

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE SECTION 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

Evaluation of the Cleveland Scholarship Program: Second-Year Report (1997-98)

18 November, 1998

Submitted by:

Kim K. Metcalf, Director
Patricia Muller
William Boone
Polly Tait
Frances Stage
Nicole Stacey

**The Indiana Center for Evaluation
Indiana University
174 Smith Research Center
Bloomington, Indiana 47408**

Evaluation of the Cleveland Scholarship Program: Second-Year Report (1997-98)

School choice continues to be the most contentious issue in U.S. education, and choice made possible through publicly-funded tuition vouchers represents the paramount of choice proposals (Witte, 1992). While 16 such programs provided funding that allowed families to choose from among either public or private schools in 1997-98, only two were supported with public funds. In Wisconsin, the Milwaukee Parental Choice Program was expanded in 1998 to include religious as well as secular private schools and supported nearly 6,000 children attending 86 private schools. In Ohio, the Cleveland Scholarship and Tutoring Grant Program completed its second year of operation, expanding from 1,801 students in grades kindergarten through three during 1996-97 to 3,000 students in grades kindergarten through four in 1997-98. Both programs are confronting challenges to their constitutionality, but each continues to grow with substantial public support (Van Dunk, 1998).

A multi-year evaluation of the Cleveland choice program was initiated during the program's first year by a team of investigators from the Indiana Center for Evaluation at Indiana University. Over a period of at least three years, the project will examine the impact of the choice program on students, families, and schools, focusing on investigating the effects of the program on students' academic achievement. During the first year, the evaluation examined two factors: (1) characteristics of students whose families chose to use a voucher to move them from a Cleveland public school to one of 41 participating private schools; and (2) the impact of participation in the voucher program on these students' academic achievement. The team analyzed pre-program data drawn from Cleveland public school records including student test data. Further, in the spring of the students' first year in the program, the team tested 94 scholarship program students attending 39 previously existing private schools and 549 non-program public school students.¹ All of the students were enrolled in third-grade during the 1996-97

¹ As noted in Metcalf, Boone, Stage, Chilton, Muller, and Tait (1997) post-hoc analyses conducted as part of the previous evaluation found that there were no reliable differences between the public schools included in the study and those excluded on any of 20 available school variables.

year. The findings were reported in Metcalf, Boone, Stage, Chilton, Muller, and Tait (1997) and included:

- Voucher students were very similar to their public school peers in terms of gender, ethnicity, family income, and family living arrangement. Approximately half of the students (51%) were females; most (88%) were non-white; most (85%) qualified for the free lunch program; and a majority (63%) lived with only their mother.
- Voucher students had been performing at slightly higher levels of academic achievement than their public school peers in second-grade before they entered their voucher schools (approximately 7.5 NCEs in vocabulary, 5.0 NCEs in comprehension).
- After controlling for demographic variables and pre-program achievement, there were no significant differences ($p \leq .05$) in third-grade achievement between voucher students and their public school peers (39.5 and 40.0 in reading, 37.3 and 39.7 in language, 37.9 and 36.5 in mathematics, 32.9 and 36.0 in science, 36.0 and 36.6 in social studies for scholarship and public school students respectively).

It was noted that these findings were preliminary, based on only the first of a multi-year evaluation, and that caution was necessary in interpreting them. Much more information was necessary before firm conclusions could be drawn about the program and its impact on students and schools. Critical to understanding the efficacy of the Cleveland choice program were data to be collected as students continue in the program over several years. Particularly important was ensuring the collection of equivalent achievement test data for students attending the two newly established private schools that were opened on an entrepreneurial basis in response to the creation of the scholarship program.² These schools represent a new and potentially unique educational environment that may be encouraged in market-driven choice programs (see Chubb & Moe, 1990).

² Tests were not administered to scholarship students enrolled in two newly established private schools during the first-year evaluation. As a result, comparisons were possible only between public schools and established private schools. However, this problem was corrected in the second year of the evaluation.

The present report documents findings from the 1997-98 school year based on continued examination of the academic progress of students who were included in the first-year evaluation.³

Specifically, the second-year evaluation was guided by the following questions:

- 1) Are there differences between students who returned to the scholarship program during their fourth-grade year and those who did not return after their third-grade year?
- 2) Are there differences between fourth-grade scholarship and public school students with regard to demographic and background characteristics or pre-program achievement?
- 3) Are there differences in classroom-relevant variables (e.g., class size, teachers' education level, and teachers' experience) between scholarship classes and public school classes?
- 4) What are the effects of the scholarship program on students' academic achievement after two years in the program and when other relevant variables are controlled?

Methodology

Year two evaluation activities focused on continuing to collect and analyze data on student performance in the scholarship program. Of primary interest were 125 scholarship students who had attended a public school as a second grader during the 1995-96 academic year.⁴ This included the 94 scholarship students who attended existing private schools and 31 scholarship students who attended the two newly created private schools. Similar attempts were made to collect data on each of 449 public school students who had continued to attend public school from second through fourth-grade and who were included in the first-year evaluation. Background and demographic data were collected on each of these students in the preceding year. In addition, and in order to address questions related to the learning

³ For the sake of brevity, the current report does not provide extensive information about the Cleveland Scholarship and Tutoring Grant Program itself or about the first-year evaluation. Those interested in learning more about either of these topics are encouraged to refer directly to the earlier report (Metcalf, Boone, Stage, Chilton, Muller, and Tait, 1997).

⁴ Examination of records from the Cleveland Scholarship and Tutoring Grant Office also revealed 41 fourth-grade scholarship students who entered the program in 1997-98. These students are not included in the primary analyses described in the body of this report. However, a summary of their demographic and background characteristics as well as their second-grade achievement levels is included in Appendix A.

environments in the scholarship and public schools during 1997-98, data were collected on class size, teacher's education level, and teacher's years of experience for each student.

Subjects and Subject Selection

As noted earlier, background demographic and pre-program achievement data were available from the first-year evaluation on 125 scholarship students. These included 31 students who had attended the two newly created private schools and 94 students attending existing private schools. Similarly, background demographic and achievement data were available on 449 public school students from the previous year. Using records maintained by the Cleveland Scholarship and Tutoring Grant Office and the Cleveland Public Schools, evaluation staff attempted to identify the school of fourth-grade enrollment for each of these students. Of the 125 scholarship students included in the study as third-graders, only 94 (66 attending established private schools, 28 attending newly created private schools) could be located in the 1997-98 scholarship records for fourth-grade. Examination of public and private school records revealed that three of the scholarship students had been retained in third-grade, but did not identify the current school of enrollment for any of the remaining 28 students. Similarly, public school records allowed identification of current school of enrollment for 343 of the original 449 public school students included in the study. As a result, the sample for the second-year scholarship evaluation included the following fourth-graders:

- 66 returning scholarship⁵ students attending existing private schools;
- 28 returning scholarship students attending the two newly established private schools; and
- 343 students who continued to attend public school.

Data Collection

As in year one, the quasi-experimental methodology employed in the evaluation was intended to examine the impact of the scholarship program on students' academic achievement after previous

⁵ In the remainder of the report, the terms *existing private schools*, *new private schools*, and *public schools* are used to describe private schools which were in existence prior to initiation of the scholarship program, private schools which were established as a result of demand created by the scholarship program, and Cleveland Public Schools, respectively.

achievement levels and non-program factors were taken into consideration. As noted above, background demographic and achievement data were available for 94 returning scholarship students and 343 returning public school students. These data included: family income (eligibility for the free lunch program); family living arrangement (whether or not the student lives with only the mother); student's race; student's gender; and student's second grade achievement test scores (vocabulary and comprehension). Student's third-grade achievement test scores (reading, language, mathematics, social studies, and science) were available from testing conducted in 1996-97 for the scholarship students attending existing private schools and the returning public school students in the sample.

Student's current academic achievement (fourth-grade) was measured using the *Terra Nova* Survey, Level 14, Form A (CTB-McGraw-Hill). Proctors who had been trained and were supervised by the evaluation team administered the *Terra Nova* in single, 3-hour sessions at each of the 35 public and 34 private schools where target students were enrolled. It is important to note that the *Terra Nova* was administered to all target students, including those attending the two newly established private schools. Proctors followed a prescribed, standardized process to ensure that testing conditions were consistent across test sites. Further, throughout each administration of the test, proctors were to maintain detailed notes regarding any aspect of the testing or test conditions that might have impacted students' performance (e.g., fire drill, ill student, etc.). Before leaving their assigned schools, proctors interviewed teachers whose students were being tested regarding class size, highest degree held, and years of teaching experience. Upon completion of each testing session, proctors prepared the test materials for scoring (e.g., erasing stray marks, ensuring complete student information, alphabetizing, etc.) and submitted the materials to evaluation staff. Tests were subsequently shipped to CTB/McGraw-Hill for scoring.

The *Terra Nova* Survey measures student learning in each of five areas. The test includes 35 reading items, 20 language items, 32 mathematics items, 25 science items, and 25 social studies items. Scores are provided in a variety of forms in each of the five areas as well as a total battery score reflecting student performance across the reading, language, and mathematics portions of the test.

Thus, the following student data were available for analyses:

- *Family income.* Student eligibility for the free lunch program during the 1995-96 school year was used as an indicator of family income. Although free lunch program eligibility is a somewhat crude proxy for socio-economic status, it does at least provide broad categories of family income. Coded as 0 = full or reduced-price lunch, 1 = free lunch program.
- *Family living arrangement.* Coded as 0 = living with person(s) other than mother only, 1 = lives with mother only.
- *Race.* Coded as non-white = 0, white = 1.
- *Gender.* Coded as male = 0, female = 1.
- *Pre-program achievement.* Second-grade California Achievement Test Form E reading vocabulary scores and second-grade California Achievement Test Form E reading comprehension scores were available for both scholarship students and public school students from the 1995-96 school year when both groups were still enrolled in the Cleveland Public School system. Both second-grade vocabulary scores and comprehension scores are reported as Normal Curve Equivalents (NCEs).
- *Third-grade achievement test scores.* Third-grade achievement test scores reported as Normal Curve Equivalents (NCEs) in reading, language, mathematics, total battery, social studies, and science (not available for scholarship students attending the two newly established private schools).
- *Fourth-grade achievement test scores.* Fourth-grade achievement test scores reported as Normal Curve Equivalents (NCEs) in reading, language, mathematics, total battery, social studies, and science.
- *Class size.* Number of children in student's fourth-grade classroom.
- *Teacher's highest degree held.* Coded as 0 = certificate/no degree, 1 = bachelors degree, 2 = bachelors degree plus additional academic coursework, 3 = masters degree, 4 = masters

degree plus additional academic coursework, 5 = education specialist, 6 = doctoral candidate, 7 = doctorate.

- *Years of teaching experience.* Number of years of teaching experience.

The second-year evaluation builds and improves upon the previous year in several ways. Unlike year one, students attending the newly established private schools were tested using the same achievement test and under the same conditions as all other targeted students. It may be remembered that differences in the test and testing conditions during year one precluded comparison of these students' performance with that of other scholarship or public school students. Test data for students from these new private schools allow for examination of a key question in the private school marketplace: As entrepreneurial enterprises, do the new private schools provide similar or better educational service when compared with existing public and/or private schools? This question is particularly important in light of the limited supply of existing private schools and the increased demand created by voucher programs.

Further, collection of data related to class size, teacher education, and teacher experience in year two allow examination and control of variables suggested by some to impact student achievement (e.g., Molnar, 1997). Perhaps most importantly, data collected during year two reflect scholarship student performance after participation in the program for a longer period of time.

Data Analysis

Data were analyzed using multivariate and univariate techniques focused on each of the evaluation questions listed above and on relevant or emerging subquestions. When appropriate, multiple comparison procedures were employed to gain a more thorough understanding of differences between or among the scholarship and public school students. All analyses of student achievement were conducted and reported as Normal Curve Equivalent (NCE) scores for ease of interpretation and discussion.

In the following section, analyses, results, and implications associated with each research question are presented and discussed.

Findings

Evaluation Question 1. Are there differences between students who returned to the scholarship program during their fourth-grade year and those who did not return after their third-grade year?

Context

Thirty-one of the third-grade scholarship students who were included in the first-year evaluation had not continued in the program as fourth-graders. The reasons for leaving the scholarship program after one year remain unclear. However, if students who discontinue in the program differ in notable ways from those who remain, potential causes may be indicated.

This evaluation question was examined using multivariate analysis of variance (MANOVA) and analyses of variance (ANOVAs) to compare students who remained in the scholarship program during their fourth-grade year with students who did not return after their third-grade year. Specifically, students were compared on all pre-program background and demographic characteristics and third-grade achievement data. Means and standard deviations for students who continued in the scholarship program and those who discontinued after their third-grade year are presented in Table 1.

Results

The MANOVA used to compare the two groups on the pre-program background and demographic data and third-grade achievement measures did not produce a significant result, $F(12,81) = .85, p \geq .05$. However, F values obtained from univariate analyses of variance revealed significant differences ($p \leq .05$) between the two groups on third-grade reading, science, and social studies achievement.⁶ Students who left the scholarship program at the end of third-grade scored lower on

⁶ MANOVA is less powerful than ANOVA. However, the use of multiple univariate alphas in a series of ANOVAs inflates Type I error. In cases of conflicting significance tests for multivariate F (non-significant multivariate F, but a significant univariate F), univariate analyses are traditionally offered as a guide for future research (Tabachnick & Fidell, 1996).

measures of reading achievement ($M = 37.19$, $SD = 20.80$), science achievement ($M = 30.73$, $SD = 19.14$) and social studies achievement ($M = 32.19$, $SD = 17.62$) than those students who remained in the scholarship program during the fourth-grade ($M_{\text{reading}} = 45.21$, $SD = 17.26$; $M_{\text{science}} = 41.49$, $SD = 19.13$; $M_{\text{social studies}} = 41.99$, $SD = 17.51$). No statistically significant differences ($p \leq .05$) were found between students who continued in the scholarship program and students who discontinued their participation with respect to race, gender, family income, family living arrangements, or second-grade achievement measures.

Discussion

This finding indicates that students who do not benefit as greatly from the scholarship program are more likely to discontinue than are students doing relatively better. This seems unrelated to the demographic characteristics or achievement of students before entering the scholarship program. In these areas, discontinuing students were very much like their continuing scholarship peers. It is particularly striking that as second-graders the discontinuing students were achieving at similar levels to those who continued in the scholarship program, but were achieving at significantly lower levels at the end of their first year in the program. The evidence strongly suggests that the impact of the scholarship program on students' academic achievement is a major factor in whether or not a given student will continue in the program.

What remains unclear is why this might be the case. It may be that parents choose to return their children to public schools because the scholarship program is not producing the academic benefits they expected. Given that improved academic quality is the primary reason parents choose to participate in voucher programs (see Greene, Howell, and Peterson, 1997; Witte, Thorn, Pritchard, and Claibourn, 1994), limited academic gains might well prompt parents to return to the public schools. In this case, leaving the program would be based on deliberate choices to seek improved academic quality. However, other explanations of this finding are similarly reasonable. For example, it is possible that students who are not performing well in the program are implicitly or explicitly encouraged to withdraw or discontinue. School and program representatives may, even with the best of intentions, convey to children and parents

that they would be better served in another educational environment. Or, life events or situations (e.g., family conflict, economic changes, changes in parental custody, etc.) that occur during the student's time in the program may be responsible for diminishing academic performance and, independently, preclude his or her continued participation in the program. In any event, the striking pattern found in this result warrants continued examination and will be included in future evaluation activities.

Evaluation question 2. Are there differences between fourth-grade scholarship students and public school students with regard to demographic and background characteristics or pre-program achievement?

Context

This question concerns whether scholarship students are different in terms of race, gender, family income, family living arrangements, and prior achievement than students who chose to remain in Cleveland public schools. The first-year evaluation addressed the question, "Is the voucher program attracting students who are different from students who remain in public schools?" The results indicated that scholarship students had been performing at slightly higher levels of academic achievement than their public school peers on second-grade measures of reading comprehension and vocabulary before they entered the scholarship program, but were similar in other respects. Given attrition rates from the scholarship program discussed above, this issue was re-examined to determine whether the fourth-grade scholarship students differ from Cleveland public school students with regard to pre-program demographic and background characteristics. Whereas last year's evaluation examined whether the scholarship program was attracting students who were different than those students who chose to remain in public schools, the current evaluation examines whether the students who returned to the scholarship program as fourth-graders are different with regard to demographic and background data from those students who remained in public schools as fourth-graders.

Results

This evaluation question was examined through analyses of variance (ANOVAs) that allowed comparison of scholarship students and public school students on the available second-grade demographic and background data. Means and standard deviations for fourth-grade scholarship and public school students on the two second-grade measures of previous achievement and demographic data are presented in Table 2. The F values obtained from univariate analyses of variance revealed no statistically significant differences ($p \leq .05$) between scholarship and public school students on any pre-program or demographic characteristics. Both fourth-grade students who have participated in the scholarship program for two years and the public school students are primarily non-white (84-85% non-white), live only with their mother (62-70%), are eligible for the free lunch program (85-87%) and slightly more are females than males (52%).

Unlike last year's findings, there were no statistically significant differences ($p \leq .05$) in pre-program (second-grade) achievement between scholarship and public school students. Students in both groups were achieving at roughly the national mean before the scholarship program began. That no pre-program achievement differences are found for students included in the second-year evaluation may be a result of the smaller sample sizes and subsequently decreased statistical power. Or, perhaps relatedly, attrition of some students from either the public or the scholarship schools may have led to greater similarity between the groups.

Discussion

It would appear that students who remain in the scholarship program are similar to their public school peers on the available demographic variables and in their level of achievement before the scholarship program began. As such, the scholarship program does not seem to be drawing away from public schools students who are more advantaged or academically able. In fact, most students who participate in the program, as well as most students in the comparison sample, possess characteristics that many would assume to place them "at-risk" educationally. In light of this, it is somewhat surprising that

both groups of students were achieving at approximately the national mean as second-graders. Nonetheless, there are no consistent differences on any of the demographic or pre-program variables between the scholarship students and the public school students who continued in their respective programs during this second year.

This finding is important for several reasons. First, the scholarship program appears to serve the population for which it was intended. Scholarship families tend to be low-income, single parent, and non-white. Second, the scholarship program, at least in its second year, does not seem to have drawn from public schools a disproportionate number of advantaged or high achieving students. Students in the scholarship program are very much like their former public school classmates. Third, the similarity of the public school students with the scholarship students on these important demographic and pre-program achievement measures suggests that they represent a reasonable comparison group on which to base analyses of programmatic effects.

Evaluation Question 3. Are there differences in available classroom variables (e.g., class size, teachers' education level, teachers' experience) between scholarship classes and public school classes?

Context

Evaluation question 3 expands the first-year evaluation by examining potential differences between scholarship classes and public school classes. Consideration of these factors was necessary for two reasons. First, some have speculated that any programmatic impact that voucher programs may be found to have might well be the result of resources available to the public or private schools, not directly to educational choice (e.g., Molnar, 1997). While research on the effects of class size, teacher experience, and teacher education level on student achievement are not consistent, these variables may enhance or diminish scholarship and/or public school student performance. Second, comparisons of public and private schools have often, though not always, found differences on these three variables. Private school

classes are often believed to be smaller than public school classes, public school teachers tend to be more experienced than private school teachers, and public school teachers often have a higher education level than private school teachers (e.g., Corwin, 1993). Thus, just as background demographic and pre-program achievement must be controlled for in the present evaluation, it was believed important to consider these classroom-relevant variables as well.

Results

Analyses of variance were conducted to examine differences in available classroom variables between scholarship classes and public school classes. For these analyses, the classroom rather than the individual student was used as the unit of analysis. As noted in Table 3, statistically significant differences ($p \leq .05$) were found between scholarship classrooms and public school classroom on all three available classroom variables: class size, teachers' education level, and teachers' years of experience. Public school teachers had significantly more years of experience in the classroom ($M = 14.19$, $SD = 10.00$) than teachers in scholarship classrooms ($M = 8.55$, $SD = 9.31$). Public school teachers also had significantly higher educational levels ($M = 2.09$, $SD = 1.03$) than teachers in the scholarship classrooms ($M = 1.52$, $SD = 1.13$). While on average both scholarship and public school teachers have a bachelors degree plus some additional coursework (0= certificate/no formal degree, 1= bachelors, 2 = bachelors plus some additional coursework, 3 = masters), more public school teachers have engaged in at least some work beyond the baccalaureate level. Statistically significant differences were also noted between public school classroom sizes and scholarship classroom sizes. Public school students who participated in the study were in slightly larger classrooms ($M = 23.60$, $SD = 3.19$) than fourth-grade scholarship students ($M = 20.62$, $SD = 4.80$).

Discussion

These differences indicate that the classroom experience of scholarship students may be different from that of public school students. Students who attend public schools work with teachers who have considerably more teaching experience than teachers in the scholarship classrooms. Both public and private school teachers tended to have completed coursework beyond their undergraduate degree, but

public school teachers had taken somewhat more graduate coursework than their private school colleagues. Of course, this may be related to public school teachers' longer tenure in the classroom and, thus, greater opportunity for taking coursework. On the other hand, private school students may receive slightly more individual attention from their teachers than public school students, because their classes have an average of three fewer students. Across the three classroom-relevant variables, and to the extent these variables impact student achievement, neither public school nor private school students appear to be significantly advantaged.

Evaluation Question 4. What are the effects of participation in the scholarship program on students' fourth-grade achievement when background demographic and previous achievement variables are controlled?

Context

This issue is fundamental to the second-year evaluation in that it addresses the impact of the scholarship program on student learning. This evaluation question concerns whether fourth-grade test performance of scholarship students and public school students is significantly different after controlling for non-program variables that might also impact achievement. Controlling for these non-program variables is particularly important in quasi-experimental studies, like the present, because the groups being compared may not be completely equivalent. For example, if students in the scholarship group were found to achieve at higher levels than students in the public school group after one year in the program, but had already been achieving at higher levels before they entered the scholarship program, it would be difficult to determine whether the program had had any effect on their performance. Thus, it is critical that these non-program factors be taken into consideration in all analyses. In the first-year evaluation, the non-program variables included background demographic characteristics (gender, race, family income, family living arrangements) and pre-program achievement test scores (second-grade

comprehension and vocabulary scores from the California Achievement Test). When these factors were controlled, no significant program effect was found.

In the present study, these same factors were again included in analyses of student achievement data. Further, because it was possible to obtain additional data on classroom-relevant data which might also impact students' learning (class size, teacher's education level, teacher experience), an additional set of analyses was conducted in which all of these factors were accounted for. Thus, the results presented in answer to this research question are presented in two ways. To allow comparison of the current findings with those from the 1996-97 analysis, a first set of results is presented to answer the question, what are the effects of the scholarship program on students' academic achievement after two years in the program and when background demographic variables and pre-program achievement are controlled? These analyses include the same set of control variables (covariates) as the first-year analyses (i.e., gender, race, family income, family living arrangements; second-grade comprehension score, second-grade vocabulary score), but do not include the classroom-relevant variables. A second set of analyses includes all of the demographic and pre-program achievement variables as well as the three classroom-relevant variables. These analyses answer a similar question, but examine the effect of the scholarship program on students' achievement after a greater range of non-program factors are controlled. In each set of analyses, student achievement was examined in five areas (language, reading, mathematics, science, and social studies) and on total battery score (integrating language, reading, and mathematics).

Before analyzing fourth-grade student achievement, the data were examined to ensure that each group (scholarship and public schools) could reasonably be assumed to consist of similar schools and students. In other words, in order for subsequent comparisons *between* scholarship and public school students to make sense, student performance *within* each of these groups should be similar. If systematic patterns are found to differentiate particular schools or students within one or the other of the groups, these represent a unique subset of the larger group. Such subsets often appear as outliers, or groups of data that are substantially higher or lower than all others within the group being studied. When such situations arise, analyses are most accurate when they treat the outliers as a separate group for

investigation (see Pedhazur & Schmelkin, 1991; Stevens, 1996). While a variety of statistical techniques exist for identifying outliers within a data set, merely plotting the data on a graph is often sufficient to allow their discovery.

In the present data set, a subset of achievement test scores (i.e., outliers) was discovered within the scholarship schools. Further examination of the data revealed that student academic achievement in the newly established private schools appeared to be systematically lower than for any other private schools. This pattern was substantial and consistent across all six fourth-grade achievement measures. Figures 1, 2, and 3 below are presented to illustrate the patterns found in social studies scores, however, similar patterns are found across the six scores.

Figure 1 depicts the mean total battery score of public school students in each public school in which at least three targeted students are enrolled.⁷ The plots reveal that the public school means vary from 29.11 to 52.51 NCE points, but also that this pattern does not seem to indicate any schools or groups of particular schools that appear to be systematically different from the others. Figure 2 depicts the mean social studies score of scholarship students in each private school in which at least three scholarship students are enrolled. As can be seen, most of the schools fall in the range of 33.71 to 48.94 and, as was seen for the public schools, variability among these schools seems to be random rather than systematic. However, one plot on the graph is notably lower than the others and does not fall within the random variability of the other scholarship schools. Such a pattern suggests that this unusual plot, the outlier, is best considered as distinct from other plots, even though all represent student performance in scholarship schools.

Figure 3 depicts mean total battery scores for both public and scholarship schools. The figure demonstrates two important characteristics of the data set. First, the plots between scholarship schools (red diamonds) and public schools (green circles) overlap. All public schools do not perform better or worse than all scholarship schools, and vice versa. Second, and with the exception of the newly established private school plot, a general pattern can be seen in which the scholarship schools are grouped

near the upper end of the graph. Thus, while it is not possible from this figure to make comparisons between scholarship and public school students' academic achievement, the data suggest that further analyses may reveal at least some systematic differences

These and similar graphs, along with additional analyses of the data, indicated that it was necessary to consider scholarship students attending the newly established private schools as a distinct group. Thus, all analyses of the effects of the scholarship program were conducted comparing students across three groups: scholarship students attending *existing private schools*, scholarship students attending *the newly established private schools*, and students attending *Cleveland Public Schools*. It should also be noted that analyses of the background demographic and pre-program data were re-analyzed to ensure that the three groups of students did not differ on these non-program variables. These analyses revealed no significant differences between or among the groups. The results of these analyses are presented in Appendix B.

Results

As noted earlier, two sets of analyses were conducted to examine the effect of the scholarship program on students' academic achievement after pre-program achievement and other relevant variables were controlled. Each set of analyses is presented below.

Set One Analyses

The first set of analyses address Evaluation Question 4 and parallel those conducted in year one, in that they compare the fourth-grade achievement of students after controlling for background demographic and pre-program achievement scores. Thus, in the present evaluation, these analyses attempt to answer the question, Are there differences in fourth-grade student achievement between scholarship students attending *existing private schools*, scholarship students attending *the newly established private schools*, or students attending *public schools*?

Multivariate analysis of covariance (MANCOVA) was conducted to determine if scholarship status for either existing or new private schools was significantly related to students' fourth-grade

⁷ School means based on fewer than three students are highly variable and difficult to interpret.

achievement after pre-program achievement, gender and race of student, family income, and family living arrangement were controlled (Standardized regression coefficients for these covariates are presented in Appendix C, Table C.1). Fourth-grade achievement scores considered in combination were significantly related ($p \leq .001$) to scholarship status after adjusting for covariates, (Wilks' Lambda = .862, $F(12, 798) = 5.12$). Table 4 reports the results of univariate F tests adjusted for covariates, as well as adjusted means, for each of the fourth-grade test scores. Table 5 presents statistically significant pairwise comparisons among means of the three groups.

As shown in Tables 4 and 5, scholarship students attending existing private schools score significantly higher ($p \leq .05$) than public school students on fourth-grade language and science achievement. In language, existing private school scholarship students' mean score ($M = 44.96$, Std. error = 1.68) was slightly more than five NCE points higher than for public school students ($M = 39.82$, Std. error = .77). Similarly, existing private school scholarship students' mean score in science ($M = 39.66$, Std. error = 1.80) was four NCE points higher than public school students' ($M = 35.66$, Std. error = .82). These differences, while significant, are relatively small, reflecting standardized effect sizes of .25 in language and .19 in science. However, there were no statistically significant differences ($p \leq .05$) between existing private school scholarship students and public school students in fourth-grade reading, mathematics, total battery or social studies achievement.

Additional comparisons reveal that scholarship students attending the newly established private schools score significantly lower ($p \leq .05$) on each measure of fourth-grade achievement than both other scholarship students and public school students. As reflected in Tables 4 and 5, these findings are consistent across all subject areas. For example, newly established private school scholarship students' adjusted fourth-grade social studies achievement scores ($M = 19.39$, Std. error = 3.13) are significantly and substantially lower than both other scholarship students ($M = 38.71$, Std. error = 1.82) and public school students ($M = 37.69$, Std. error = .83). Mean differences across the six scores between new private school scholarship students and public school students was 12.57 NCE points (standardized effect size = .61), ranging from a difference of 9.69 points in language to 18.30 points in social studies. Mean

differences between the new and existing private school scholarship students was 14.95 NCE points (standardized effect size = .73), ranging from a difference of 10.44 points in mathematics to 19.32 points in social studies.

Set Two Analyses

The second set of analyses directed at examining the effects of the scholarship program on students' achievement control for a broader set of variables. In these analyses, classroom-relevant variables, such as class size, teacher education level, and teacher experience, are included and controlled in addition to the background demographic and pre-program achievement variables (Standardized regression coefficients for these covariates are presented in Appendix C, Table C.2). Thus, these analyses address the question, Are there differences in fourth-grade student achievement between scholarship students attending *existing private schools*, scholarship students attending the *newly established private schools*, or students attending *public schools* after background demographic, pre-program achievement, and classroom-relevant variables are controlled?

Multivariate analysis of covariance was conducted to determine if a main effect of scholarship status existed after classroom-relevant variables, pre-program achievement, gender and race of student, family income, and family living arrangement were controlled. The fourth-grade achievement scores considered together were significantly related ($p \leq .001$) to voucher status after adjusting for all background demographic, pre-program, and classroom variables (Wilks' Lambda = .88, $F(12, 786) = 4.18$).

Table 6 reports results of univariate F tests and means adjusted for all covariates for each of the fourth-grade test scores and Table 7 presents pairwise mean differences that are significant between or among the three groups. As indicated in Tables 6 and 7, existing private school scholarship students score significantly higher ($p \leq .05$) than public school students on fourth-grade language achievement ($M = 45.29$, Std. error = 1.85; $M = 39.67$, Std. error = .80, respectively). Again, while this difference is statistically significant, it is relatively small, representing a standardized effect size of .27. There are no

statistically significant differences between existing private school scholarship students and public school students in reading, mathematics, total battery, science, and social studies achievement.

As in the earlier set of analyses, pairwise comparisons indicate that new private school scholarship students score significantly lower ($p \leq .05$) on all measures of fourth-grade achievement than both other scholarship students and public school students. As shown in Table 6 and 7, these results are consistent across all subject areas. Differences between new scholarship school and public school students ranged from 7.91 NCE points in language (standardized effect size = .38) to 18.27 points in social studies (standardized effect size = .87). Mean differences across the six scores between new private school scholarship students and public school students was 12.40 NCE points (standardized effect size = .60). Mean differences between the new and existing private school scholarship students was 13.83 NCE points (standardized effect size = .67), ranging from a difference of 9.99 points in mathematics to 17.89 points in social studies.

Discussion

It should be reiterated, before discussing these results, that students in both the scholarship and public school groups are very much alike in terms of gender, race, family income, family living arrangements, and second-grade (pre-program) achievement. However, significant differences between the groups exist in terms of class size (scholarship classes were smaller than public school classes), teacher experience (public school teachers were more experienced), and teacher education level (public school teachers had somewhat more education beyond the undergraduate level). These factors represent potential influences on student achievement gains and, as a result, must be taken into account before comparing fourth-grade student achievement between the groups. The influence of these variables is reflected in the slightly different results produced by each of the two sets of analyses described above. Nonetheless, the results of these analyses present an interesting and relatively consistent pattern of student performance.

After non-program variables are taken into account, scholarship students who attend a private school that had been in existence prior to the scholarship program do as well as their public school peers

in most subjects, and do significantly better at least in language and possibly in science. Thus, after two years in the scholarship program, the effects of participation for these students seem to be positive, though the magnitude of the effect is small. Because the effects of schooling are cumulative and generally long-term, this small but significant effect may represent a trend toward gradually improving academic achievement for the scholarship students. If so, the small differences found in language and science will be found to grow and positive differences in other academic areas will emerge over time. However, it is equally possible that the small difference in language achievement favoring existing private school scholarship students is the result of factors unique to the 1997-98 academic year and will not be found in future years. In research of the Milwaukee voucher program, no effect was found in years one or two, but significant, though mixed and inconsistent effects appeared in years three and four (see Greene, Peterson, and Du., 1996; Rouse, 1996; Witte, Thorn, Pritchard and Claibourn, 1994). Thus, only collection of additional student achievement data in future years will allow more definitive conclusions to be drawn about the impact of the scholarship program.

The performance of scholarship students attending newly established rather than existing private schools presents a much more consistent, though surprising result. The impact of attending these existing private schools appears to be negative, at least at the end of the students' second year in the school. While not differing significantly from other scholarship and public school students on any of the pre-program or non-program variables, these students' fourth-grade performance is significantly and dramatically lower than both public school and other scholarship students'. Unlike the relatively small positive effect in language and science found for scholarship students attending existing private schools, the mean effect of the program for these students is $-.60$ across the six achievement scores. The cause or causes of these results is unclear and will only be explained through the collection of additional data. It may be that the instructional or curricular approaches in these schools are less effective than in other schools and lead to reduced student performance. However, because these schools were established specifically to serve the increased market for private schooling created by the scholarship program, an additional set of explanations are also reasonable.

For example, the proportion of scholarship to non-scholarship students is much greater in the newly established schools than in other private schools. Scholarship students who moved from public schools into existing private schools could rely on a substantial group of classmates to help them adjust to the unique expectations and norms of their new school. In contrast, scholarship students in the newly established schools had no such peer support because virtually all of them were new to the schools. Similarly, the faculty and staff in the newly established schools were completely new to their professional surroundings. There was no existing group of colleagues who could provide support or guidance in learning the culture of the school. In fact, the expectations, and norms of the school were being established by the students, teachers, and staff themselves as they sought to create formal and informal approaches to their work. To the extent these factors influenced students' academic performance in these first two years, the schools are likely to become more effective over time. However, and as noted above, additional data collected over the next several years are critical to understanding this finding.

Summary

The results of this evaluation provide an indication of the impact of the Cleveland Scholarship Program after its first two years of operation. As such, they provide somewhat more thorough answers to questions about the effect of the program on students' learning, the types of students and families who are served by the program, and the classrooms in which these students are enrolled. Of course, as with any evaluation, the findings also raise a number of questions as well. For example, questions remain about the variety of instructional approaches used in the schools, parents' involvement in their children's education, impacts on students that cannot be measured by standardized achievement tests, and the effect of the program on the schools that accept voucher students. These questions are important and will be addressed in future years of the evaluation. Nonetheless, the current evaluation reveals several interesting early findings about the scholarship program in its second year and adds to the information gained in year one.

Students who did not return to the scholarship program from third to fourth-grade had been achieving at lower levels than those who remained in the program. While there are no significant

differences in background variables between these two groups, non-returning students' third-grade achievement was significantly lower than that of returning students in reading, science, and social studies. Intuitively, it seems reasonable that students who were not benefiting as much from the program would be more likely to leave. However, there are no data yet available to support this and alternative explanations are likely.

Scholarship students who continued in the scholarship program from third to fourth-grade continue to be very much like their public school counterparts. There are no significant differences between the returning scholarship and public school students in any of the background demographic or second-grade achievement variables. Both groups of students had been achieving at or near the national mean in comprehension and vocabulary as second graders, about 85% were non-white, 85-87% had family incomes that qualified them for the free lunch program, about half (52%) were female, and between 62% and 70% lived only with their mother. Thus, the scholarship program appears to serve those students for whom it is intended: minority children, living in single parent households, of low-income.

The limited classroom variables examined in the present study differed significantly between scholarship and public schools. Scholarship students were in smaller classes than public school students. However, public school teachers were more experienced and more likely to have completed graduate coursework than teachers of scholarship students. These findings support earlier research of these issues (e.g., Golhaber, 1996; Lankford & Wyckoff, 1992). However, our analyses suggest that these three classroom variables alone and in combination do not completely explain differences in students' academic achievement.

The effects of the program on scholarship students' academic performance are slightly positive, but are mediated by the schools they attend. When previous achievement and demographic variables are controlled in the analyses, scholarship students attending existing private schools demonstrate significantly greater achievement in language and science than their public school peers. When classroom variables are also controlled in the analyses, existing private school scholarship students'

performance in language remains significantly greater than for public school students. However, the existing private school scholarship students do not differ significantly from their public school peers on any of the other achievement measures.

In contrast, in all analyses in which background, previous achievement, and/or classroom variables are controlled, scholarship students attending the newly established private schools consistently and significantly achieve at lower levels on each of the six achievement measures than either the other scholarship or public school students.

The results suggest that the scholarship program may be successful in promoting at least slightly greater achievement in the private schools that were in existence prior to the scholarship program. In year two, these schools helped students achieve at the same or slightly higher levels than they are likely to have achieved had they continued to attend public schools. In contrast, the newly established schools did not seem to promote greater student achievement and may, in fact, have diminished students' academic achievement.

In all, the second year findings add to information collected during the first-year evaluation and, at the same time, raise additional questions. Considerable additional work remains to be done before definitive conclusions about the effectiveness of the scholarship program can be drawn. Future evaluation efforts will be focused not only on continuing to examine the effects of the program on students' learning, but also the family, school, and student characteristics that may explain them. During the 1998-99 year, the evaluation will collect data on first-grade students in both scholarship and public schools and will continue to collect these data over multiple years. This project includes examination of academic achievement, demographic, and classroom variables used in the evaluation to date, and adds collection of additional data from families, teachers, administrators, and students.

Undoubtedly, voucher programs that allow families to use public funds to enroll their children in private schools present a unique alternative to traditional ways of thinking about public education. The constitutionality of publicly-funded voucher programs is yet to be established and they serve only a very small percentage of school-aged children. Still, publicly-funded voucher programs in Milwaukee and

Cleveland represent an important experiment in providing greater educational choice to low-income, urban families. The impact of such programs, not just for children, but for families and schools themselves remains to be determined. However, thorough, objective, on-going evaluation of these and other choice experiments, many in the public sector, have the potential to provide information that can help all schools better meet the needs of the students and families they serve.

References

- Chubb, J. and Moe, T. (1990). Politics, markets, and America's schools. Washington, DC: The Brookings Institution.
- Corwin, R. (1993). Private schools and parental choice: Dubious assumptions, frail claims, and excessive hyperbole. Los Alamitos, California: Southwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 345 612).
- Golhaber, D. (1996). Public and private high schools: Is school choice an answer to the productivity problem? Economics of Education Review, 15, 91-109.
- Green, J., Howell, W., and Peterson, P. (1997). Lessons from the Cleveland Scholarship Program. Paper presented before the Association of Public Policy and Management, Washington, DC.
- Greene, J., Peterson, P., and Du, J. (1996). The effectiveness of school choice in Milwaukee: A secondary analysis of data from the program's evaluation. Paper presented at the August-September meetings of the American Political Science Association, San Francisco.
- Lankford, H. and Wyckoff, J. (1992). Primary and secondary school choice among public and religious alternatives. Economics of Education Review, 11, 351-337.
- Metcalf, K., Boone, W., Stage, F. Chilton, T., Muller, P., and Tait, P. (1998). A comparative evaluation of the Cleveland Scholarship and Tutoring Grant Program. Bloomington, Indiana: Indiana University, The Indiana Center for Evaluation.
- Molnar, A. (1997, August 6). The real lesson of Milwaukee's voucher experiment: It's political, not educational. Education Week, pp.76, 56.
- Pehazur, E. and Schmelkin, L. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Rouse, C. (1996). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. Unpublished manuscript, Princeton, New Jersey: Princeton University.
- Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- VanDunk, E. (1998). Choice school accountability: A consensus of views in Ohio and Wisconsin.. Milwaukee, Wisconsin: The Public Policy Forum.
- Witte, J. (1992). Public subsidies for private schools: What we know and how to proceed. Education Policy, 6, 206-227.
- Witte, J., Thorn, C., Pritchard, K., Claibourn, M. (1994). Fourth-year report: Milwaukee Parental Choice Program. Madison, Wisconsin: University of Wisconsin, Department of Political Science and The Robert LaFollette Institute of Public Affairs.

Table 1. Comparison of Returning Scholarship and Non-Returning Scholarship Students

| | Returning Scholarship Students (N=68) | | Non-Returning Scholarship Students (N=26) | | Mean Square (MS error) | F df = (1,92) |
|---|--|-------|--|-------|-------------------------------|----------------------|
| | Mean | S.D. | Mean | S.D. | | |
| 2 nd -Grade Comprehension | 52.81 | 20.59 | 47.23 | 19.15 | 585.22 (408.45) | 1.43 |
| 2 nd -Grade Vocabulary | 59.50 | 23.88 | 56.92 | 26.19 | 124.90 (601.64) | 0.21 |
| Race | 0.19 | 0.40 | 0.12 | 0.33 | 0.11 (0.14) | 0.76 |
| Family Income | 0.87 | 0.34 | 0.92 | 0.27 | 0.06 (0.11) | 0.55 |
| Gender | 0.54 | 0.50 | 0.58 | 0.50 | 0.02 (0.25) | 0.08 |
| Living Arrangements | 0.71 | 0.46 | 0.65 | 0.49 | 0.05 (0.22) | 0.23 |
| 3rd-Grade Achievement | | | | | | |
| Reading | 45.21 | 17.26 | 37.19 | 15.28 | 1207.83 (280.43) | 4.31* |
| Language | 43.91 | 17.96 | 37.96 | 19.78 | 665.92 (341.20) | 1.95 |
| Mathematics | 40.85 | 16.47 | 33.38 | 17.77 | 1049.06 (283.46) | 3.70 |
| Total Battery | 42.62 | 16.75 | 35.08 | 17.40 | 1069.50 (286.54) | 3.73 |
| Science | 41.49 | 19.13 | 30.73 | 19.14 | 2175.39 (365.89) | 5.95* |
| Social Studies | 41.99 | 17.51 | 32.19 | 17.62 | 1803.79 (307.69) | 5.86* |

*p ≤ .05

Table 2. Means and Standard Deviations of Univariate F Tests of Background Demographic and Pre-Program Achievement Measures for Scholarship and Public School Students

| Covariate | Scholarship (N=93) | | Public (N=343) | | Mean Square (MS error) | F |
|--------------------------------------|-----------------------|-------|-------------------|-------|---------------------------|------|
| | Mean | SD | Mean | SD | | |
| 2 nd -Grade Comprehension | 52.58 | 21.67 | 48.01 | 20.32 | 1529.27 (425.00) | 3.60 |
| 2 nd -Grade Vocabulary | 57.30 | 22.48 | 52.87 | 23.31 | 1439.18 (535.28) | 2.70 |
| Race | 0.15 | 0.36 | 0.14 | 0.35 | .00 (.124) | .035 |
| Family Income | 0.85 | 0.36 | 0.87 | 0.34 | .03 (.12) | .23 |
| Gender | 0.52 | 0.50 | 0.52 | 0.50 | .00 (.25) | .003 |
| Living Arrangements | 0.70 | 0.46 | 0.62 | 0.49 | .44 (.23) | 1.92 |

* $p \leq .05$

Table 3. Means and Standard Deviations of Univariate F Tests of Classroom-Relevant Variables for Scholarship and Public School Students (by classroom)

| | Combined | | Public | | Mean Square (MS error) | F |
|--|-------------|------|--------|-------------------|---------------------------|----------------------------|
| | Scholarship | | School | | | |
| | Mean | SD | Mean | SD | | |
| Teacher Years of Experience df = (1, 139) | 8.55 | 9.31 | 14.19 | 10.00 | 939.48 (96.13) | 9.77** |
| Teacher Highest Degree Held df = (1, 136) | 1.52 | 1.13 | 2.09 | 1.03 | 9.47 (1.12) | 8.47** |
| Class Size df = (1, 140) | 20.62 | 4.80 | 23.60 | 262.83 (13.94) | 262.83 (13.94) | 18.85*** |
| | | | | | | ** p ≤ .01 *** p ≤ .001 |

Table 4. Fourth-Grade Scores for Existing Scholarship, Newly Established Scholarship, and Public School Students
Observed and Adjusted for Background Demographic and Pre-Program Achievement Variables

| | Existing Scholarship | | Newly Established Scholarship | | Public School | | Mean Square (MS error) | F df = (2, 404) |
|--|-----------------------|-------------------------------|-------------------------------|-------------------------------|-----------------------|-------------------------------|-------------------------------|---|
| | N=68 | | N=23 | | N=322 | | | |
| | Observed Mean (SD) | Adjusted Mean (Std. error) | Observed Mean (SD) | Adjusted Mean (Std. error) | Observed Mean (SD) | Adjusted Mean (Std. Error) | | |
| Reading | 46.46 (16.42) | 44.44 (1.59) | 29.48 (9.30) | 28.69 (2.73) | 40.83 (15.85) | 41.31 (0.72) | 2082.90 (167.46) | 12.44*** |
| Language | 46.94 (19.22) | 44.96 (1.68) | 30.22 (12.18) | 30.13 (2.90) | 39.39 (15.92) | 39.82 (0.77) | 1886.88 (187.71) | 10.05*** |
| Mathematics | 41.37 (16.70) | 38.85 (1.91) | 28.48 (15.64) | 28.41 (3.30) | 40.03 (19.05) | 40.57 (0.87) | 1575.48 (243.26) | 6.48** |
| Total Battery | 44.56 (16.76) | 42.22 (1.57) | 28.17 (10.92) | 27.81 (2.70) | 39.22 (16.27) | 39.74 (0.71) | 1769.15 (162.90) | 10.86*** |
| Science | 42.50 (19.53) | 39.66 (1.80) | 24.61 (12.07) | 24.70 (3.10) | 35.07 (18.88) | 35.66 (0.82) | 1855.80 (214.76) | 8.64*** |
| Social Studies | 41.06 (19.93) | 38.71 (1.82) | 19.65 (12.37) | 19.39 (3.13) | 37.18 (17.59) | 37.69 (0.83) | 3627.61 (219.96) | 16.49*** |
| | | | | | | | | * p ≤ .05 ** p ≤ .01 *** p ≤ .001 |
| Note: See next table for results of significant pairwise comparisons of group means. | | | | | | | | |

Table 5. Significant Mean Differences⁸ Between Existing Private Schools, Newly Established Private Schools, and Public Schools
Means Adjusted for Background Demographic and Pre-Program Achievement Variables

| Subject Measure | | Group Means Compared (Differences) | | |
|-----------------|------------------|------------------------------------|-----------------------------------|----------------|
| | | Existing Private Schools | Newly Established Private Schools | Public Schools |
| Reading | Existing Private | ----- | 15.75 | n.s. |
| | New Private | 15.75 | ----- | 12.62 |
| | Public | n.s. | 12.62 | ----- |
| | | | | |
| Language | Existing Private | ----- | 14.83 | 5.14 |
| | New Private | 14.83 | ----- | 9.69 |
| | Public | 5.14 | 9.69 | ----- |
| | | | | |
| Mathematics | Existing Private | ----- | 10.44 | n.s. |
| | New Private | 10.44 | ----- | 12.16 |
| | Public | n.s. | 12.16 | ----- |
| | | | | |
| Total Battery | Existing Private | ----- | 14.41 | n.s. |
| | New Private | 14.41 | ----- | 11.93 |
| | Public | n.s. | 11.93 | ----- |
| | | | | |
| Science | Existing Private | ----- | 15.59 | 4.00 |
| | New Private | 15.59 | ----- | 10.96 |
| | Public | 4.00 | 10.96 | ----- |
| | | | | |
| Social Studies | Existing Private | ----- | 19.32 | n.s. |
| | New Private | 19.32 | ----- | 18.30 |
| | Public | n.s. | 18.30 | ----- |
| | | | | |

⁸ Paired differences significant at $p \leq .05$.

Table 6. Fourth-Grade Scores for Existing Scholarship, Newly Established Scholarship, and Public School Students

Observed and Adjusted for Background Demographic, Pre-Program Achievement, and Classroom-Relevant Variables

| | Existing Scholarship | | Newly Established Scholarship | | Public School | | Mean Square (MS error) | F df = (2, 398) |
|---|-----------------------|-------------------------------|-------------------------------|-------------------------------|-----------------------|-------------------------------|---------------------------|---|
| | N=68 | | N=23 | | N=319 | | | |
| | Observed Mean (SD) | Adjusted Mean (Std. error) | Observed Mean (SD) | Adjusted Mean (Std. Error) | Observed Mean (SD) | Adjusted Mean (Std. error) | | |
| Reading | 46.46 (16.42) | 43.14 (1.75) | 29.48 (9.30) | 28.36 (3.00) | 40.83 (15.85) | 41.65 (0.76) | 1734.38 (166.54) | 10.41*** |
| Language | 46.94 (19.22) | 45.29 (1.85) | 30.22 (12.18) | 31.76 (3.18) | 39.39 (15.92) | 39.67 (0.80) | 1552.73 (187.00) | 8.30*** |
| Mathematics | 41.37 (16.70) | 38.38 (2.10) | 28.48 (15.64) | 28.39 (3.60) | 40.03 (19.05) | 40.34 (0.91) | 1290.21 (239.63) | 5.38** |
| Total Battery | 44.56 (16.76) | 41.52 (1.72) | 28.17 (10.92) | 28.09 (2.95) | 39.22 (16.27) | 39.96 (0.75) | 1415.76 (160.90) | 8.80*** |
| Science | 42.50 (19.53) | 38.30 (1.96) | 24.61 (12.07) | 24.97 (3.36) | 35.07 (18.88) | 36.07 (0.85) | 1360.54 (208.92) | 6.51** |
| Social Studies | 41.06 (19.93) | 37.63 (2.00) | 19.65 (12.37) | 19.74 (3.43) | 37.18 (17.59) | 38.01 (0.87) | 2888.46 (217.74) | 13.27*** |
| | | | | | | | | * p ≤ .05 ** p ≤ .01 *** p ≤ .001 |
| Note: See next table for significant pairwise comparisons of group means. | | | | | | | | |

Table 7. Significant Mean Differences⁹ Between Existing Private Schools, Newly Established Private Schools, and Public Schools
Means Adjusted for Background Demographic, Pre-Program Achievement, and Classroom-Relevant Variables

| Subject Measure | | Group Means Compared (Differences) | | |
|-----------------|------------------|------------------------------------|-----------------------------------|----------------|
| | | Existing Private Schools | Newly Established Private Schools | Public Schools |
| Reading | Existing Private | ----- | 14.78 | n.s. |
| | New Private | 14.78 | ----- | 13.29 |
| | Public | n.s. | 13.29 | ----- |
| | | | | |
| Language | Existing Private | ----- | 13.53 | 5.62 |
| | New Private | 13.53 | ----- | 7.91 |
| | Public | 5.62 | 7.91 | ----- |
| | | | | |
| Mathematics | Existing Private | ----- | 9.99 | n.s. |
| | New Private | 9.99 | ----- | 11.95 |
| | Public | n.s. | 11.95 | ----- |
| | | | | |
| Total Battery | Existing Private | ----- | 13.43 | n.s. |
| | New Private | 13.43 | ----- | 11.87 |
| | Public | n.s. | 11.87 | ----- |
| | | | | |
| Science | Existing Private | ----- | 13.33 | n.s. |
| | New Private | 13.33 | ----- | 11.10 |
| | Public | n.s. | 11.10 | ----- |
| | | | | |
| Social Studies | Existing Private | ----- | 17.89 | n.s. |
| | New Private | 17.89 | ----- | 18.27 |
| | Public | n.s. | 18.27 | ----- |
| | | | | |

⁹ Paired differences significant at $p \leq .05$.

Figure 1. Social Studies Aggregated Public School Scores

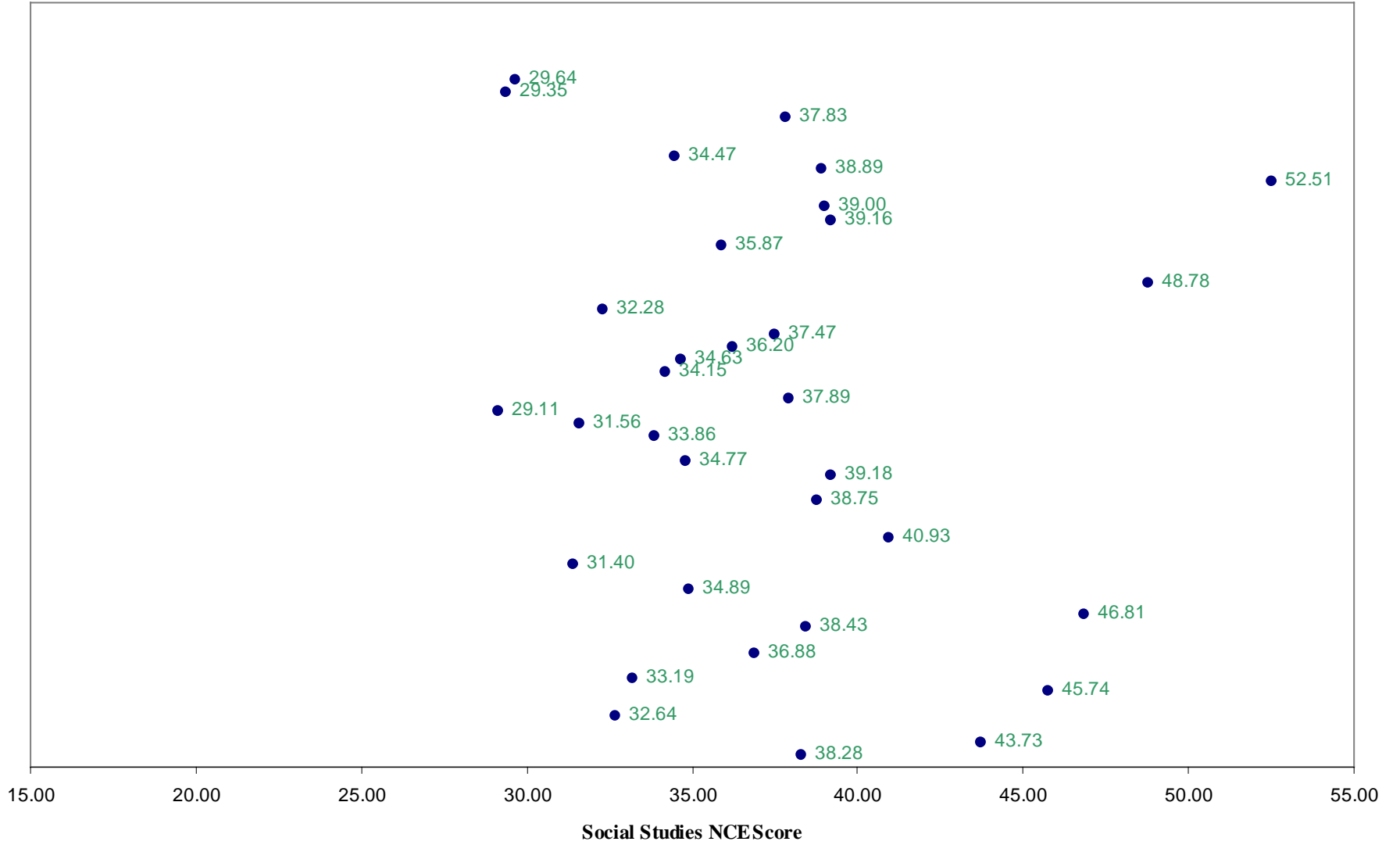


Figure 2. Social Studies Aggregated Scholarship School Scores

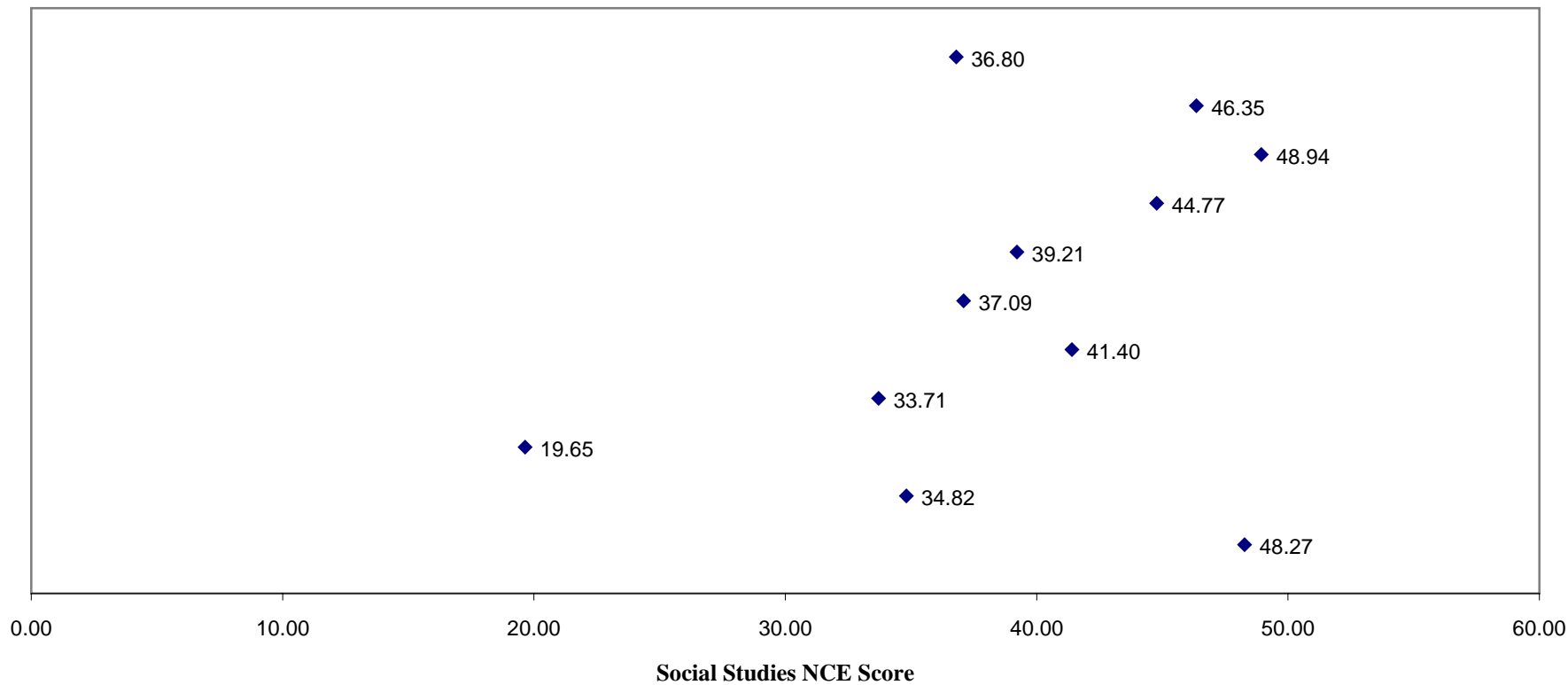
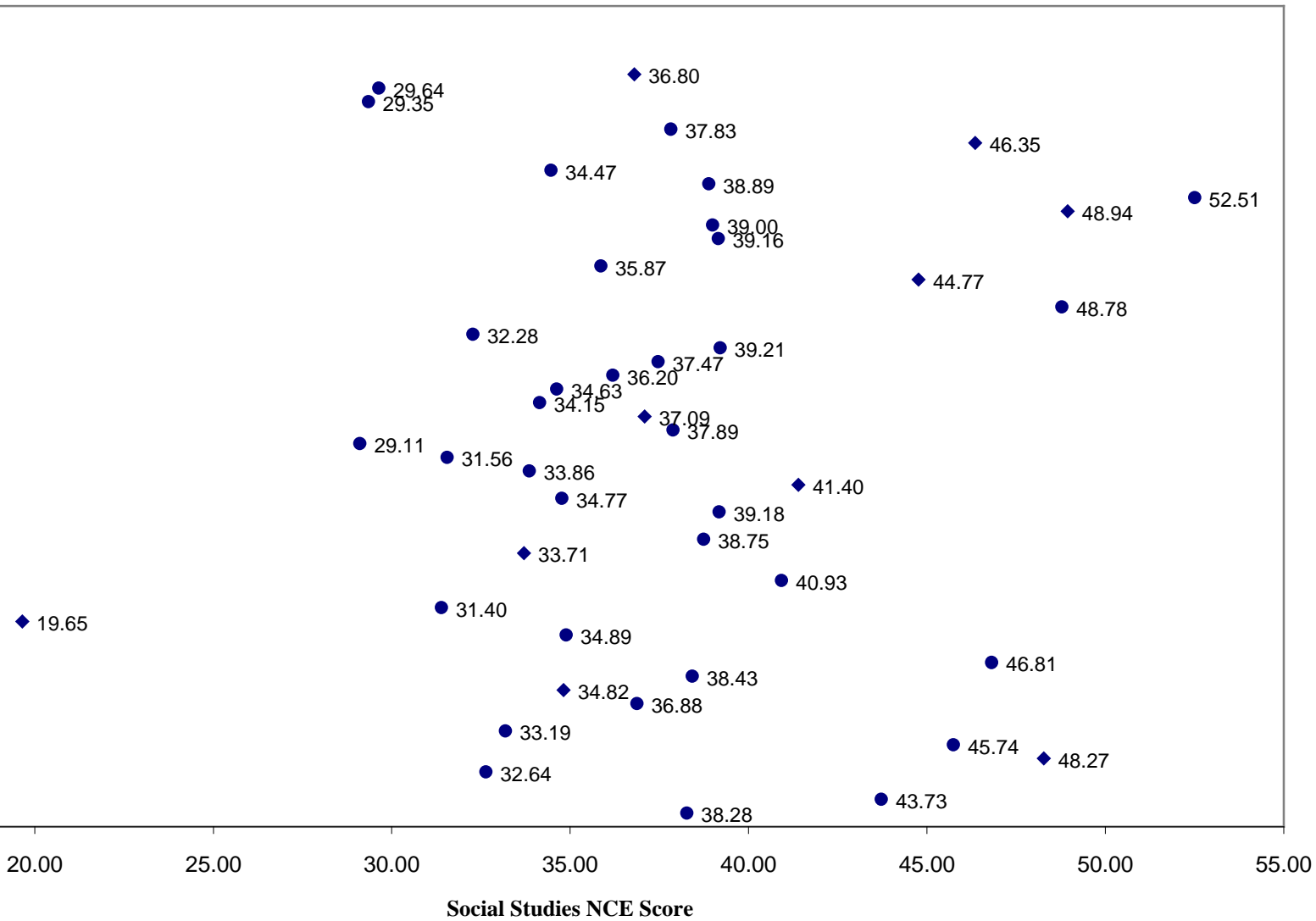


Figure 3. Combined Social Studies Aggregated School Scores



Appendix A

Examination of Students Entering the Scholarship Program as Fourth-Graders

Descriptive statistics are presented in Table A.1 for the new fourth-grade scholarship students, specifically those scholarship students who attended public school in the third-grade, but entered the scholarship program during the fourth grade. Given the small sample size (complete background data were available for only 22 of the 41 students), these new students were not included in the comparative analyses detailed in the body of the report. However, descriptive background and previous achievement data are presented here.

Students who entered the scholarship program as fourth-graders from public schools (during the second year of the program) were somewhat different from those who had entered during the first year. New scholarship students were substantially more likely to be non-white (94%), males (59%), live only with their mothers (78%), and come from a low-income household (91%). The reasons for this are unclear. It is possible that the scholarship program was more successful in disseminating information during the second year to families of students who might be considered more difficult to reach. Similarly, the visibility of the scholarship program and opportunities for participation may have promoted greater discussion and thus awareness of the program in neighborhood, religious, or other social networks. It may also be that these families were not sufficiently interested in participating in the scholarship program during its first year, but were moved to do so by year two. In any event, students who entered the program as fourth-graders can be considered to have been somewhat more “at-risk” than even their peers who entered the program in year one. Whether this finding represents a trend that might continue in future years or merely a one-year aberration remains to be seen.

Table A.1. Means and Standard Deviations of Background Demographic and Pre-Program Achievement Measures for New Scholarship Students

| | New Fourth-Grade Scholarship Students | |
|---|--|-----------|
| | Mean | SD |
| 2 nd -Grade Comprehension (N=22) | 44.41 | 19.82 |
| 2 nd -Grade Vocabulary (N=22) | 47.45 | 20.48 |
| Race (N=32) | 0.06 | 0.25 |
| Family Income (N=32) | 0.91 | 0.30 |
| Gender (N=39) | 0.41 | 0.50 |
| Living Arrangements (N=32) | 0.78 | 0.42 |

Appendix B

Table B.1. Means and Standard Deviations of Univariate F Tests of Background Demographic and Pre-Program Achievement Variables for Newly Established Scholarship, Existing Scholarship, and Public School Students

| | Newly Established Scholarship (N=23) | | Existing Scholarship (N=70) | | Public School (N=343) | | Mean Square (MS error) | F |
|--|--|-------|-----------------------------------|-------|--------------------------|-------|----------------------------------|----------------|
| | Mean | SD | Mean | SD | Mean | SD | | |
| 2 nd -Grade Comprehension df = (2, 433) | 52.09 | 24.66 | 52.74 | 20.79 | 48.01 | 20.32 | 768.36 (425.97) | 1.80 |
| 2 nd -Grade Vocabulary df = (2, 433) | 53.04 | 16.42 | 58.70 | 24.07 | 52.87 | 23.31 | 996.55 (535.24) | 1.86 |
| Race df = (2, 433) | 0.00 | 0.00 | 0.20 | 0.40 | 0.14 | 0.35 | .35 (.12) | 2.84 |
| Family Income df = (2, 416) | 0.74 | 0.45 | 0.88 | 0.32 | 0.87 | 0.34 | .19 (.12) | 1.60 |
| Gender df = (2, 427) | 0.48 | 0.51 | 0.53 | 0.50 | 0.52 | 0.50 | .02 (.25) | .09 |
| Living Arrangements df = (2, 433) | 0.61 | 0.50 | 0.73 | 0.45 | 0.62 | 0.49 | .35 (.23) | 1.50 |
| | | | | | | | | * $p \leq .05$ |

Appendix C

Table C.1. Standardized Regression Coefficients for Background Demographic and Pre-Program Achievement Variables by Fourth-Grade Achievement Score (NCE)

| | Standardized Regression Coefficients (Significance of coefficient in parentheses) | | | | | |
|---------------------------------|--|----------------|-------------------|---------------------|---------------|--------------------------------------|
| | Reading Score | Language Score | Mathematics Score | Total Battery Score | Science Score | Social Studies Score |
| Adjusted R ² | .32 | .28 | .27 | .35 | .37 | .28 |
| N | 434 | 434 | 434 | 434 | 434 | 434 |
| Covariates | | | | | | |
| Vocabulary score (Grade 2) | .21*** | .16** | .27*** | .24*** | .17*** | .16** |
| Comprehension score (Grade 2) | .38*** | .35*** | .23*** | .36*** | .37*** | .35*** |
| White (Yes/No) | .05 | .13** | .16*** | .13** | .19*** | .17*** |
| Female (Yes/No) | .05 | .12** | .01 | .06 | -.08* | .02 |
| Lives with mother only (Yes/No) | .02 | -.01 | .01 | .00 | .03 | .01 |
| On free lunch program (Yes/No) | .01 | .00 | -.07 | -.03 | -.08* | -.05 |
| | | | | | | *p ≤ .05 **p ≤ .01 ***p ≤ .001 |

Table C.2. Standardized Regression Coefficients for Background Demographic, Pre-Program Achievement, and Classroom-Relevant Variables by Fourth-Grade Achievement Score (NCE)

| | Standardized Regression Coefficients (Significance of coefficient in parentheses) | | | | | |
|---------------------------------|--|----------------|-------------------|---------------------|---------------|----------------------|
| | Reading Score | Language Score | Mathematics Score | Total Battery Score | Science Score | Social Studies Score |
| Adjusted R ² | .32 | .29 | .27 | .36 | .39 | .29 |
| N | 431 | 431 | 431 | 431 | 431 | 431 |
| Covariates | | | | | | |
| Vocabulary score (Grade 2) | .22*** | .19** | .28*** | .26*** | .37*** | .17** |
| Comprehension score (Grade 2) | .39*** | .34*** | .23*** | .35*** | .20*** | .35*** |
| White (Yes/No) | .06 | .13** | .16*** | .14*** | .26*** | .18*** |
| Female (Yes/No) | .04 | .12** | .00 | .06 | -.08* | .01 |
| Lives with mother only (Yes/No) | .03 | -.01 | .00 | .01 | .04 | .02 |
| On free lunch program (Yes/No) | .00 | .00 | -.07 | -.03 | -.08* | -.05 |
| Class size | -.07 | -.01 | .00 | -.04 | -.04 | -.03 |
| Teacher education level | -.10* | -.12** | -.02 | -.09* | -.12** | -.08 |
| Teacher experience | .12** | .07 | .08 | .10* | .08 | .13** |