

## **Prospective Teachers as Tutors:**

### **Measuring the Impact of a Service-Learning Program on Upper Elementary Students**

With the recent proliferation of college service-learning programs such as America Reads and tutoring programs such as Howard Street and Book Buddies, college students now have more opportunities to serve as tutors in elementary schools than ever before. A number of reports have appeared in the research literature describing tutoring programs that could serve as models for tutoring initiatives (Morrow & Woo, 2001; Fitzgerald, 2001; Juel, 1993; Prosser & Levesque, 1997; Wasik & Slavin, 1993). Based on this research, we developed a service-learning tutoring program for our undergraduate prospective teachers designed to provide support for them in what was for many their first experience as elementary school teachers. Our evaluation of their experiences indicated that the undergraduate tutors in our program were transformed in significant ways and developed new perspectives and attitudes in such areas as: identity and personal development; teaching and learning; and service and responsibility to the community (Malone, Jones, & Stallings, 2001). The undergraduate tutors also enjoyed the opportunity to help others and to interact with the children.

We were also interested in whether the tutoring was having a positive effect on the elementary school children. As recently as 1997 there was “very little work documenting the *effectiveness* of using volunteers as tutors” (Wasik, 1997, p. 283, emphasis added). The structure of our program gave us an opportunity to help to fill this gap in the research record. The purpose of this article is to: 1) describe the service-learning tutoring program that we developed for prospective teachers; 2) assess the impact of the tutoring program on the tutees (elementary school children); and 3) discuss the implications of the results within the context of the growing body of research literature on tutoring programs.

### **Literature Review**

One of the first studies on the effectiveness of tutoring programs was conducted by Cohen, Kulik, and Kulik (1982). These researchers concluded that tutoring programs that are able to demonstrate

effectiveness using statistical data generally: are more structured; provide training; target basic skill acquisition; use achievement in basic skills as outcome measurements; use evaluation instruments that are locally developed specifically to measure the effect of tutoring, as opposed to more commercial standardized tests; and focus on more well-defined academic areas such as math rather than on more complex tasks such as reading. Cohen, Kulik, and Kulik further concluded that tutoring programs, in general, do have a positive effect on academic achievement, attitudes about learning, and student behavior.

Many of the findings of the seminal Cohen, Kulik, and Kulik study have been supported by more recent research. For example, Wasik (1998a) reviewed four tutoring programs that had been shown to demonstrate effectiveness. In examining the programs, Wasik found several common programmatic elements necessary to ensure effectiveness, including: structured tutoring sessions; ongoing training; regular participation; and supervision of tutors by a qualified professional. In another study, however, Wasik (1997) pointed out that conclusions drawn by researchers concerning the effectiveness of tutoring programs are frequently based on inadequate program evaluation. In examining 17 tutoring programs, Wasik found that a majority of programs claiming to be effective did not utilize comparison groups, lacked well-conceived evaluation design, failed to account for the achievement that may have resulted from maturation or classroom instruction, and in general lacked experimental rigor. Wasik concluded that while the potential for well-designed tutoring programs is significant, more attention and resources needed to be focused on evaluating existing programs.

Since the appearance of the Wasik articles, researchers have begun to address this need. In one such study, Elbaum, Vaughn, Hughes, and Moody (2000) conducted a meta-analysis of 29 studies of supplemental adult-delivered one-to-one instruction for elementary school students with low reading skills. Each of the 29 studies used in the meta-analysis “contrasted one or more groups of students who participated in a supplemental one to one instructional intervention in reading with a group of students who did not receive any one to one instruction” (p. 608). Almost all of the children represented in the

meta-analysis were first graders (n = 1, 164); the meta-analysis included only five samples of students in higher grades (n = 130). Elbaum et al. indicated that many factors appeared to have an impact on the reported effects of tutoring documented in the 29 studies. The researchers concluded: 1) children's grade level was significantly associated with the variation in effect sizes (The effects of tutoring in studies of students in grades four to six varied greatly and the overall mean effect was not significant; the mean effect in lower grades was significant.); 2) the qualifications of the tutor/instructor was significant (College students who served as a tutor had the greatest effect; paraprofessionals and community volunteers had a smaller but significant impact.); 3) training of tutors was positively related to achievement; 4) treatment fidelity was significant (e.g., tutors who came to each tutoring session and tutored for the entire time had a more positive effect); 5) research design and rigor varied among the 29 studies (In some studies, students in the control group received supplemental academic support in after school programs, thus perhaps confounding the results.); 6) the focus of the intervention had an impact on the overall effect (Focusing on reading comprehension appeared to have a greater effect than a focus on phonics or a mixed focus.); 7) the type of outcome measure utilized in a study was related to effect size (Studies that utilized non-standardized measures that were more closely aligned with interventions yielded larger effects than studies that utilized standardized tests as outcome measures.); 8) the particular aspect of reading that was measured after the intervention was associated with variation in effect sizes (Studies that used measures of listening comprehension and writing vocabulary as outcomes tended to report larger effects than studies that used reading comprehension or composites scores based on subtests of different skills; 9) the same amount of instructional time delivered over a shorter rather than a longer period of time appeared to have more powerful effects (Total instructional time by itself or duration of tutoring did *not* appear to have an effect.); and 10) the method of assigning children to comparison groups had an impact on effect size (Studies that used random assignment or matching yielded significantly larger effect sizes than studies that assigned children based on teacher judgment or convenience.). In sum, Elbaum et al. concluded that one-to-one reading interventions that use trained college students and

volunteers “can make a significant contribution to improved reading outcomes for many students whose poor reading skills place them at-risk for academic futures” (p. 617).

Over the past five years, research on the effectiveness of tutoring programs has increased dramatically. This phenomenon is perhaps the result of several factors: the advent of the America Reads Challenge (Adler, 1999); the proliferation of nationally recognized one-on-one reading interventions such as Reading Recovery (Clay, 1993), Success for All (Success for All Foundation, 2000), Wallach Tutoring Program (Wallach & Wallach, 1976), Book Buddies (2000), and Howard Street (Morris, 1999); a greater utilization of service learning as a pedagogical approach; an increase in volunteerism and civic engagement; low scores on reading achievement by fourth graders on the National Assessment of Educational Progress; and the growing use of standardized tests in the early grades as methods of ensuring accountability. Given that the recent No Child Left Behind Act requires even more rigorous evaluation of students, tutoring will continue to play an important role in supporting student achievement, as will efforts to address the deficiencies in past evaluations of tutoring programs noted by Wasik.

As Elbaum et al. (2000) have indicated, much of the recent research on tutoring has focused on improving reading scores in the early grades. Although some studies have examined the effects of tutoring on children in upper elementary grades (Butler, 1991; Graves, 1986; Hedrick, 1999; McGrady, 1984; Ramey, 1991; Topping & Whitley, 1990) and a few studies have examined broader academic skills rather than focusing solely on reading (Cavanaugh, Johnston, Kitay, & Yuratovac, 1997; Wasik, 1998a), research in these areas is limited. One of the purposes of the current study is to examine what impact, if any, trained college students could have on the academic skills of upper elementary school children who are at-risk for school failure.

### **Description of Our Service-Learning Tutoring Program**

Partners for Success (PFS) is a service-learning tutoring program that matches prospective undergraduate teachers with children in the fourth and fifth grades who are at risk of school failure. The

program is named Partners for Success to highlight its cooperative and reciprocal design and structure. In fact, PFS is more than just a tutoring program; it is also a bridge between neighborhood schools and the university. School input was critical in the formation of the program and continues to be an important aspect of evaluation and revision.

The program originated as a collaborative project involving a university education department and a local public school that served as a placement site for undergraduate student teachers. A large percentage of children at this school received free or reduced lunch (an indication of low family income) and the school itself had been identified by the state as near low-performing based on the end-of-grade tests that are part of the statewide accountability program. A planning committee consisting of the school principal, teachers, school counselor, university faculty, undergraduate student teachers, and administrators from the school system met over a period of several months to design the program. In designing the program, three factors were taken into account: the needs of the school and the children; the goals of the university education department; and the aspects of effective tutoring programs as identified in the research literature. The program contained many of the components that Wasik (1998a) described as being essential to making tutoring programs effective, including: ongoing training and feedback for tutors; tutoring sessions that were structured and contained basic elements (e.g., writing activities, reading familiar and new stories); intensive and consistent tutoring; quality materials to facilitate the tutoring model; ongoing assessment of students; and regular tutor attendance.

After successfully launching PFS at one school, the program was taken to three other local urban elementary schools. On each occasion, principals and teachers were involved in setting up the program for each new school. PFS now involves four schools and nearly one hundred undergraduates each semester, each paired with a single fourth or fifth grader.

### Participants

Undergraduate participants in the tutoring program were enrolled in one of three courses: an educational psychology course; a course in motivating at-risk students; or a course in tutoring and

literacy. Some tutors completed only one of these courses, while others completed all three. There is no education major at this university (participants represented a wide array of majors, including psychology, English, public policy, history, sociology, biology, and biological anthropology, among others), but there is a teacher preparation program that allows students to receive their state teaching certificate in elementary education or in secondary English, mathematics, social studies, or general science. Students complete five specific education courses and an internship in addition to coursework in their majors to become eligible for their state teaching certificate. Of the three courses in this study, only the educational psychology course was required for the state teaching certificate.

Elementary students who participated in the program were all fourth and fifth graders, a small fraction of whom were repeating a grade. Every effort was made to work with the same fourth graders from their fourth grade fall semester all the way through their fifth grade spring semester, but very few students ( $n = 19$ ) were able to participate in the program for two years, and an even smaller number participated for four full semesters (mainly due to students' families moving and extremely poor school attendance).

Children were not randomly selected for the program because schools required some discretion in selecting students who received the service; however, students were selected with the aid of a standard guideline: they had demonstrated a proficiency level of 2 (out of a possible 4 levels; 3 is a passing score) on at least one of the two North Carolina End-of-Grade (EOG) tests from the previous school year. Students with scores of 2 were preferred over students with scores of 1 because the undergraduates participating in the program were not trained educators and could not be expected to provide the support and coaching necessary to benefit the lowest-scoring students as effectively as professional educators could.

Any bias in selection of students for the program on the part of one school representative was mitigated to some extent by the fact that there were over 20 different teachers at four different schools making selections independently. In addition, teachers were asked not to select students who received any

services outside of the regular classroom that might confound results (such as learning disabled assistance or additional tutoring). Over the course of the three years of this study, the program served 139 African-American students, 17 Hispanic students, 12 Caucasian students, and one student of Middle Eastern descent; 90 students were females and 79 were males. The majority of students served scored in the 2 range on the previous year's EOG tests. On the whole, though not random, the experimental population was relatively representative of level 2 students in a typical fourth or fifth grade class in our local school system.

#### Service-Learning Coordinator and Site-Based Assistants

Two critical elements of the program were the feedback and support tutors received on-site and the maintenance of a communication link between the schools and the university. These services were provided by a university service-learning coordinator whose responsibilities included training tutors, making frequent visits to each of the four schools, meeting with principals and staff, responding to reflection journals in writing and through conversation, observing and providing constructive feedback for tutors, working closely with university professors, and gathering data. One site-based assistant was assigned to each of the four participating schools. An assistant's responsibilities included responding to reflection journals, keeping tutors abreast of changes to the normal school schedule, solving minor tutoring problems, and keeping the coordinator aware of major issues in need of attention. The assistants were almost exclusively tutors from previous years and therefore brought firsthand experience to their responses to reflection journals. Funding from local grant sources and university support made these positions possible.

#### Tutoring Sessions

PFS tutors visited the schools twice weekly for about eleven weeks each semester. Each tutor worked with the same tutee for the entire semester. Tutors were required to prepare a tutoring plan prior

to each session. Tutors based their plans on diagnostic information about their child provided by the service-learning coordinator (e.g., previous test scores, specific skills teachers indicated that the child needed to work on). Tutoring sessions ranged from 30 to 45 minutes in length, providing enough time for beneficial tutoring to take place without removing tutees from the regular classroom for too long. Tutors concluded each visit by reflecting on the lesson in writing, and frequently they were able to talk about their sessions with a site-based assistant or the service-learning coordinator. Undergraduates also subsequently reflected on their tutoring experiences in their college classes, thus forming a loop of learning that allowed them to better understand the relationship between theory and practice.

### Lessons

Before beginning the tutoring program, professional elementary educators and reading specialists were hired to create over 100 self-contained math and reading lessons designed specifically for fourth and fifth graders who were at-risk of school failure. These lessons were designed to allow the tutors to teach specific skills and concepts from the North Carolina Standard Course of Study (SCOS) Goals and Objectives. Lessons are revised each year based on comments from tutors and the service learning facilitator and assistants. PFS now uses over 180 lessons, many of which have replaced the lessons used during the time of this study.

To provide consistency for the tutors and to ensure that critical elements of instruction were covered during the lesson, each lesson was based on a lesson plan format that included: Goal(s) & Objective(s); Overview; Materials; and Procedure. Many of the lessons incorporated aspects of the language and format of the EOG tests to buttress the process of preparing students for the tests themselves.

Several lessons used a multidisciplinary, constructivist approach to help tutees make connections between math, reading, and the rest of the curriculum. For instance, several fourth grade lessons were set against a backdrop of Native American history, a major component of the state's fourth grade social

studies curriculum. Several math lessons also incorporated topics and themes from across the curriculum and from outside the elementary school experience as well. For example, one series of lessons required tutees to master the concept of scale by applying it to the design and layout of a room. The content of these lessons was not unique; rather, the most important element was that each lesson provided structure for tutee and tutor alike against a backdrop of skills and topics necessary for the success of the tutee on the state EOG tests.

### Training

Training was ongoing in the sense that the Education courses in which the undergraduates were enrolled focused on topics directly related to the problems experienced in the tutoring sessions. In fact, one of the goals of PFS was to help undergraduates join education theory and practice throughout the courses. Course topics included best practices for tutoring, principles of learning and motivation, fundamentals of child development, and reflection techniques. All undergraduate participants were provided with a handbook, as well as ongoing on-site support from the service-learning coordinator and site-based assistants. Some areas covered in training included how to prepare and teach a lesson appropriate for a tutoring session, how to reflect on the tutoring session, and principles of teaching children reading and math skills. Tutors were also taught how to read and understand the North Carolina EOG results and the North Carolina SCOS. Armed with this knowledge, they were able to select appropriate lessons for their tutees based on the results of pre-EOG tests. These tests – called Benchmark tests – were developed and administered by the local school system and provide diagnostic information on each child.

### Reflection and Feedback

After every session, tutors wrote about their experiences on an open-ended form that asked tutors to: circle a word from a list that best described their feelings about the session; write 7 to 10 sentences

reflecting on their session; ask questions; and describe any concerns. These reflection journals were read regularly by the site-based assistants and/or the service-learning coordinator, who provided feedback in the form of questions for further consideration, strategies to try, and connections to concepts from their college courses. When schedules coincided, tutors met and talked to the service-learning coordinator or the assistants at the school and engaged in informal conversation about tutoring issues.

In addition to the regular structured mechanisms for feedback, some students chose to participate in monthly group reflection sessions outside of class led by trained and experienced undergraduate peers. This environment allowed tutors to exchange ideas, learn from each other's experiences, and discuss problems and issues in a forum unhindered by concerns about reactions from their professors.

Tutors received feedback about tutoring through written responses to reflection journals, meetings with the service-learning coordinator and assistants, individual conferences with professors, conversations with teachers, and weekly classroom discussions that linked class material with tutors' experiences.

### Course Concepts

Because the tutoring was a component of undergraduate *education* courses, there was ample opportunity to connect theory to practice. The courses participating in the program reserved portions of each class for discussion of tutoring issues and how they related to the class, and vice-versa. As the semester progressed, course concepts began to enter into the reflection journal entries and tutors became more adept at recognizing ways to make the formal lessons more appropriate for their tutees. Some tutors even began to write their own lessons, and several of these were incorporated into the lesson collection.

### Emphasis on Differentiation of Instruction/Intervention

Because tutors worked with the same tutees all semester (and sometimes across semesters), they were able to individualize instruction for their tutees. An important key to this individualization was the

availability of quarterly EOG Benchmark test results, which helped the tutors to identify specific SCOS goals and objectives with which the students were struggling. Providing lessons that taught specific skills also aided in this individualization. Teacher feedback and suggestions were also helpful, but one of the strengths of the program was that tutors could craft individualized lessons (with guidance from the service-learning staff) even when already-busy teachers were not able to meet with them.

### **Research Procedures**

The effects of the tutoring on the tutees were measured through: 1) an examination of tutees' end-of-grade standardized test scores in math and reading; and 2) essays written by the tutors near the end of the tutoring semester.

#### End-of-Grade (EOG) Test Score Analysis

The EOG tests are state-generated, standardized, multiple-choice tests administered statewide every year to students in grades three through eight that measure competence in the North Carolina SCOS objectives and goals for math and reading. Testing has occurred a few weeks before the end of each school year.

We compared the results of the tutees' math and reading test scores with the scores of non-tutored children (the comparison group). In doing so we only included the tutees who were tutored for an entire school year (two college semesters). Students in the comparison group were chosen independently from the same schools and paired with tutored children based on ethnicity, gender, grade, and EOG test scores from the school year before tutoring. By measuring the differences in score growth between the comparison students and tutees, we estimated the amount of tutee score growth attributable to the program as opposed to maturation or regular school instruction. The strength of the pairs was tested by measuring the similarity of their reading and math pre-scores. For instance, in the second year of the program, score correlation for reading and math at both grade levels (fourth reading  $r = 0.6478$ , fourth math  $r = 0.7028$ ; fifth reading  $r = 0.7205$ , fifth math  $r = 0.6996$ ) indicated that the pairs were relatively

good fits for the assessment. Dubious pairs in which the discrepancy between pre-scores was greater than one grade level of growth were dropped from the evaluation.

To assess the significance of EOG score gains for tutees, their score improvements (the difference between the current year's score and the previous year's score) were compared to the score improvements of the paired comparison students using a paired-differences t-test. A paired differences t-test (which assumes the two variables are not independent) was used to measure the significance of any differences in growth because tutees and controls attended the same schools and learned from the same teachers. For a complete analysis, a total of four separate test scores per school subject were required for each member of a pair (reading and math scores from the previous and the current year); as a result, several tutee-comparison pairs were eliminated from the final evaluation each year because of at least one missing score.

We were also interested in analyzing the performance improvements of several subgroups (e.g., tutees who were tutored for all four semesters, tutees who worked with those who we considered to be our best tutors, tutees who worked with the same tutor for two or more semesters, etc.), but our analyses of these subgroups were either hampered by sample sizes that were too small to perform meaningful analysis or by inconsistent and generally uninformative results. We are attempting to address data analysis for these subgroups again as our database of tutees grows.

### Essay Analysis

Near the end of each semester of tutoring, we asked the tutors to write a short essay about how the students they tutored grew and changed as a result of the tutoring experience. We coded the essays by reading them for regularities and patterns and represented these patterns with words and phrases that became the coding categories (Bogdan & Biklen, 1998).

## Results

### EOG Test Scores

Our analysis of the EOG test scores over three years generated ten group comparisons (see Table 1). In five instances, tutees outperformed the comparison group; in four instances the comparison group outperformed the tutees; and in one instance both groups scored the same. Neither group outpaced the other at a statistically significant level of  $p < 0.10$ , with two exceptions: the fifth grade reading comparison group for school year 1998-99 and the fourth grade reading tutee group for school year 1999-2000. The number of valid student pairs available for both of these analysis (4 and 12, respectively) is, however, too low to assume any actual programmatic deficiencies or successes with confidence, especially since these two instances do not appear to represent a recurring pattern.

-----  
 Insert Table 1 here  
 -----

Because the number of students in each group was low when we evaluated the scores on a yearly basis (as is shown in Table 1), we also combined outcomes from all three years. The results show that the fourth grade math tutees outperformed the comparison group; however, there were no significant differences between the groups in the other grades or subjects (see Table 2).

-----  
 Insert Table 2 here  
 -----

### Essays

When asked how they felt tutoring had benefited their tutees, the most common response from undergraduates was that the tutees gained more confidence. One tutor noted, "I truly believe that his confidence in his ability to competently solve math problems has increased due to our sessions together."

Tutors also listed other things that the tutees gained from the experience besides the academic

skills measured by the EOG tests, including: social skills, personal growth, self-esteem, knowledge of how to deal with social issues, an understanding that learning can relate to life, an interest in a subject, and an understanding of the value of education. As one student commented, “Regardless of the result of any EOG test, I know that Renee has grown tremendously during this semester, academically, socially, and personally.”

## **Discussion**

### EOG Test Results

Why were there so few significant gains in tutees’ EOG achievement test scores as compared to the comparison students? After all, one of the purposes of the service-learning tutoring program was to help at-risk students show improvement in school achievement. We have identified several possible limitations of this study that might have contributed to the lack of significant difference between the tutees and the comparison groups on the EOG tests.

Treatment fidelity and focus. Several tutors reported in their reflections and essays that during tutoring sessions they focused on issues they thought were more important than EOG preparation. We allowed and encouraged tutors to be creative and to design their own lessons if they so wished, but it appears that allowing the tutors to have this freedom often led them to teach lessons that were consistent with their own philosophies of education but that were not necessarily consistent with the topics on the EOG tests. As a result of their own personal philosophies of education, the tutors often decided to adjust the content of the tutoring sessions and teach the things that they deemed “more important.” For example, students reported that even though they prepared a lesson plan, after entering the tutoring session and encountering tutee resistance, they might abandon the lesson in favor of a less structured activity.

Several students specifically noted that, while their tutee’s EOG test scores might not show improvement, they were convinced that the tutees learned things that would help them over time or in other areas not tested on the EOG tests. For example, one student wrote, “The end result is that Charles

seems to have a higher self-esteem now, as well as being more intrinsically motivated. While his EOG scores may not rise significantly on account of my interactions with him, Charles has acquired some qualities that are going to serve him well in the long haul.” This tutor seemed to think that he knew the important qualities that this student would need later in life. Similarly, another student noted, “Hakim may not score any higher on the EOG test, [but] I hope that he will. Yet, if he’s grown more comfortable, more confident, or more prepared to deal with the world in any way, then I’ll consider it a success.” This tutor had her own ideas about what it means to be a successful tutor. This sentiment seemed to be shared by another tutor who wrote:

As the EOG tests approach and my tutoring sessions come to an end, I can only wonder if Maya’s scores will improve as a result of our work together. Unfortunately, I am not sure if I successfully “taught the test” during the past semester, and although this could have a negative effect on Maya in the short-run, I feel as if it will ultimately benefit her in the future. Looking back on the past few months, I realize that I have helped Maya to become a more confident and eager learner, which I believe has more beneficial long-term effects than simply teaching her how to take a standardized test.

Once again, this tutor thought that she knew what would benefit this student in the future; therefore, she was willing to sacrifice the tutee’s EOG scores this year for perceived long-term benefits. From an educational standpoint, these tutoring sessions provided vital support for the tutees, but not the kind of support that could necessarily be measured by the standardized EOG test. One implication of this finding is that if the service-learning experience is designed specifically to improve tutees’ scores on standardized tests, steps need to be taken to ensure that tutors spend most of the tutoring session time on the specific skills measured by the test.

Quality of prepared lessons. Another complicating factor for the results of this study could have been the lessons themselves. Because we wanted tutors to work with their tutees on the specific SCOS skills identified as weaknesses by the Benchmark tests and by their teachers, we chose to write our own lessons instead of borrowing lessons from other tutoring programs and converting them to meet our specific needs. While each lesson was written by a certified teacher and was keyed directly to the SCOS goals and objectives for 4<sup>th</sup> and 5<sup>th</sup> grades, there was not much lesson variety in terms of approach or structure (especially among the reading lessons). Tutors and tutees alike reported dissatisfaction with the content of some lessons, and this dissatisfaction led several tutors to develop their own lessons. Some tutors developed very appropriate lessons that were later included in the PFS collection, but PFS did not provide lesson-writing training beyond providing a template, so the overall quality of tutor-produced lessons was mixed.

Age of tutees. Several studies have reported the positive effects of tutoring programs, but in most cases, these studies have involved children in the early elementary grades, not children in the fourth and fifth grades. There are fewer studies of children in grades four and five and these studies have reported mixed results (Elbaum et al., 2000). In this respect, our results are consistent with those of many previous studies. Perhaps more research on the dynamic intersection of tutoring programs and developmental learning stages could help us understand better the varying degrees of impact such programs can have across all elementary grades.

Number of tutoring sessions. PFS required students to tutor twice a week, which might have limited the achievement gains, since some research has indicated that three sessions a week has a greater effect on tutoring outcomes (Reisner, Petry, & Armitage, 1990), and two hours a week has been recommended (Wasik, 1998b). As one tutor observed, “Sometimes I question whether my being here has really made a difference. I feel I am just now learning how to be a good tutor, and now it is over.”

Confounding factors in assignment to tutee group. As noted earlier, we were not able to randomly assign students to the tutee group, and this lack of random assignment might have reduced the effect of

the treatment. Also, despite having selection guidelines in place (only those students who scored in a certain range on the state EOG test and who did not receive any additional tutoring services outside the regular classroom were eligible), the method of assigning children to the tutee group was sometimes based on teacher judgment and convenience, and not on our guidelines. As a result, our tutee demographics reflected a wider array of abilities and experiences than we intended.

Comparison group confounding. In post-analysis interviews, we learned that many of the children who were in our comparison group had received individualized instruction from other sources such as America Reads or after-school programs. The fact that at least some of the comparison group received some type of additional academic support similar to the tutee group confounded our results. While we applaud the effort of the schools who adopted these programs and the “no child left behind” philosophy, it made it difficult for us to conduct a more purely controlled study.

Type of outcome measure used. The single-administration, statewide EOG test is the object of much scrutiny in North Carolina. In fact, the state Department of Public Instruction admitted that a grading glitch in the 2001 math test allowed too many third through eighth graders to pass the test (Silberman, 2001). Cases such as this remind us that it is difficult if not impossible to assess the knowledge and skills of students based on the results of a single test that may or may not be reliable and valid. Elbaum et al. (2000) found that studies with non-standardized measures that were more closely aligned with the intervention yielded larger effects than studies using standardized test measures. Based on this information, we might have found more significant differences between the groups if we had created our own assessment measure.

Quantity of tutee data. Once ill-matched tutee pairs and pairs in which one or both students changed schools were eliminated and sub-groups were established by grade, the total number of tutees or tutee-comparison pairs available for any single statistical test was relatively small. Therefore, detecting significant differences between the comparison group and the tutees was unlikely, especially since standard deviations were often quite large (usually between four and eight points – see Tables 1 and 2).

Students with family backgrounds that are believed to contribute to risk level (such as poverty) and students in urban school settings typically are more mobile than their peers (cf. Kerbow, 1996), and this factor is likely to hamper evaluation of many tutoring programs designed to work with at-risk students.

Type and timing of lessons. Another flaw in our evaluation methods is that these methods do not take into account the *types* of lessons that tutors deliver or the *timing* of the lessons over the course of the semester. For example, a tutor who delivers 20 math-only sessions in the Spring (right before the test) might be more effective than a tutor who splits 36 sessions evenly between math and reading across the year. As one student commented: "I am somewhat disappointed because of the lack of progress that I have seen Cedrick make through working with me this semester. In retrospect, I would advise working on a single subject (e.g., arithmetic) for the whole semester in order to see student improvement demonstrated. I feel like because Cedrick and I skipped around, that he didn't really improve on anything." Tutoring might have been more effective if tutors had only worked on either literacy or math.

### Tutors' Essays

While PFS has not been shown to have a large impact on students' EOG test scores, the essay data indicate that the program did have an impact on the children in other ways. Specifically, tutors reported that it increased students' confidence, among other things. As one student noted, "Devon may do no better on his EOGs than he did before I began to tutor him, but I have gauged his progress during our sessions and I have seen evidence of his becoming a stronger, more confident, and more motivated person." These types of comments from tutors suggest that the tutoring did make a difference in these children's lives, just not the type of difference that was measured by the EOG standardized tests.

This increase in tutee confidence is important because individuals with high self-efficacy (a self-evaluation of one's abilities on a specific task) have been shown to set higher goals (Locke & Latham, 1990, 1994; Zimmerman & Bandura, 1994), choose more difficult tasks (Sexton & Tuckman, 1991), exert more effort and persist longer at tasks (Bandura & Cervone, 1983, 1986; Bouffard-Bouchard, 1990;

Zimmerman, 1995), use more cognitive and meta-cognitive strategies (Pintrich & De Groot, 1990); and perform at higher levels (Bandura, 1982, 1986). Many tutors noticed that their tutees exerted more effort and persisted longer; one noted: “I feel like if there were one thing that I have changed in Rebecca, it would be her increased confidence in reading, and her willingness to push herself.” Because of all of the positive outcomes associated with high self-efficacy, it is significant that the tutors often perceived their biggest triumph as that of increasing students’ confidence.

In general, we find tutors’ reports of increased tutee confidence encouraging; however, we also recognize the limitations in relying on tutor perceptions. That is, it is difficult to assess the reliability and validity of a tutor’s perception of his tutee’s development. For instance, the tutor might have thought that his tutee developed more confidence in mathematics based on observations during tutoring sessions, when in fact, the student was confident in attempting a certain type of math problems with the tutor sitting close by, but not as confident about her ability in her regular classroom.

Although we recognize the limitations of using tutors’ perceptions in our analyses, we believe that such reports are generally useful and accurate for a variety of reasons. First, tutors’ reports are helpful in identifying untestable effects of the tutoring such as improved social skills. Second, the consistency and frequency of tutor responses lends credence to the results reported in this paper. Third, many of the tutors had taken college course work in education or child psychology and all of the tutors were concurrently enrolled in a course that included child psychology topics such as development and motivation. As such, these tutors were receiving specialized training in these areas and were likely more qualified to make these types of judgments than untrained tutors.

### Effects on Tutors

One of the main purposes of this article was to assess the impacts of a service-learning tutoring experience on the *tutees*; however, we have also studied the effects of the tutoring experience on the *tutors*. We briefly highlight a few of these findings in this section. A complete description of these effects

can be found in a previous article (see Malone, Jones, and Stallings, 2002) in which we analyzed questionnaire data and structured essays written by the tutors about how their perspectives and attitudes were changed as a result of the tutoring experience. Through our analyses, we found that the tutoring experience transformed the tutors in many ways, which we grouped into one of four categories. The first category, “Perspectives on Identity and Personal Development,” included three themes that focused on issues related to the tutors’ self-growth, including: clarification of career and life goals, personal growth, and increased confidence. We labeled the second category “Perspectives on teaching and Learning” because it included tutor comments about how their perspectives on teaching and learning were transformed as a result of their tutoring experience. The five themes in this category included: learning course concepts, awareness of learning from tutee, a more holistic view of learning, changed perspectives on teaching or teachers, and applications of course concepts. The third category included tutor comments that related to how they began to see their roles in the community or as service providers in a different light and was titled “Perspectives on Service and Responsibility to Community.” The fourth category was titled “Personal Satisfaction” and included comments related to the sense of satisfaction and enjoyment that the tutors received from helping and interacting with their tutee. In sum, the tutors reported significant positive changes in their perspectives and attitudes in several areas based on their service-learning tutoring experience.

### **Implications**

Based on the results of our prior research (Malone, Jones, & Stallings, 2002) and that of other researchers (McKenna, 2000), there is little doubt that the type of service-learning tutoring experience described in this article can have positive effects on the tutors. It can help candidates for teacher licensure develop skills and attitudes that can be beneficial to them as teachers. When used as an early field experience in a teacher preparation program, this type of service-learning experience also has implications for many of the key issues faced by teacher educators, including: helping to recruit prospective teachers,

improving teacher quality by connecting theory to instructional practice, helping to prepare teachers to teach children from diverse backgrounds, and showing prospective teachers the importance of community involvement in supporting instruction (Malone, Jones, & Stallings, 2002). Moreover, we believe that the tutoring provided tutors firsthand experience in reflecting on their experience (see the “Reflection and Feedback” section in this paper). In this way, the undergraduate tutors learned to carve out time in their lives for reflection, a disposition that, as teacher educators, we expect them to maintain throughout their teaching careers.

The fact that we found so few significant gains in tutees’ achievement test scores suggests that teacher educators must be sensitive to the needs of the tutees participating in these types of programs and that the dictum “First, do no harm” should be obeyed. We are not suggesting that the tutors would inflict any emotional or physical harm, but rather, that the tutees are harmed by missing regular class instruction during the time of the tutoring. If the tutoring is not providing the tutee at least a similar amount of instruction as he would receive in the classroom, then he is at risk of falling behind his classmates. Even within the rigorous, structured tutoring program described in this paper, the tutees were unable to show consistent significant improvements in student achievement. This leads us to wonder whether tutees in less-structured programs could learn less from the tutoring than from remaining in the classroom (if they were removed from an instructional activity). Of course, there are many goals of tutoring besides increased achievement, such as developing students’ confidence, motivation, and social skills. Therefore, the goals related to student outcomes need to be clear to the teacher educator, tutor, classroom teacher, and parent of the tutee. We can envision many different types of goals that might be appropriate for a tutoring program, including increasing student achievement, providing a role model for the student, and developing interpersonal skills.

### **Summary and Conclusions**

The prospective teachers in the Partners for Success tutoring program are making some positive strides with fourth and fifth grade children. An analysis of qualitative data revealed that tutees had grown

in important ways from the tutoring experience. Tutors have noted positive development in such areas as self-esteem, self-efficacy, self-confidence, knowledge of how to deal with social issues, social interaction, an understanding that learning can relate to life, an interest in a particular subject, and an understanding of the value of education.

However, while qualitative indicators have suggested that the children are benefiting from tutoring, the quantitative support for the success of PFS is not as substantial. We outlined several possible reasons for this lack of significant quantitative differences between the tutee and comparison groups in the previous section. This list of limitations is lengthy because we wanted to not only provide a comprehensive list of the deficiencies of our study, but also, and perhaps more importantly, because we wanted to draw attention to what may be an important issue with respect to the manner in which evidence from research studies on tutoring effectiveness is interpreted in general.

The results of this study lead us to believe that researchers who are examining the effectiveness of tutoring programs need to be keenly aware of issues that might influence research results. Results must be interpreted within the rich context in which a particular tutoring program takes place. For example, research findings from studies that examine first grade tutoring programs most likely are not generalizable to other grade levels. Before widely touting research findings as having demonstrated the effectiveness of tutoring programs, researchers need to examine closely the limitations and factors discussed above. Differences in outcome measures, research design, and/or focus of intervention strongly impact the usefulness and implications of research findings. As researchers who are interested in determining the effectiveness of tutoring programs, we need to do a better job of designing research protocols, choosing appropriate outcome measures, ensuring rigor, and qualifying our claims. As a case in point, our own study would have been stronger if we had addressed the limitations discussed above prior to implementing the program.

Even given these concerns, though, the notion that tutoring can have a positive effect on children is not in question. What *is* in question is the degree to which researchers have underestimated the

complexity of research necessary in this area. And as a result, reports proclaiming the significant impact of tutoring should be read critically and with caution.

## References

- Adler, M. (1999). The America Reads Challenge: An analysis of college students' tutoring. Ann Arbor: Center for the Improvement of Early Reading Achievement, University of Michigan.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37(2), 122-147.
- Bandura, A. (1986). Social Foundations of Thought and Action. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. Journal of Personality and Social Psychology, 45, 1017-1028.
- Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. Organizational Behavior and Human Decision Processes, 38, 92-133.
- Bogdan, R. C., & Biklen, S. K. (1998). Qualitative Research for Education. Boston: Allyn and Bacon.
- Book Buddies (2000). Book Buddies. Retrieved April 15, 2003 from <http://curry.edschool.virginia.edu/reading/projects/bookbuddies/home.html>
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. Journal of Social Psychology, 139, 353-363.
- Butler, S.R. (1991). Reading program--Remedial, integrated and innovative. Annals of Dyslexia, 41, 119-127.
- Cavanaugh, M. P., Johnston, D., Kitay, N., & Yuratovac (1997). Learning from "the big kids." Educational Leadership, 54(7), 53-55.
- Clay, M.M. (1993). Reading Recovery: A guidebook for teachers in training. Portsmouth, NH: Heinemann.
- Cohen, P., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. American Educational Research Journal, 19, 237-248.

- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (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.
- Fitzgerald, J. (2001). Tutoring programs for struggling readers: The America Reads Challenge. (140-158). New York: Guilford Press.
- Graves, A. W. (1986). Effects of direct instruction and metacomprehension training on finding main ideas. Learning Disabilities Research *2*, 90-100.
- Hedrick, W. B. (1999). Pre-service teachers tutoring 3rd, 4th, and 5th graders one-on-one within the school setting. Reading Research and Instruction, *38*(3), 211-219.
- Juel, C. (1993). What makes literacy tutoring effective? Reading Research Quarterly, *31*, 268-289.
- Kerbow, D. (1996). Patterns of urban student mobility and local school reform: A technical report. Center for Research on the Education of Students Placed At Risk. Retrieved August 23, 2002 from <http://www.csos.jhu.edu/crespar/Reports/report05entire.html>
- Locke, E., & Latham, G. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice Hall.
- Locke, E., & Latham, G. (1994). Goal setting theory. In H. O'Neil & M. Drillings (Eds.), Motivation: Theory and research (13-29). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Malone, D., Jones, B., & Stallings, D. T. (2001). Perspective transformation: Effects of a service-learning tutoring experience on prospective teachers. Teacher Education Quarterly, *29*, 61-81.
- McGrady, D.S. (1984). The effect of programmed tutoring upon the reading comprehension of fourth-grade students enrolled in a Chapter 1 reading program. Paper presented at the annual meeting of the Association for Educational Communication and Technology, Dallas, TX. (ERIC Document Reproduction Service No. ED 243-430)
- McKenna, M. J. (2000). Students, agencies, and faculty sign on for academic service-learning: A natural partnership for teacher education. Teacher Education Quarterly, *27*(3), 89-102.

- Morris, D. (1999). The Howard Street tutoring manual: Teaching at-risk readers in the primary grades. New York: Guilford Press.
- Morrow, L.M., & Woo, D. G. (2001). Tutoring programs for struggling readers: The America Reads Challenge. New York: Guilford Press
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82, 32-40.
- Prosser, T. M., & Levesque, J. A. (1997). Supporting literacy through service learning. The Reading Teacher, 51(1), 32-38.
- Ramey, M. (1991). Compensatory education sustained gains from Spring 1988 to Spring 1990. Seattle Public Schools. (ERIC Document Reproduction Service No. ED 243-430)
- Reisner, E.R., Petry, C.A., & Armitage, M.(1990). A review of programs involving college students as tutors or mentors in Grades K-12. Washington, D.C.: U.S. Department of Education.
- Sexton, T., & Tuckman, B. (1991). Self-beliefs and behavior: The role of self-efficacy and outcome expectation over time. Personality and Individual Differences, 12, 725-736.
- Silberman, T. (2001, May22). High passing rate on math tests a mistake, officials say. The News & Observer, pp. 1A, 10A.
- Success for all Foundation (2000). Research. Retrieved April 15, 2003, from <http://www.successforall.net/resource/researchpub.htm>
- Topping, K., & Whiteley, M. (1990). Participant evaluation of parent-tutored and peer-tutored projects in reading. Educational Research, 32(1), 14-32.
- Wallach, M.A. & Wallach, L. (1976). Teaching all children to read. Chicago: University of Chicago Press.
- Wasik, B.A., & Slavin, R.E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. Reading Research Quarterly, 28, 178-200.

- Wasik, B. A. (1997, December). Volunteer tutoring programs: Do we know what works? Phi Delta Kappan, 282-287.
- Wasik, B. A. (1998a). Using volunteers as reading tutors: Guidelines for successful practices. The Reading Teacher, 51(7), 562-570.
- Wasik, B.A. (1998b). Volunteer tutoring programs in reading: A review. Reading Research Quarterly, 33, 266-292.
- Zimmerman, B. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), Self-efficacy in changing societies (pp. 202-231). New York: Cambridge University Press.
- Zimmerman, B., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. American Educational Research Journal, 31, 845-862.

**Table 1. Yearly Tutee and Comparison Group EOG Test Score<sup>a</sup> Growth<sup>b</sup>**

	1998-1999			1999-2000			2000-2001 <sup>d</sup>		
	Tutee Growth	Comp. Growth		Tutee Growth	Comp. Growth		Tutee Growth	Comp. Growth	
<b>4<sup>th</sup> Grade</b>	N=10 <sup>c</sup>			N=12			N=24		
Reading	3.00	1.10	p=0.208	3.30	4.00	p=0.371	2.59	2.41	p=0.455
SD	7.79	6.01		5.96	5.12		6.18	4.43	
Math	13.70	10.20	p=0.152	<u>11.00</u>	7.40	<b>p=0.055</b>	---	---	
SD	7.89	7.58		6.01	9.01				
<b>5<sup>th</sup> Grade</b>	N=4			N=16 (N=15 for Math)			N=43		
Reading	4.25	<u>6.00</u>	<b>p=0.094</b>	7.60	8.81	p=0.263	7.74	7.67	p=0.469
SD	3.10	2.45		3.67	5.59		6.51	4.06	
Math	12.75	12.75	p=0.500	10.60	11.27	p=0.381	---	---	
SD	7.54	5.91		8.59	4.77				

<sup>a</sup> Scaled scores. The North Carolina Department of Public Instruction calculates expected yearly growth at the school level for each grade and subject based on the previous year's school-level and state average scores. The expected growth number varies by school, subject, grade, and year. For example, expected growth was between 4.1 and 5.9 points for 4<sup>th</sup> and 5<sup>th</sup> reading for students at our participating schools in 1998-1999 and between 7.8 and 8.1 points for math.

<sup>b</sup> Differences in growth were tested using a paired-differences t-test, which does not assume independence.

<sup>c</sup> Several values for N are very low throughout this table for two reasons: 1) Each year's data is disaggregated by grade; and 2) Many tutees were eliminated from the study each year because of incomplete test data or incomplete test data for matched comparison students. The one comparison that exhibits statistical significance - 5<sup>th</sup> grade reading scores for School Year 1998-1999 - is itself of little value in evaluating PFS, given the extremely low N (5). See Discussion section, above.

<sup>d</sup> The format and scoring for the math component of the EOG changed for the 2000-2001 school year, making analysis of changes in math scores for this year impossible.

**Table 2. Tutee and Comparison Group EOG Growth, 1999-2001, Combined<sup>a</sup>**

<b>1998-2001</b>			
	Tutee Growth	Comp. Growth	
<b>4<sup>th</sup> Grade</b>	N=44 (N=22 for Math <sup>b</sup> )		
Reading	2.864	2.545	p=0.387
SD	6.363	4.990	
Math	<u>12.136</u>	8.682	<b><i>p=0.032</i></b>
SD	6.910	8.317	
<b>5<sup>th</sup> Grade</b>	N=62 (N=19 for Math <sup>b</sup> )		
Reading	7.484	7.855	p=0.315
SD	5.745	4.424	
Math	11.053	11.579	p=0.401
SD	8.229	4.891	

<sup>a</sup> Differences in growth were tested using a paired-differences t-test, which does not assume independence.

<sup>b</sup> The format and scoring for the math component of the EOG changed for the 2000-2001 school year, making analysis of changes in math scores for this year impossible. Therefore, the number of tutee-comparison scores analyzed is much smaller for math than for reading.