



An Examination of the Effect of a Pilot of the National Institute for School Leadership's  
Executive Development Program on School Performance Trends in Massachusetts

John A. Nunnery, Ed.D.  
Executive Director, The Center for Educational Partnerships  
Old Dominion University

Steven M. Ross, Ph.D.  
Professor  
Johns Hopkins University

Cherng-jyh Yen, Ph.D.  
Assistant Professor  
Old Dominion University

August, 2010

The Center for Educational Partnerships at Old Dominion University

**The Center for Educational Partnerships**  
**Old Dominion University ▪ Darden College of Education ▪ Room 135 ▪ Norfolk, VA**  
**23529**  
**Phone: (757) 683-5449**

The Center for Educational Partnerships establishes collaborative educational enterprises with schools and school divisions that support dissemination of proven practices, rigorous field trials of promising models, and development and testing of innovative research-based models in collaboration with Old Dominion University's primary educational partners.

The Center for Educational Partnerships focuses its efforts on the following activities:

- Developing broad-based partnerships with school divisions in the Hampton Roads area and greater Virginia.
- Serving as a conduit through which Old Dominion University will endeavor to make available the best local, state, national, and international resources to the primary partners to fulfill our mutual obligation to assure academic achievement, development of responsible citizenship, and self-fulfillment of the youth served by our partner schools.
- Pursuing selected inter-institutional and international educational partnerships that capitalize on identified areas of strength at ODU or enhance strategic relationships between these partners and the Hampton Roads community.

### **Values**

The guiding principle behind all of the Center's decisions is to develop the best possible partnership arrangements to enhance the educational experiences and future prospects of children and young adults in the Hampton Roads community and across greater Virginia.

As a result, the Center will:

- Place special emphasis on partnership arrangements that seek to improve the education of youth from historically underserved ethnic or geographical backgrounds.
- Submit its actions and staff to the highest possible standards of professional conduct, accountability, and transparency of decision-making and operations.
- Respect the dignity, rights, and essential humanity of all persons with whom we may interact, regardless of their station in life or background.

## Table of Contents

<b>Executive Summary</b> .....	<b>3</b>
<b>Introduction</b> .....	<b>6</b>
<b>Research Questions</b> .....	<b>8</b>
<b>Method</b> .....	<b>9</b>
<b>NISL Schools</b> .....	<b>9</b>
<b>Dependent Measures</b> .....	<b>9</b>
<b>Matched Comparison Sample</b> .....	<b>10</b>
<b>Analyses</b> .....	<b>12</b>
<b>Results</b> .....	<b>14</b>
<b>Math</b> .....	<b>14</b>
<b>English/Language Arts</b> .....	<b>15</b>
<b>State-Wide Comparison Results</b> .....	<b>17</b>
<b>Math</b> .....	<b>17</b>
<b>English/Language Arts</b> .....	<b>18</b>
<b>Findings and Discussion</b> .....	<b>21</b>
<b>Study Overview</b> .....	<b>21</b>
<b>Findings</b> .....	<b>21</b>
<b>Summary Effects</b> .....	<b>21</b>
<b>Discussion</b> .....	<b>22</b>
<b>References</b> .....	<b>24</b>

**List of Tables**

**Table 1.....12**  
**Table 2.....17**  
**Table 3.....20**

**List of Figures**

**Figure 1.....15**  
**Figure 2.....16**  
**Figure 3.....18**  
**Figure 4.....19**  
**Figure 5.....22**

## EXECUTIVE SUMMARY

This study examined the impact of a pilot of NISL's Executive Development Program for principals on student achievement in Massachusetts schools between 2006 and 2009. Of 131 Massachusetts principals who began the pilot program in 2006, 65 both completed the program and served as principals of the same school through 2009. Only 2 high schools participated, so program effects were not estimated at this level. Schools served by principals participating in the Executive Development Program were matched 1:4 to comparison schools with similar school performance and demographic profiles in 2006, which served as the baseline year. Comparisons were also made to all schools statewide. For elementary schools, there were 45 NISL schools and 180 comparison schools, for middle schools there were 13 NISL schools and 52 comparison schools, and for combined elementary-middle schools there were 6 NISL schools and 24 comparison schools. The study sample reflects the small portion of the 1,150 school leaders who began participation in NISL training between 2006 and 2009 who completed the program prior to 2009 and who remained as principal of the same school.

Individual student raw scores in mathematics and English/Language Arts (ELA) on the Massachusetts Comprehensive Assessment System tests (MCAS) were converted to standardized (Z-scores) based on the statewide means and standard deviations for each grade level and year. The mean Z-score for each test was computed for each school as an overall index of student performance. Repeated-measures analyses were performed on these overall index scores to determine whether there were differences in school performance trends between schools served by NISL-trained principals and matched comparison schools and pilot NISL schools and all schools in the Commonwealth. NISL schools tended to serve higher proportions of economically-disadvantaged students and had substantially lower performance than even the

matched comparison group, as many of the schools that were selected to participate had been identified as low-performing schools.

The repeated-measures analyses indicated that math effects associated with NISL were consistent across grade levels, so combined effects were estimated. Statistically significant differences in school performance trends were observed on mathematics between NISL and comparison schools and between NISL schools and all schools in the state. In 2006, the students in schools identified for the NISL program scored, on average, -0.17 standard deviation units below students in matched comparison schools in 2006. This difference shrunk to -0.05 standard deviation units by 2009. Adjusting for aggregate student demographics, Cohen's  $d$  estimate of effect size was +0.10. Results obtained by comparing NISL schools to all schools in Massachusetts's yielded similar results—i.e., NISL status was associated with a statistically significant positive trend in aggregate achievement, yielding an effect size estimate of  $d = +0.08$ .

As with math effects, reading effects were consistent across grade levels so combined effects were estimated. No statistically significant differences in ELA performance trends were observed between NISL and the matched comparison sample. Overall performance on ELA remained virtually unchanged each of the four years in both NISL and comparison schools, yielding an effect size estimate of  $d = +0.00$ . Likewise, performance on ELA tests in NISL schools remained constant relative to all schools in Massachusetts, yielding an effect size estimate of  $d = +0.00$ . This finding was not surprising given that a substantial body of prior research suggests that math achievement is more responsive than reading achievement to interventions in the near term because English/language arts and reading achievement are more influenced by home experiences and out-of-school opportunities. Moreover, the principals in the

NISL pilot sample had only one year post-completion prior to 2009 testing, which is probably insufficient time to observe full impact of the program on test scores.

## INTRODUCTION

The establishment of the National Institute of School Leadership's (NISL's) Executive Development Program stemmed from the need to assist school leaders in their ability to promote high performance for all students in their schools. The program emphasizes the role of principals as strategic thinkers, instructional leaders, and creators of a just, fair, and caring culture in which all students meet high standards. Its primary goal is to ensure that the participating school leaders have the knowledge, skills, and tools to effectively set direction for teachers, support their staff in improving instructional practices, and design an efficient organization that becomes a professional learning community. Professional development goals are to provide high-quality instruction (both online and face-to-face), an advanced research-based curriculum, and an interactive approach to learning that includes self-assessments simulations, case studies, school evaluations, and online activities.

Key expectancies for NISL-trained principals include:

- Formulating a clear vision to inspire others in the school communities;
- Implementing fully-aligned, standards-based instructional systems;
- Building effective instructional programs in the core academic subjects, particularly math, language arts and science;
- Using data to produce continuous improvements in instruction and student achievement;
- Providing effective training programs to build a professional learning community for school faculty and staff, and;
- Creating integrated school improvement plans that reflect strategic and systemic thinking.

The curriculum, designed by experts on leadership training across a number of fields, reflects an \$11 million investment and five years of research and piloting. The curriculum is organized into



four courses: World-Class Schooling (Principal as a Strategic Thinker and School Designer, Standards-Based Instruction); Teaching and Learning; Developing Capacity and Commitment; and Driving for Results. Designed to be highly interactive, training sessions use simulations and assignment of “pre-work” and applications (“homework”) to participants.

Prior evaluations of the Executive Development Program prove that the NISL program can be implemented economically and with high fidelity (Meristem Group, 2009). Perhaps more importantly, the research indicates that positive student achievement patterns have been associated with program participation by school leaders. However, these prior studies have used descriptive or correlational designs lacking comparison groups or strong controls over sample selection bias.

More recently, Nunnery, Ross, and Yen (2010) conducted a carefully matched comparison-group ex post facto design to examine NISL program effects in Pennsylvania. Their findings indicate that program participation by school leaders was associated with statistically significant improvement in student achievement for both mathematics and reading over a four-year period. This current study represents a further enhancement in the rigor of the evidence regarding potential effects of the NISL program, as it also is based on an ex post facto, matched comparison design. This interim report provides a preliminary estimate of NISL program effects, with a more sophisticated multi-level modeling of program effects to be completed during the course of 2010.

*Research questions*

The research questions that guided the analyses and results reported herein were:

1. How did 2006-2009 trends in school level performance in mathematics differ between schools served by NISL-trained principals and matched comparison schools at the elementary and middle school levels?
2. How did 2006-2009 trends in school level performance in English/Language Arts (ELA) differ between schools served by NISL-trained principals and matched comparison schools at the elementary and middle school levels?
3. How did trends in math and ELA performance differ between NISL schools and the Commonwealth as a whole?

## METHOD

### *NISL schools*

A total of 131 principals in Massachusetts participated in the NISL program. The analysis sample was restricted to include schools whose principal both completed the NISL program and remained at the same school from 2006 through the end of the 2009 school year. Of the 131 principals who started the program, about 18% ( $n = 23$ ) did not complete the program due to retirement or transfer, and about 27% ( $n = 35$ ) changed schools or positions, thus yielding a potential analysis sample of 73 schools. Of these, two were high schools that were eliminated due to the small school-level sample size of NISL schools at this level, and one was a primary school (K-3) which was excluded because test scores were not available for second grade. Finally, five of the remaining schools were not included because complete test score and demographic data were not available for all years for various reasons (being a new school started after 2006, being a school that was closed during the course of the study, etc.). The final sample for the analysis included 65 NISL schools.

### *Dependent measures*

The dependent measures employed in the study were standardized scores (Z-scores) computed from raw scores on the Massachusetts Comprehensive Assessment Program tests in English/Language Arts (ELA) and mathematics. Z-scores were computed separately for each grade level by subtracting the state-mean from each individual student score, then dividing the difference by the state-wide standard deviation. Individual Z-scores were then aggregated across grade levels served by each school, yielding a single school performance index that reflected the mean Z-score for all tested students in each school.

Schools were classified into grade-level types on the basis of the lowest and highest grades served. Schools serving grades three to four, three to five, or three to six were classified as elementary schools. Schools serving grades five-, six-, or seven- to eight were classified as middle schools, and schools serving grades three- or four- to eight were classified as elementary-middle schools.

#### *Matched comparison school sample*

To facilitate construction of a matched comparison sample, a principal components analysis was performed on the following variables measured in 2006: mean Z-score in English, mean Z-score in Mathematics, the percentage of students receiving free or reduced-price lunch (FRL), the percentage of students receiving special education services (SPED), and the percentage of students with limited English proficiency (LEP). A single component accounted for 63.3% of the variance in these variables across schools. Item loadings were -0.95 for mean English Z-score, -0.94 for mean Math Z-score, 0.83 for FRL, 0.67 for LEP, and 0.50 for SPED. A regression-based factor score was computed from the principal components analysis to yield a composite index to use for matching purposes. Based on the item loadings, higher composite index scores are associated with lower mean achievement and higher rates of FRL, LEP, and SPED.

A matched comparison school sample was constructed by selecting four non-NISL pilot schools with the same grade-level type as each NISL school—two with the closest higher composite index scores and two with the closest lower composite scores. For instances in which composite index scores for NISL schools were within this span, comparison schools were selected to yield a total of four comparison schools per NISL school by alternately selecting a comparison school with composite index scores higher, then lower, than those already included.

To assess the adequacy of the comparison school selection process, a 2 (NISL, non-NISL) X 3 (Level) X 5 (dependent variables) multivariate analysis of variance (MANOVA) was performed with mean 2006 English Z-score, mean 2006 math Z-score, percentage of students receiving free or reduced-price lunch (FRL), percentage of LEP students, and percentage of SPED students as dependent variables. Wilk's lambda was used as the criterion of multivariate significance. The analysis indicated no significant multivariate effect for NISL program status ( $\lambda = 0.986$ ;  $F_{5,310} = 0.91$ ,  $p = 0.48$ ) and no significant NISL X Level interaction effect ( $\lambda = 0.967$ ;  $F_{10,620} = 1.04$ ,  $p = 0.41$ ). This indicates that, across the set of matching variables, there was no multivariate difference between pilot NISL and comparison schools. Likewise, univariate tests on each matching dependent variable indicate no statistically significant differences between NISL schools and comparison schools: English Z-score ( $F_{1,314} = 2.44$ ,  $p = .12$ ), math Z-score ( $F_{1,314} = 1.85$ ,  $p = .17$ ), FRL ( $F_{1,314} = 0.92$ ,  $p = .33$ ); LEP ( $F_{1,314} = 2.41$ ,  $p = .12$ ), or SPED ( $F_{1,314} = 0.12$ ,  $p = .73$ ). As shown in Table 1, although there were no statistically significant differences, NISL schools tended to have lower Z-scores than comparison schools in both English and math at the elementary and middle school levels, although these were relatively equal in the elementary-middle grade configurations. In addition to the matched comparison school analyses, NISL school performance was compared to performance in all Massachusetts' schools as a sensitivity test. Furthermore, 12 of the comparison schools were Cohort 2 NISL schools that began the program in 2008, but had not completed the program by the time 2009 test scores were available. A separate sensitivity test was performed that excluded these schools to determine if their inclusion as comparison schools affected the results of the inferential tests.

Table 1. Mean Scores on Matching Variables by NISL Status and School Level, 2006

Level	NISL program				Special	
	completion	English <sup>1</sup>	Math <sup>1</sup>	Free lunch <sup>2</sup>	LEP <sup>2</sup>	Education <sup>2</sup>
Elementary	Comparison	-.1780	-.1549	.5003	.1028	.1946
	NISL	-.3453	-.3474	.5905	.1130	.1667
Middle	Comparison	-.3964	-.3567	.5093	.0839	.2043
	NISL	-.5715	-.5275	.5650	.1604	.2406
Elementary- middle	Comparison	-.3867	-.3951	.6287	.1052	.2350
	NISL School	-.4278	-.3624	.6334	.1201	.2007

<sup>1</sup>Mean Z-scores. <sup>2</sup>Mean proportions. Sample sizes for Comparison schools (180, 52, 24) and NISL schools (45, 13, 6) for Elementary, Middle, and Elementary-Middle, respectively.

### *Analyses*

Two 2 (NISL status) X 3 (School Level) X 4 (Outcomes from 2006-2009) repeated-measures analyses of variance were performed for both English and math outcomes, with mean Z-scores serving as the dependent measures. In cases where no NISL status X School Level interactions were observed, the School Level variable was dropped. Box's test of the equality of variance-covariance matrices and Levene's tests of homogeneity of variance were performed to test model assumptions. Where these were violated, the Greenhouse-Geiser correction was performed to yield a conservative inferential test of program effects. Cohen's *d* effect size estimates were derived by computing the NISL-Comparison schools differences in 2009 mean Z-scores adjusted for 2006 mean Z-scores:  $d = (2009_{\text{NISL}} - 2006_{\text{NISL}}) - (2009_{\text{Comp}} - 2006_{\text{Comp}})$ . Because the outcome is already expressed in standard deviation units derived from population values, this difference directly provides Cohen's *d*. To control for experiment wise alpha

inflation, Holme's sequential Bonferroni procedure was employed, rendering the criterion  $\alpha = 0.025$ . Similar procedures were employed for the state-wide comparisons.

## RESULTS

### *Matched-samples Comparison School Results*

#### *Math*

Preliminary analyses indicated no NISL status X School Level interaction effects for within-subjects effects ( $F_{6,942} = 0.18, p = .98$ ) or between-subjects effects ( $F_{2,314} = 0.63, p = .53$ ), so analyses were performed using only NISL status as an independent variable. Levene's test indicated that the equality of variance assumption held for all four testing periods. Box's test indicated a possible violation of the equality of covariance matrices assumption ( $F_{10,59173} = 2.63, p = .003$ ), so the Greenhouse-Geiser correction was performed. A test of the within-subjects effects revealed a statistically significant interaction of trends in mean math Z scores and NISL program status ( $F_{2,4,773.3} = 6.32, p = .001$ ). Tests of within-subject contrasts revealed only a statistically significant linear component to the interaction ( $F_{1,318} = 10.78, p = .001$ ). As shown in Figure 1, in 2006 students in comparison schools scored an average of +0.17 standard deviation units higher than students in pilot NISL schools, whereas by 2009 that difference was cut to +0.05 standard deviation units. Follow-up multivariate analysis of variance indicated that, in 2006, comparison schools scored statistically significantly higher than pilot NISL schools ( $F_{1,318} = 7.76, p = .006$ ), but by 2009 there was no statistically significant difference in the performance of pilot NISL and comparison schools ( $F_{1,318} = 0.66, p = .42$ ), yielding a Cohen's  $d$  estimate of effect size of  $d = +.10$ , which was statistically significant at  $p < .01$ . Analyses excluding the 12 cohort 2 NISL schools from the comparison sample yielded similar results ( $F_{3,918} = 0.678, p < .001$ ), although the effect size increased to  $d = +0.12$ , indicating that the



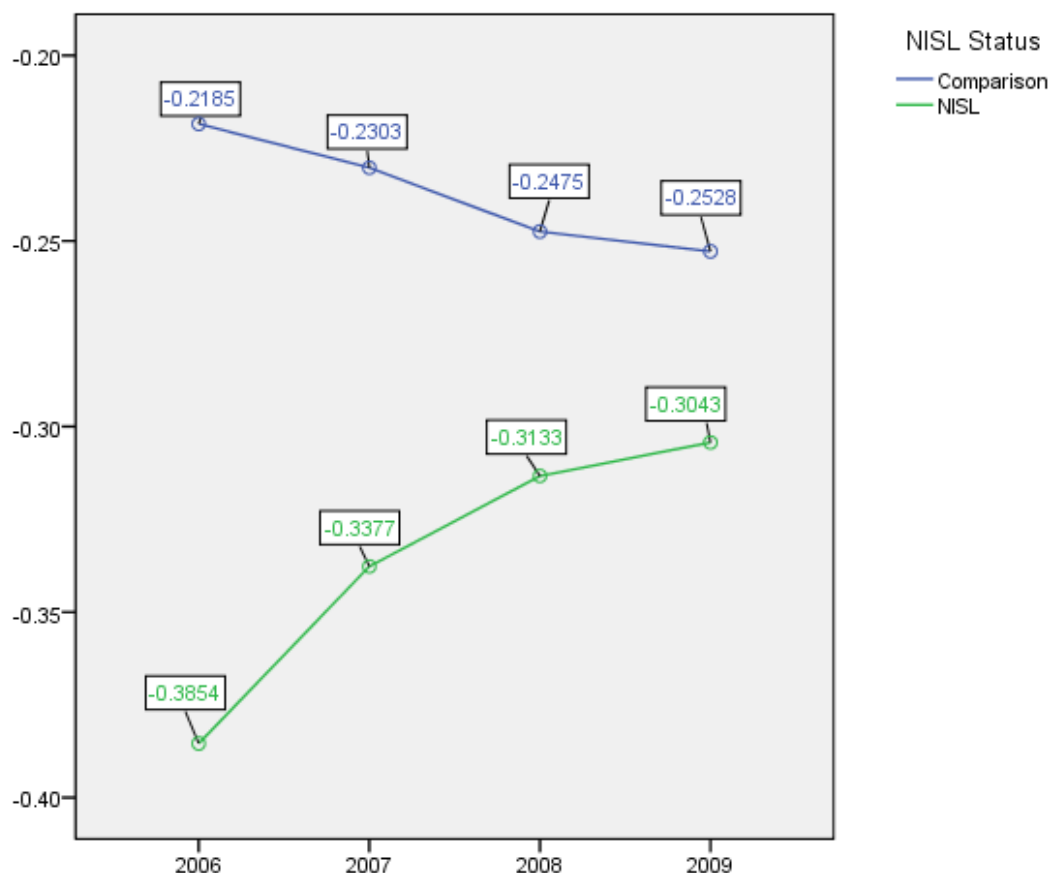


Figure 1. Trends in Mean Math Z-scores in pilot NISL Schools versus Matched Comparison Schools by Year, 2006-2009.

### *English/Language Arts*

As with math, preliminary analyses indicated no NISL status X School Level interaction effects for within-subjects effects ( $F_{6,942} = 0.34, p = .91$ ) or between-subjects effects ( $F_{2,314} = 0.35, p = .70$ ), so analyses were performed using only NISL status as an independent variable. Box's test ( $F_{10,59173} = 1.10, p = .354$ ) and Levene's tests indicated that the assumptions of the model were tenable. Tests of the within-subjects interaction between trends in English/Language Arts performance and NISL status showed no statistically significant difference between performance trends in NISL versus comparison schools ( $F_{3,954} = 0.60, p = .61$ ), yielding an effect size estimate of  $d = +0.00$ . Analyses excluding the 12 cohort 2 NISL schools from the

comparison sample yielded similar results ( $F_{3,918} = 0.74, p = .53$ ), although the effect size increased to  $d = +0.01$ , indicating that the Cohort 2 schools that were excluded from performed somewhat better than the other included comparison schools.

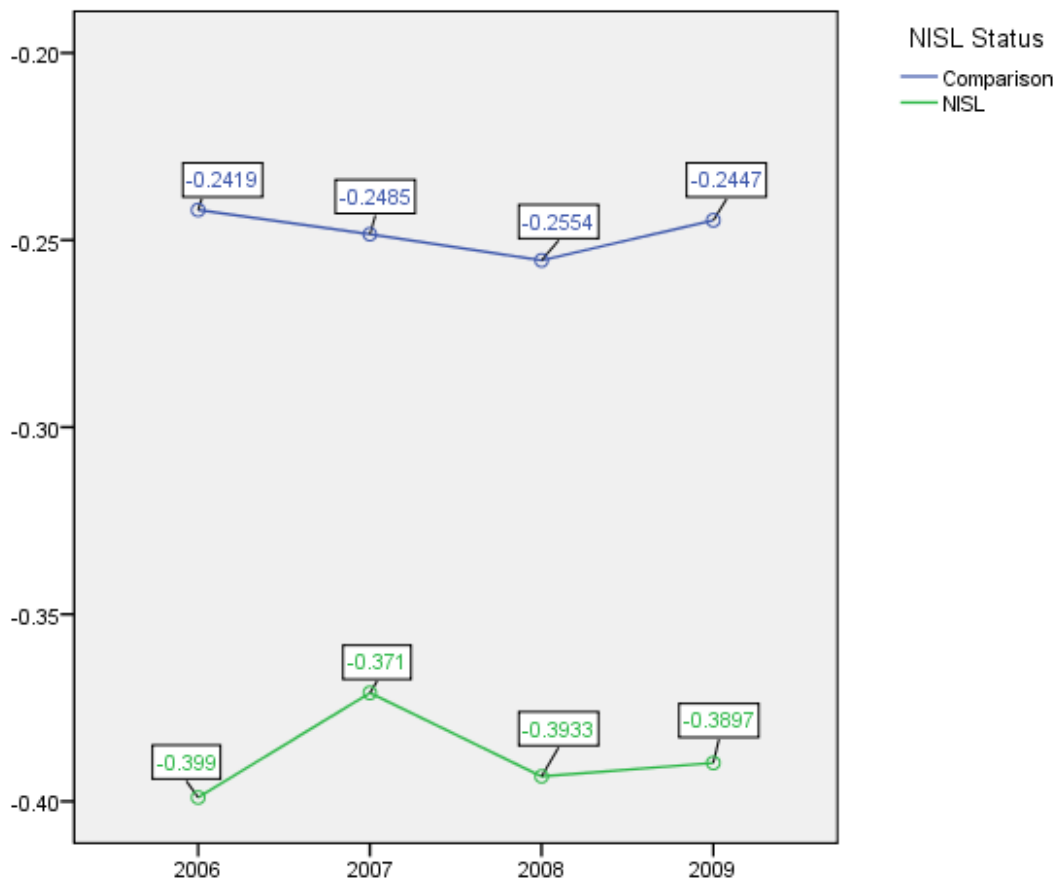


Figure 2. Trends in Mean English/Language Arts Z-scores in NISL versus Matched Comparison Schools by Year, 2006-2009.

Table 2. Mean Scores by NISL Status and Subject Area with Effect Size

	2006	2007	2008	2009	Effect Size
<b>Math</b>					
NISL Schools	-.3854	-.3377	-.3133	-.3043	+.10 (+.12 <sup>a</sup> )
Comparison Schools	-.2185	-.2303	-.2475	-.2528	
<b>English</b>					
NISL Schools	-.3990	-.3710	-.3933	-.3897	+.00 (+.01 <sup>a</sup> )
Comparison Schools	-.2419	-.2485	-.2554	-.2447	

Comparison School  $n = 256$ ; NISL School  $n = 64$ . <sup>a</sup>Excluding 12 Cohort 2 NISL schools from comparison sample.

#### *State-wide Comparisons*

##### *Math*

Preliminary analyses indicated no NISL status X School Level interaction effects for within-subjects effects ( $F_{6,942} = 0.18, p = .98$ ) or between-subjects effects ( $F_{2,314} = 0.63, p = .53$ ), so analyses were performed using only NISL status as an independent variable. Levene's test indicated that the equality of variance assumption held for all four testing periods. Box's test indicated a possible violation of the equality of covariance matrices assumption ( $F_{10,50895} = 3.04, p = .001$ ), so the Greenhouse-Geiser correction was performed. A test of the within-subjects effects revealed a statistically significant interaction of trends in mean math Z scores and NISL program status ( $F_{2.5,3192.3} = 6.91, p < .001$ ). Tests of within-subject contrasts revealed only a statistically significant linear component to the interaction ( $F_{1,1277} = 12.06, p = .001$ ). As shown in Figure 3, in 2006 students in comparison schools scored an average of +0.38 standard deviation units higher than students in NISL schools, whereas by 2009 that difference was cut to +0.30 standard deviation units. Adjusted 2009 means were -0.04 for comparison schools and

+0.04 for NISL schools, yielding a Cohen's  $d$  estimate of effect size of  $d = +.08$ , which was statistically significant at  $p < .01$ .

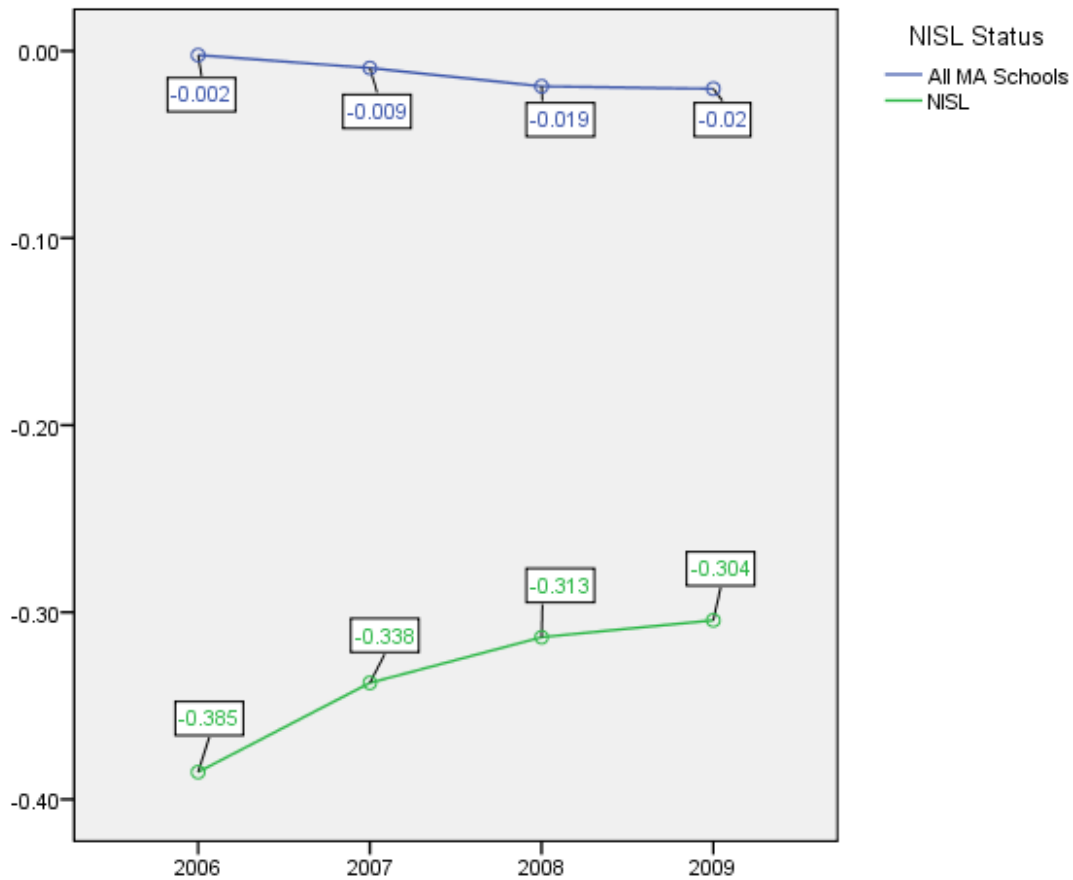


Figure 3. Trends in Mean Math Z-scores in NISL Schools versus All Massachusetts Schools by Year, 2006-2009.

#### *English/Language Arts*

As with math, preliminary analyses indicated no NISL status X School Level interaction effects for within-subjects effects ( $F_{6,3819} = 0.26, p = .96$ ) or between-subjects effects ( $F_{2,1273} = 2.39, p = .09$ ), so analyses were performed using only NISL status as an independent variable. Box's test ( $F_{10,50895} = 2.25, p = .013$ ) and Levene's tests indicated that the assumptions of the model were tenable. (Note: A conservative  $\alpha$  value of .01 is typically used for Box's test). Tests

of the within-subjects interaction between trends in English/Language Arts performance and NISL status showed no statistically significant difference between performance trends in NISL versus comparison schools ( $F_{3,3831} = 0.69, p = .56$ ). Adjusted 2009 means were -0.04 for NISL schools and -0.04 for all Massachusetts' schools, yielding an effect size estimate of  $d = +0.00$ . As shown in Figure 4, performance in NISL schools remained relatively consistent at about -0.39 standard deviation units below the state-wide average.

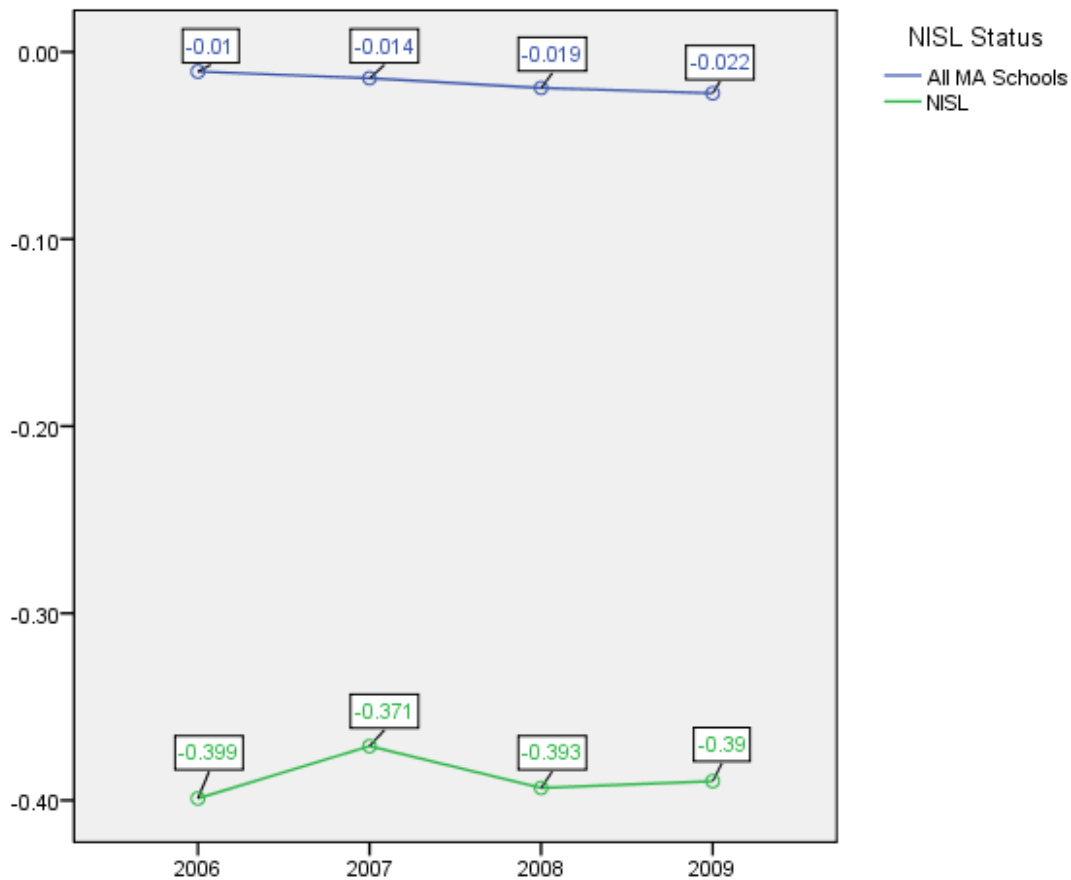


Figure 4. Trends in Mean English/Language Arts Z-scores in NISL Schools versus All Massachusetts Schools by Year, 2006-2009.

Table 3. Comparison of Mean English and Math Scores of NISL Schools versus All Massachusetts Schools by Year

	2006	2007	2008	2009
<b>Math</b>				
NISL Schools	-.3854	-.3377	-.3133	-.3043
Massachusetts	-.0022	-.0091	-.0189	-.0202
<b>English</b>				
NISL Schools	-.3990	-.3710	-.3933	-.3897 <sup>a</sup>
Massachusetts	-.0104	-.0140	-.0192	-.0220

All MA schools,  $n = 1215$ ; pilot NISL Schools,  $n = 64$ . <sup>a</sup>The gap is decreasing; however, it is not statistically significant.

Note. Statewide trends are expected to be relatively flat and levels close to zero because the outcome has been standardized each year to a mean of 0 based on all tested students in the state. They do not equal zero exactly because the means reported are based on school-level aggregates.

## FINDINGS AND DISCUSSION

### *Study Overview*

This study examined the impact of a pilot of NISL's Executive Development Program for principals on student achievement in Massachusetts schools between 2006-2009. The 2006 school year was treated as the baseline year for the analysis. Sixty-five elementary and middle schools served by principals participating in the Executive Development Program were matched on a 1:4 ratio to comparison schools that had similar school performance and demographic profiles in 2006. High schools were not included because only two high schools were led by a NISL-trained principal who remained principal of the school throughout the 2006-2009 period. Mean Z-scores in English/Language Arts (ELA) and Math across all grade levels were computed to create aggregate school performance indices for each year 2006-2009. Repeated-measures analyses were performed to determine whether there were differences in school performance trends between schools served by NISL-trained principals and matched comparison schools and NISL schools and all schools in the Commonwealth.

### *Findings*

#### *Summary effects*

As shown in Figure 5, NISL schools had higher-than-expected performance in 2009 relative to comparison schools ( $d = +0.10$ ) and relative to all schools ( $d = +0.08$ ) in mathematics. On unadjusted results, the NISL schools reduced the overall gap in math performance relative to comparison schools by 0.12 standard deviation units. Results for ELA indicated no difference in aggregate achievement trends between NISL schools and either the matched comparison sample or the population of elementary and middle schools in the Commonwealth.

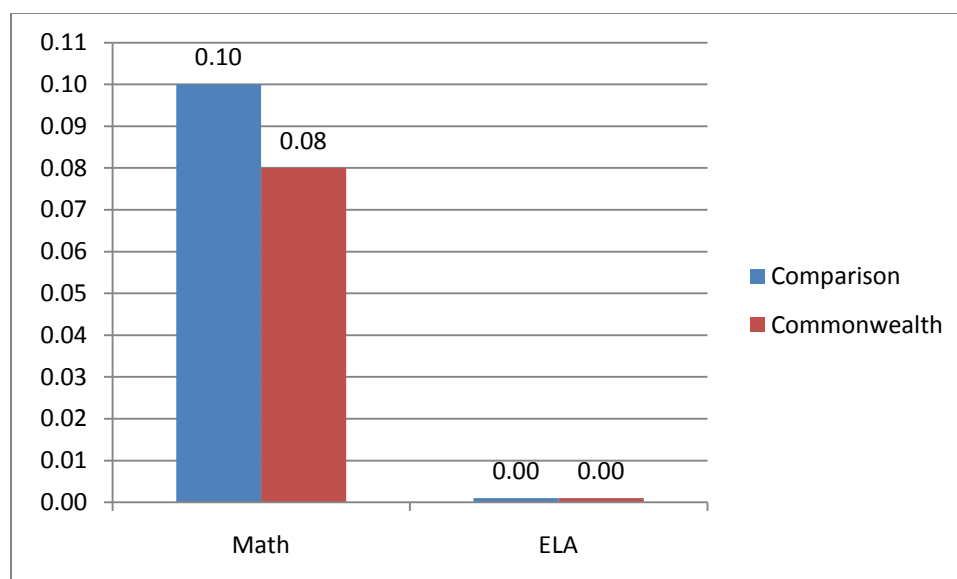


Figure 5. *NISL Effects: 2009 Covariate-adjusted School-level Effect Size Estimates in Math and ELA Relative to Comparison Schools and to All Schools in the Commonwealth.*

### *Discussion*

NISL schools consistently surpassed the comparison schools in math achievement gains at a statistically significant level from the baseline year of 2006 to 2009, with an adjusted effect size of  $d = +0.10$  at the school level, which is quite large relative to results observed in similar studies of comprehensive school reform effects or Title 1 program effects (see, for example, Borman et al., 2003). No statistically significant effects were observed for ELA performance trends, which is not consistent with findings from the Pennsylvania study (Nunnery et al., 2010). However, reading/ELA effects observed in the Pennsylvania study were somewhat smaller than the effects observed on school-level trends in mathematics performance, so in that sense the findings are consistent. Prior research suggests that math achievement is more responsive to interventions in the near term than reading achievement because time and quality of instruction in school matters more for math achievement, whereas English/language arts and reading achievement are more influenced by home experiences and out-of-school opportunities (Bryk &



Raudenbush, 1988; Harris, Kelly, Valentine, & Muhlenbruck, 2000; Rimm-Kaufman, Fan, Chiu, and You, 2007).

In a sense, the lack of a relationship between NISL status and reading trends in the short term lends validity to the observed math effects. The contrast militates against attributing the math effects to selection bias or regression to the mean, both of which potentially are salient threats to the internal validity of the ex post facto design employed in this study. A randomized experiment was not feasible given state and district policies for program implementation. However, the present ex post facto design appears highly rigorous, particularly in minimizing validity threats frequently associated in evaluations of leadership programs with sampling bias. Specifically, participants were described by the state and districts as being mixed in their experiences, success rates, and skills, with some targeted due to demonstrating strong potential for leadership and others due to needing professional development to address weaknesses. Also, the repeated-measures design treated nearly all principals as their own controls in analyzing school achievement patterns over time. Further refinement of the findings, such as examination of differential impact on student subgroups, is expected when the final report is produced. Given the relatively large size of the impact on math achievement observed in this interim study, it is very likely that this finding will be corroborated in the final report, which will be based on a more precise multilevel model of individual and school effects. Further, when Cohort 2 NISL schools were excluded from the comparison sample, effect size estimates increased for both reading and mathematics, suggesting a possible positive effect of the program in these schools although their principals had not yet completed the program.

## References

- Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125-230.
- Bryk, A., & Raudenbush, S. (1988). Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model. *American Journal of Education, 97*(1), 65-108. Retrieved from Education Research Complete database.
- Harris, C., Kelly, C., Valentine, J., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development, 65*(1), v-118. Retrieved from PsycINFO database.
- Meristem Group (2009). *National Institute for School Leadership (NISL): Massachusetts program implementation 2005-2008*. Boston, MA: Meristem Group.
- Nunnery, J., Ross, S.M., & Yen, C. (2010). *The Effect of the National Institute for School Leadership's Executive Development Program on School Performance Trends in Pennsylvania*. Norfolk, VA: The Center for Educational Partnerships at Old Dominion University.
- Rimm-Kaufman, S., & Chiu, Y. (2007). Promoting social and academic competence in the classroom: An intervention study examining the contribution of the Responsive Classroom approach. *Psychology in the Schools, 44*(4), 397-413. doi:10.1002/pits.20231.