Can Assessment Programs Reduce the Achievement Gap between Low and High Achievers?

Spyros Konstantopoulos
Michigan State University

Wei Li
University of Missouri

Shazia Miller
American Institutes for Research

Arie van der Ploeg

Study Rationale

- In the U.S. the No Child Left Behind (NCLB) Act about school accountability resulted in a abundance of assessment-based school interventions that aim to improve student performance.

- Interim assessments are school interventions designed to help teachers use evidence about student performance to modify and differentiate instruction to facilitate learning.

- These assessments are administered three or four times during the school year (certain time periods) and are viewed as promising mechanisms to increase student achievement.

Hypothesis

- Interim assessments are designed to help teachers identify areas for instructional need by providing immediate, detailed insight on student strengths and weaknesses.

- They are data-driven interventions that are likely to improve ongoing classroom instruction, and provide useful feedback to students, teachers, and administrators.

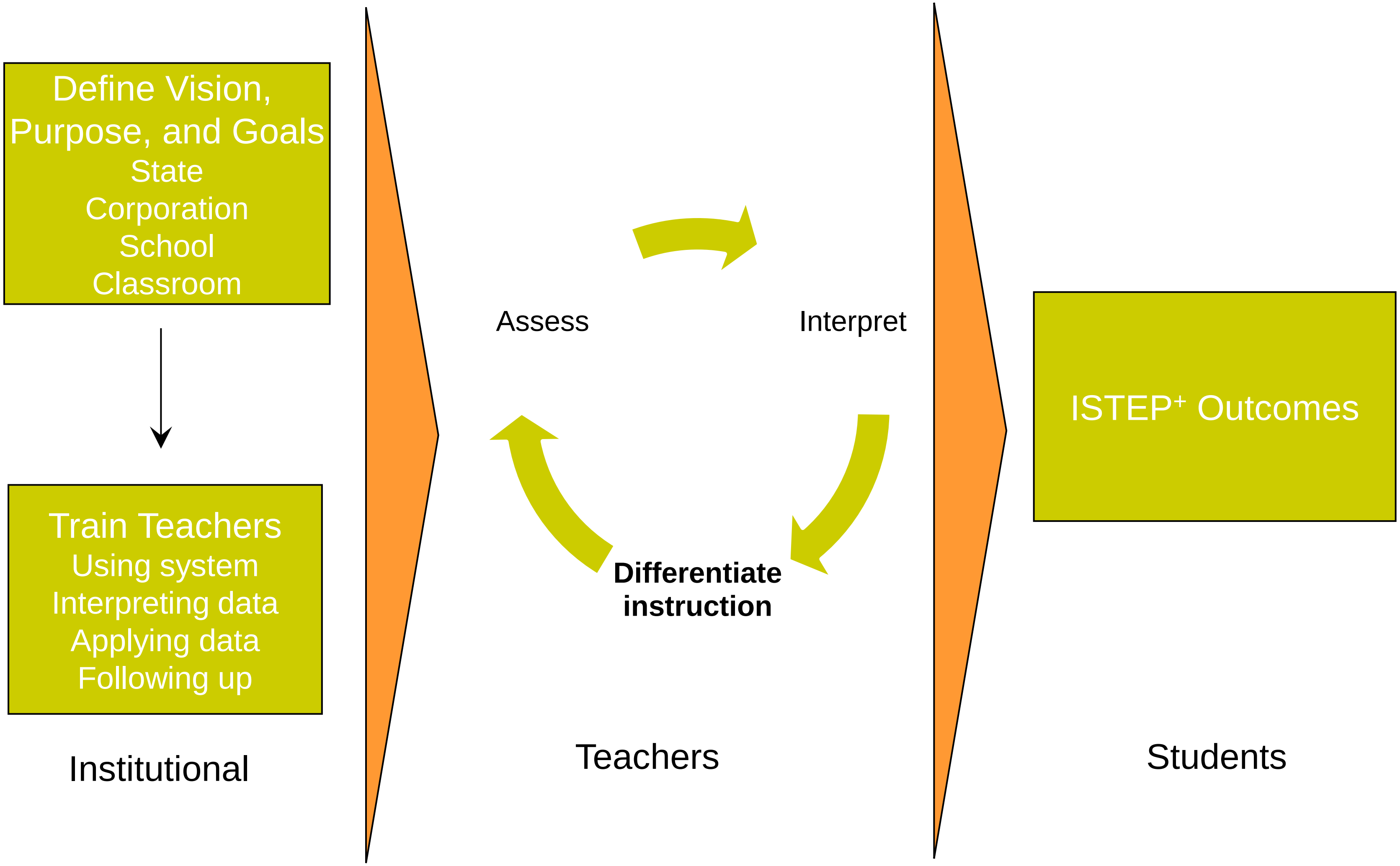
- Therefore, our research hypothesis is that students who have teachers who work with interim assessments will have different average achievement than those who did not receive these assessments.

Interim Assessments

- Involve an iterative process designed to aid teachers use assessment-based evidence to modify instruction and facilitate learning.

- These assessments should result in student achievement improvements. Constructive feedback and differentiated instruction should meet students’ learning needs and improve student learning further.

Logic Model


Define Vision, Purpose, and Goals
State
Corporation
School
Classroom
Train Teachers
Using system
Interpreting data
Applying data
Following up
Institutional
Assess
Interpret
Differentiate instruction
Teachers
ISTEP+ Outcomes
Students

Interim Assessments in the State of Indiana in the U.S.

- The state of Indiana has been implementing a voluntary interim assessment program since 2008-2009.

- An important objective was to help students perform well in state accountability tests (Indiana's state tests).

- Indiana expects teachers to use the constantly updated diagnostic information about student learning to improve ongoing instruction for individual students and ultimately increase student achievement.

Interim Assessments in the State of Indiana in the U.S.

- In 2008 the Indiana Department of Education (IDOE) began the roll-out of its system of interim assessments.

- The plan required that assessments be voluntary. In schools that chose to use them, IDOE would cover costs. The plan also tasked IDOE to ensure alignment of test content to Indiana standards and grade-level expectations.

- Two commercial products were identified. The first program was Wireless Generation's *mCLASS* as the K–2 solution, and the second program was CTB/McGraw-Hill's *Acuity* product for Grades 3–8.

Related Literature

- Research evidence on the effects of interim assessments on student achievement is by and large mixed.

- Earlier studies have reported effect sizes of 0.40 standard deviations for formative assessments (Black & William, 1998). Some evidence that formative assessments decrease the achievement gap.

- Meta-analyses have reported effect sizes of smaller magnitude for formative assessments (between 0.20 and 0.30 standard deviations) (e.g., Kingston & Nash, 2011).

- However, recent evidence from quasi-experiments points to marginal significant or insignificant effects (Faria et al., 2012; Quint, Sepanik, & Smith, 2008).

Related Literature

- Large-scale experiments have reported effects of smaller magnitude (< 0.10 standard deviations).
- Carlson, Borman, & Robinson (2011) have reported small significant effects in mathematics in grades 3-8 (but not in reading).
- Slavin et al. (2011, 2013) however, have reported significant effects on both mathematics and reading.
- Cordray, Pion, Brant, & Molefe (2012) found no effects on reading achievement in grades 4-5.
- Konstantopoulos, Miller, and van der Ploeg (2013) have reported significant effects mainly in mathematics in grades 5 and 6
- Black and William (1998) have documented that assessments can benefit low-achievers.

Study Objective

- Examine the effectiveness of interim assessments across the distribution of student achievement (mathematics or reading) in K-8 public schools in Indiana U.S. Focus is on low-achievers (lower tail of achievement distribution). Emphasis on student performance on the state's annual Indiana Statewide Testing for Educational Progress–Plus (ISTEP^{+}) measure.
- Interim assessments should help teachers identify areas for instructional need for low-achievers and then modify their instruction accordingly to improve their performance.
- Use data from a randomized control trial (RCT) or experiment. The duration of the experiment was one year (2009-2010).

The Mechanism

- Interim assessments provide detailed insight on students' learning (strengths and weaknesses)

- Teachers use these assessment data to change their instruction (differentiated instruction) and maximize learning.

- Through interim assessments low-achievers' instructional needs should be easily indentified and via appropriate differentiated instruction low-achievers should benefit more than other students.

The Mechanism

- Indiana expected these assessments to vary by achievement level. We expected a decrease in the achievement gap under the assumption that teachers may have more chances to help low-achievers. That is, we expected greater gains in the lower tail of the achievement distribution (i.e., effect varies across the achievement distribution).

- There is no evidence thus far in the literature except for the study by Black and Wiliam (1998). We intend to fill in that gap.

Research Design

- We designed a large-scale cluster randomized experiment where schools (the clusters) were assigned randomly to an intervention/treatment (interim assessment) or to a control group (business as usual).

- The experiment took place in Indiana in 2009-2010 and included K-8 public schools that had volunteered to implement diagnostic assessments in the Spring of 2009.

Research Design

- Initial sample was 57 schools that were randomly assigned to control or treatment conditions (35 treatment and 22 control). Unbalanced design to facilitate school participation in the experiment.

- Potential contamination, district requirements, school closures and refusal to participate reduced the pool of schools to 50 (32 treatment and 18 control). These schools participated in the study the whole academic year.

- Over 25,000 students participated in the study in 2009-2010.

Statistical Analysis

- The statistical method we employed was chosen carefully to address the research question of interest (is the treatment effect uniform across the achievement or more pronounced in the lower tail?).

- Quantile regression is appropriate because it produces estimates in different quantiles (or percentiles) of the achievement distribution. We used quantile regression to estimate treatment effects at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles (or percentiles).

Statistical Analysis

- We conducted analyses using data across all grades (i.e., k through 8), lower grades (i.e., k to 2), or upper grades (i.e., 3 to 6). The latter analyses allowed us to determine the effects of *mCLASS* or *Acuity* separately. We also conducted within grade analyses for each grade separately to determine whether effects varied by grade. Finally, we conducted analyses in grades 4, 5, and 6 with prior scores as an additional covariate.

- Standard errors of estimates were corrected for clustering and heteroscedasticity.

Intention to Treat (ITT) or Treatment on the Treated (TOT)

- First we conducted analysis on the sample of schools that were randomly assigned to either the treatment or the control condition (ITT analysis). This analysis should produce an unbiased treatment effect.

- We also conducted sensitivity analysis on the sample of participating schools. (TOT analysis). This analysis provides evidence about the robustness of the estimates.

Statistical Model for ITT or TOT Analysis

- At each quantile the regression model for student *i* is

$$y_i = \beta_0 + \beta_1 Treatment_j + \boldsymbol{X}_i \boldsymbol{B}_2 + \boldsymbol{Z}_j \boldsymbol{B}_3 + \boldsymbol{G}_i \boldsymbol{B}_4 + \varepsilon_i$$

- where *y* is mathematics or reading scores, ***X*** represents student predictors, ***Z*** represents school predictors, ***G*** represents grade fixed effects, and ε is a student error.
- The within-grade models were similar, only grade effects were no longer in the regression equation.

Potential Caveats with ITT or TOT Analyses

- ITT analyses assume that all treatment schools assigned randomly to treatment actually received treatment (but some treatment schools and some control schools did not participate in the experiment). That is, not all schools complied with their assignment.

- TOT analyses may produce a biased treatment effect estimate if nonparticipation is not random (i.e., the remaining treatment and control schools are not equal on average). These analyses do not take into account possible selection due to nonparticipating schools

Potential Caveats with ITT or TOT Analyses

- One solution is to conduct and IV (Instrumental Variables) analysis that eliminates selection due to unobserbvables (unmeasured variables).

- This analysis address potential weaknesses of the ITT or TOT analyses.

IV Analysis

- In step 1 we used logistic regression where a dummy for treatment actually received is regressed on random assignment to treatment and other covariates. In this stage the probabilities that schools have complied with their assignment is computed (i.e., schools assigned to treatment actually receive treatment).

- In step 2 these probabilities computed in step 1 are used to create weights to model the propensity of an assigned treatment school actually receiving the treatment (see Abadie, Angrist, & Imbens, 2002). In step 2 we used quantile regression (with weights). The command *ivqte* was used in STATA.

Variables

- Dependent variables: Mathematics and reading scores (Terra Nova for grades K-2, ISTEP+ for grades 3-8).

- Independent variables:
  - Treatment (receive *mCLASS and Acuity* or not).
  - Student level: gender, race, low SES, age, special education, and limited English proficiency (LEP).
  - School level: percent of minority, female, low SES, and LEP students.

Results

Table 1. Random Assignment Check Using Observed Variables

Variable	M_d	SE_d	P-value	ES
Grades K to 2: 57 Schools				
Proportion of Female Students	-0.008	0.010	0.442	-0.233
Proportion of Minority Students	0.009	0.080	0.914	0.029
Proportion of Disadvantaged Students	0.007	0.053	0.897	0.033
Proportion of Special Education Students	0.016	0.030	0.603	0.117
Proportion of Limited English Proficiency Students	0.025	0.014	0.084	0.391
Grades 3 to 8: 57 Schools				
Proportion of Female Students	-0.006	0.010	0.569	-0.168
Proportion of Minority Students	0.016	0.078	0.839	0.054
Proportion of Disadvantaged Students	0.009	0.052	0.856	0.046
Proportion of Special Education Students	0.011	0.031	0.730	0.078
Proportion of Limited English Proficiency Students	0.025	0.014	0.079	0.399
Spring 2009 Math Scores	2.672	6.298	0.673	0.115
Spring 2009 ELA Scores	0.343	5.453	0.950	0.017

Note: Md = Difference Between Treatment and Control Group School Means;
SEd = Standard Error of the Mean Difference; ES = Effect Size Reported is Hedges g.

Table 2. Desriptive Statistics of Variables of Interest

	N	Mean	SD
TerraNova Reading Score	7640	535.4	59.2
TerraNova Mathematics Score	7667	566.5	55.3
ISTEP English Language Arts Score	13287	481.6	64.4
ISTEP Mathematics Score	13307	495.8	72.2
Age (months)	25477	112.3	25.30
Female	12339	0.48	0.50
Male	13185	0.52	0.50
Race			
White	19470	0.77	0.42
Black	3776	0.15	0.36
Latino	1123	0.04	0.21
Other	1058	0.04	0.20
Limited English Proficiency - LEP	701	0.03	0.16
Non-LEP	24899	0.97	0.16
Free or Reduced-Price Lunch - FRPL	13501	0.53	0.50
Non-FRPL	12067	0.47	0.50
Special Education - SE	1580	0.06	0.24
Non-SE	24020	0.94	0.24

Table 3. Quantile Regression Estimates of Treatment Effects in Mathematics and Reading Achievement: ITT Analysis

	Mathematics					Reading				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Grades K to 8										
Treatment Effect	0.149*	0.136	0.112	0.118	0.117	0.067	0.058	0.056	0.021	0.007
SE	0.073	0.070	0.062	0.073	0.087	0.060	0.048	0.053	0.054	0.047
Number of Schools	57					57				
Number of Students	20792					20795				
Grades 3 to 8										
Treatment Effect	0.203*	0.194*	0.175*	0.157*	0.179*	0.114*	0.065	0.060	0.028	0.014
SE	0.074	0.075	0.074	0.076	0.061	0.057	0.046	0.042	0.049	0.050
Number of Schools	57					57				
Number of Students	13274					13254				
Grades K to 6										
Treatment Effect	0.155	0.136*	0.116	0.113	0.099	0.076	0.067	0.064	0.036	0.024
SE	0.084	0.065	0.080	0.067	0.085	0.041	0.047	0.040	0.059	0.060
Number of Schools	57					57				
Number of Students	20107					20107				
Grades 3 to 6										
Treatment Effect	0.214*	0.205*	0.176*	0.155*	0.172*	0.128*	0.082	0.072	0.047	0.030
SE	0.063	0.071	0.063	0.065	0.081	0.060	0.056	0.047	0.043	0.047
Number of Schools	57					57				
Number of Students	12589					12566				

*p ≤ .05; Note: SE = Standard Error

Table 4. Quantile Regression Estimates of Treatment Effects in Mathematics and Reading Achievement: TOT Analysis

	Mathematics					Reading				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Grades K to 8										
Treatment Effect	0.158*	0.144	0.118	0.121	0.111	0.071	0.057	0.053	0.023	0.017
SE	0.065	0.077	0.082	0.071	0.085	0.055	0.062	0.057	0.055	0.068
Number of Schools	50					50				
Number of Students	19859					19861				
Grades K to 2										
Treatment Effect	0.023	0.027	0.019	0.016	-0.029	-0.018	-0.010	0.026	-0.022	0.001
SE	0.093	0.097	0.078	0.101	0.130	0.088	0.103	0.074	0.072	0.101
Number of Schools	44					44				
Number of Students	7517					7540				
Grades 3 to 8										
Treatment Effect	0.227*	0.206*	0.189*	0.176*	0.185*	0.135*	0.075	0.062	0.028	0.024
SE	0.070	0.064	0.068	0.079	0.086	0.067	0.044	0.039	0.055	0.057
Number of Schools	50					50				
Number of Students	12342					12321				
Grades K to 6										
Treatment Effect	0.171*	0.145*	0.123*	0.119	0.104	0.079	0.071	0.063	0.041	0.034
SE	0.067	0.065	0.062	0.080	0.075	0.054	0.051	0.049	0.059	0.054
Number of Schools	50					50				
Number of Students	19174					19173				
Grades 3 to 6										
Treatment Effect	0.239*	0.220*	0.192*	0.175*	0.183*	0.150*	0.093	0.072	0.055	0.047
SE	0.078	0.062	0.065	0.066	0.076	0.059	0.051	0.046	0.048	0.051
Number of Schools	50					50				
Number of Students	11657					11633				

*p ≤ .05; Note: SE = Standard Error

Table 5. Quantile Regression Estimates of Treatment Effects in Mathematics and Reading Achievement: IV Analysis

	Mathematics					Reading				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Grades K to 8										
Treatment Effect	0.156*	0.136*	0.113	0.118	0.116	0.064	0.056	0.051	0.023	0.024
SE	0.078	0.069	0.067	0.064	0.078	0.058	0.054	0.051	0.054	0.060
Number of Schools	57					57				
Number of Students	20792					20795				
Grades 3 to 8										
Treatment Effect	0.218*	0.203*	0.177*	0.160*	0.185*	0.107	0.062	0.058	0.031	0.030
SE	0.081	0.073	0.072	0.063	0.078	0.061	0.058	0.047	0.051	0.049
Number of Schools	57					57				
Number of Students	13274					13254				
Grades K to 6										
Treatment Effect	0.165*	0.143*	0.112	0.114	0.104	0.071	0.066	0.059	0.038	0.038
SE	0.073	0.061	0.064	0.068	0.077	0.056	0.055	0.054	0.049	0.058
Number of Schools	57					57				
Number of Students	20107					20107				
Grades 3 to 6										
Treatment Effect	0.235*	0.210*	0.177*	0.159*	0.178*	0.124*	0.078	0.071	0.052	0.055
SE	0.081	0.063	0.070	0.065	0.076	0.060	0.057	0.047	0.048	0.050
Number of Schools	57					57				
Number of Students	12589					12566				

*p ≤ .05; Note: SE = Standard Error

Findings Across Grades

- ITT, TOT, and IV analyses produced similar results. Evidence is robust.
- Treatment effect is positive, but not systematically significant (statistically).
- Treatment is small and statistically insignificant in early grades (k-2). Treatment is larger in upper grades (3-8).
- Treatment is larger and statistically significant in upper grades (3-8) in mathematics.
- Treatment is typically larger for low-achievers. Some weak evidence to support our hypothesis.

Table 6. Grade Specific Quantile Regression Estimates of Treatment Effects in Mathematics an Reading Achievement: IV Analysis

	Mathematics					Reading				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Grade K										
Treatment Effect	0.092	0.020	0.032	0.008	0.003	-0.037	-0.087	-0.041	-0.117	-0.140
SE	0.132	0.126	0.125	0.176	0.227	0.138	0.124	0.133	0.140	0.199
Number of Schools	39					39				
Number of Students	2386					2404				
Grade 1										
Treatment Effect	0.045	0.031	-0.028	-0.014	0.037	-0.067	0.004	0.020	0.020	0.080
SE	0.150	0.098	0.088	0.093	0.114	0.149	0.121	0.089	0.093	0.117
Number of Schools	38					38				
Number of Students	2473					2471				
Grade 2										
Treatment Effect	0.137	0.010	0.027	0.023	-0.064	0.073	0.059	0.103	0.009	-0.092
SE	0.101	0.094	0.083	0.078	0.129	0.095	0.072	0.059	0.077	0.095
Number of Schools	44					44				
Number of Students	2658					2665				
Grade 3										
Treatment Effect	0.208*	0.176	0.157	0.101	0.131	0.215*	0.109	0.095	0.047	0.030
SE	0.093	0.093	0.092	0.090	0.104	0.092	0.063	0.054	0.067	0.085
Number of Schools	57					57				
Number of Students	3741					3737				
Grade 4										
Treatment Effect	0.128	0.093	0.091	0.045	0.087	0.115	0.123	0.093	0.060	0.043
SE	0.121	0.083	0.101	0.108	0.138	0.093	0.067	0.063	0.059	0.057
Number of Schools	57					57				
Number of Students	3739					3730				
Grade 5										
Treatment Effect	0.279*	0.262*	0.240*	0.290*	0.392*	0.107	0.059	0.087	0.102	0.154
SE	0.116	0.102	0.086	0.105	0.139	0.098	0.075	0.072	0.077	0.086
Number of Schools	56					56				
Number of Students	3509					3502				
Grade 6										
Treatment Effect	0.276	0.313	0.287	0.198	0.117	0.005	-0.080	-0.109	-0.203	-0.135
SE	0.253	0.179	0.151	0.154	0.172	0.115	0.103	0.121	0.148	0.134
Number of Schools	26					26				
Number of Students	1600					1597				

*p ≤ .05; Note: SE = Standard Error

Table 7. Grade Specific Quantile Regression Estimates of Treatment Effects in Mathematics and Reading Achievement: IV Analysis that Controls for Prior Scores

	Mathematics					Reading				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Grade 4										
Treatment Effect	0.107	0.056	0.060	0.037	0.082	0.139*	0.111	0.111*	0.092*	0.080
SE	0.100	0.087	0.092	0.109	0.113	0.065	0.066	0.037	0.043	0.054
Number of Schools	57					57				
Number of Students	3604					3572				
Grade 5										
Treatment Effect	0.251*	0.184*	0.209*	0.203*	0.264*	-0.033	-0.028	-0.028	-0.029	0.032
SE	0.081	0.082	0.082	0.087	0.111	0.071	0.051	0.049	0.055	0.073
Number of Schools	56					56				
Number of Students	3375					3357				
Grade 6										
Treatment Effect	0.062	0.091	0.119	0.114	0.187	-0.052	-0.095	-0.024	-0.076	0.009
SE	0.168	0.152	0.161	0.168	0.217	0.119	0.104	0.081	0.091	0.100
Number of Schools	26					26				
Number of Students	1505					1498				

*$p \leq .05$; Note: SE = Standard Error

Findings Within Grades

- ITT, TOT, and IV analyses produced similar results. Evidence is robust.
- Evidence of positive *Acuity* effects in grades 3-6.
- Effects were larger for low-achievers in grades 3-6.
- The evidence is not systematic nor conclusive.
- Perhaps interim assessments are more effective in later grades and in some cases for low-achievers.
- Estimates in lower tail are typically larger than those in upper tails, but not significantly different.

Limitations

- Terra Nova test is not used for accountability purposes (teachers may be focusing mostly on performance on the state test, ISTEP+).

- Due to budget constraints we had no sophisticated implementation plan. Unclear how teachers used the assessment data in treatment or control conditions.

- Due to lack of data we could not capture teacher and district effects.

Acknowledgement

- This research was supported by a grant from the Institute of Education Sciences, U.S. Department of Education (R305E090005). The opinions expressed are those of the authors and do not represent the position or policies of the U.S. Department of Education, or of the Indiana Department of Education.

Thank You!

Paper Is in Press in

Educational and Psychological Measurement

Contact Info:

Spyros Konstantopoulos

College of Education

Michigan State University

spyros@msu.edu

http://spyros.wiki.educ.msu.edu/