CPRE

CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

The Impact of the GE Foundation *Developing Futures*™ in Education Program on Mathematics Performance Trends in Four Districts

RESEARCH REPORT

Philip Sirinides

Jonathan Supovitz

Namrata Tognatta

Henry May

GE Foundation

GE Foundation *Developing Futures*™ in Education
EVALUATION SERIES

April 2013
RR-74

About GE Foundation's *Developing Futures*™ in Education Program

For more than 50 years, GE Foundation has invested in education programs based on a fundamental premise: A quality education ushers in a lifetime of opportunity, which helps build a strong and diverse citizenry to work and live in an increasingly competitive world. The GE Foundation believes that a quality education can help prepare young Americans — especially those in underserved urban districts — for careers in a global economy.

The GE Foundation is addressing this education imperative by supporting high-impact initiatives that improve access to, and the equity and quality of, public education. The *Developing Futures*™ in Education program is one such endeavor, created to raise student achievement through improved mathematics and science curricula and management capacity in schools. The program has been expanded with a grant investment of over $200 million in seven targeted U.S. school districts.

School districts use their grants to develop a rigorous, system-wide mathematics and science curriculum and provide comprehensive professional development for their teachers. Working with the GE Foundation, districts have made more efficient management of human resources using GE's Six Sigma, developing educational leaders to coach others and model best practices, implementing GE's process management tools, and developing IT systems and capacity to use data to better inform decision making. More recently, with GE Foundation leadership, partner districts have increasingly focused on implementation of the new Common Core State Standards.

About the Consortium for Policy Research in Education (CPRE)

Since 1985, the Consortium for Policy Research in Education (CPRE) has brought together renowned experts from major research universities to improve elementary and secondary education by bridging the gap between educational policy and student learning. CPRE researchers employ a range of rigorous and innovative research methods to investigate pressing problems in education today.

Having earned an international reputation for quality research and evaluation, policy design and technical assistance, and dissemination and training, CPRE is a premier source of advice for education policymakers and practitioners. CPRE is known for its work in developing theory and evidence through studies of standards-based reform, education finance and resource allocation, educational leadership, assessment and data use, and instructional improvement initiatives. CPRE researchers have extensive experience conducting experimental studies, large-scale quasi-experimental research, qualitative studies, and multi-state policy surveys.

CPRE's member institutions are the University of Pennsylvania, Teachers College Columbia University, Harvard University, Stanford University, University of Michigan, University of Wisconsin-Madison, and Northwestern University.

Since 2010, CPRE has conducted the external evaluation of the *Developing Futures*™ in Education program for the GE Foundation. In addition to this report on the impact of the *Developing Futures*™ in Education program on mathematics performance trends in four districts, look for forthcoming reports on district support for the improvement of teaching and learning in the *Developing Futures*™ districts, as well as Common Core implementation and impacts.

About the Authors

Phil Sirinides, Ph.D., is a statistician and researcher with expertise in the application of quantitative research methods and the development and use of integrated data systems for public sector planning and evaluation. His interests include early childhood programing, educational leadership, accountability policy, and comprehensive school reform. As a Senior Researcher at CPRE, Sirinides plans and conducts studies in the areas of estimating intervention impacts and measuring teacher effectiveness. Formerly, as Director of Research and Evaluation for Early Childhood at the Pennsylvania Department of Education, he developed and implemented a broad research program to inform and promote the effectiveness of Pennsylvania's children and family services. He was instrumental in development of Pennsylvania's unified early childhood data system, the Early Learning Network, which is nationally recognized as a leader in generating and using evidence to promote accountable and higher quality early childhood systems. He has regularly prepared reports for policy makers, practitioners & the public, as well as for national conferences and peer-reviewed journals.

Jonathan Supovitz, Ph.D., is an Associate Professor at the University of Pennsylvania's Graduate School of Education and Co-Director of the Consortium for Policy Research in Education (CPRE). He has published findings from numerous educational studies and evaluations of school and district reform efforts and the effects of professional development on teacher and leader practice. Supovitz is an accomplished mixed-method researcher and evaluator, employing both quantitative and qualitative techniques. He has published findings from a number of educational studies, including multiple studies of programmatic effectiveness; the relationship between data use and professional development, teacher and leadership practice, and student achievement; studies of educational leadership; research on efforts to develop communities of instructional practice in schools; an examination of the equitability of different forms of student assessment; and the use of technology for evaluative data collection. His current research focuses on how schools and districts use different forms of data to support the improvement of teaching and learning. He currently directs an experimental study on the utility to teachers of linking practice data to student performance data and a study of distributed leadership. He also leads the Evidence-Based Leadership strand of the Mid-Career Leadership Program at the University of Pennsylvania.

Henry May, Ph.D. is Associate Professor of Education and Human Development at the University of Delaware. May's primary areas of expertise include methods for program evaluation, experimental and quasi-experimental design, multilevel modeling, longitudinal analysis, item response theory (IRT), and missing data theory. His current and recent research projects include the evaluation of the Reading Recovery i3 Scale-Up, a randomized evaluation of the National Institute for School Leadership, a randomized evaluation of the Ohio Personalized Assessment Reporting System, a regression discontinuity study of the America's Choice Ramp-Up to Mathematics program, and a longitudinal study of the International Baccalaureate Students' access, persistence, and performance in postsecondary education. May has extensive experience collecting and analyzing survey and activity log data from large samples of teachers and schools. He has published numerous articles in peer-reviewed journals including Educational Evaluation and Policy Analysis, School Effectiveness and School Improvement, Education Finance and Policy, Education Administration Quarterly, the Elementary School Journal, the Journal of School Leadership, the American Journal of Evaluation, and the Journal of Educational and Behavioral Statistics. He was also the primary author on an NCEE Technical Methods report from the Institute of Education Sciences on the use of state test scores in education experiments. Over the past decade, he has taught advanced statistics courses to graduate students at the University of Pennsylvania and the University of Delaware.

Namrata Tognatta is a Ph.D. candidate in the Policy, Measurement & Evaluation Division at the Graduate School of Education at the University of Pennsylvania where she has been awarded a Predoctoral Fellowship in Advanced Quantitative Methods from the Institute of Educational Sciences. Her research interests include international education development, issues in vocational education and training, program evaluation, and causal inference. Namrata has presented her work at annual meetings of the Comparative and International Education Society, the Association for Public Policy and Management, and the American Educational Research Association. Previously, Namrata has worked at the Educational Testing Service conducting research on the validity and fairness of educational assessments and at the Azim Premji Foundation on initiatives in rural elementary schools focusing on improving quality and capacity-building of the system.

Table of Contents

Executive Summary

Beginning in 2005, the GE Foundation initiated a commitment of expertise and financial resources to a set of urban school districts to improve public education and enhance student achievement in mathematics and science. With strong emphasis on stakeholder engagement, the GE Foundation's *Developing Futures*™ in Education program pursued a strategy of: (1) facilitating school board, union, and district leaders to work together to articulate system goals and priorities; (2) helping district leaders to build systemic change processes and develop internal-management capacity; and (3) supporting district science and mathematics initiatives through materials alignment, coaching, professional development, and other capacity-building measures. This report analyzes the impacts of the GE Foundation commitment to the partner districts by examining trends in student performance in mathematics over time in four districts. We hypothesized that the GE Foundation's collaborative efforts with the district educators would produce detectable and significant improvements in student outcomes.

This report analyzes the longitudinal impact of *Developing Futures*™ in four urban school districts that have worked with the GE Foundation for at least four years, including Cincinnati, Ohio; Erie, Pennsylvania; Jefferson County (Louisville), Kentucky; and Stamford, Connecticut. Using individual student records over a period of up to 10 years, we analyzed performance trends both before and after the GE Foundation began working with the districts to assess how student achievement in mathematics changed during the introduction of *Developing Futures*™. This report provides details of an interrupted time series analysis that was used to isolate the impacts of district reform efforts, as well as explore differential effects by grade level. In a separate report, CPRE researchers provide a detailed analysis of the processes that each district employed to produce these results.

Overall, we found strong evidence that the GE Foundation's efforts significantly contributed to improvements in student mathematics test performance across the partner districts. In Cincinnati, Jefferson County, and Stamford, the introduction of GE Foundation support marked the beginning of statistically significant gains on end-of-year state test performance. The initial effects in the Jefferson County Public Schools were notably large, while students in Cincinnati and Stamford had smaller immediate impacts but demonstrated increased rates of learning over time. In Erie, the introduction of these initiatives marked the stabilization of prior negative trends in mathematics performance in the district.

Introduction

A rigorous evaluation of the impact of any intervention requires an equivalent comparison group. By contrasting the results of one group against another, we can address the central question of whether the introduction of a reform produced better outcomes than what would be experienced in its absence. Finding a counterfactual for district-wide initiatives typically presents a challenge. It is often impractical for researchers to conduct a controlled experiment in which districts are randomly selected to enact a given reform due to the scope of such an endeavor. One common solution is to identify a set of comparison districts against which to compare the reformed districts. However, methodologists critique this approach primarily because they are suspicious of estimated impacts, in part, due to differences in their propensity to enact the reform or in the composition and context of the sites, rather than to the reform itself.

A reasonable alternative is to compare the district to itself. In this way, we can ask whether performance trends in a district have shifted in conjunction with the introduction of a reform effort. In this study, we employed just such a longitudinal evaluation approach, called an *interrupted time series design*, which is a particularly strong quasi-experimental alternative to a randomized design when randomization is not feasible and when longitudinal data are available (Bloom, 2003; Quint, Bloom, Black, & Stephens, 2005; Shadish, Cook, & Campbell, 2002). An interrupted time series compares the trend in performance before the introduction of a reform to the trend in performance after the reform is in place. The "interruption" is the introduction of the reform and the central question this design addresses is whether performance (i.e., the level and slope of the trend line) is significantly improved after the introduction of the reform.

This report describes the results of four interrupted time series analyses conducted in four districts that adopted the GE Foundation's *Developing Futures*™ program. The sections that follow detail the data used for these analyses, the test measures, the analytic approach, and statistical models. Study findings are presented for each district as we answer the evaluative questions about the program's impact in each district.

District Data

Each of the four districts provided student records, including individual state test scores over time and demographic information. The availability of student data varied by district and year but typically contained mathematics scores on state assessments for each year, student ethnicity, gender, English proficiency status, eligibility for free/reduced-price lunch, and special needs status. Unique student identifiers were also provided by the districts and used to link student records over time. Using unique IDs, we can follow students from year to year across grades and schools within each district. Moreover, the IDs benefited analysis by allowing the removal of possible duplicate records for a single student in a given year.

To give the reader a sense of the sizes and student compositions in each district, Table 1 presents the number of years of data that were analyzed, and the district and student demographic characteristics. In each of the districts, we analyzed at least three years of student performance data prior to beginning to work with the GE Foundation, as well as at least four years following. Jefferson County was the largest of the four districts, with 135 schools, while Stamford had the fewest (17) schools.

Table 1. District Size and Demographics[a]

Indicator	Stamford, CT	Jefferson County, KY	Erie, PA	Cincinnati, OH
Years of Data Analyzed	10 years (2002-2011)	10 years (2002-2011)	7 years (2005-2011)	9 years (2003-2011)
Baseline Years	5	4	3	4
Years District was Working with the GE Foundation	5	6	4	5
Number of Schools	17	135	23	74
Average Number of Students per Grade[b]	1,083	6,552	870	2,385
Average Percent Female	49	49	49	50
Average Percent White	42	56	51	23
Average Percent of Students Receiving Lunch Assistance	41	65	62	69
Percent of Students Classified as Limited English Proficient	11	5	8	3
Average Percent of Students Identified as Special Education	-	18	16	18

Notes: [a]Numbers reported in Table 1 may be different than those publicly reported due to pooling of the data over multiple years. [b]The average number of students in the same grade with completed test scores in 2007.

The gender breakdown was fairly similar across the districts, while Cincinnati and Stamford were majority minority districts. Erie and Jefferson County were the lowest-income districts in our sample, with over 60% of the students in the study receiving free or reduced-price lunch assistance in at least one year (these data were not provided by Cincinnati). About 16% to 18% of the students in each district were identified in at least one year as special education students (these data were not reported for Stamford).

We employed several techniques to handle missing data and other discrepancies. Some student records, for example, did not contain complete demographic information (e.g., in one of five records for a given student, ethnicity was not reported). Also, for a few students, demographic information changed from year to year (e.g., in one of five available records for a student, ethnicity was reported as white whereas the student was reported as Hispanic in the other four records). Rather than remove these students from the study, we addressed missing and/or conflicting demographic values using the preponderance of evidence from multiple records for a given student, thus ensuring the completeness and consistency of information. For students with conflicting race or gender information in multiple records, the most common value for that student was used for all years. In the rare event where conflicting information for a student was equally represented (which occurred for 0.95% of students), the value that was more prevalent in the school was selected from the conflicting values. Poverty status, English language proficiency, and special needs status were each reported as binary indicators. To handle missing or conflicting data for these demographics, new variables were created indicating if a given student was ever identified as such. The data in Table 1 reflect the percentages of students in the analytic sample who were identified as economically disadvantaged, limited in English proficiency, or receiving special education services in any of the years of the study. Because these statistics are an aggregate of only the student records, pooled over multiple years, that were included in the analysis, the rates may differ slightly from the rates that are reported by the districts in any given year.

Student Performance Measures

The outcome measure used to evaluate the impacts of *Developing Futures*™ was student performance on each district's end-of-year state mathematics assessment. Not all grades were assessed in all years of this study; generally, before 2007, testing was more sporadic. After 2007, when the annual testing provision of the federal No Child Left Behind (NCLB) Act in grades 3-8 went into effect, testing became more regular. Appendix A presents the mathematics testing schedules in each district by grade.

An additional challenge we faced in analyses was changes in state test metrics over time. In each district, more than one state test was used during the period of the study. This occurred because some states use different tests in different grades or because a state may have revised its test instruments during the period of the study. To properly account for test differences in the longitudinal analysis, individual student outcomes were benchmarked to produce a new standard score. This approach converts student test scores to a relative ranking (i.e., z-score) within test and grade, and is congruent with the recommendations from *Using State Tests in Education Experiments: A Discussion of the Issues* (May, Perez-Johnson, Haimson, Sattar, & Gleason, 2009). Standardization of test scores ensured that student outcomes are calibrated such that test scores can be compared from one year to the next within each district. Consequently, the effects are relative to the districts' distribution of scores and cannot be compared across districts because of possible differences in the amount of preexisting variation in student performance. We used this within-district standardization procedure for Cincinnati, Erie, and Stamford.

In the case of Jefferson County, student performance was provided as performance levels, and not as continuous test scores. Due to differences in the number and labeling of performance categories over time,[1] we created a binary indicator of proficiency for each student (i.e., proficient or not proficient). The statistical models were modified to correctly account for this type of outcome data. Therefore, the results for Jefferson County Public Schools are interpreted as odds ratios.

Table 2. Number of Students by the Number of Years for which Mathematics Scores are Available

Years	Cincinnati, OH	Stamford, CT	Erie, PA	Jefferson County, KY
1	14,689	5,232	10,189	55,132
2	7,700	3,123	2,470	27,988
3	6,708	2,831	2,497	20,797
4	5,458	2,043	1,527	20,498
5 and More	3,283	3,175	1,072	486
Unduplicated Total	37,838	16,404	17,755	73,173

[1] Four categories of performance levels were typically used to express achievement (i.e., distinguished, proficient, apprentice, and novice). In some years, however, the novice and apprentice categories were further divided into "high" and "low" groups.

A key requirement for longitudinal analysis is repeated measures for students across time. It is important for the stability of estimated effects to have a large proportion of students who are tracked over multiple years. Table 2 presents the number of students in each district who can be tracked over several years. The numbers are a function of district size, the years for which data were available for each district, and the number of grades tested in each district. In each of the four districts, a large proportion of students contain multiple years of data. Table 2 also shows the total number of students in each district who were included in the impact analysis.

Analytic Methods

To investigate the impact of *Developing Futures*™ on student mathematics achievement, we used a multilevel interrupted time series framework. This section describes this technique, as well as the limitations to inferences that can be made using this approach.

Time Series Model

Our approach models the repeated student measures and school-level achievement trajectories prior and subsequent to the *Developing Futures*™ intervention. Essentially, we compare rates of learning at the school level, before and after a selected point in time (e.g., the start of GE Foundation support), in order to isolate the program impacts. The benefit of using this approach is that we can leverage the rich, longitudinal, individual student-level data to assess how much student performance in mathematics changed, if at all, as a result of district-wide efforts supported by *Developing Futures*™.

The interrupted time series model uses observations over several points in time, before and during an intervention, to model its impact. The districts began working with the GE Foundation in different years and so the interruption was modeled at the beginning of the appropriate school year (i.e., 2005-2006 for Jefferson County, 2006-2007 for Cincinnati and Stamford, and 2007-2008 for Erie). Achievement trajectories in each of the partner districts were based on at least three years of data before and after introduction of GE Foundation reform efforts, thereby mitigating the potential of natural student maturation as a threat to internal validity (Campbell & Stanley, 1962). Use of longitudinal student-level data further reduced the possible influence of changes in student populations over time by modeling learning trajectories using individual student data instead of comparing cohort trends where students routinely enter and leave the cohorts. For these reasons, the analytic approach provides strong evidence of the relationship between GE Foundation support and changes in student mathematics performance.

Another advantage of our approach to this analysis is the ability to include students who have more or fewer years of available data. Trajectories are based on all student data provided by the districts, which may begin or end at different points in time for different students. Therefore, each student's test scores contribute to information about school and district performance only for those years in which the student was enrolled in the district. The impact results from this type of model are robust to missing data, provided that the data are missing at random (Little & Rubin, 1987; Schafer, 1997).

An additional step was taken to further ensure that our statistical models properly accounted for the longitudinal nature of the data. To account for the repeated measures for each student and resulting lack of independence among errors, a variance components error covariance matrix was used to allow for the correlation among errors between lagged repeated measures. This structure of errors relaxes the independence assumption by allowing errors of measures within an individual to be correlated. Likelihood ratio chi-square tests were used to verify the fit of this model relative to simpler covariance structures.

Multilevel Analyses and Contextual School Effects

The statistical model is also multilevel in recognition of the contextual influences on student achievement that exist within schools. To account for the resulting lack of independence of observations between students within schools, the analysis included random effects for schools (Raudenbush & Bryk, 2002). By including random effects for the intercept and slope of each school in the model, the multilevel approach allows us to model the impacts on mathematics achievement trends across schools within a district.

Student mobility and the natural progression through grades resulted in students attending more than one school over the duration of the study. This adds some complexity to the estimation of school-level trends. However, student attendance in more than one school over time can be handled within a mixed-effect model by specifying multiple membership cross-classified random effects. This conventional approach to the nesting of students across multiple schools is useful for longitudinal education studies, and is the basis of value-added models, which use lagged student gains. To account for the somewhat more complex data structure, we allowed for cross-classification in which lower-level units could be nested within two or more higher-level units.

All student demographic data were also aggregated to the school level for use in the statistical analysis. School aggregate data were included to understand how schools within a district differ from each other on key student characteristics, and how program impacts may be related to those contextual differences.

Estimating Impacts

Employing an interrupted time series model allowed us to test whether there were significant changes in mathematics achievement trends in the district from before to after the introduction of the *Developing Futures*™ program. The main impact model contains three predictors: time, GE Foundation support, and their interaction. The fixed effect for continuous time gives us an estimate of the average growth rate in achievement scores over the entire time series, while the fixed effect for the GE Foundation indicator provides an estimate of the average shift in student achievement trajectories during the years of GE Foundation support. The interaction of the two provides an estimate of the average change in student achievement growth rates during the implementation period of the *Developing Futures*™ initiatives. The specification of the statistical model used in these analyses is shown in Appendix B.

In addition to main effects, we also examined the extent to which measurable impacts persist after controlling for student characteristics and school contexts. For this analysis, available student demographic information (i.e., gender, ethnicity, poverty status, special needs status, and English language proficiency) was added to the base model as predictors of both student outcomes and school-level growth trajectories. These variables were also expressed in terms of school means or rates (i.e., the percent of students for that school). Expressing student characteristics in terms of school rates allowed us to better interpret the variance around schools in a given district, as well as adjust school-level growth trajectories (i.e., slopes as outcomes). Note that the program effects and the degree of change from the base impact model are not directly comparable across districts because slightly different student data were available in each of the districts (i.e., individual student poverty and special education status were unavailable in Cincinnati and Stamford respectively).

Finally, we explored variation in student achievement at different levels of schooling where elementary level was defined as grades 3-5, middle level was defined as grades 6-8, and high level was defined as grades 9-12. Including indicators for the grade levels of students, and their interaction with all of the parameters in the base model, we then compared performance trends for students in elementary, middle, and high school grades. These are not intended to conform to the different grade configurations in schools across the four districts, of which there were many, but rather to indicate differences in performance at different grade ranges. When examined along with details on grade-specific reform emphases in each of the districts, these findings can provide potentially useful additional information on the variation in effectiveness of different grade-level reform efforts.

Limitations to Inference

Because these results only look within each district, they do not capture major external environmental changes like state or federal policy changes, test revisions, or other events that may affect the entire district in other years. Moreover, while this approach is robust to changes in student populations, we do not have the necessary historic implementation data to understand how specific program activities rolled out and evolved over time. Given that those data are unavailable, the analyses in this study must assume that GE Foundation-supported district improvement efforts were implemented consistently in all years following rollout. Finally, while our approach can estimate the overall impact on performance trends, it cannot isolate one or more specific components of the intervention (i.e., aligning district support components versus teacher professional development) or distinguish between GE Foundation-supported efforts and other major reforms coincidental to the *Developing Futures*™ reform efforts. For these reasons, significant trends (either positive or negative) cannot be attributed exclusively to impacts of GE Foundation initiatives, nor can they be attributed solely to the districts' instructional improvement efforts. Despite these limitations, it is nonetheless appropriate to test whether there was a statistically significant change in the typical achievement trajectory after new GE Foundation-supported instructional programs were implemented and to plausibly attribute the GE Foundation support to these changes. Further, if we are able to replicate this pattern across multiple districts, our confidence in attributing these effects to the GE Foundation's efforts becomes increasingly stronger.

Results

If *Developing Futures*™ had an impact, the causal hypothesis is that the student performance trend will have a change in the level and/or slope that is coincident with the time of its introduction, and we can describe the effects in terms of immediacy and persistence respectively (Shadish, Cook, & Campbell, 2002). The immediacy of the program impact is observed as a discontinuity of performance levels at the point of interruption. The persistence (or permanence) of the impact speaks to the difference between the slopes of the trend line before and after the point of interruption.

Table 3. GE Foundation Support Impact Estimates in Four Districts

		Cincinnati, OH Base Model	Stamford, CT Base Model	Erie, PA Base Model	Jefferson County, KY[a] Base Model	Cincinnati, OH Full Model	Stamford, CT Full Model	Erie, PA Full Model	Jefferson County, KY[a] Full Model
Program Attributes	Year	-.016 (.01)	.016* (.01)	-.024* (.01)	1.192*** (1.72,1.21)	-.008 (.01)	.056*** (.01)	-.026** (.01)	1.233*** (1.22,1.25)
	GE Foundation Support	.031** (.01)	-.014 (.02)	-.024 (.02)	1.319*** (1.28,1.36)	.025* (.01)	-.046*** (.01)	-.011 (.02)	1.424*** (1.38,1.47)
	Year* GE Foundation Support	.025*** (.01)	.039*** (.01)	-.024* (.01)	.897*** (.88,.92)	.025** (.01)	-.014*** (.01)	-.018 (.01)	.878*** (.86,.90)
Student Attributes	Female					-.047*** (.01)	-.023*** (.01)	-.076*** (.01)	.900*** (.88,.92)
	White					.501*** (.01)	.491*** (.01)	.314*** (.01)	2.168*** (2.15,2.19)
	Lunch Assistance					N/A	-.417*** (.01)	-.250*** (.01)	.468*** (.45,.49)
	Special Assistance					-.678*** (.01)	N/A	-.785*** (.01)	.314*** (.29,.34)
	English Language Learner					-.135*** (.02)	-.710*** (.01)	-.473*** (.02)	1.009 (.97,.1.05)
School Attributes	Percent Female					-.045*** (.01)	-.026 (.01)	-.071*** (.02)	.919*** (.88,.96)
	Percent White					.503*** (.01)	.491*** (.01)	.327*** (.01)	2.150*** (2.13,2.17)
	Percent on Lunch Assistance					-.373* (.18)	.420*** (.01)	-.241*** (.01)	.456*** (.43,.48)
	Percent Special Education					-.706*** (.01)	N/A	-.816*** (.02)	0.330*** (.29,.37)
	Percent English Language Learner					.142*** (.02)	.697*** (.02)	-.461*** (.02)	1.018 (.97,1.06)

*p ≤ .05, **p ≤ .01, ***p ≤ .001; standard errors shown in parentheses. a Estimates for Jefferson County Public Schools are expressed as odds ratios with 95% confidence intervals.

A summary of the effects is shown in Table 3, in which we present two models for each district. The first model is a base impact model, which shows the effects associated with the GE Foundation efforts. This model provides an estimate of the full program effects experienced in the district, irrespective of demographic shifts that may have occurred during the time period. The second model is a full model that includes additional parameters for student and school attributes. The full model for each district shows the adjusted effects of the GE Foundation support on student performance trends, and also illustrates differences in mathematics performance by student and school attributes.

For each model, at least three effects are reported. The "year" effect represents the slope of performance trend during the years preceding GE Foundation support. The "GE Foundation support" effect represents the impact at the point of interruption, or the impact in the year in which the GE Foundation began working with the district. The "year by GE Foundation" effect is the interaction of the "year" and "GE Foundation support" variables and represents the change in the overall trend associated with the intervention.

We must be careful making comparisons across districts because the model in each district was slightly different. For example, while we had a measure of student poverty — whether a student received free or reduced-price lunch — for individual students in Stamford, Erie, and Jefferson County, we had only a school-level indicator in Cincinnati. Likewise, we had no indicator for a student's special education status in Stamford. Also in Jefferson County, student outcomes were modeled as a proficiency indicator and as standard scores in the other three districts. Furthermore, each state context was different, and their tests might measure different aspects of mathematics achievement.

Base Model Impacts

Focusing on the base models, we observed several characteristics of the effect of *Developing Futures*™ on mathematics performance in the four districts. First, it is important to note the performance trends prior to GE Foundation support. In Stamford and Jefferson County, mathematics performance was slowly but significantly improving, while in Erie, it was slowly declining. There was no significant prior trend in mathematics performance detected in Cincinnati. Using interrupted time series analysis, we then tested for program effects and found significant impacts in the four districts.

In Cincinnati, there was no significant change in mathematics performance in the three years prior to the beginning of GE Foundation support in 2006-2007. After the GE Foundation began to work with the district, there was a statistically significant increase in student mathematics performance of three-tenths of a standard deviation. This small but significant jump in district-wide mathematics performance continued over the next four years, increasing on average by .022 standardized units per year. Thus, the trend of Cincinnati's mathematics performance was essentially flat in the three years prior to *Developing Futures*™, significantly increased in the year in which *Developing Futures*™ began, and sustained increases in the four years that the GE Foundation continued to support district efforts.

The findings in Stamford were slightly different. No significant changes in students' mathematics performance were found in the four years of data that we analyzed prior to the beginning of GE Foundation support (2003 to 2006). There were also no significant changes in the year that the GE Foundation support was initiated (2006-2007), which the district reports was largely a planning year. However, the trend for the four years following the initiation of GE Foundation support showed a statistically significant and positive increase in mathematics test score performance. Thus, like Cincinnati, the trend in Stamford student mathematics performance was significant and positive over the years of GE Foundation support.

In Erie, similar positive effects of GE Foundation support on district mathematics performance were found, but within a different context. Student performance prior to GE Foundation support was declining in a significant downward trend. However, we find that GE Foundation support significantly altered the trend line in an equal and opposite direction. Beginning in 2007-2008, the trend during the next three years of GE Foundation support was equal to zero. We also note that this was the smallest district, making it relatively hard to detect statistical significance of program effects.

The findings in Jefferson County are of explosive growth in the first year of GE Foundation support, followed by a slight average decline in subsequent years that nonetheless substantially exceeds pre-GE Foundation expected performance. Because the data from Jefferson County are modeled as proficiency rather than standard scores, we report Jefferson County results as odds ratios. An odds ratio greater than one is a positive effect and an odds ratio of less than one is a negative effect. The statistically significant effect for year in the base mode can be interpreted as modest increase in the odds of a student's proficiency in the three years prior to the inception of GE Foundation support. In 2005-2006, the year GE Foundation support began in Jefferson County, there was a statistically significant 32% increase in the odds of students achieving proficiency. In the five subsequent years (through 2011), there was a slight but statistically significant average decline in performance. However, this average decline was small compared to the large boost in mathematics performance associated with the year GE Foundation support began in Jefferson County. Thus, the overall trend in Jefferson County shows an initial improvement that was so large that even though the rate of growth slowed in subsequent years, student performance under GE Foundation exceeded what was predicted by the baseline district trajectory in every year of the study.

Full Model Impacts

The four columns on the right of Table 3 present results of impact models for each of the four districts that include a series of control variables for student and school attributes. We examine the full models in two ways. First, we considered how program impacts have changed with the addition of covariates into the models. We then examined the additional estimates to explore trends in student and school attributes. All the control variables (gender, ethnicity, poverty status, English proficiency, and special needs status) are significantly correlated with student mathematics achievement across the districts, justifying their inclusion in the model. Results indicate that even after controlling for student attributes and school contexts, mathematics achievement in the four school districts continued to show program impacts during the period of GE Foundation support. Despite the fact that student background characteristics were found to be highly predictive of mathematics performance, the positive trends associated with GE Foundation support persisted, demonstrating that impacts were not confounded with shifts in the demographic contexts of the districts that would explain effects on student performance.

In three districts, there was essentially no change between the base and full models. In Cincinnati, we found no change in the main effect for the year after the introduction of the student and school attributes. The magnitude of the average effect on mathematics performance associated with the GE Foundation's support remained the same, even after controlling for student and school attributes. The full model results from Erie indicate that while the estimate for GE Foundation support was no longer significant, the trend in student mathematics performance prior to GE Foundation support remained significant and negative, while the trend line beginning in 2007-2008 was essentially flat.

The Jefferson County findings were also consistent across models. There was a significant and positive trend of improving mathematics performance in the three years of data prior to the introduction of *Developing Futures*™; the trend was of a dramatic upward surge in performance in 2005-2006, the year that the *Developing Futures*™ in Education program was implemented in the district, and a slight decline in performance thereafter that did little to mitigate the effects of the initial boost.

In Stamford, findings were consistent across the base and full models in terms of the direction of program effects; however, the magnitude and significance of the findings are slightly different. The trends before and after introduction of GE Foundation support remain positive and significant, with a post-intervention increase in rate of growth. A noticeable difference between models is that the effect associated with the 2006-2007 introduction of *Developing Futures*™ support, which was negative and non-significant in the base model, becomes statistically significant in the full model. Despite the one-year drop, the annual program effects more than compensated for the initial loss and by the last year of analyzed data, the net influence of the GE Foundation's work was significant and positive.

Examining the effects for student attributes and school contexts, we see largely similar patterns across the districts. In all four districts, boys performed better than girls in mathematics. In Jefferson County, the coefficient of .887 indicates that the odds of a student being proficient on the mathematics assessment were 11% lower for girls than for boys. We also see from these models that white students significantly outperformed minority students in all four districts.

In each of the districts, socioeconomic status was indicated by the student's receiving assistance to purchase school lunch. Students who received lunch assistance performed significantly worse than students who did not receive lunch assistance. In Cincinnati, Erie, and Jefferson County, where we had an indicator of the special education status of students, these students performed significantly worse than regular education students. Finally, in all four districts, non-native English speakers performed significantly worse than non-native English speakers on their state test.

The school attributes showed similar patterns. There was a small negative effect associated with each increasing percentage of female students a school had. Similarly, schools with higher percentages of white students outperformed schools with higher percentages of minority students. In Stamford and Erie, where we included school-level lunch assistance data, schools with a higher percentage of students receiving lunch assistance performed significantly less well than schools with lower percentages of students receiving lunch assistance. As is common in most cases, schools with higher percentages of special needs and English language learning students scored less well on their average mathematics performance than did schools with lower percentages of these students.

Impacts by Grade Level

To examine how program impacts varied by grade levels in each of the districts, we modified the base model to include student grade level. In these models, we defined elementary grades as grades 3-5, middle grade as grades 6-8, and high school grades as grades 9-12. Indicators for middle grades and high grades were included in the analysis with the elementary grades serving as the reference category (thus not shown in Table 4). Because elementary grades were held as a reference category, the reader must interpret estimates for middle and high grades in relation to the main effects. Also, all two- and three-way interactions were included. Therefore, the reader must be careful to interpret individual effects because the interaction terms must be combined with the main effects to produce estimates of overall performance by year and grade level. The results for the grade-level models for the four districts are shown in Table 4.

Table 4. Analysis of School Growth Trajectories by Grade Bands

	Cincinnati, OH	Stamford, CT	Erie, PA	Jefferson County, KY
Year	-.015 (.01)	.016~ (.01)	-.085*** (.02)	1.197*** (1.18,1.22)
GE Foundation Support	-.008 (.01)	-.037 (.02)	-.075** (.03)	1.482*** (1.42,1.54)
Year*GE Foundation Support	.015~ (.01)	.033** (.01)	.085*** (.02)	.866*** (.85,.89)
Middle (6-8) Grades	-.044 ** (.01)	.051 (.06)	.035 (.03)	.338*** (.18,.50)
High (9-12) Grades	-.063* (.02)	.063 (.10)	.283* (.14)	.491*** (.35,.63)
Middle Grades* Year	-.020** (.01)	-.006 (.01)	.098*** (.02)	.996 (.96,1.04)
High Grades* Year	.135*** (.03)	.033~ (.02)	.198*** (.05)	.968 (.93,1.01)
Middle Grades* GE Foundation Support	.079** (.02)	.046 (.03)	.065~ (.04)	1.021 (.92,1.12)
High Grades* GE Foundation Support	-.040** (.03)	.032 (.05)	.111 (.09)	.590 (.48,.70)
Middle Grades* Year* GE Foundation Support	.020 (.01)	.024~ (.01)	-.083** (.03)	1.094*** (.95,1.24)
High Grades* Year* GE Foundation Support	-0.080** (.03)	-.035* (.02)	-.260*** (.05)	1.091*** (1.05,1.13)

There was no consistent testing before grade 3 in the districts.

~ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; standard errors shown in parentheses. a Estimates for Jefferson County Public Schools are expressed as odds ratios with 95% confidence intervals.

The results indicate that, after controlling for student grade level, overall program impacts remained consistent with the full models in the direction and significance of impacts. In Cincinnati, Stamford, and Erie, there was overall statistically significant positive growth in the period after the GE Foundation-supported work was initiated. In Jefferson County, we see a large initial effect followed by slight decline in the growth rate in subsequent years, although the trend was still positive. When looking at program impacts by grade levels, we see some variation in the timing of the impacts, with some grades showing larger initial gains and others more gradual. Across districts in the elementary grades, we see that by the final year of the study, student performance in all four districts exceeded what was expected based on the pre-GE Foundation trend. This suggests that GE Foundation-supported district efforts may have focused on the early grades in terms of implementation, effectiveness, or both.

Program impacts were less consistent in the upper grades, and in Erie we find a dramatic decline during the period of GE Foundation support. This finding for Erie schools helps interpret the overall effects that stabilized a downward trend. Here, we see that the positive and significant impacts in the elementary grades for Erie (β=0.085) were counteracted by declining performance in high school (β=-0.26). In the case of Jefferson County, the grade-specific trends of mathematics performance during GE Foundation support are consistent with the overall effects and found to be focused on the elementary and middle school levels, with all grade levels outperforming the baseline trends.

To illustrate both the overall and grade-level effects in each district, we produced graphical representations of the performance trajectories for each district. These trends show the model-implied values by year and grade level. It is important to note that trajectories are less stable in districts, years, and grades that have relatively fewer tested students. This is the case for smaller districts (i.e., Erie), years with fewer tested students (i.e., typically prior to 2007, as shown in Appendix A), and in grade levels with fewer tested students (i.e., high schools in the districts typically have only one tested grade). Grade-level trends are presented separately by district and discussed along with overall program impacts.

Figures 1 to 4 show the adjusted performance trend overall (the bold line) and for each grade level by year for each district. The figures represent predicted values based upon the models in Table 4, and are only interpretable by combining the main effects and interactions. The trends in Cincinnati, Stamford, and Erie are presented as standardized effect sizes, which equates for both changes in state tests across time and for year-over-year comparisons (see the section on "Student Performance Measures"). The results of Jefferson County are presented as the model-adjusted percent proficient in the district each year.

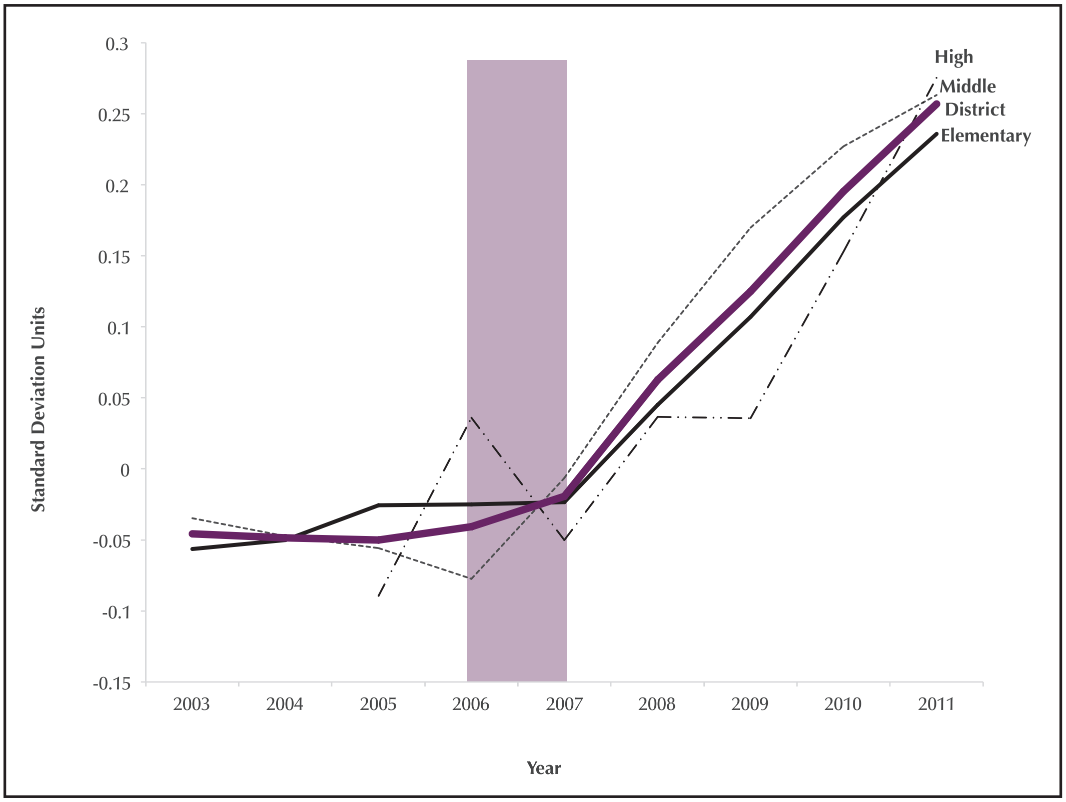
Each figure includes a vertical indicator of the year in which the GE Foundation introduced *Developing Futures*™ in Education to the district. The indicator spans the period of a year to represent that the GE Foundation program implementation occurred not at a moment in time, but rather unfolded from that school year forward. Also note that the grade-level trends presented here are based on the grade bands of students estimated from the impact analysis, where elementary represents test performance in grades 3-5, middle represents test performance in grades 6-8, and high represents test performance (where available) in grades 9-12.

Performance Trends in Cincinnati, OH

The trends in Cincinnati from 2003 to 2007, before the introduction of *Developing Futures*™, were generally flat. During these years, there was some variation by grade level, like the jump in high school grade-level performance in 2006, but performance was fairly stable.

As Figure 1 shows, the introduction of *Developing Futures*™ in the 2006-2007 school year was coincident with a statistically significant increase in overall mathematics performance, which was consistent across all the grade levels assessed. This increase in performance continued from 2008 through 2011, the last year for which we analyzed data. Impressively in Cincinnati, these year-over-year gains in performance were fairly consistent at the elementary, middle, and high school grade levels.

Figure 1. Cincinnati Mathematics Performance Trends by Grade Level

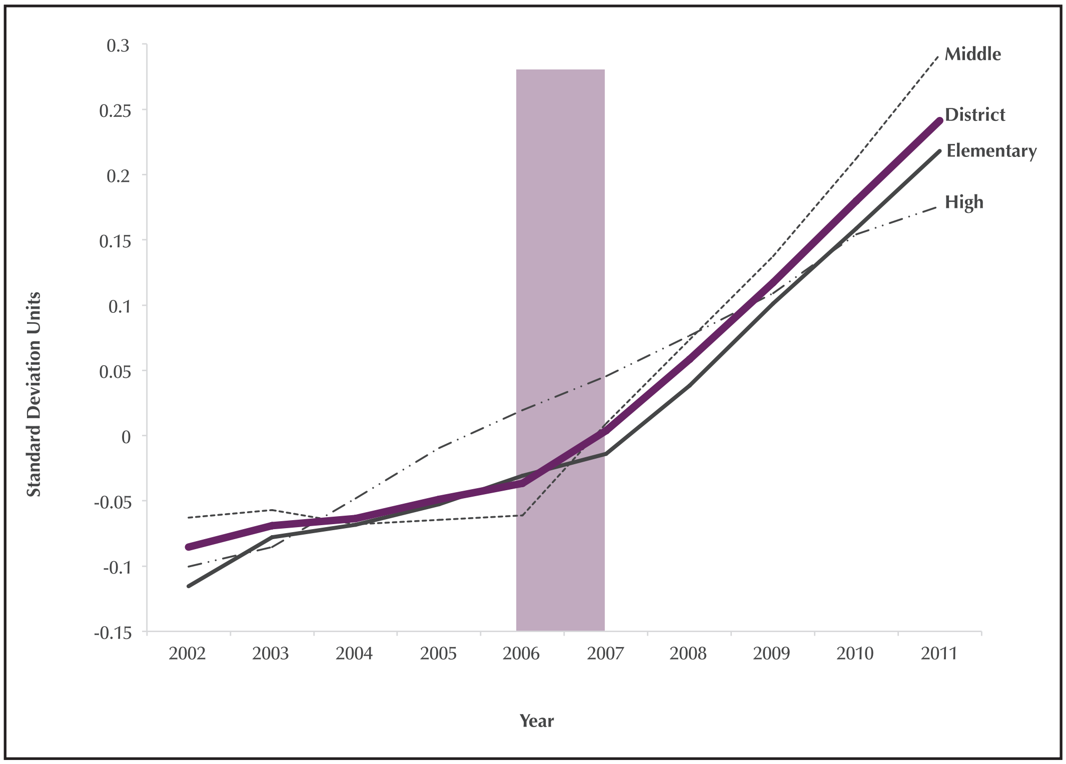


Note: Shaded bar represents the school year within which GE Foundation support began.

Performance Trends in Stamford, CT

Student mathematics performance in Stamford showed no major changes in the four years prior to the introduction of *Developing Futures*™ (2003 to 2006). In 2006-2007, the year in which *Developing Futures*™ began in the district, which the project reports as primarily a planning year, there was a slight increase in overall mathematics performance that was driven largely by middle and high school grade performance. The increased slope of performance from 2007 through 2011 shows a steady increase in performance at all three grade ranges. Notably, the three grade levels are tightly clustered with consistent upward trends. Stamford showed consistent improvements in overall mathematics performance with positive and statistically significant improvements following the introduction of GE Foundation support. (See Figure 2.)

Figure 2. Stamford Mathematics Performance Trends by Grade Level



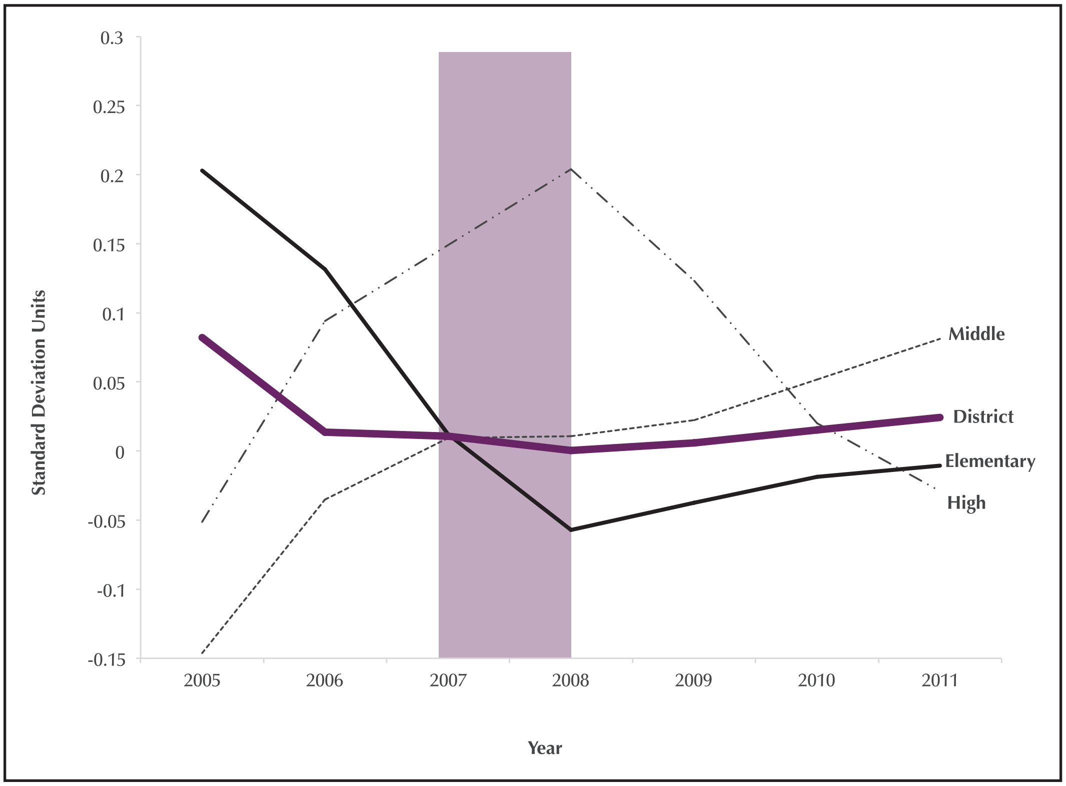
Note: Shaded bar represents the school year within which GE Foundation support began.

Performance Trends in Erie, PA

Mathematics performance in Erie declined from 2005 to 2007, the period prior to the district's initiation of work with the GE Foundation. As seen in Figure 3, the overall decline in mathematics performance from 2005 to 2007 was driven by a decline in performance in the tested elementary grades (grades 3-5), while the tested grades in middle and high schools increased in this pre-*Developing Futures*™ period. The increase in high school performance in the three years from 2005 to 2008 may have been related to the Pennsylvania High School Coaching initiative, an intensive teacher professional development and coaching program focused on high schools across the state. Notably, that program ended in 2008, as the GE Foundation support was beginning. In 2007-2008, with the inception of its GE Foundation grant, Erie's efforts were focused on elementary and middle schools, not on high schools.

In 2007-2008, the beginning of the district's work with *Developing Futures*™, overall district performance began a period of stabilization, which persisted over the course of the next three years, from 2008 to 2011. The stabilization in performance was a pattern mirrored in trends in both the elementary and middle grades mathematics performance over the period from 2008 to 2011. Perhaps not coincidentally, these were also the grade levels at which the district reported focusing its *Developing Futures*™ resources in that period. High school mathematics performance in Erie peaked in 2008 and showed a striking decline in the following years, corresponding to the end of the High School Coaching initiative in 2008, which seemingly initiated a sharp decline in high school performance. As previously stated, *Developing Futures*™ did not focus on high schools in Erie from 2008-2011, the period analyzed.

Figure 3. Erie Mathematics Performance Trends by Grade Level

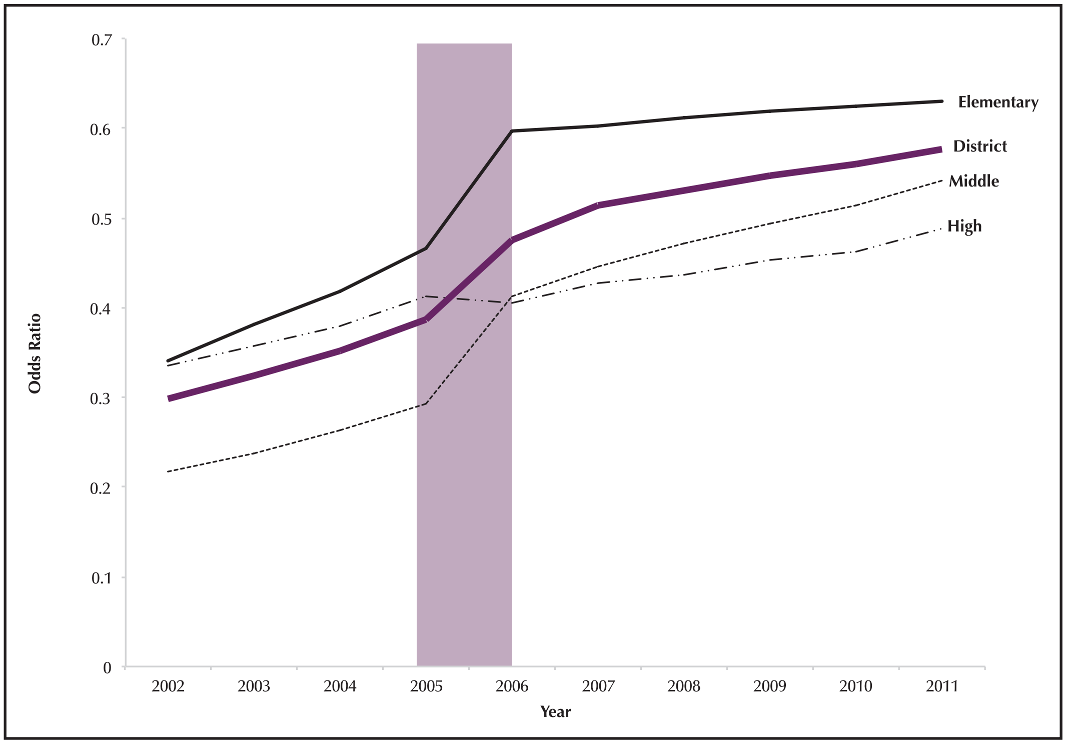


Note: Shaded bar represents the school year within which GE Foundation support began.

Performance Trends in Jefferson County, KY

In the four-year period before the introduction of GE Foundation support in Jefferson County, student mathematics achievement was experiencing a significant upward trend. In 2005-2006, the year *Developing Futures*™ started to work with the district, there was a surge in mathematics achievement, particularly in the elementary and middle schools. In the five-year period that followed, from 2006 to 2011, the early overall significant gains were sustained. This overall persistence in performance mirrors the stable trend in elementary schools. Particularly in the middle grades of 6-8, and to a lesser extent in the high school grade tested (grade 11), there was slow but steady growth in the years following the introduction of *Developing Futures*™ in the district. (See Figure 4.)

Figure 4. Jefferson County Mathematics Performance Trends by Grade Level



Note: Shaded bar represents the school year within which GE Foundation support began.

Summary

This report looked retrospectively at up to 10 years of student mathematics performance trend data in each of four districts that have had a long-standing engagement with the GE Foundation's *Developing Futures*™ in Education program. The central question that the report focused on was: were there detectable effects in students' mathematics performance associated with the introduction and ongoing work of *Developing Futures*™ in the four districts? The rigorous, longitudinal analyses presented in this report provide strong evidence that the GE Foundation's *Developing Futures*™ in Education program produced improvements in mathematics performance in each of the four districts. In all four districts, there are statistically significant and positive changes in student mathematics performance associated with the efforts of the GE Foundation. The contours of the effects were different in each district, which reflects the different contexts and coincident work occurring in each location. In Cincinnati and Stamford, the effects were both significantly positive and sustained over the period examined. In Jefferson County, the initial impact was substantial, with trends in subsequent years maintaining the initial boost. In Erie, the introduction of *Developing Futures*™ arrested and stabilized a notable decline. While the stories from each district were different, the larger picture shows a clear and reinforcing pattern of positive student mathematics outcomes associated with the work in the districts during the time of their partnership with the GE Foundation.

The cumulative portrait of positive impacts across the four districts is particularly important because of an inherent constraint in the analytical method used in these analyses. By examining within-district trends over time, and comparing districts against themselves, this approach cannot account for simultaneous, but independent, influences in the districts. Therefore, by examining each district alone, it is possible that the impact that we associate with the GE Foundation's efforts may be attributable to some simultaneous event, like a notable shift in state policy or adjustments in district resources or composition. However, by looking not only at longitudinal within-district trends, but also by examining the accumulated pattern across the four districts in four different states in different regions of the United States, we reduce the likelihood of any alternative district or state explanations. Put simply, the pattern of positive effects across four disparate districts in four states together make a compelling case that the results are attributable to the good work catalyzed by the GE Foundation's *Developing Futures*™ in Education program.

References

Bloom, H. S. (2003). Using "short" interrupted time-series analysis to measure the impacts of whole-school reforms with applications to a study of accelerated schools. *Evaluation Review*, 27(1), 3-49.

Campbell, D. T., & Stanley, J. C. (1962). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Quint, J., Bloom, H. S., Black, A. R., & Stephens, L. (2005). *The challenge of scaling up educational reform*. New York: MDRC.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.).* Thousand Oaks: Sage.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Appendix A. Annual Testing Schedules by District

Testing Schedule in Cincinnati, OH from 2003 to 2011

Grade	2003	2004	2005	2006	2007	2008	2009	2010	2011
3			✔	✔	✔	✔	✔	✔	✔
4	✔	✔	✔	✔	✔	✔	✔	✔	✔
5				✔	✔	✔	✔	✔	✔
6	✔	✔	✔	✔	✔	✔	✔	✔	✔
7			✔	✔	✔	✔	✔	✔	✔
8			✔	✔	✔	✔	✔	✔	✔
9								✔	
10	✔	✔	✔	✔	✔	✔	✔	✔	✔
11	✔	✔	✔	✔	✔	✔	✔	✔	✔
12	✔	✔	✔	✔				✔	

Testing Schedule in Erie, PA from 2005 to 2011

Grade	2005	2006	2007	2008	2009	2010	2011
3		✔	✔	✔	✔	✔	✔
4			✔	✔	✔	✔	✔
5	✔	✔	✔	✔	✔	✔	✔
6			✔	✔	✔	✔	✔
7			✔	✔	✔	✔	✔
8	✔	✔	✔	✔	✔	✔	✔
9							
10							
11	✔	✔	✔	✔	✔	✔	✔
12						✔	✔

Testing Schedule in Jefferson County, KY from 2002 to 2011

Grade	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
3						✔	✔	✔	✔	✔
4						✔	✔	✔	✔	✔
5	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
6						✔	✔	✔	✔	✔
7						✔	✔	✔	✔	✔
8	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
9										
10										
11	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
12										

Testing Schedule in Stamford, CT from 2002 to 2011

Grade	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
3					✔	✔	✔	✔	✔	✔
4	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
5					✔	✔	✔	✔	✔	✔
6	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
7					✔	✔	✔	✔	✔	✔
8	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
9										
10	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
11										
12										

Note: ✔ denotes year and grade tested.

Appendix B. Statistical Model

The statistical model used to evaluate the significance of the impact of the GE Foundation support in each district was a multi-level interrupted time series model, with annual repeated measures of student performance in mathematics, and annual test proficiency rates nested within schools. The random effect terms for the hierarchical linear modeling are included as alpha and gamma. The functional form of the model is:

$$Y_{jt}=\beta_{0j}+\beta_{1j}\ (\text{Year})_{ij}+\beta_2\ (\text{GE})+\beta_3\ (\text{Year*GE})+\alpha_j+\gamma_j\ (\text{Year})+\varepsilon_{jt}$$

Where:

Y_{tj} is the student outcome in school j in year t

β_0 is the average student outcome in year GE Foundation first implemented

β_1 is the average annual change in student outcome

β_2 is the initial shift in student outcome in year GE Foundation first implemented

β_3 is the adjustments to average annual change in student outcome in GE Foundation implementation years

a_j is the mean deviation for percent proficient in school j in year first implemented

γ_j is the mean deviation in the annual change in proficient in school j

ε_{tj} is the difference between predicted and observed percent proficient (i.e., residual) in school j in year t

CPRE

Consortium for Policy Research in Education

University of Pennsylvania

Teachers College, Columbia University

Harvard University

Stanford University

University of Michigan

University of Wisconsin-Madison

Northwestern University