Evaluation of the East Side Learning Center tutoring program

APRIL 2007

Evaluation of the East Side Learning Center tutoring program

April 2007

Prepared by:
Jennifer Lee Schultz and Dan Mueller

Wilder Research
1295 Bandana Boulevard North, Suite 210
Saint Paul, Minnesota 55108
651-647-4600
www.wilder.org

Contents

Figures

Acknowledgments

Wilder Research Center staff who contributed to the analysis and production of this report include:

Mark Anton
Marilyn Conrad
Louann Graham
Heather Johnson

Appreciation is also extended to the East Side Learning Center and Saint Paul Public School staff members who helped collect and mine the data, especially:

Sister Audrey Lindenfelser, ESLC
Meredith Tessier, ESLC
Syed Rahman, SPPS

Summary

The East Side Learning Center (ESLC) contracted with Wilder Research to conduct an initial evaluation of the impact of the ESLC tutoring program. The evaluation used existing program data and Saint Paul Public Schools' student record data to assess the effectiveness of tutoring services on the reading skills of student participants at two schools, John A. Johnson Achievement Plus Elementary and Bruce F. Vento Elementary. Skills were measured at pretest (prior to the first relevant term of tutoring) and at two posttests (one and two years after pretest). The program's impact was measured by comparing participants' gains in reading skills to gains achieved by a comparison group of students who were matched on the basis of initial reading skills and demographic characteristics. Because of sample and data limitations, this evaluation should be considered a pilot study. Results should be taken as more suggestive than conclusive.

Progress while in tutoring

Students' progress while in tutoring was measured using the Diagnostic Reading Assessment (DRA), a one-on-one assessment that measures children's reading instructional level, fluency level, and comprehension level.

DRA results: Pretest to first posttest

- DRA results indicate that ESLC students made significant gains in reading between pretest and first posttest, an average of 10.65 points, exceeding the program goal of eight points.
- Students tutored at Johnson significantly outperformed students tutored at Vento on the DRA at both pretest and first posttest. Nevertheless, Vento students made slightly larger gains than did Johnson students, narrowing the achievement gap between the schools.

Factors contributing to success

- There is a strong positive correlation between days tutored and DRA gains, with larger gains for students tutored more days.
- Students need 40 days of tutoring in order to make an expected gain of eight DRA points, which is one of the program goals.

Program impact

The program's impact was estimated by comparing progress made by ESLC students to progress made by their matched Saint Paul Public Schools (SPPS) counterparts in

vocabulary, reading comprehension, and total reading, as measured by the Stanford Achievement Test, Tenth Edition (SAT10). Since SAT10 results are grade-standardized, no change indicates normative progress, positive change indicates accelerated progress, and negative change indicates slower progress in comparison to one's peers. In what follows, "progress" or "gains" means *accelerated* progress or gains.

SAT10 results: Pretest to first posttest

- Both the ESLC students and their matched SPPS counterparts made significant progress between pretest and first posttest on all three SAT10 measures – total reading, vocabulary, and reading comprehension.
- ESLC tutoring had a significant effect on students' reading comprehension gains between pretest and first posttest, as evidenced by the significantly larger gains made by ESLC students in comparison to their SPPS counterparts. The size of the program's effect on reading comprehension was determined to be small to medium.
- On the other hand, pretest to first posttest gains in vocabulary and total reading did not significantly differ between the ESLC students and their SPPS counterparts.

SAT10 results: Pretest to second posttest

- ESLC students made significant progress on all three reading measures over a two-year period, while their SPPS counterparts made significant progress on only one.
- In comparison to their SPPS counterparts, ESLC students made significantly larger gains in total reading between pretest and second posttest. The program had a medium size effect on total reading.
- ESLC students appeared to make substantially larger gains in reading comprehension between pretest and second posttest. However, the result was not statistically significant, likely due to variability within the small sample.
- Vocabulary results indicate that the gains made by ESLC students between pretest and second posttest did not significantly differ from the gains made by their SPPS counterparts.

Factors contributing to success

- Among the ESLC students, those with good school attendance (absent less than 11 days) made significantly larger gains on SAT10 total reading and vocabulary ($p<0.05$) than did those who were frequently absent (11+ days).
- Results varied depending upon the year in which the student was tutored. Gains were largest for students tutored in 2003-04, followed by students tutored in 2004-05. Results

for 2005-06 indicated that the tutored students made below normative progress, while their SPPS counterparts continued to make accelerated gains. These results are based on small numbers of students and should be interpreted with caution.

- Males appeared to benefit more from ESLC tutoring than females, especially with regard to gains made in total reading and vocabulary.

Future directions and recommendations

The present study represents the first look at the ESLC's impact on students' reading skills. The scope of the study was limited by data already available on students, and as a result, some issues, such as the effects of ESLC tutoring on the reading skills of first and second graders, could not be tested. In addition, the small sample size made it difficult to find statistically significant results, and the lack of variability in the characteristics of ESLC students made it difficult to examine who benefits most and least from tutoring. Because of these limitations, the conclusions that can be drawn from this study are limited, and the reader should be cautious in drawing any strong conclusions about the future of the program based on this study alone. It should also be noted that this study did not examine which aspects or components of the tutoring program contribute to students' success. Therefore, conclusions cannot be drawn regarding the specific strengths of the program or components that need to be changed.

Consider options for further evaluation. In order to address some of the unanswered questions, it may make sense for this initial study to be followed by a more extensive prospective study – that is, a study with a larger sample that would collect data on the full range of program participants (grades 1-4) before, during, and after they are in the program. Because this type of study requires a large investment, program staff may want to consider joining with other tutoring programs as part of a larger evaluation effort so that the costs can be shared and minimized. It may be worthwhile to discuss this possibility with SPPS staff members, who have expressed an interest in investigating the impact of tutoring programs that are currently being used in the district.

Determine the reasons for why female students are benefiting less from tutoring than male students, and consider strategies for improving their gains. Results suggest that female students benefited less from ESLC tutoring than did males, and this difference was especially apparent in the areas of total reading and vocabulary. One staff member noted that several tutors were particularly successful at working with boys. It would be helpful to gain a better understanding of why females are making smaller gains and if there are things that ESLC staff could do to help increase their gains.

Consider expanding program efforts beyond tutoring to address other factors associated with success, including student attendance. Results indicate that students with good school attendance achieved larger gains than students with poor attendance. In order to maximize the impact on students' reading skills, ESLC may want to consider expanding its efforts to help in addressing the attendance issue, perhaps by collaborating with other school staff or organizations that are currently working to improve attendance and stability. Additional efforts could include monitoring and encouraging good attendance and finding ways to help break down barriers that are keeping students from attending regularly. In addition, program staff may want to consider ways to involve parents more in their children's learning if possible, including working with their children on their reading skills at home. Although parental involvement was not examined in this study, other research suggests that parental involvement contributes to children's success in school.

Consider reasons why the DRA results differ from the SAT10 results and the usefulness of each of these assessments. Analyses based on the DRA produced results that differed, in some cases, from the results obtained from the SAT10 analyses. For example, the SAT10 results indicated significantly larger total reading and vocabulary gains for students tutored at Johnson versus Vento, whereas the DRA results showed no significant differences in gains based on school. Another example is the finding that the number of days tutored was associated with gains on the DRA, but not on the SAT10. In interpreting the results of this study, it is worthwhile to consider why the results may have differed based on the measure used. Some possible explanations regard the skills that are measured and the ways in which the tests are administered. The skills tested by the DRA are more closely tied to the lessons covered during tutoring than are those tested by the SAT10, and research suggests that assessments which measure skills directly addressed in tutoring may exaggerate program impact, given that tutored students tend to be more familiar with these skills and assessments than are their non-tutored peers (Wasik and Slavin, 1993; Elbaum et al., 2000). The timing of SAT10 testing was not in as close proximity to the beginning and end of tutoring as was DRA testing. In addition, the posttest DRA score was an estimated rather than actual score, and consequently, its accuracy may be questionable. It should also be noted that the SAT10 has been used across the country for over 80 years and has well-established validity and reliability. On the other hand, the DRA, and in particular the correlation charts used to estimate the DRA score, do not appear to be as well-established. Given these differences, ESLC staff may want to consider the usefulness of each of these assessments as they interpret the results of this study and as they consider future data collection efforts.

Introduction

The East Side Learning Center (ESLC) is a one-on-one tutoring program that serves students in grades K-4 who are below grade level in reading and who live on the East Side of Saint Paul or attend school there. The program is administered by the School Sisters of Notre Dame. Tutoring is provided after school by part-time professional tutors (licensed elementary school teachers) and a staff of over 100 trained volunteer tutors.

This report presents results from an initial evaluation of the program's impact. The study compared the progress made by ESLC students to progress made by matched Saint Paul Public Schools (SPPS) comparison students in the areas of vocabulary, reading comprehension, and total reading.

Background

Research has consistently shown that tutoring programs can effectively improve students' reading skills. Tutoring also appears to prevent reading failure, as demonstrated through reductions in grade retentions and special education referrals (Wasik and Slavin, 1993). In the current political climate of standards-based education and accountability, tutoring is often recommended as a remediation strategy for students whose literacy development is behind or delayed. Under the Bush administration's No Child Left Behind (NCLB) legislation, underperforming Title I schools are required to provide supplemental services, including free tutoring, or the option for eligible students to transfer to higher-performing local schools. Earlier policy initiatives, such as the Clinton administration's America Reads Challenge, also encouraged the proliferation of tutoring programs.

Tutoring programs in reading typically target students in the early elementary grades, and particularly in first grade, when students learn to read for the first time. The reason behind targeting young learners is to intervene at the first sign of reading difficulty in order to prevent reading failure. Not only does it make sense to target young learners from a prevention standpoint; research also indicates that students in early elementary grades (1-3) tend to benefit more from tutoring than do students in later elementary grades (4-6) (Elbaum et al., 2000).

Although the structure of tutoring programs varies greatly, most successful programs include the following activities: (a) reading of new material by the student, (b) reading books with familiar words or stories, (c) actively emphasizing word analysis and letter-sound relationships, and (d) writing activities emphasizing composing (Wasik, 1998, p.282). In addition, successful tutoring tends to focus more time on activities that better engage the child as an active participant and that emphasize vital skills (Juel, 1996). Rather

than addressing or focusing on only a few components, programs that produced the largest effects tended to be those that were based on more comprehensive models of reading and consequently had more complete instructional interventions (ibid.). In addition, successful tutoring tends to use a larger variety of delivery methods (Wasik and Slavin, 1993) and more frequent use of two strategies: walking a child through a process (modeling) and segmenting the task into smaller, clearer ones (scaffolding) (Juel, 1996, p.286).

While some research has shown that program using certified teachers produced larger impacts than program using paraprofessionals (Wasik and Slavin, 1993), other research has shown that programs using college students as tutors produced the largest impacts, followed by programs that used paraprofessionals, certified teachers, and community volunteers, respectively (Elbaum et al., 2000). Given the high cost of hiring certified teachers, the evidence in favor of adult volunteers is promising. Programs that use volunteer tutors can help ensure success by providing tutors with extensive training, including more time spent on training prior to tutoring and ongoing training and feedback during the course of tutoring (Wasik and Slavin, 1993; Abt Associates, 2001). In addition, several sources cite the importance of having a qualified supervisor or site coordinator to design lesson plans, coordinate tutoring dyads, and provide tutors with feedback and advice (e.g., Wasik, 1998).

In addition to the skills and literacy knowledge of the tutor, the ability to develop a quality relationship with the student is a factor for success. In a qualitative examination of video-taped tutoring sessions, Juel (1996) found that successful tutoring dyads shared "obvious affection, bonding, and verbal and nonverbal reinforcement of children's progress" (pg. 282). The successful tutors identified with the children and made special efforts to communicate to the children that they could succeed in reading and writing. The commitment and consistency of the tutor also appear to be a factor for success (Elbaum et al., 2000).

It is commonly cited as essential that tutoring programs be integrated with regular classroom instruction. However, there does not appear to be enough research evidence to support this claim, and the issue needs further exploring (Wasik and Slavin, 1993; Wasik, 1998; Abt Associates, 2001). With regard to the amount or dosage of tutoring needed to make an impact, studies have produced differing results (Wasik, 1998). Overall, the research suggests that it is *intensity* of the intervention that tends to produce more powerful effects. In other words, the same amount of instructional time delivered over a shorter period appears to produce the largest impact (Elbaum et al., 2000).

It is important to keep in mind that many of the components of successful programs are likely to be interdependent. For example, the need for structure and the amount of

training needed likely depend upon the abilities and expertise of the tutor, the materials used, and the literacy skills addressed.

Program description

The East Side Learning Center tutoring program has many of the key components that the research literature suggests are associated with success.

The East Side Learning Center (ESLC) is a one-on-one tutoring program for students in grades K-4 who are below grade level in reading and who live on the East Side of Saint Paul or attend school there. The program targets students in early grades in an effort to help them reach grade level before they get caught in the cycle of academic failure. This strategy is consistent with research showing that tutoring can prevent reading failure (Wasik and Slavin, 1993) and that students in early elementary grades (1-3) tend to benefit more from tutoring than do students in later elementary grades (4-6) (Elbaum et al., 2000). In addition, almost all of the students served by ESLC are from low-income families, most of which could not afford private tutoring on their own.

The goals of the program are:

- To help children who are below grade level in reading reach their grade level no later than the end of fourth grade
- To empower the children served to be successful in school and society

ESLC is administered by the School Sisters of Notre Dame as part of their educational ministry. The licensed staff prepares detailed, individualized lesson plans so that each student receives systematic skilled reading instruction through one-on-one tutoring. The use of reading specialists who develop individualized lesson plans is a common feature of successful volunteer tutoring programs (Wasik, 1998).

ESLC tutoring is provided by part-time professional tutors (licensed elementary teachers) and a staff of over 100 trained volunteer tutors. Tutoring sessions last 45-50 minutes and occur three times per week during after school hours. The number of times a student is tutored depends upon the student's needs and progress. On average (median), students received 60 days of tutoring over approximately 20 weeks. This intensity seems consistent with what is provided in other tutoring programs shown to be effective at improving reading performance (Elbaum et al., 2000).

ESLC also operates a summer session in which students can participate in one hour of tutoring four days a week for five weeks. The program keeps track of student progress in reading through individual assessments completed at the beginning and end of each term.

ESLC began tutoring students in spring of 2001 at John A. Johnson Achievement Plus Elementary (Johnson). The program was expanded to Trinity Catholic School (Trinity) in fall 2002 and to Bruce F. Vento Elementary (Vento) in fall 2004. Most ESLC students are referred to the program by one of their teachers. By the end of the 2005-06 school year, the program had served a total of 400 children (see Figure 1). The Johnson site has served the largest number of students (254), followed by the Vento (86) and Trinity (61) sites, respectively. Likewise, the majority of ESLC students attended Johnson (52%), with smaller percentages attending Vento (21%) and Trinity (15%). The program has also served a small percentage of students (12%) who attended other schools.

1. Numbers served

Site where tutored	School attended	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	Total
Johnson	Johnson	19	40	53	69	62	59	**254**
	Other	0	4	6	12	16	13	
	Total	**19**	**44**	**59**	**81**	**78**	**72**	
Trinity	Trinity	0	4	20	23	20	17	**61**
	Other	0	0	1	1	0	0	
	Total	**0**	**4**	**21**	**24**	**20**	**17**	
Vento	Vento	0	0	0	4	47	52	**86**
	Other	0	0	0	0	1	0	
	Total	**0**	**0**	**0**	**4**	**48**	**52**	
Total	Johnson	19	40	53	69	62	59	209
	Trinity	0	4	20	23	20	17	60
	Vento	0	0	0	4	47	52	84
	Other	0	4	7	13	17	13	47[a]
	Total	**19**	**48**	**80**	**109**	**146**	**141**	**400**

Note: *These numbers differ slightly from those reported in the ESLC annual report because these numbers are unduplicated.*

[a] *Includes one student who attended Johnson and Vento and another student who attended Johnson and Trinity.*

ESLC works in collaboration with other efforts to improve student achievement on the East Side. Collaborative partners include the East Side Neighborhood Development Company and the Saint Paul Public Schools (SPPS). Other sources of support include contributions from foundations, individual donations, in-kind contributions, contract income, volunteerism, school district funding, and private sector contributions.

Assessing progress while in tutoring

Research methods

Students' progress in reading was measured using the Diagnostic Reading Assessment (DRA). The DRA is a one-on-one assessment that measures children's reading instructional level, fluency level, and comprehension level. While the child reads a passage, the teacher keeps a running record of mispronounced words and errors in phonics. The child's fluency level, or reading speed, is also measured. After reading the passage, the child is asked questions that gauge his/her level of reading comprehension.

The DRA is administered by school teachers at Johnson and Vento in the spring and fall of each school year. ESLC requests teachers to provide the DRA level of each tutored child at the beginning of each term. ESLC program staff also complete the DRA with some children who are not assessed at school (e.g., Trinity students). In order to keep a record of student progress, ESLC staff estimate the students' DRA level at the end of each term based on how far the student reached in the tutoring curriculum (a correlation chart is used). The student's estimated DRA level at the end of the tutoring term was used as the posttest measure for this analysis.

There is some variation among schools as to how the DRA levels are interpreted. According to the correlation chart used by ESLC, DRA levels correspond to grade levels as follows: levels 1-4 correspond to kindergarten, levels 5-16 correspond to Grade 1, levels 18-28 correspond to Grade 2, levels 30-38 correspond to Grade 3, level 40 corresponds to Grade 4, and levels 44 and above correspond to Grade 5. There are some gaps in the correlation chart, so if the child scores at levels 17, 29, 39, or 41-43, the teacher uses discretion and moves the child up or down a level depending on whether or not the child is a strong reader. DRA data were used to examine progress made by ESLC students while they were in the tutoring program, but could not be used to determine whether gains could be attributed to the program's impact since data were not available for the comparison group of non-tutored students.

Results

One of the program goals is to help students increase their reading skills by eight DRA points. An eight-point gain is the amount of progress needed for third graders to maintain average reading growth with peers.

ESLC provided data on students' DRA results at the beginning and end of each term of tutoring. DRA scores at pretest (i.e., prior to first relevant term of tutoring) and first

posttest were compared in order to examine students' progress in reading. The analysis includes students who were third graders between pretest and first posttest. Students who were fourth graders were excluded from the analysis because expected gains vary by grade level and the fourth grade sample was too small to examine separately.

The results indicate that ESLC students made significant progress in reading between pretest and first posttest (Figure 2). On average, ESLC students scored 18.42 points on the DRA at pretest. By the first posttest, the average score was 29.07, or 10.65 points higher. These gains are statistically significant and substantial, exceeding the program goal of eight points.

DRA results were also examined separately by school. Johnson students performed significantly better than Vento students at pretest (20.47 vs. 14.90, $p<0.01$). Students at both schools made significant progress between pretest and first posttest. Vento students made slightly more progress (+11.76 DRA points) than Johnson students (+10.00 DRA points), but the difference was not statistically significant. Johnson students continued to outperform the Vento students at posttest (30.47 vs. 26.67), but the achievement gap was no longer statistically significant.

2. DRA results from pretest to posttest 1

Sample[a]	Test	Mean	Change[b]	Sig.[c]
Total sample (N=57)	Pretest	18.42	+10.65	p<0.001
	Posttest 1	29.07		
Johnson (N=36)	Pretest	20.47	+10.00	p<0.001
	Posttest 1	30.47		
Vento (N=21)	Pretest	14.90	+11.76	p<0.001
	Posttest 1	26.67		

[a] *The samples include students who were third graders between pretest and posttest 1.*

[b] *Change from pretest to posttest 1 (change score = posttest – pretest).*

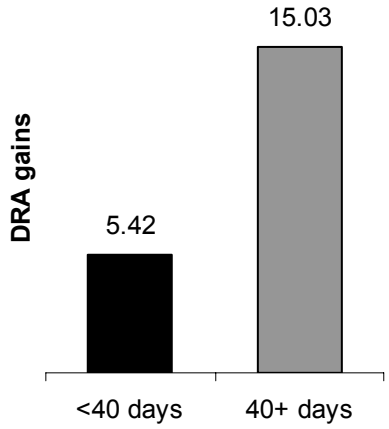
[c] *One-tailed t-tests were used to determine whether pretest and posttest means within sample, and change score means between samples, were significantly different (i.e., exceeding variation expected by chance): p<0.05 means there is only a 5% probability at most that the finding resulted by chance, p<0.01 means there is only a 1% probability at most that the finding resulted by chance, p<0.001 means there is only a 0.1% probability at most that the finding resulted by chance, and ns means that the finding was not statistically significant.*

Factors contributing to success

Days tutored

There is a strong positive correlation ($r=0.68$, $p<0.001$) between days tutored and DRA gains. Students who were tutored 40 or more days made significantly larger gains on the DRA than students who were tutored less than 40 days ($p<0.001$) (Figure 3).

3. Average DRA gains by days tutored

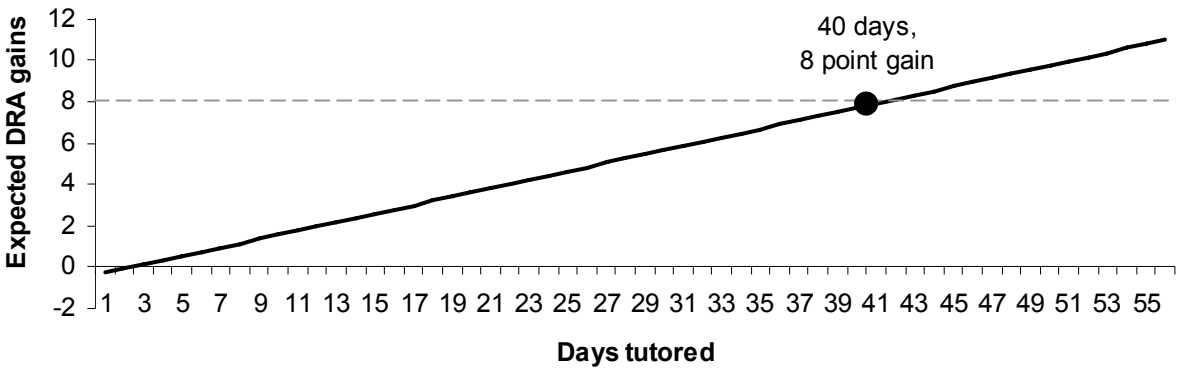


Note: *The sample includes students who were third graders between baseline and posttest 1.*

As mentioned previously, one of the program goals is to help students increase their reading skills by eight DRA points. An analysis (linear regression) was conducted to estimate the number of days of tutoring needed in order to meet this goal. The analysis was based on ESLC students who were third graders between baseline and first posttest. In addition to days tutored, the analysis took into account several other factors, including baseline DRA score, student mobility, attendance, school, school year, special education status, free/reduced price lunch status, Limited English Proficiency status, and race/ethnicity. The linear regression model was statistically significant and fit the data well, accounting for 73 percent of the variance in DRA gains. The results indicate that students would need 40 days of tutoring in order to make an expected gain of eight DRA points (Figure 4).

Results of the DRA analyses should be viewed with caution because they are based on an estimated rather than actual posttest score.

4. Expected DRA gains by days tutored



Note: *The sample includes students who were third graders between baseline and posttest 1.*

Assessing program impact

Research methods

Matched comparison group

The ESLC sample consisted of students who were tutored at the Johnson and Vento sites and who were third and/or fourth graders during the 2003-04, 2004-05, and 2005-06 school years. The ESLC students were individually matched with non-tutored students from Saint Paul Public Schools (SPPS) on the basis of several variables. All pairs were matched on grade level at baseline (or pretest), reading skills at baseline (Stanford Achievement Test, Tenth Edition total reading stanine score), and special education status.

Comparison students were drawn from the same schools as participants when possible. However, as expected, it was difficult to identify enough comparison students at Johnson because most of the students with very low reading scores were already participating in the program. About half of the ESLC Johnson students were matched with SPPS students who also attended Johnson. When same school matches could not be identified, Johnson students were matched with students from Dayton's Bluff elementary school, which is also an Achievement Plus school, and consequently has classroom, extended learning, and student and family support programs that are similar to Johnson, but does not have the ESLC tutoring program. In four cases neither a Johnson nor a Dayton's Bluff match could be found, so the ESLC Johnson student was matched with a SPPS Vento student. The pool of students at Vento who had similar initial reading skills and did not participate in ESLC tutoring was larger, and thus it was easier to find same school matches for ESLC students who received tutoring at Vento. All but one of the ESLC Vento students were matched with SPPS students who also attended Vento. The one ESLC Vento student for whom a SPPS Vento match could not be found was instead matched with a SPPS Dayton's Bluff student.

In addition to school, grade level, initial reading skills, and special education status, students were matched on as many other variables as possible, including baseline school year, Limited English Proficiency (LEP) status, free/reduced price lunch status, ethnicity, and gender. The variables were organized into a hierarchy of importance, and a multi-stage matching process based on the hierarchy was used to identify the best possible matches. Each stage of the process had a set of criteria for matching that was predetermined by the variable hierarchy. The process began with the strictest criteria, and the strictness declined with each subsequent stage. After all possible matches were identified in a stage, they were removed from the sample, and the matching process continued for the remaining cases using the next stage of criteria (i.e., either a variable was removed from

the selection criteria or the categories of a variable were simplified). In cases where more than one possible match was identified, the computer program randomly selected one of the possibilities using a random number generator.

Twenty-two percent of the sample was matched exactly on all nine matching variables. Another 23 percent was matched exactly on eight of the nine variables, and 38 percent was matched exactly on seven of the nine variables, resulting in strong matches for 83 percent of the sample. Of the remaining 11 matches, seven were matched on six of the nine variables, three were matched on five of the nine variables, and one was matched on four of the nine variables. The total sample included 65 ESLC students matched with 65 SPPS comparison students, for a total of 130 students.

Measures

The Stanford Achievement Test, Tenth Edition (SAT10) was used to match ESLC and SPPS students based on their initial reading skills. The SAT10 is a standardized multiple-choice assessment test that is administered by SPPS in the spring of each school year to students in grades two and above. ESLC and SPPS students were matched on SAT10 total reading stanine score at baseline. Stanine scores range from one to nine. Scores of one to three are considered below average, scores of four to six are considered average, and scores of seven to nine are considered above average. Nationally, 77 percent score average or above.

The SAT10 was also used to measure ESLC students' gains in reading and to compare their gains to those achieved by the SPPS comparison group. The analysis examined students' scores on the vocabulary and reading comprehension subscales as well as their total reading composite scores. SAT10 results are presented as Normal Curve Equivalent (NCE) scores. NCE score range from 0-99, and the national average (norm) is 50.0. NCE scores are standardized, reflecting skills in relation to one's peers. If a student maintains the same NCE score for two years in a row (change=0), this does not indicate that the student did not make progress, but rather indicates that the student is progressing at the same rate as his or her peers. Likewise, a positive change in NCE score indicates accelerated progress or growth relative to peers, and a negative change indicates slower progress relative to peers. Hence, in what follows, "progress" or "gains" should be interpreted as accelerated progress relative to a student's peers.

Data analysis

Students' SAT10 reading test scores were examined at baseline (pretest) and at each following year for up to three posttests (results are reported for first and second posttest only, as too few students were tutored for three years to include). Baseline test scores were taken from the spring prior to the start of tutoring when possible. In other words,

baseline scores were taken from the spring of second grade for students who were tutored for the first time as third graders and from the spring of third grade for students who were tutored for the first time as fourth graders. Some students also received tutoring prior to the third grade. However, since second grade is the first year in which the SAT10 is administered, the baseline data were taken from the spring of second grade, and the term in which the student was a third grader was counted as the first relevant term of tutoring.

Data analyses examined whether the change in test scores from pretest to posttest was statistically significant. This was examined separately for ESLC and matched SPPS students in order to determine whether tutored and non-tutored students made significant progress. The second stage of the analysis involved comparing the pre-post change demonstrated by the ESLC students to that made by their SPPS matches and determining whether the comparative change was statistically significant. If significant differences were found between ESLC and SPPS students, the size of the program's effect was calculated.

Student characteristics

Information on the characteristics of the study sample, including ESLC students and their matched SPPS counterparts, is presented in Figure 5. Sixty-three percent of the ESLC participants were tutored at Johnson, and 37 percent were tutored at Vento. The percentage of comparison students at Johnson was smaller (29%) since most of the students with low reading skills were participating in tutoring. Due to this situation, some of the comparison students (29%) were drawn from Dayton's Bluff, and the rest (42%) attended Vento. Most of the participants (88%) were in second grade at baseline, while a small percentage (12%) was in third grade. The grade level distribution was identical among the comparison students, as all pairs were matched exactly on this variable. An equal percentage of ESLC and SPPS students were female (62%). In regard to racial/ethnic background, the largest group of tutored students was Black (46%), followed by Asian (19%), Hispanic (17%), White (17%), and American Indian (2%). This distribution was similar among the SPPS comparison students, but a smaller percentage was Hispanic (9%) and a larger percentage was Asian (26%). An equal percentage of ESLC and SPPS students were identified as Limited English Proficiency (LEP, 39%), as all pairs were matched exactly on this variable. Among the LEP students, the most common home languages were Hmong and Spanish. Almost all of the participants (95%) and comparison students (99%) were receiving free/reduced price lunch, indicating that they came from low-income families. The percentage of students receiving special education services was small (9%) and identical among ESLC and SPPS students since all pairs were matched exactly on this variable.

5. Participant characteristics (N=65 ESLC and 65 SPPS, 130 total)

Characteristics		ESLC students		SPPS matched comparison students	
		Number	Percent	Number	Percent
School	Johnson	41	63%	19	29%
	Vento	24	37%	27	42%
	Dayton's Bluff	0	0%	19	29%
Grade at baseline	Second grader	57	88%	57	88%
	Third grader	8	12%	8	12%
Gender	Female	40	62%	40	62%
	Male	25	39%	25	39%
Race/ethnicity	Black, African American, or African	30	46%	32	49%
	Asian or Pacific Islander	12	19%	17	26%
	Hispanic, Latino, or Chicano	11	17%	6	9%
	White or Caucasian	11	17%	9	14%
	American Indian or Native American	1	2%	1	2%
Limited English Proficiency	Yes	25	39%	25	39%
	No	40	62%	40	62%
Home language	English	40	62%	40	62%
	Hmong	12	19%	16	25%
	Spanish	9	14%	6	9%
	Creolized English	2	3%	1	2%
	Somali	1	2%	0	0%
	Tigrinya	1	2%	1	2%
	Vietnamese	0	0%	1	2%
Free/reduced price lunch	Yes	62	95%	64	99%
	No	3	5%	1	2%
Special education	Yes	6	9%	6	9%
	No	59	91%	59	91%

Results

In order to examine the impact of ESLC on students' reading skills, analyses were conducted to compare gains made by ESLC students to gains made by their matched SPPS comparisons. The analyses were based on students' SAT10 scores in total reading, vocabulary, and reading comprehension. First, the amount of progress made between pretest and posttest was examined separately for tutored and non-tutored students. Then analyses were conducted to determine whether ESLC students made significantly larger gains in comparison to their matched SPPS counterparts.

Change from pretest to first posttest

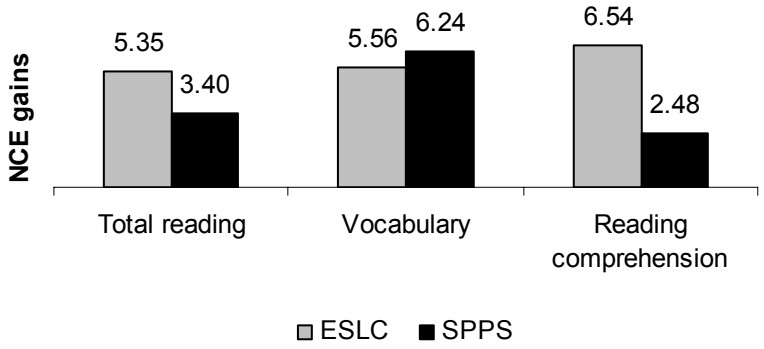
The first set of analyses examined the gains made between pretest and first posttest (Figures 6 and 7). Results indicate that both the ESLC students and their matched SPPS counterparts made significant progress between pretest and first posttest on all three SAT10 measures.

ESLC students made an average gain of 5.35 points in total reading, while their matched SPPS counterparts made an average gain of 3.40 points. A comparison of the gains made by the two groups shows that the gain for ESLC students was 1.94 points higher than the gain made by their SPPS counterparts. However, this difference was not statistically significant.

Turning to vocabulary, the SPPS students appeared to make a slightly larger gain (6.24 points) than the ESLC students (5.56 points), but this difference (0.68 points) was not statistically significant.

On the other hand, larger differences were observed between the groups in reading comprehension. While ESLC students made an average gain of 6.54 points, their matched SPPS comparisons made smaller gains, an average of 2.48 points. This represents a substantial and statistically significant difference of 4.06 points in favor of ESLC students. This difference is larger than could be expected by chance, and therefore, it should be attributable to some difference between the two groups. Since the SPPS comparison students were similar to the ESLC students on many of the variables thought to correlate with test scores but did not receive tutoring, these results suggest that ESLC tutoring had a beneficial impact on students' reading comprehension gains. Statistical calculations show that the size of the program's effect on reading comprehension gains is 0.36, indicating a small to medium effect (small effect = 0.2, medium effect = 0.5, large effect = 0.8).

6. Average SAT10 NCE gains from pretest to posttest 1



Note: *The difference in gains between ESLC students and SPPS comparison students was statistically significant ($p<0.05$) for reading comprehension, but not for total reading or vocabulary.*

7. Results from pretest to first posttest, total sample

Measure	Sample	N	Mean Pretest	Mean Posttest 1	Change[a]	Sig.[b]	Comparative change[c]	Sig.[b]	Effect size[d]
SAT10 total reading NCE	ESLC	54	32.77	38.11	+5.35	p<0.01	+1.94	ns	-
	SPPS	54	31.68	35.08	+3.40	p<0.01			
SAT10 vocabulary NCE	ESLC	54	31.42	36.98	+5.56	p<0.01	-0.68	ns	-
	SPPS	54	27.93	34.16	+6.24	p<0.001			
SAT10 reading comprehension NCE	ESLC	61	35.23	41.78	+6.54	p<0.001	+4.06	p<0.05	0.36
	SPPS	61	37.38	39.86	+2.48	p<0.05			

[a] *Change from pretest to posttest 1 (change score = posttest – pretest).*

[b] *One-tailed t-tests were used to determine whether pretest and posttest means within sample, and change score means between samples, were significantly different (i.e., exceeding variation expected by chance): p<0.05 means there is only a 5% probability at most that the finding resulted by chance, p<0.01 means there is only a 1% probability at most that the finding resulted by chance, p<0.001 means there is only a 0.1% probability at most that the finding resulted by chance, and ns means that the finding was not statistically significant.*

[c] *Comparative change is computed by subtracting the mean change for SPPS from the mean change for ESLC (difference in change scores = ESLC mean change score – SPPS mean change score). This provides an indication of the amount of progress ESLC students made in comparison to their SPPS counterparts. A negative score indicates they made less than normative progress. A positive score indicates they made more than normative progress, or accelerated progress.*

[d] *Small effect = 0.2, medium effect = 0.5, large effect = 0.8.*

Change from pretest to second posttest

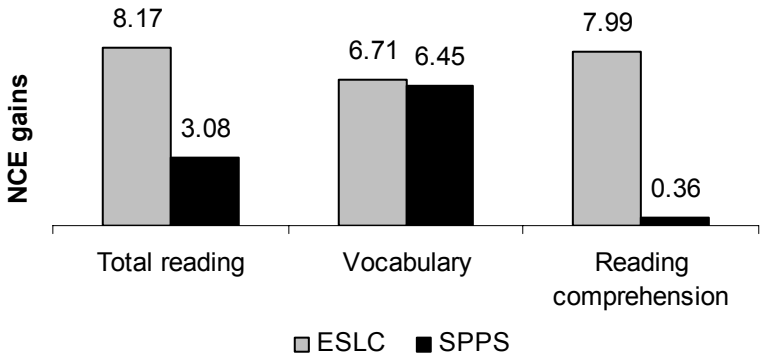
SAT10 test score data were examined at a second posttest (i.e., two years after pretest) for students who continued to receive tutoring and their matched SPPS counterparts (Figures 8 and 9). These data were available for a smaller sample of students. Analyses examined the gains made over this two-year period. Results indicate that ESLC students made significant gains between pretest and second posttest on all three SAT10 measures, whereas their matched SPPS counterparts made significant gains on only one of the measures (vocabulary). This finding suggests that the ESLC students made more consistent progress between pretest and second posttest than did their SPPS counterparts.

Results for total reading show that ESLC students made an average gain of 8.17 points (significant), while their matched SPPS counterparts made an average gain of only 3.08 points (not significant). In other words, the average gain made by tutored students was 5.09 points larger than the gain made by their non-tutored counterparts. This difference was large and statistically significant, and the corresponding effect size of 0.51 suggests that the tutoring program had a medium effect on students' total reading gains between pretest and second posttest.

Vocabulary results show that ESLC students made significant gains between pretest and second posttest (+6.71 on average). Likewise, the SPPS comparison students made similar gains (+6.45 on average). The difference in gains made by the two groups (0.26 points in favor of ESLC students) was small and not significant.

Once again, large differences were observed between ESLC students and their SPPS counterparts on reading comprehension. ESLC students made large and significant gains between pretest and second posttest (+7.99 on average), while the gains made by their SPPS counterparts were small and not significant (+0.36 on average). Hence, gains were an average of 7.63 points larger for ESLC students as compared to their SPPS counterparts. Although this difference appears to be substantial, it was not statistically significant, likely due to variability within the small sample.

8. Average SAT10 NCE gains from pretest to posttest 2



Note: *The difference in gains between ESLC students and SPPS comparison students was statistically significant ($p<0.05$) for total reading, but not for vocabulary or reading comprehension..*

9. Results from pretest to second posttest, total sample

Measure	Sample	N	Mean Pretest	Mean Posttest 2	Change[a]	Sig.[b]	Comparative change[c]	Sig.[b]	Effect size[d]
SAT10 total reading NCE	ESLC	19	31.14	39.31	+8.17	p<0.01	+5.09	p<0.05	0.51
	SPPS	19	31.64	34.72	+3.08	ns			
SAT10 vocabulary NCE	ESLC	19	29.55	36.26	+6.71	p<0.01	+0.26	ns	-
	SPPS	19	25.79	32.24	+6.45	p<0.01			
SAT10 reading comprehension NCE	ESLC	20	36.22	44.21	+7.99	p<0.05	+7.63	ns	-
	SPPS	20	39.45	39.81	+0.36	ns			

[a] *Change from pretest to posttest 2 (change score = posttest 2 – pretest).*

[b] *One-tailed t-tests were used to determine whether pretest and posttest means within sample, and change score means between samples, were significantly different (i.e., exceeding variation expected by chance): p<0.05 means there is only a 5% probability at most that the finding resulted by chance, p<0.01 means there is only a 1% probability at most that the finding resulted by chance, p<0.001 means there is only a 0.1% probability at most that the finding resulted by chance, and ns means that the finding was not statistically significant.*

[c] *Comparative change is computed by subtracting the mean change for SPPS from the mean change for ESLC (difference in change scores = ESLC mean change score – SPPS mean change score). This provides an indication of the amount of progress ESLC students made in comparison to their SPPS counterparts. A negative score indicates they made less than normative progress. A positive score indicates they made more than normative progress, or accelerated progress.*

[d] *Small effect = 0.2, medium effect = 0.5, large effect = 0.8.*

Factors contributing to success

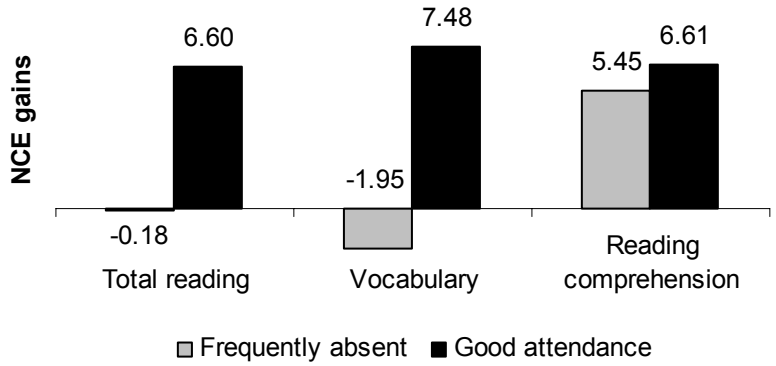
Additional analyses were conducted to examine the impact of other factors that could contribute to a students' success. In order to detect significant differences between groups of students with different characteristics, it is necessary to have a sufficiently large number of students in each group. Because the sample of students served by ESLC is very homogenous, it was difficult to detect statistically significant differences based on their characteristics. It is possible that there are differences between students based on income, special education status, and race/ethnicity, but there was not enough variability among the tutored students to detect significant differences based on these characteristics.

SAT10 gains did not significantly differ by Limited English Proficiency (LEP) status, grade level at baseline, and days tutored.

Attendance

Among the ESLC students, those with good school attendance (absent less than 11 days) made significantly larger gains on SAT10 total reading and vocabulary ($p<0.05$) than did those who were frequently absent (11+ days). In fact, ESLC students who were frequently absent made less than normative progress on both measures. On the other hand, the results for reading comprehension showed no significant difference in ESLC students' gains based on attendance (Figure 10).

10. Average SAT10 NCE gains by attendance for ESLC students

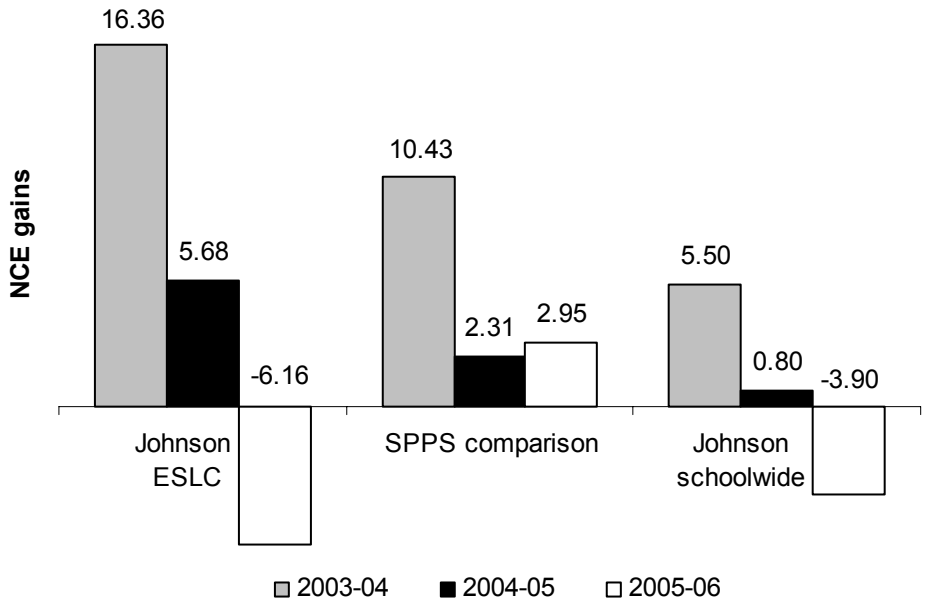


Note: *The differences in gains between ESLC students who were frequently absent and ESLC with good attendance were statistically significant ($p<0.05$) for total reading and vocabulary, but not for reading comprehension.*

School year

SAT10 reading results varied by school year. Figure 11 presents average SAT10 reading gains by school year (year between pretest and posttest) for ESLC students tutored at Johnson, their SPPS counterparts, and Johnson school as a whole. Comparable data was not available for the Vento students. The largest gains were achieved in 2003-04, with smaller gains in 2004-05. In both 2003-04 and 2004-05, ESLC students tutored at Johnson made larger gains than their SPPS counterparts. However, this pattern was reversed in 2005-06, when the ESLC students experienced below normative gains in reading scores while their SPPS counterparts continued to make accelerated gains. The overall trend across years for students tutored at Johnson reflects the trend observed for Johnson school-wide. Caution should be exercised in interpreting these results since they are based on small numbers of students (n=8-16 each year).

11. Average SAT10 total reading NCE gains by school year



Note: *SPPS comparison includes the comparison students who were individually matched with the Johnson ESLC students. The largest proportion of the SPPS comparisons also attended Johnson (46%), but 44% attended Dayton's Bluff and 10% attended Vento.*

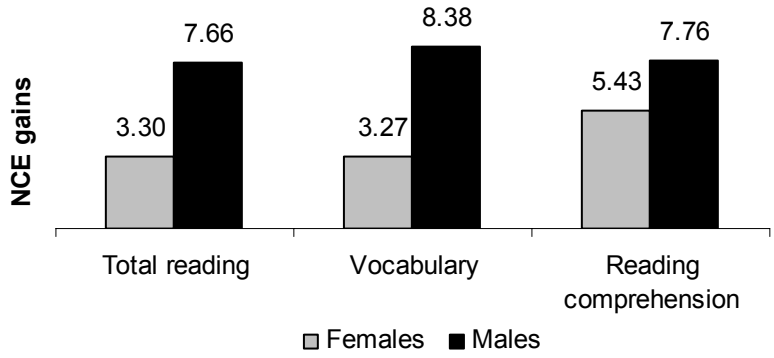
Gender

ESLC males compared to ESLC females

Among the ESLC students, both the males and the females made gains between pretest and first posttest on all three SAT10 measures (Figure 12). However, ESLC males appeared to make more consistent progress, with large, significant gains on all three measures. In contrast, ESLC females made a significant gain on only one of the

measures, reading comprehension. Although the gains made by ESLC females appeared to be considerably smaller than the gains made by ESLC males, the differences between genders were not statistically significant.

12. Average SAT10 NCE gains by gender for ESLC students



Note: *The differences in gains between ESLC males and ESLC females were not statistically significant.*

ESLC students compared to SPPS students, by gender

On the other hand, gains made by ESLC students in comparison to their SPPS matched comparisons did significantly differ by gender (Figures 13-15). This analysis included only pairs that were matched on gender.

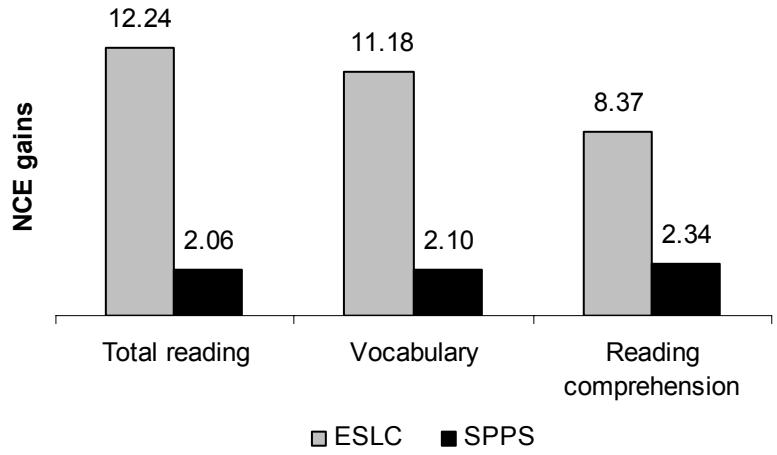
ESLC males made significantly greater progress in total reading than did their SPPS counterparts (+10.17 points), whereas ESLC females appeared to progress slower than their SPPS counterparts on this measure (-3.67 points, not significant).

Vocabulary results also show that ESLC males made significantly larger gains in comparison to their SPPS counterparts (+9.08 points). In contrast, ESLC females made considerably less progress than their SPPS counterparts did in vocabulary (-6.22 points).

Results for reading comprehension show that ESLC males and females both appeared to make greater progress than their SPPS counterparts, and the comparative gain was much larger for the ESLC males (+6.03 points) than for the ESLC females (+0.63 points). However, the difference in gains made by ESLC versus SPPS comparison students was not statistically significant for either gender on this measure.

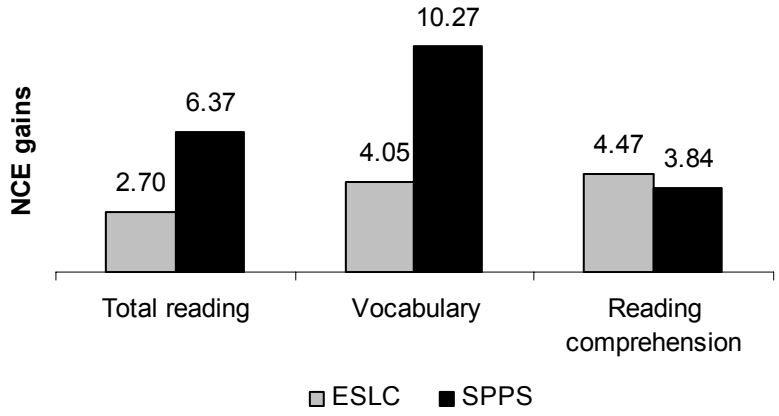
Overall, these results suggest that males benefited more from ESLC tutoring than did females, especially with regard to gains made in total reading and vocabulary.

13. Average SAT10 NCE gains for males



Note: *The differences in gains between ESLC males and SPPS comparison males were statistically significant ($p<0.05$) for total reading and vocabulary, but not for reading comprehension.*

14. Average SAT10 NCE gains for females



Note: *The differences in gains between ESLC females and SPPS comparison females were not statistically significant. The difference between the groups on vocabulary would have been statistically significant had we used a two-tailed t-test (used for non-directional hypotheses). However, we used a one-tailed t-test, given our directional hypothesis that ESLC students would outperform their SPPS counterparts. Hence, the result is not significant because it is counter to our hypothesis.*

15. Results from pretest to posttest 1, by gender

Measure	Gender	Sample	N	Mean Pretest	Mean Posttest 1	Change score[a]	Sig.[b]	Comparative change[c]	Sig.[b]	Effect size[d]
SAT10 total reading NCE	Female	ESLC	26	34.78	37.48	+2.70	ns	-3.67	ns	-
		SPPS	26	33.23	39.60	+6.37	p<0.01			
	Male	ESLC	11	34.05	46.28	+12.24	p<0.01	+10.17	p<0.05	0.84
		SPPS	11	32.86	34.93	+2.06	ns			
SAT10 vocabulary NCE	Female	ESLC	24	33.25	37.30	+4.05	ns	-6.22	ns[e]	-
		SPPS	24	28.97	39.23	+10.27	p<0.01			
	Male	ESLC	14	30.38	41.56	+11.18	p<0.05	+9.08	p<0.05	0.69
		SPPS	14	27.66	29.76	+2.10	ns			
SAT10 reading comprehension NCE	Female	ESLC	29	37.16	41.63	+4.47	p<0.05	+0.63	ns	-
		SPPS	29	39.37	43.21	+3.84	p<0.01			
	Male	ESLC	15	35.03	43.40	+8.37	p<0.05	+6.03	ns	-
		SPPS	15	34.98	37.32	+2.34	ns			

[a] *Change score = posttest – pretest; this provides an indication of change from pretest to posttest 1.*

[b] *One-tailed t-tests were used to determine whether pretest and posttest means within sample, and change score means between samples, were significantly different (i.e., exceeding variation expected by chance): p<0.05 means there is only a 5% probability at most that the finding resulted by chance, p<0.01 means there is only a 1% probability at most that the finding resulted by chance, p<0.001 means there is only a 0.1% probability at most that the finding resulted by chance, and ns means that the finding was not statistically significant.*

[c] *Comparative change is computed by subtracting the mean change for SPPS from the mean change for ESLC (difference in change scores = ESLC mean change score – SPPS mean change score). This provides an indication of the amount of progress ESLC students made in comparison to their SPPS counterparts. A negative score indicates they made less than normative progress. A positive score indicates they made more than normative progress, or accelerated progress.*

[d] *Small effect = 0.2, medium effect = 0.5, large effect = 0.8.*

[e] *Although p<0.05, this result is not significant because it is counter to our one-tailed hypothesis.*

References

Abt Associates Inc. (2001). *AmeriCorps tutoring outcomes study.* Cambridge, MA: Abt Associates Inc.

Elbaum, B., S. Vaughn, M.T. Hughes, and S.W. Moody. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92, 4,* 605-619.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31, 3,* 268-289.

Wasik, B.A. and R.E. Slavin. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28, 2,* 179-200.

Wasik, B.A. (1998). Volunteer tutoring programs in reading: A review. *Reading Research Quarterly, 33, 3,* 266-291.