

Building on the Basics:

The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects

Marcus A. Winters
Manhattan Institute for Policy Research

Jay P. Greene
Manhattan Institute
University of Arkansas

Julie Trivitt
Arkansas Tech University

Accountability Testing

- General strategy is to hold schools accountable for teaching basic skills
 - Widespread belief that students in worst schools don't have even basic proficiency
- Students are tested in basic subjects (most often math and reading) and someone is held accountable for the results through sanctions or rewards
- Most states had adopted some sort of high-stakes testing regime before NCLB makes it universal

Effect on High-Stakes Subjects

- Previous research has focused on the impact of accountability testing on the high-stakes subjects
 - The subjects most intended to improve
 - Test scores are available
- Evidence has been generally positive, though with some mixed results

Potential Negative Externalities

- Even if it improves average proficiency in high-stakes subjects, there are many potential limitations to accountability testing
- One particularly strong criticism is that it tends to narrow the curriculum
 - Focus on particular students
 - Low-achieving students (Fordham Foundation, 2008)
 - Regular education students
 - Focus on particular subjects

What About the Low-Stakes Subjects?

- There is a clear incentive to focus on the tested subjects and ignore the rest
 - What about science, social studies, and the arts?
- Many worry that accountability testing in some subjects harms proficiency in others
 - Barack Obama has explicitly made this case against NCLB
 - Groups as different as the teacher unions and the Fordham Foundation
- Anecdotes that schools spend less time in low-stakes subjects
- Rouse, et al. (2007) find that 86% of schools facing sanctions under Florida's policy require minimum time spent on tested subjects, while only 61% of them have such a policy for social studies

Goal of the Paper

- Evaluate the impact of the threat of sanctions under Florida's high-stakes testing policy in math and reading on student proficiency in science
 - Florida's program one of the most aggressive in the nation, and a template for NCLB
- Adopt a proficiency output approach rather than a reported input approach
 - Interested in changes in low-stakes proficiency rather than changes in school behavior

Florida's A+ Program

- Schools receive points based primarily on the results of student scores on the state's math and reading exams in grades 3 – 10, which determines a grade it receives from A-F
 - Schools that earn fewer than 280 points earn an F-grade
 - Grades were given prior to 2001-02, but with no point system
- Two potential sanctions for F-grade
 - Shame of being deemed “failing”
 - If receive two F's in four-year window students offered vouchers
 - The voucher provision is no longer in effect
- Several prior studies have found F-schools have made gains in math and reading
 - Though, there is some disagreement as to the cause

Science Exam

- Students in grades 5, 8, and 10 are also administered a science test
- During year studied here (2002-03), the results of the science exam were not used for any accountability purpose
 - Science exam is now used in the school grading system
 - This is similar to NCLB, which requires students be tested in science but does not use the scores in the calculation of AYP
- No prior research of which we are aware has evaluated the impact of A+ on science scores

Data

- Individual student data for the universe of fifth grade students in 2002-03
- Dataset contains test scores in math, reading, and science as well as demographic information about the student
- Supplement this with information about the school the student attended
 - In particular, number of points and grade school earned at end of 2001-02

Strategy

- Adopt the regression-discontinuity strategy used by Rouse et al. (2007) to study effect of F-sanction on math and reading proficiency
 - Found students in F-graded schools made improvements in math and reading
- Takes advantage of the known point cut-off at which schools receive particular grades

Econometric Approach

- Estimate a cross-sectional regression
 - Dependent variable: test score in 02-03
 - Independent variables include a function of the number of points earned by student's school at end of 2001-02 and grade it earned due to the points
 - Also control for student demographics and prior measure of proficiency
 - Estimate models in math, reading, and science
- The primary assumption of the model is that the grade itself contains no additional information about the quality of the school that is not contained in the number of points earned
 - If true, then the estimate of the grade-level variables can be interpreted as a causal effect of the grade itself

Intuition of Approach

- When studying impact of any program, must account for heterogeneity of schools
- The point system is a continuous scale that measures several aspects of the school's achievement, while the grade is a dichotomous measure
- Imagine 2 schools – One that earns 279 points (F) and one that earns 281 points (D) under the system
 - Schools are nearly identical in productivity
 - They face very different incentives under the policy

Model

$$T_{ist} = \delta_0 + \delta_1 f(T_{ist-1}) + \delta_2 \text{Grade}_{st} + \delta_3 X_{ist} + \delta_4 f(\text{POINTS}_{st-1}) + \rho_{ist}$$

T_{ist} = score of student i , in subject s , year t (02-03)

$f(T)$ = cubic function of prior score

Grade = school grade at end of 2001-02

X = observed characteristics

$f(\text{POINTS})$ = cubic function of school points, 01-02

ρ = stochastic term, clustered by school

Difficulty in Science Measurement

- Can estimate equation directly in math and reading
- However, we do not observe prior student science score
 - 4th grade students not tested in science
- Instead, we replace prior science score with prior scores in math and reading as proxy
 - Assume no differential relationship in knowledge in these subjects across schools prior to sanction
 - Results are similar when we calculate prior proficiency score in a way that includes other observed characteristics as well
 - Results are similar when use only levels and do not include measure of prior proficiency

Results

- Confirm Rouse et al.'s prior results in math and reading in F-schools
 - Reading improved 0.086 standard deviations
 - Math improved 0.175 standard deviations
 - Both results significant at 1% level
 - Very similar to those reported by Rouse et al.
- Science proficiency improved 0.087 standard deviations in F-schools
 - Significant at 5% level

How Could HS Sanctions Improve LS Proficiency?

- Positive science effect is puzzling
 - Schools clearly have an incentive to ignore science in favor of math and reading
- Two potential explanations:
 - Schools could adopt school-wide reforms (systemic effect)
 - Science proficiency could be correlated with learning in math and reading (correlation effect)

Systemic or Correlation Effect?

- Ran a similar regression model as before
 - Dependent variable: science score
 - Independent variables now include student's 2002-03 test score gain in math and reading (a measure of the correlation effect)
 - Can now think of the estimate of the school grade as the systemic effect

Model

$$Science_t = \psi_0 + \psi_1 Science_{t-1} + \psi_2 X_{ist} + \psi_3 [Math_t(F) - Math_{t-1}] + \psi_4 [Read_t(F) - Read_{t-1}] + \psi_5 F_{ist-1} + \mu_{ist}$$

- Model incorporates fact that 2002-03 math and reading scores are a function of F grade
- Again, $Science_{t-1}$ is a proxy using prior math and reading
- Incorporate a cubic function of math and reading gains

Results

- Appears to be strong correlation effect
 - All levels of cubic functions for math and reading gains are positive and statistically significant
- No real evidence of a systemic effect
 - Coefficients on each of the grades (A-F) are statistically insignificant

Limitations

- We are limited to evaluating science proficiency in elementary school only
 - Perhaps a different result in other grades
- Science may be more susceptible to a correlation effect than other subjects
 - Jacob (2005) found positive result in social studies as well as science in Chicago
 - The impact in the arts is particularly unclear

Summary

- Evaluated the impact of an aggressive high-stakes testing policy on student proficiency in an important low-stakes subject
- Evidence indicates that HS testing in math and reading could improve student science proficiency by improving the basic skills