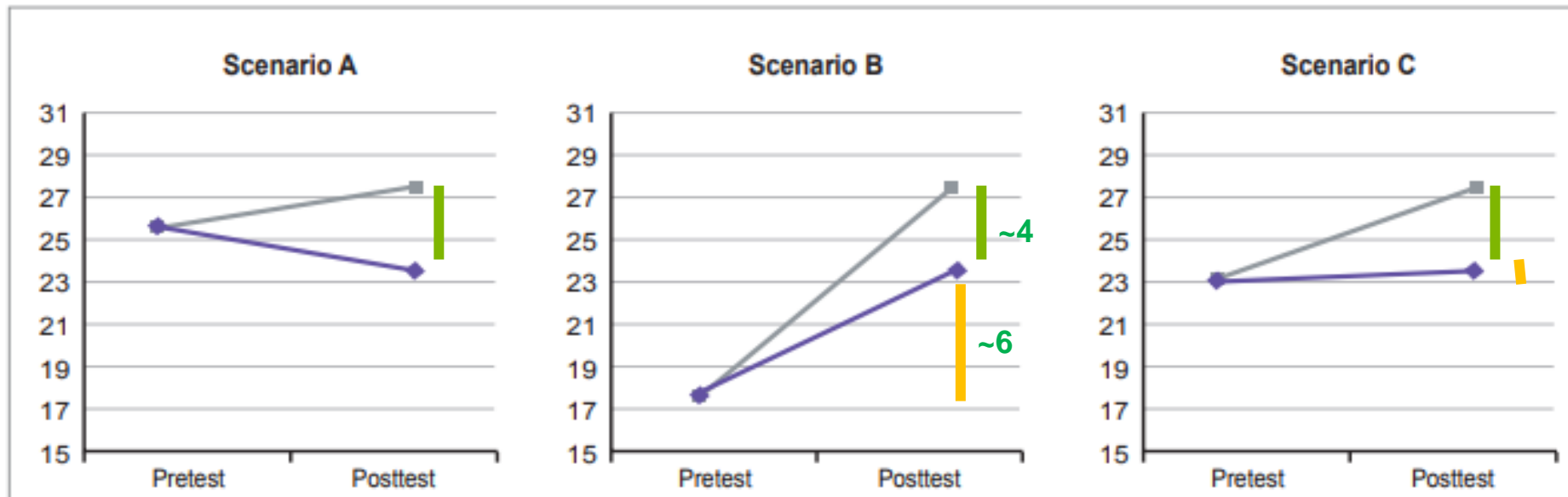
4/6= 66% improvement

4/.01= 1000% improvement

Table 5. Annual achievement gain: Mean effect sizes across seven nationally-normed tests

Grade Transition	Reading	Math	Science	Social Studies
Grade K - 1	1.52			
Grade 1 - 2	0.97			
Grade 2 - 3	0.60			
Grade 3 - 4	0.36			
Grade 4 - 5	0.40			
Grade 5 - 6	0.32			
Grade 6 - 7	0.23			
Grade 7 - 8	0.26			
Grade 8 - 9	0.24			
Grade 9 - 10	0.19			
Grade 10 - 11	0.19			
Grade 11 - 12	0.06			

NOTES: Adapted from Bloom,
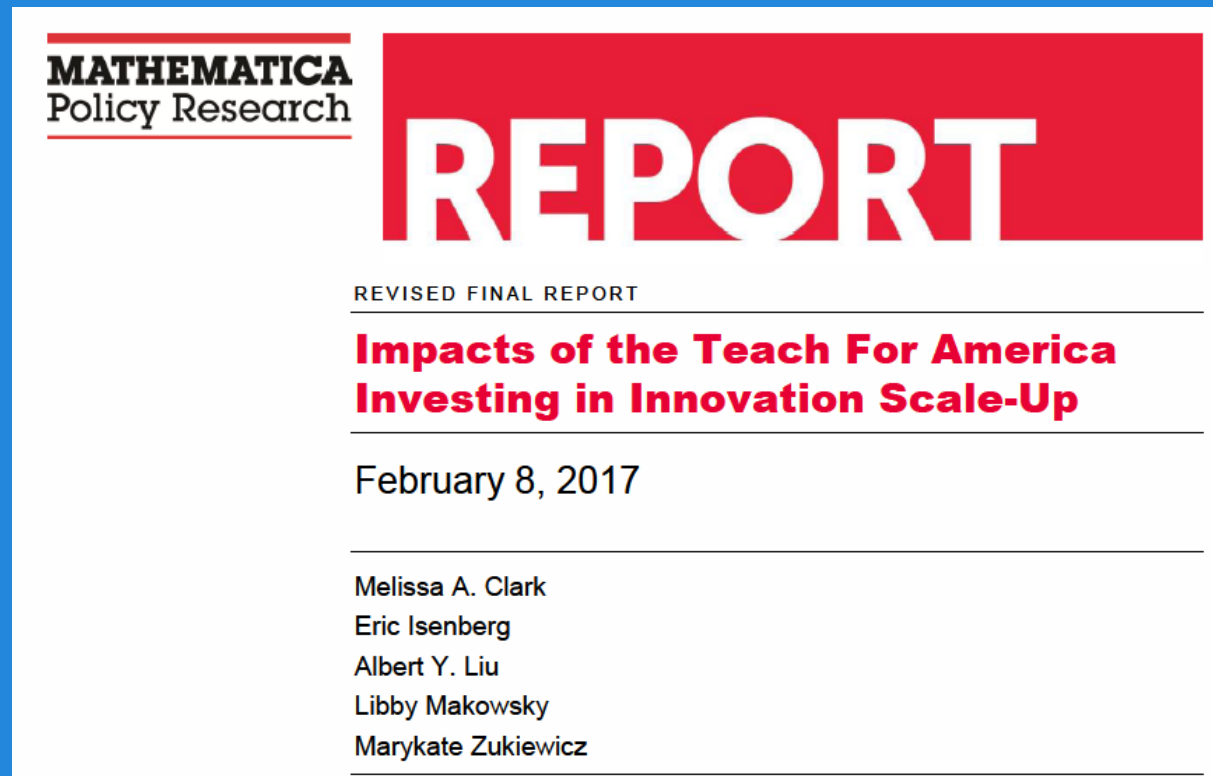The means shown are the simpl
tests: CAT5, SAT9, Terra Nova-C

.15/.60=25%

.15/.19=~100%

On standardized tests of achievement, the amount learned each year goes down so small effect sizes are mor meaningful.

"An effect size of, say
intervention will, therefore,
annual gain otherwise expec
judged to have practical s
same effect size for tenth
their annual gain and, by t
stupendous effect" – Lipsey, et al. 2013

"We also found that TFA teachers in grades 1 and 2 had a positive effect on student math achievement of 0.16 standard deviations, or about 1.5 additional months of learning."

MATHEMATICA
Policy Research

REPORT

REVISED FINAL REPORT

Impacts of the Teach For America Investing in Innovation Scale-Up

February 8, 2017

Melissa A. Clark
Eric Isenberg
Albert Y. Liu
Libby Makowsky
Marykate Zukiewicz

Take Aways…

Anytime anyone tells you just an effect size, ask them to apply the IES Lipsey et al., report to interpret the effect size. How much growth would you normally get in a year?

Ask about the measures and how aligned they are. Experimenter created measures will tend to report bigger effect sizes.
Ask what grade level.

All metrics are gamable, but Lipsey's recommendation are more likely to lead to sounder public policy