# The Center for Educational Partnerships

## Table of Contents

*Appendix: Supplementary Tables*



*List of Tables*

**Table**

**List of Figures**

# School Performance Trends in Pennsylvania

# School Performance Trends in Pennsylvania

## INTRODUCTION

established to train school leaders to drive their schools to high performance. 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 staffs, and design an efficient organization. The curriculum,

e).ET B4(ff)6hva(sylva9s. su4(d)dBT 1 0 0 1 72.024 5)adoT(e)417.55vanT 1 0 0 1 175.58 610.660 0700417.55uler

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. Training sessions are designed to be highly

-

participants.

Prior evaluations of the Executive Development Program have shown that the NISL program can be economically implemented with high fidelity (Meristem Group, 2009). Importantly, 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. Accordingly, to provide more rigorous evidence and support causal conclusions regarding program impacts, the present longitudinal study of student achievement in Pennsylvania schools, from 2006-2009, was conducted . A carefully matched comparison-group ex post facto design was employed in which schools served by principals participating in the program were individually matched to control schools with similar school performance and demographic profiles in the baseline (pre-program) year of 2006. The specific research questions addressed were:

1. How do the trends in school level performance in mathematics differ between schools served by NISL-trained principals and matched comparison schools at the elementary, middle, and high school levels?
2. How do the trends in school level performance in reading and English/Language Arts (ELA) differ between schools served by NISL-trained principals and matched comparison schools at the elementary, middle, and high school levels?

## METHOD

*Sample*

Data from all Pennsylvania elementary schools with complete test score data from 2005-2006 through 2008-2009 were initially considered for inclusion in the analyses. There were a total of 70 NISL elementary schools, 19 NISL middle schools, and 12 NISL high schools. As explained below, 36 of the NISL elementary schools were included in a within-district matched samples analysis, and 32 were included in a separate set of analyses based on an out-of-district matched comparison sample. In the elementary school sample, 19 of the NISL principals

and others regarded as needing professional development support to improve instructional leadership skills. Actual applicants were selected by regional coordinators using an evaluation

percentages of students with limited English proficiency (18.4% versus 14.8%), and a lower percentage of students who were proficient in math (78.0% versus 80.5%). For the out-of-district matches, all matching variables were within 0.1%, except percentage of limited English proficient students served (1.1% in comparison sites versus 0.6% in NISL sites). Table A1.A in the appendix provides 2006 characteristics for each pair in the within-district matched samples, and Table A1.B provides this information for the out-of-district matched pairs.

*Secondary school matching procedure*. It was not possible to individually match middle and high school NISL schools to a comparison school within the same school district in many cases, the NISL secondary school was the only school at that level within the district. At the middle and high school levels, an out-of-district match was made to each NISL school by matching the NISL school to a comparison school with the closest factor score. There were 19 NISL middle schools and 14 NISL high schools. As shown in Table 1, the matching process led to relatively well-matched samples for the middle school analyses, although the NISL sample

*Measures*

*School demographics*. The proportions of students in tested grade levels (3-8 and 11) who were economically disadvantaged, received special education services as evidenced by the existence of an individualized education plan (IEP), or who were classified as having limited English proficiency (LEP) were computed for each school. In addition to be used to establish

**Table 1**

Selected 2006 Characteristics of NISL and Comparison Schools

| School Type | Economically Disadvantaged % | IEP % |
|---|---|---|

matches, these variables were incorporated as covariates in the inferential statistical models.

*School performance indices.* Summary indices of school performance were constructed for both reading/ELA and mathematics by computing the proportion of students at all tested grade levels (grades 3 to 8 and grade 11) who scored proficient or higher on the Pennsylvania State Assessment. Note that for high schools, test scores were available only for eleventh grade.

*Analyses*

At each grade-level configuration (elementary, middle, and high school) and for each subject area (reading and math), a repeated-measures analysis were conducted with 2006, 2007, 2008, and 2009 school performance index values as the dependent variable. NISL status was treated as the independent variable. Covariates included the 2006 proportions of students who were economically disadvantaged, had IEPs, or who were designated as limited English proficient. School performance index values in 2006 were also included as covariates to facilitate comparison of trend lines. The analyses were weighted based on the average number of students tested per year for each school between 2006 and 2009 in each respective subject. Importantly, the repeated-measures analyses employed and the fact that there was low principal mobility (only 4 NISL principals, or about 4%, were re-assigned to different schools at any point from 2006-2009) during the period of the study provided direct control over sampling bias, which is often a strong validity threat in evaluating school and teacher leadership programs. That is, the achievement trajectories analyzed for NISL and non-NISL schools reflected pre- and post-program outcomes almost exclusively associated with the same school leaders. Thus, essentially, each school leader served as his/her own control for analyzing longitudinal achievement patterns.

**Table 2**

Results of Levene s tests of equal variances

| Subject and Year of Testing | | | | | | | |
|---|---|---|---|---|---|---|---|
| School Level and Results | Mathematics | | | | Reading | | |
| Elementary (w/in district) | 2007 | 2008 | 2009 | | 2007 | 2008 | 2009 |
| df | 1, 10467 | 1, 10467 | 1, 10467 | | 1, 10455 | 1, 10455 | 1, 10455 |
| F | 228.5 | 239.6 | 114.5 | | 58.8 | 104.7 | 3.39 |
| Prob. | <.001 | <.001 | <.001 | | <.001 | <.001 | .06 |
| Elementary (out of district) | 2007 | 2008 | 2009 | | 2007 | 2008 | 2009 |
| df | 1, 10088 | 1, 10088 | 1, 10088 | | 1, 10077 | 1, 10077 | 1, 10077 |
| F | 129.6 | 0.5 | 0.4 | | 344.7 | 724.6 | 655.8 |
| Prob. | <.001 | .50 | .55 | | <.001 | <.001 | <.001 |
| Middle | 2007 | 2008 | 2009 | | 2007 | 2008 | 2009 |
| df | 1, 16128 | 1, 16128 | 1, 16128 | | 1, 16416 | 1, 16416 | |

schools ($M$ = 81.7) had surpassed comparison schools ($M$ = 81.5). Figure 1 displays the observed trend lines in math school performance index values for each group. As depicted, the trend lines cross, indicating greater gains over time for NISL schools.

**Table 3**

Mean Percentage of Students Scoring Proficient or Higher in Mathematics by School Type and Year, Elementary Schools (Within-district Matched Samples)

| School Type | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| NISL | 78.8 | 80.8 | 80.3 | 81.7 |
| Comparison | 80.8 | 82.0 | 81.1 | 81.5 |

*Note.* Weighted by number of students tested, so figures may not correspond to unweighted means presented in Table 1.

*Inferential tests.* Tests of within-subjects effects revealed a statistically significant trend in math school performance index values over time ($F_{2.7,\ 23659} = 620.2$, $p < .01$), suggesting that

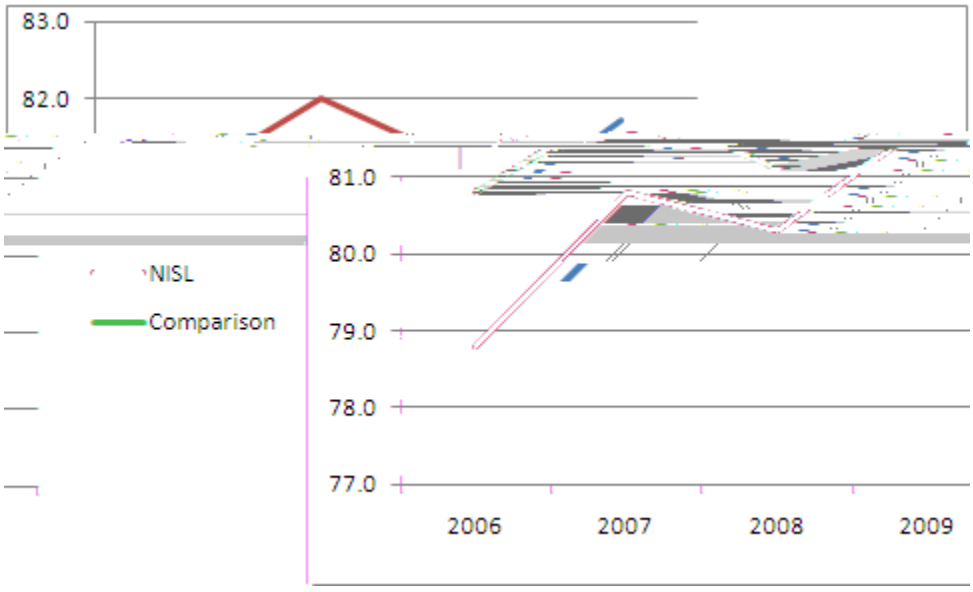


Figure 1. Unadjusted Mean Percentage of Students Scoring Proficient or Higher in Mathematics by School Type (NISL, comparison) and Year, Elementary Schools (Within-district Matched Samples)

subjects interaction effect was observed ($F_{2.7,\ 23659}$ = 119.3, $p$ < .001), indicating that the trend lines in school-level math performance were not equal in NISL and comparison sites. In addition to the significant difference in the linear component of the trend lines, within-subjects contrasts indicated a statistically significant quadratic effect of NISL status on math score trends ($F_{1,10463}$ = 68.2, p <.001). As shown in Figure 2, the significant linear effect indicates that, across the time period, NISL schools gained at a greater average rate than comparison schools. The statistically significant quadratic effect indicates an acceleration in the rate of gain over time for NISL schools.

Figure 2. Covariate Adjusted Trend Lines in Math School Performance Index Values by School Type, 2006-2009 Elementary Schools (Within-district Matched Samples).

*Reading*

*Descriptive results*. The unadjusted mean school performance index values in 2006 were 70.0 for comparison schools and 69.5 for NISL schools. As shown in Table 4, by 2009 NISL schools ($M$ = 73.6) had virtually equal reading performance as comparison schools ($M$ = 73.7).

**Table 4**

Mean Percentage of Students Scoring Proficient or Higher in Reading by School Type and Year, Elementary Schools

| School Type | [redacted] |
|---|---|

time across schools in the analysis sample. The economically-disadvantaged ($F_{2.4,\ 25135}$ = 226.0, $p$ < .001),), LEP ($F_{2.4,\ 25135}$ = 261.9, $p$ < .01), and IEP ($F_{2.4,\ 25135}$ = 65.9, $p$ < .01) covariates were also statistically significant predictors of trend. A statistically significant *NISL status X reading/ELA* within-subjects interaction effect was observed ($F_{2.4,\ 25135}$ = 28.7, $p$ < .01), indicating that the trend lines in school-level reading/ELA performance were not equal in NISL and comparison sites. In addition to the significant difference in the linear component of the trend lines, within-subjects contrasts indicated a statistically significant cubic effect of NISL status on math score trends ($F_{1,10451}$ = 42.5, p <.001). As shown in Figure 4, the significant linear effect indicates that, across the time period, NISL schools gained at a greater average rate than comparison schools after controlling for school demographics. The statistically significant cubic effect was produced because NISL schools had a statistically significantly lower rate of decline from 2007 to 2008 relative to comparison schools.

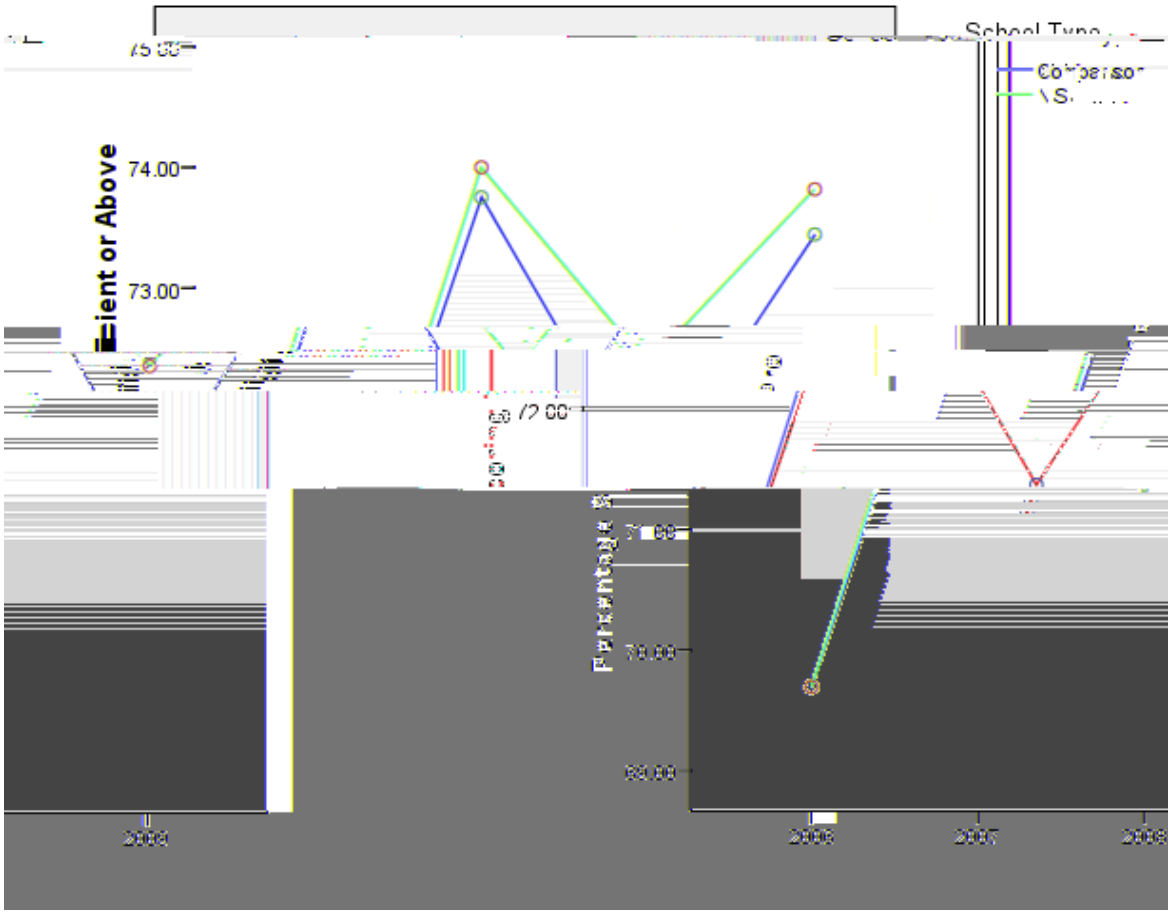


Figure 4. Covariate-adjusted Trend Lines in Reading/ELA School Performance Index Values by School Type, 2006-2009 Elementary Schools (Within-district Matched Samples).

## Elementary Schools: Out-of-district Matched Samples

*Mathematics*.

*Descriptive results*. The unadjusted mean school performance index values in 2006 were 78.6 for comparison schools and 78.5 for NISL schools. As shown in Table 5, by 2009 NISL schools ($M$ = 81.4) had surpassed comparison schools ($M$ = 77.9). Figure 5 displays the observed trend lines in math school performance index values for each group.

**Table 5**

Mean Percentage of Students Scoring Proficient or Higher in Mathematics by School Type and Year, Elementary Schools (Out-of-district Matched Samples).

| School Type | 2006 | 2007 |
|---|---|---|

schools. The economically-disadvantaged ($F_{2.8, 28499} = 74.2$, $p < .001$), LEP ($F_{2.8, 28499} = 260.3$, $p < .001$), and IEP ($F_{2.8, 28499} = 27.1$, $p < .001$) covariates were also statistically significant predictors of trend. A statistically significant *NISL status X math* within-subjects interaction effect was observed ($F_{2.8, 28499} = 596.1$, $p < .001$), indicating that the trend lines in school-level math performance were not equal in NISL and comparison sites. In addition to the significant difference in the linear component of the trend lines, within-subjects contrasts indicated a statistically significant quadratic effect of NISL status on math score trends ($F_{1,10084} = 203.8$, $p < .001$). As shown in Figure 6, the significant linear effect indicates that, across the time period, NISL schools gained at a greater average rate than comparison schools. The statistically significant quadratic effect indicates an acceleration in the rate of gain over time for NISL schools.

**Table 6**

Mean Percentage of Students Scoring Proficient or Higher in Reading by School Type and Year, Elementary Schools (Out-of-district Matched Samples).

| School Type | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| NISL | 68.1 | 71.8 | 70.8 | 72.0 |
| Comparison | 70.0 | 69.9 | 70.1 | 70.4 |

*Note.* Weighted by number of students tested, so figures may not correspond to unweighted means presented in Table 1.

Figure 7. Unadjusted Mean Percentage of Students Scoring Proficient or Higher in Reading by School Type (NISL, comparison) and Year, Elementary Schools (Out-of-district Matched Samples).

*Inferential tests.* Tests of within-subjects effects revealed a statistically significant trend in reading/ELA school performance index values over time ($F_{2.8,\ 27774} = 283.7$, $p < .001$), suggesting that the percentage of students achieving proficiency in reading/ELA increased over time across schools in the analysis sample. The economically-disadvantaged ($F_{2.8,\ 27774} = 25.4$, $p$

# School Performance Trends in Pennsylvania

Figure 9. Unadjusted Mean Percentage of Students Scoring Proficient or Higher in Mathematics by School Type (NISL, comparison) and Year, Middle Schools

*Inferential tests.* Tests of within-subjects effects revealed a statistically significant trend 5(r in Ma)4(ther

**School Performance Trends in**

**Table 8**

Mean Percentage of Students Scoring Proficient or Higher in Reading/ELA by School Type and Year, Middle Schools

| School Type | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| NISL | 72.2 | 71.9 | 73.8 | 75.2 |
| Comparison | 69.7 | 69.9 | 69.0 | 70.7 |

*Note.* Weighted by number of students tested, so figures may not correspond to unweighted means presented in Table 1.

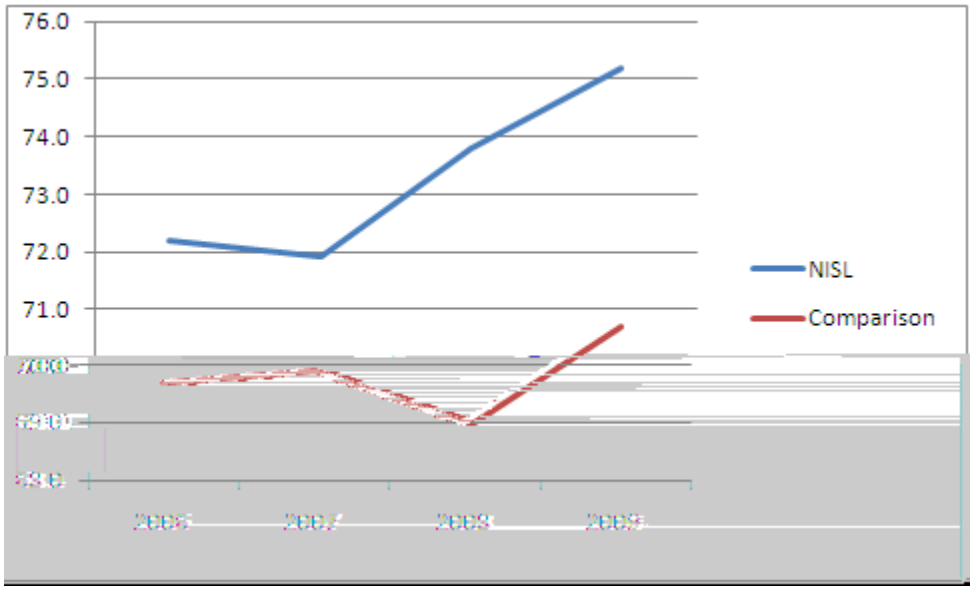


Figure 11. Unadjusted Mean Percentage of Students Scoring Proficient or Higher in Reading by School Type (NISL, comparison) and Year, Middle Schools

*Inferential tests.* Tests of within-subjects effects revealed a statistically significant trend in reading/ELA school performance index values over time ($F_{2.9,\ 47970} = 1759.4$, $p < .001$), suggesting that the percentage of students achieving proficiency in reading/ELA increased. The economically-disadvantaged ($F_{2.9,\ 47970} = 1010.9$, $p < .001$), LEP ($F_{2.9,\ 47970} = 422.3$, $p < .001$),

and IEP ($F_{2.9, 47970} = 687.0$, $p < .01$) covariates were also statistically significant predictors of trend. A statistically significant *NISL status X reading/ELA* within-subjects interaction effect was observed ($F_{2.9, 47970} = 1196.2$, $p < .001$), indicating that the trend lines in school-level reading/ELA performance were not equal in NISL and comparison sites. In addition to the significant difference in the linear component of the trend lines, within-subjects contrasts indicated a statistically significant quadratic effect of NISL status on reading/ELA score trends ($F_{1,16412} = 43.05$, $p < .001$). As shown in Figure 12, the significant linear effect indicates that, across the time period, NISL schools gained at a greater average rate than comparison schools. The statistically significant quadratic effect indicates an acceleration in the rate of gain over time for NISL schools.

Figure 12. Covariate-adjusted Trend Lines in Reading/ELA School Performance Index Values by School Type, 2006-2009 Middle Schools.

## High Schools

*Mathematics*

*Descriptive results*. The unadjusted mean school performance index values in 2006 were 44.9 for comparison schools and 46.2 for NISL schools. As shown in Table 9, by 2009 NISL schools ($M$ = 52.1) had increased proficiency rates by about 5%, whereas comparison schools declined by about 0.5% ($M$ = 44.4). Figure 13 displays the observed trend lines in math school performance index values for each group.

**Table 9**

Mean Percentage of Students Scoring Proficient or Higher in Mathematics by School Type and Year, High Schools

| School Type | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| NISL | 46.2 | 49.4 | 53.6 | 52.1 |

= 168.4, *p*

*Reading*

*Descriptive results*. The unadjusted mean school performance index values in 2006 were 64.9 for comparison schools and 59.6 for NISL schools. As shown in Table 10, by 2009 NISL schools (*M* = 60.2) had increased proficiency rates by 0.6%, whereas comparison schools declined by about 3% (*M* = 61.8). Figure 15 displays the observed trend lines in reading/ELA school performance index values for each group.

**Table 10**

Mean Percentage of Students Scoring Proficient or Higher in Reading/ELA by School Type and Year, High Schools

| School Type | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| NISL | 59.6 | 62.0 | 61.2 | 60.2 |
| Comparison | 64.9 | 60.0 | 63.6 | 61.8 |

*Note.* Weighted by number of students tested, so figures may not correspond to unweighted means presented in Table 1.

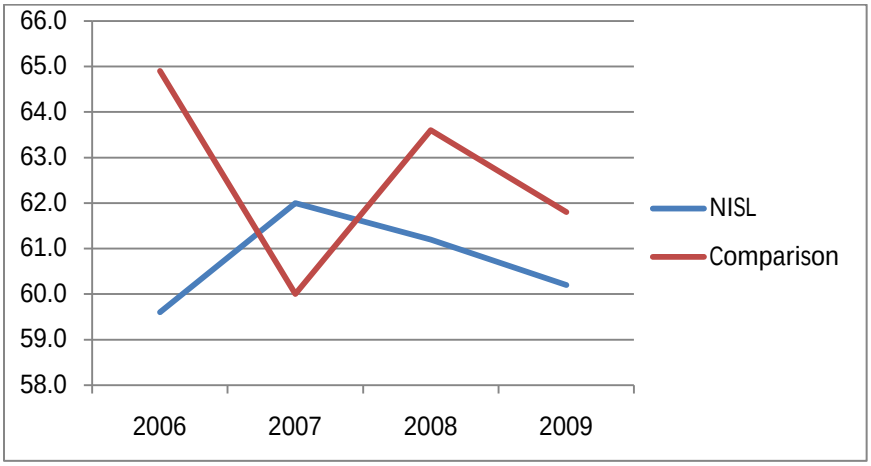


Figure 15. Unadjusted Mean Percentage of Students Scoring Proficient or Higher in Reading/ELA by School Type (NISL, comparison) and Year, High Schools

*Inferential tests.* Tests of within-subjects effects revealed a statistically significant trend in reading/ELA school performance index values over time ($F_{2.4,\ 14121} = 403.3$, $p < .001$), suggesting that the percentage of students achieving proficiency in reading/ELA declined over time across schools. The economically-disadvantaged ($F_{2.4,\ 14121} = 40.1$, $p < .001$), LEP ($F_{2.4,\ 14121} = 354.2$, $p < .001$), and IEP ($F_{2.4,\ 14121} = 32.5$, $p < .001$) covariates were also statistically significant predictors of trend. A statistically significant *NISL status X reading/ELA* within-

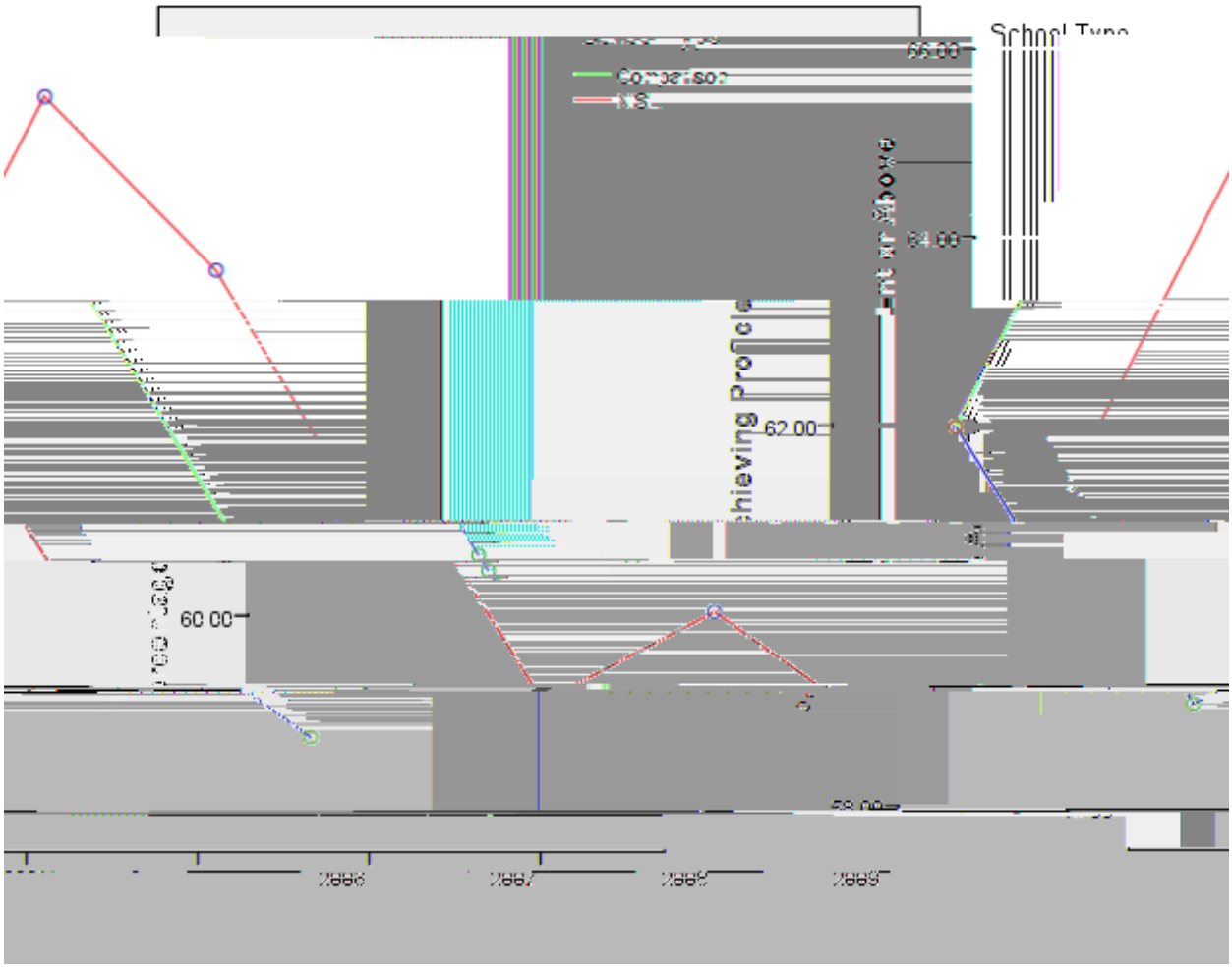


Figure 16. Covariate-adjusted Trend Lines in Reading/ELA School Performance Index Values by School Type, 2006-2009 High Schools.

## FINDINGS AND DISCUSSION

*Study Overview*

principals on student achievement in Pennsylvania schools between 2006-2009. The 2006 school year was treated as the baseline year for the analysis. Roughly half of the NISL-trained principals started the program during the 2007 school year and finished in the 2008 school year, whereas the other half started during the 2008 school year and finished in the 2009 school year. Schools served by principals participating in the Executive Development Program were individually matched to comparison schools with similar school performance and demographic profiles in 2006. For elementary schools, it was possible to make individual school matches within the same school district for 36 school pairs. An additional 32 elementary schools were included in an out-of-district matched comparison sample. For all middle and high schools, it was necessary to match outside the school district. The percentages of students achieving proficient or above in mathematics and reading or English/Language Arts (ELA) across all grade levels were used 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.

*Findings*

*Summary effects*. As shown in Figure 17, NISL schools had higher-than-expected performance in 2009 relative to comparison schools at all grade levels in both subject areas. The largest differences between the percentages of NISL and comparison students achieving proficiency were in mathematics: +2.69%, +3.71%, +1.70%, and +5.52% for elementary within-

district, elementary out-of-district, middle schools, and high schools, respectively. Smaller, but statistically significant gains were observed for reading/ELA: +0.37%, +2.55%, +1.63%, and +1.89%, the four school cohorts, respectively.

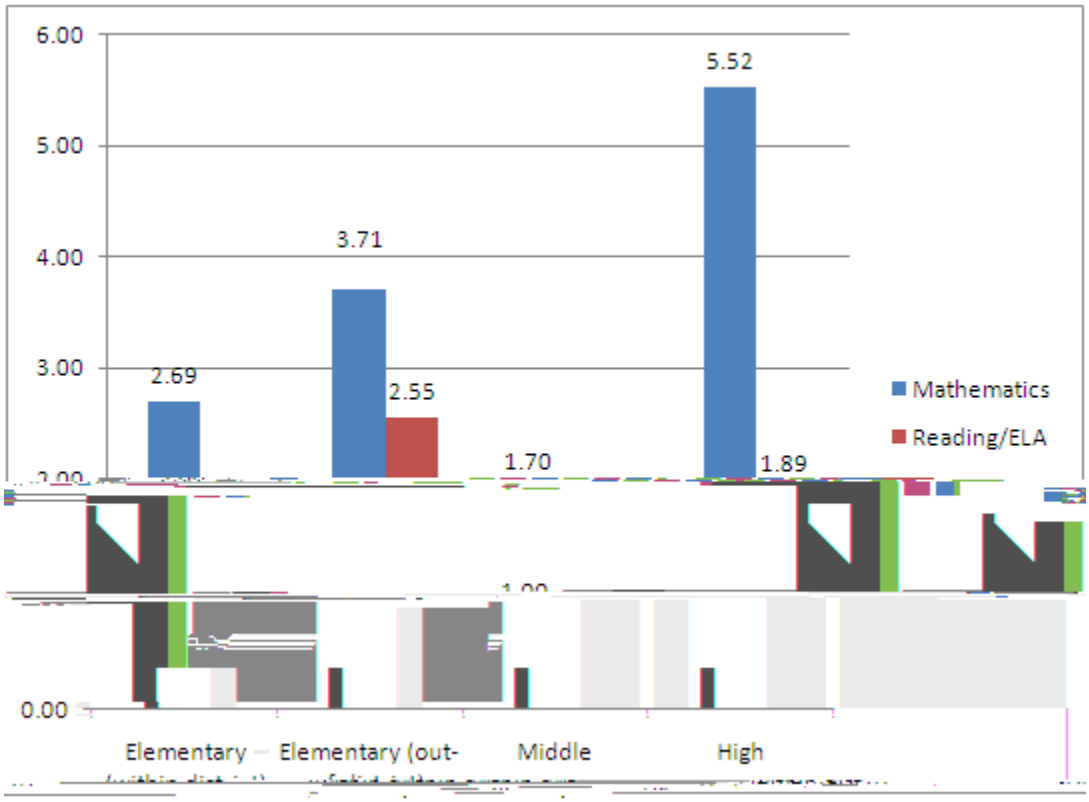


Figure 17. NISL Effects: 2009 Covariate-adjusted Differences in Percentages of Students Achieving Proficiency Relative to Comparison Schools .

*Elementary Schools*. For the matched within-district elementary school sample, statistically significant differences in school performance trends were observed between NISL and comparison schools in both mathematics and reading/ELA. In both cases, NISL schools had statistically significantly higher rates of improvement in school performance than did comparison schools. After controlling for differences in school demographics and 2006 school performance, NISL elementary schools had about 2% more students achieve proficiency in math than did comparison schools in 2009. The adjusted effects on reading, although statistically

significant, were smaller (about +0.5% difference favoring NISL schools). In terms of unadjusted results, NISL schools gained 3.9% versus 0.7% for comparison sites in mathematics, and 4.1% versus 3.7% in reading/ELA. For the matched out-of-district elementary schools, significant trends favoring NISL over comparison schools were also indicated in both mathematics and reading/ELA. Significant quadratic effects in mathematics further revealed acceleration in the growth rate over time for NISL schools.

*Middle schools*. As with the elementary school results, statistically significant positive effects of NISL status were observed for both mathematics and reading in the middle school sample. On an adjusted basis, NISL middle schools had about 2% more students scoring proficient or better in both mathematics and reading/ELA than comparison schools. As with elementary schools, a significant quadratic effect on mathematics school performance trends revealed that the rate of improvement was accelerating over time in NISL schools.

*High schools*. In mathematics, NISL schools had statistically significantly higher rates of improvement, with adjusted differences in 2009 performance equal to 5%. In reading/ELA, covariate adjusted differences significantly favored NISL over comparison sites by about 1.5%, although overall adjusted performance declined between 2007 and 2009 for both groups of schools.

*Discussion*

NISL schools consistently surpassed the comparison schools in achievement gains at a statistically significant level from the baseline year of 2006 to 2009. A randomized experiment was not feasible given state and district policies for program implementation (e.g., see Appendix B). 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.

Predictably, the achievement gains for NISL principals were strongest in 2008 and 2009 as levels of participation in the NSL program (both number of principals and exposure) increased. The significant quadratic effects obtained in several of the analyses reflected this trend for program effects to accelerate over time. As depicted in Figure 17, across the four school cohorts examined (within-district-matched elementary schools, out-of-district matched elementary schools, middle schools, and high schools), NISL schools surpassed comparison schools in the percentage of students achieving proficiency in mathematics by 2.69%, 3.71%, 1.70%, and 5.52%, respectively; and in reading/ELA by .37%, 2.55%, 1.63%, and 1.89%. Given that approximately 40,000 students were included in the combined samples, these advantages appear highly meaningful. For example, across the high school subsample alone, replication of the present NSL effects in mathematics and reading/ELA in similar schools would result in about 275 and 103 more students achieving proficiency on the respective tests. Given that half of the principals began the program in 2007 and the other half not until 2008, the present usage of 2009 as the most distant assessment year certainly seems likely to under-estimate potential program impacts. An additional consideration is that principals need time to implement new strategies in ways that impact teachers, who in turn, need time to improve instruction, learning, and achievement. Follow-up evaluation research of the present 2007 and 2008 cohorts, therefore, is strongly encouraged to determine post-program effects over a longer time period.

## References

The Meristem Group (2009). *National Institute for School Leadership (NISL): Massachusetts program implementation 2005-2008*. Boston, MA: Meristem Group.

# Appendix A: Supplementary Tables

**Table A1.A**

**Selected 2006 Characteristics of Mat**

**Table A1.A Continued**

**Selected 2006 Characteristics of Matched Comparison and NISL Schools in 2006: Elementary Within-district Matches**

| | | | | |
|---|---|---|---|---|
| 14 | Control | .1222 | .0333 | .0000 |

**Table A1.A Continued**

**Selected 2006 Characteristics of Matched Comparison and NISL Schools in 2006: Elementary Within-district Matches**

| | | | | | | |
|---|---|---|---|---|---|---|
| 30 | Control | .1166 | .0429 | .0061 | 79.9667 | 73.8000 |
| | Treatment | .1495 | .0935 | .0000 | 79.4333 | 74.6333 |
| 31 | Control | .2111 | .0889 | .0000 | 84.0000 | 67.6667 |
| | Treatment | .1333 | .0571 | .0000 | 73.1333 | 61.0000 |
| 32 | Control | .2222 | .1111 | .0000 | 88.1000 | 77.3000 |
| | Treatment | .1714 | .1286 | .0000 | 90.9000 | 80.8500 |

**Table A1.B**

**Selected 2006 Characteristics of Matched Comparison and NISL Schools in 2006: Elementary Out-of-district Matches**

| Pair | Program | FRL2006 | IEP2006 | LEP2006 | MATH 2006 | READING 2006 |
|---|---|---|---|---|---|---|
| 101 | Control | .3545 | | .0000 | 71.1000 | 67.1250 |
| | Treatment | | | | | |

**Table A1.B**

**Table A1.B Continued**

**Table A3 Continued**

**School Performance Trends in**