Relative Efficacy of Reciprocal and Nonreciprocal Peer Tutoring for Students At-Risk for Academic Failure

Keri F. Menesses and Frank M. Gresham
Louisiana State University

This study directly compared the academic gains of reciprocal peer tutoring, nonreciprocal peer tutoring, and a waiting-list control group. Participants included 59 elementary students from second-, third-, and fourth-grade classrooms who performed below average on curriculum-based measurement (CBM) math probes. Students involved in peer tutoring were trained to tutor basic math facts using a constant time delay procedure. Results indicated that the two types of peer tutoring produced comparable gains in basic math facts. Furthermore, both types of peer tutoring produced substantially larger academic gains than the waiting-list control group, demonstrating at-risk students can successfully tutor each other.

Keywords: peer tutoring, reciprocal peer tutoring, math intervention, curriculum-based measurement

Peer tutoring is the process by which a student helps one or more students learn a skill or concept (Thomas, 1993). A substantial advantage of peer tutoring is the decreased amount of teacher responsibility in implementing an intervention. Because students are the primary change agents, peer tutoring is an efficient method for providing individualized instruction to many students simultaneously. Previous research has documented the benefits of peer tutoring, including the acquisition of academic skills (Cohen, Kulik, & Kulik, 1982), enhancement of peer relations (Greenwood, Carta, Kamps, & Hall, 1988), and improved classroom behavior (Fuchs, Fuchs, Phillips, Hamlett, & Karns, 1995). Additional benefits for tutees include increased academic engaged time, increased opportunities to respond, (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986), immediate performance feedback (Topping, 2005), and continuous progress monitoring (Greenwood et al., 1988). Additional benefits for tutors include positive attitudes toward subject matter (Cohen et al., 1982), improved self-esteem (Cardenas, Harris, del Refugio, & Supik, 1991), and improved attitude toward school (Cardenas et al., 1991).

Fantuzzo and colleagues developed a procedure known as Reciprocal Peer Tutoring (RPT), in which students alternate between roles of tutor and tutee so that both students have access to all of the advantages of peer tutoring (Fantuzzo, King, & Heller, 1992). The tutoring pair works together to prompt, monitor, and evaluate each other while learning a specific academic skill (Fantuzzo, Polite, & Grayson, 1990). Greenwood and colleagues utilized a similar method in their Classwide Peer Tutoring (CWPT) approach, in which all students within a classroom participate in RPT simultaneously (Delquadri, Greenwood, Stretton, & Hall, 1983). RPT and CWPT have been effective with elementary students ranging from first to sixth grade (Barley et al., 2002) across various academic subjects, including mathematics, reading, spelling, vocabulary, social studies, and science (Greenwood et al., 1991).

Robinson and colleagues conducted a meta-analysis of peer tutoring programs that primarily focused on mathematics, but also incorporated peer tutoring studies involving spelling, reading, and science when research on math was scarce for a particular outcome (Robinson, Schofield, & Steers-Wentzell, 2005). These authors reviewed 28 studies published after 1988 and found peer tutoring had positive effects on a variety of math skills, including computation,
conceptual understanding, and problem-solving skills. Effect sizes of academic achievement ranged from $d = .48$ to $1.17$ for tutees with the binomial effect size display (BESD; Rosenthal, Rosnow, & Rubin, 2000) ranging from 62% to 76% improvement rate for the treatment group. Effect sizes ranged from $d = .48$ to $1.37$ for tutors, with the success rate for the treatment group ranging from 62% to 79%.

The Robinson et al. review also examined how specific features of tutoring programs influenced outcomes (Robinson et al., 2005). The authors discovered shorter tutoring programs resulted in larger effect sizes ($d = 1.01$, BESD = 73%) than longer programs ($d = .38$, BESD = 60%), and proposed the “newness” of brief programs as a possible explanation. The importance of tutor training was underscored, as tutees showed more academic improvement when their tutors had been trained as opposed to having nontrained tutors. Rewards that are contingent on performance also influenced results, with the combination of tutoring and rewards producing better academic performance than either condition alone. When comparing outcomes of reciprocal and nonreciprocal programs, the authors concluded that RPT resulted in larger effect sizes, ranging from medium ($d = .5$) to large ($d = .96$) and a BESD ranging from 62% to 72%, whereas nonreciprocal tutoring showed small ($d = .2$) to medium ($d = .48$) effect sizes and a BESD ranging from 55% to 62%.

Most research on peer tutoring involves a higher-achieving peer as the tutor. However, Telecsan, Slaton, and Stevens (1999) and Hughes and Fredrick (2006) integrated constant time delay procedures into peer tutoring to enable students with learning disabilities to tutor other students with learning disabilities. Both of their procedures proved extremely effective in teaching spelling and vocabulary facts and demonstrated a higher-performing peer is not necessary to produce positive results.

Purpose of the Present Study

Although the Robinson et al. (2005) review reported larger effect sizes for reciprocal programs than nonreciprocal programs, these two intervention strategies have not been directly compared within the same experiment. Also, Robinson et al. (2005) note that many of the reciprocal programs included a reward contingency, whereas the nonreciprocal programs often did not, which possibly skewed the results. As such, the major purpose of the current investigation was to directly compare reciprocal and nonreciprocal peer tutoring (NPT) to determine which program results in greater academic gains. It was hypothesized that RPT would produce larger academic gains relative to NPT.

Another purpose of this study was to add to the literature on the effectiveness of peer tutoring as a remedial intervention when dyads are comprised of only at-risk students rather than high- and low-performing students. There is a growing concern about students’ underachievement in the domain of mathematics (Smith, 2005); however, research on remediating low math achievement is limited relative to that of reading (Mazzocco & Myers, 2003). Therefore, basic math facts, which are prerequisite to complex, higher-order concepts, were chosen as tutoring material. Math facts have not yet been investigated in the context of peer tutoring using a constant time delay procedure, which is an effective method for teaching basic academic facts (Hughes, Fredrick, & Keel, 2002). Also, CTD with flashcards enables at-risk students to reciprocally tutor each other, because correct answers can be provided without understanding how to reach an answer. It was hypothesized that both NPT and RPT would increase students’ mathematics performance to an average range, thereby demonstrating at-risk students are capable of remediating other at-risk students’ performances.

Method

Setting, Screening, and Participants

This study was conducted at an elementary school located in a large district in the southeastern United States. The school serves approximately 620 students in Pre-K through fifth grade; 53% of students are girls, 47% are boys; 77% of students are African American, 19% are Caucasian, 2% are Hispanic, 2% are Asian, and 83% receive free or reduced lunch. Only seven general education classrooms (two second grade, three third grade, and two fourth grade) were nominated for participation because other classrooms at the school were involved in a program designed to improve classroom be-
behavior, and the school's administrative staff did not want them involved in the current study also. The seven classrooms contained 143 students; however, two gates of screening (described below) resulted in 59 students (32 boys, 27 girls; 76% African American, 12% Caucasian, 9% Asian American, 3% Hispanic) serving as participants.

The seven nominated teachers agreed to have their classrooms screened for participants. All screening procedures were conducted by the primary investigator and took place during the school day in late January. The seven classrooms, which had an average of 22 students, were administered computer-generated CBM math probes (Math Computation Probes, available from www.interventioncentral.org) with 60 addition, subtraction, multiplication, or division problems during the first gate of screening, a classwide screening. The types of problems distributed to each classroom were determined during a teacher interview in which each teacher reported the two types of math problems that should have been mastered most recently by their students. Therefore, this procedure screened for students who did not acquire or maintain facts that were no longer being directly taught by the teacher.

Math probes were administered and scored according to the Aimsweb scoring instructions for CBM math probes (Math-CBM, available from www.aimsweb.com). A student in second or third grade scoring 0 through 19 digits correct within two minutes is considered at a "frustrational" level as determined by national benchmarks (Deno & Mirkin, 1977), whereas a fourth-grade student who produces 0 through 39 digits correct falls in the frustrational range. According to these benchmarks, 67 students fell within the frustrational range and entered the second gate of screening, which was administered individually. Written parental consent and student assent were obtained before the individual screening procedures began, and students had the option of dropping out of the study at any point.

To meet criteria for inclusion, students were required to name numbers 0 to 18 for addition and subtraction tutoring and 0 to 81 for multiplication and division tutoring. Also, participants had to read aloud 10 examples of completed mathematical equations that were to be included in their tutoring program (e.g., $7 \times 5 = 35$). These screening procedures ensured students were at an appropriate skill level to participate. Students were then given two additional probes with the same type of problem on which they had scored in the frustrational range during the first gate of screening. The median of these three nonidentical probes served as the preintervention score for each participant. If that score fell above the frustrational range, the student was no longer eligible for participation. Finally, students could not be receiving any remedial instruction in mathematics. Sixty-two students met criteria and were included in the study. Throughout the course of the study, no students dropped out, and 3 students moved to different schools. As a result, a total of 59 participants were included in all data analyses. All of these students received general education instruction; however, 6 students received small group, pull-out instruction in reading.

All individual screenings, tutor training sessions, and peer tutoring sessions took place in a quiet hall of the school containing tables and chairs with either the primary investigator and/or a research assistant present. Times for these sessions were determined by each teacher based on the daily classroom schedule.

**Dependent Measures**

CBM probes were used to measure each participant’s fluency level in mathematics by scoring the number of digits correctly produced within two minutes. CBM was chosen because of extensive research establishing its psychometric properties, as well as its ability to monitor progress (Foegen, Jiban, & Deno, 2007; Thurber, Shinn, & Smolkowski, 2004). The probes were administered immediately before and after the intervention to determine the acquisition of math facts, as well as 3 weeks after termination of the intervention to assess maintenance of facts. As previously discussed, the preintervention score was the median of three different probes, which removed the possibility of practice effects across the three probes that were administered within several days. On the other hand, the probe used to determine postintervention and follow-up scores was identical to ensure equal difficulty level and score compatibility across the two times included in data analyses. Therefore, four different worksheets were generated for all participants. To eliminate
experimental bias, the primary investigator and a research assistant, who both scored each CBM probe, were blind to condition during scoring.

**Experimental Design and Power Analysis**

Two $4 \times 2$ (Group $\times$ Time) completely randomized factorial mixed-model experimental designs were used to determine the effectiveness of each intervention in increasing basic math skills (Kirk, 1995). Each design consisted of a between-subjects factor (Group) with four levels (reciprocal tutoring, tutees in nonreciprocal tutoring, tutors in nonreciprocal tutoring, and control) and a within-subjects factor (Time) with two levels (postintervention and follow-up, with preintervention as a covariate). The unit of randomization was classrooms, which were randomly assigned to one of the three conditions using a table of random numbers: RPT ($n = 15$), NPT ($n = 28$; 14 tutees, 14 tutors), and control ($n = 16$).

A power analysis was computed to determine an adequate sample size for the statistical analyses used in the current study. Sample size was determined using G*Power Version 3.0.5 (Faul, 2006). For an alpha level of .05 and a standardized effect size of $f = .25$, a total of 48 participants was required to reach a power level of .80.

**Procedure**

Three types of instruction (NPT, RPT, and standard classroom instruction) were randomly assigned to classrooms. All teachers were asked to avoid the use of peer tutoring during math instruction for the duration of this study. The NPT condition consisted of one-way peer tutoring in which one student was always the tutor and the other student was always the tutee. In RPT, students switched roles between tutee and tutor within the same session. Finally, students in the control group received conventional classroom instruction and had no knowledge of peer tutoring until the study was complete, at which point they participated in peer tutoring.

The three different instructional conditions were randomly assigned to classrooms so that all eligible students within a classroom received the same treatment in order to control for preference effects. Random assignment resulted in the control condition being assigned to one second-grade classroom and one third-grade classroom; NPT to one second-grade, one third-grade, and one fourth-grade classroom; and RPT to one third-grade and one fourth-grade classroom. Each student in the NPT and RPT conditions was randomly assigned to another eligible student within the classroom to form a dyad.

**Tutor Training**

Tutors were individually trained by the primary investigator using a “tell,” “show,” “do” approach modified from the Telecsan et al., (1999) study. Following an explanation of constant time delay (CTD) and peer tutoring, the process of presenting index cards using CTD was demonstrated by the primary investigator. Tutors then participated in role-play with the primary investigator who acted as a tutee. Training trials were identical to the 3-min tutoring sessions that took place during the intervention. Tutors' performance was measured using a treatment integrity checklist, which listed the six required tutoring behaviors: (a) begin when timer is set, (b) present each card for 3 seconds, (c) provide correction when necessary, (d) provide praise, (e) shuffle cards, and (f) continue until timer sounds. Positive and corrective feedback was given throughout all training trials, and tutors were trained until they reached 100% accuracy on three consecutive training trials.

Tutors were also trained to measure independent performance of the tutees in a progress monitoring session that took place immediately after each practice session. Progress monitoring sessions required four behaviors: (a) present each card for 3 seconds, (b) no talking, (c) sort cards into correct and incorrect piles, and (d) count and record number of correct cards with tutee. Tutors were again trained until they reached 100% accuracy on three consecutive training trials.

**Tutoring Procedure**

Students assigned to the tutoring conditions (NPT and RPT) participated in an average of three tutoring sessions per week until they reached the session requirement of 15 sessions total. If one student in a dyad was absent, his or
her partner also did not participate that day. Each CTD tutoring session was 3 min in duration and included 10 different math problems printed on 3” × 5” index cards. The tutoring program was explained to the students as a game, with the tutor described as the “coach” and the tutee as the “player,” as adopted from CWPT (Delquadri et al., 1983).

Two to three pairs (four to six students) were removed from their classroom at the time designated by the teacher and brought to a quiet hallway. Each pair had a pocket folder stored in their classroom that contained all the materials needed for tutoring. Once all pairs were ready to begin tutoring, the primary investigator or research assistant started the timer and said “Begin.” Each coach presented 10 cards to each player using 3-s CTD. If the player provided the correct response within 3 seconds, the coach praised the player and presented the next card. If the player answered incorrectly or did not provide an answer within 3 seconds, the coach said the correct answer, which the player had to repeat before being presented with the next card. Once all 10 cards were presented, the coach shuffled them and continued the process until the timer sounded.

Progress-Monitoring Procedure

Immediately following the tutoring procedure, coaches presented the 10 cards once again using 3-s CTD but without providing any verbal feedback. Coaches sorted correct and incorrect answers on a poster board that had a green square and red square. Once all 10 cards were sorted, the player and coach counted the correct amount of responses on the green square, which was then recorded on the progress monitoring chart kept in the pair’s folder. Once the player provided 10 correct responses on two consecutive progress monitoring sessions, a new set of 10 different cards was introduced in the next session.

Interdependent Contingencies

Based on RPT and CWPT, both members of all dyads were responsible for working toward their goal and earning points for their “team.” Because the accountability of achieving the goal does not rest with just one student, group contingencies have been shown to increase student cooperation and performance (Fantuzzo et al., 1992). During each session, every pair of students had the opportunity to earn four points. In the NPT condition, coaches were able to earn two points by tutoring correctly and players could earn two points by beating their previous score on their progress-monitoring chart. In RPT, the student who coached first could earn one point for tutoring correctly and the first player could earn one point for exceeding his or her previous score. When they switched roles, they each had the opportunity to earn another point, summing to four possible points. Each pair tracked their points on a chart, and both members were allowed to choose a reward from a box containing candy, pencils, stickers, and small toys for every five points they earned.

Treatment Integrity

Coaches monitored their own treatment integrity each day using the same treatment integrity checklist previously discussed in the tutor training section. The primary investigator and research assistants closely monitored the integrity of all tutors and provided corrective prompts whenever necessary. Coaches placed a check by the steps they completed correctly and did not check any step for which they required a prompt; the primary investigator and research assistants observed each coach as he or she filled out the checklist to ensure it was completed accurately. Booster training sessions were provided when tutors had to be prompted more than once during the procedure, which only occurred during two sessions. Data from these sessions were not included in any analyses because integrity fell below 90%, and maintaining high procedural integrity in all tutoring conditions was crucial in order to legitimately compare their outcomes. Tutors who fell below 90% integrity were retrained and required to correctly complete three consecutive tutoring sessions with the primary investigator before tutoring again. Out of the 660 integrity checklists completed, 578 (88%) had 100% integrity and 82 (12%) had 90% integrity, for a total of 98.8% integrity across conditions. The RPT group had an average of 98.6% integrity, whereas the NPT group had an average of 99.1% integrity.
Results

An alpha level of .05 was used for all statistical tests. A 4 × 2 mixed-model analysis of covariance (ANCOVA) was conducted using the preintervention score as the covariate. The two independent variables were Group and Time, and the dependent variable was the number of digits correct within two minutes on a CBM math probe. The main effect for Group was significant, $F(3, 54) = 11.41, p < .001$, partial $\eta^2 = .39$. A Bonferroni post hoc test revealed reciprocal tutoring produced higher scores ($M = 34.88$) than the control group ($M = 19.72$) and the tutees produced higher scores ($M = 32.90$) than the control group ($M = 19.72$). The main effect of Time was not significant, $F(1, 54) = .04, p = .842$, partial $\eta^2 = .001$, meaning there was no significant difference between postintervention and follow-up scores. The interaction between Group and Time was also not significant, $F(3, 54) = .75, p = .526$.

Table 1 displays group means and SDs for the CBM math probes. The adjusted means of both the postscores and follow-up scores were ordered as expected across the four instructional groups: the reciprocal group had the largest mean, tutees had a slightly smaller mean, tutors had an even smaller mean, and the control group had the smallest mean. The Bryant-Paulson post hoc procedure (Bryant & Paulson, 1976) was conducted to evaluate differences among the postintervention means. Both the reciprocal and tutee groups had significantly larger postscore means than the control group. The Bryant-Paulson post hoc procedure indicated both the reciprocal and tutee groups had significantly larger follow-up means than the control group, and the reciprocal group also had a larger follow-up mean than the tutor group. Effect sizes were computed for each tutoring group using Cohen’s (1988) $d$, which was then converted into a binomial effect size display (BESD; see Table 2).

Paired samples $t$ tests were conducted to determine change in CBM scores over time for each group. The control group showed no significant change between pre and postscores, $t(15) = -1.58, p = .136$, no significant difference between pre and follow-up scores, $t(15) = -1.12, p = .282$, and no significant difference between post and follow-up scores, $t(15) = .06, p = .953$. The reciprocal group showed a significant difference between pre and postscores, $t(14) = 7.17, p < .001$, a significant difference between pre and follow-up scores, $t(14) = 5.51, p < .001$, and no significant change between post and follow-up scores, $t(14) = .417, p = .683$. The tutee group produced a significant change between pre and postscores, $t(13) = -5.06, p < .001$, as well as between pre and follow-up scores, $t(13) = -4.80, p < .001$, and no significant change between pre and follow-up scores, $t(13) = -4.80, p < .001$, and no significant change between pre and follow-up scores, $t(13) = -4.80, p < .001$, and no significant change between pre and follow-up scores, $t(13) = -4.80, p < .001$, and no significant change between pre and follow-up scores, $t(13) = -4.80, p < .001$.

Table 1
Means, SDs, Adjusted Means, and Ranges for Curriculum-Based Measurement (CBM) Scores for Group Across Time

<table>
<thead>
<tr>
<th>Group</th>
<th>Reciprocal ($N = 15$)</th>
<th>Tutees ($N = 14$)</th>
<th>Tutors ($N = 14$)</th>
<th>Control ($N = 16$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>19.20</td>
<td>17.21</td>
<td>17.00</td>
<td>15.81</td>
</tr>
<tr>
<td>(SD)</td>
<td>(9.34)</td>
<td>(8.79)</td>
<td>(8.15)</td>
<td>(4.07)</td>
</tr>
<tr>
<td>Range</td>
<td>24</td>
<td>23</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>Post</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>37.33</td>
<td>34.07</td>
<td>25.71</td>
<td>17.50</td>
</tr>
<tr>
<td>(SD)</td>
<td>(16.33)</td>
<td>(19.05)</td>
<td>(12.94)</td>
<td>(6.96)</td>
</tr>
<tr>
<td>Adjusted $M$</td>
<td>34.04</td>
<td>34.18</td>
<td>26.16</td>
<td>19.76</td>
</tr>
<tr>
<td>Range</td>
<td>54</td>
<td>62</td>
<td>37</td>
<td>23</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>38.27</td>
<td>31.50</td>
<td>25.43</td>
<td>17.44</td>
</tr>
<tr>
<td>(SD)</td>
<td>(20.01)</td>
<td>(17.73)</td>
<td>(11.49)</td>
<td>(7.19)</td>
</tr>
<tr>
<td>Adjusted $M$</td>
<td>35.35</td>
<td>31.61</td>
<td>25.87</td>
<td>19.69</td>
</tr>
<tr>
<td>Range</td>
<td>58</td>
<td>59</td>
<td>42</td>
<td>24</td>
</tr>
</tbody>
</table>
whereas there was no significant change between post and follow-up scores, \(t(13) = 1.85, p = .087\).

The tutor group showed a significant change between pre and postscores, \(t(13) = -4.10, p = .001\), a significant difference between pre and follow-up scores, \(t(13) = -4.38, p = .001\), and no significant change between post and follow-up scores, \(t(13) = .157, p = .877\).

When individual scores were compared with national benchmarks, 12 of the 15 reciprocal students and 10 of the 14 tutee students fell within the “instructional” range rather than the “frustrational” range at both post and follow-up scores. In contrast, only 6 of the 14 tutors and 5 of the 16 control students reached the instructional range at post and follow-up measurements.

Discussion

This study was designed to directly compare RPT and NPT to determine which produces larger academic gains. When examining the differences between groups at post and follow-up, the group means were ordered as predicted. Although reciprocal students had slightly higher scores than tutees, the two groups of students being tutored did not significantly differ in their academic gains. In contrast to the Robinson et al. (2005) review, which reported larger effect sizes for RPT, the findings of this investigation suggest the two different forms of peer tutoring are comparable in effectiveness.

As hypothesized, all students receiving the intervention (reciprocal and tutee groups) produced a significant increase in math performance, which contributes to the empirical data establishing the effectiveness of peer tutoring (Barley et al., 2002; Fantuzzo et al., 1992; Greenwood et al., 1991). It is important to note that nonreciprocal tutors produced significantly more digits correct as compared with their precourse. Previous researchers (Britz et al., 1989; Cohen et al., 1982) have reported similar gains in tutors and have attributed the results to practice effects; however, this finding was unanticipated in the current study because tutors were only able to see the answers on the back of the flashcards and as a result, were never exposed to math problems. Perhaps tutoring another student led tutors to develop more positive attitudes toward math (Cohen et al., 1982) or improved self-estees (Cardenas et al., 1991), which could result in more effortful performances.

Although the reciprocal, tutee, and tutor groups all showed significant increases in their academic scores, the control group did not. It is important to note that control students only received conventional teacher-led instruction, which did not focus on the math facts they had not previously acquired. As a result, only minimal gains in basic facts were achieved across four school months (mean increase of 1.69 digits and 1.63 digits at post and follow-up measurements, respectively). Hence, a considerable amount of the academic increases produced by all other participants can be attributed to their involvement in peer tutoring rather than to the usual math instruction they received in the classroom.

When evaluating the maintenance of facts over time, results indicate that there were no significant differences between post and follow-up scores in any of the groups. Thus, students did not improve significantly within the 3 weeks after the intervention ended, nor did they show a significant decrease in maintenance of the facts.

Although all students involved in peer tutoring outperformed the control group, pairwise comparisons indicated that only the reciprocal and the tutee groups produced significantly more gains than the control group. In other words, although tutors had a significant increase from pre to postscores, their gains were not significantly different than those of the control group. These findings somewhat support the

<table>
<thead>
<tr>
<th>CBM</th>
<th>Effect size expressed as Cohen’s (d)</th>
<th>Success rate as percentage of subjects showing improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td>Post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocal</td>
<td>1.58</td>
<td>81</td>
</tr>
<tr>
<td>Tutee</td>
<td>1.16</td>
<td>75</td>
</tr>
<tr>
<td>Tutor</td>
<td>.79</td>
<td>69</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocal</td>
<td>1.39</td>
<td>79</td>
</tr>
<tr>
<td>Tutee</td>
<td>1.05</td>
<td>73</td>
</tr>
<tr>
<td>Tutor</td>
<td>.83</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 2: Overall Effect Size Display for Postintervention and Follow-Up Scores on Curriculum-Based Measurement (CBM) Probes
hypothesis that nonreciprocal tutors would not show substantial increases in math scores. The results imply reciprocal tutoring successfully remediates more at-risk students concurrently as opposed to nonreciprocal tutoring, which only produced significant gains in tutees.

When considering the capability of peer tutoring in remediating at-risk students with other at-risk peers as math tutors, the results indicate an efficacious intervention, which was previously found with spelling and vocabulary tutoring (Hughes & Fredrick, 2006; Telecsan et al., 1999). When individuals’ post and follow-up scores were compared with national benchmarks, the majority of students in the reciprocal and tutee groups fell within the “instructional” range rather than the “frustrational” range, whereas only a handful of tutors and controls reached the instructional range. It is evident that at-risk students being tutored by at-risk peers are capable of reaching an average range of math performance.

**Implications for Practice**

These findings inform practice in several ways. First, because no significant differences were detected between reciprocal and nonreciprocal tutoring programs, it can be inferred that either program can be applied in school settings. However, it is important to note that although they appear to be equally effective in teaching basic facts, reciprocal tutoring is more efficient by teaching twice the number of students simultaneously as nonreciprocal tutoring.

Second, this study contributes to the effectiveness of using low-achieving students to tutor other low-achieving peers (Hughes & Fredrick, 2006; Telecsan et al., 1999). Incorporating constant time delay procedures into peer tutoring allows students to correct others’ answers without knowing the answers on their own. This results in a more equal pairing and may reduce any frustration or ridicule that may take place with a higher-achieving student tutoring an at-risk peer.

Lastly, it is important to note the relatively low frequency and duration of this intervention: 15 sessions at 3 min each equates to a total of 45 minutes of tutoring. Therefore, an intervention that does not require significant time or resources is quite effective at increasing students’ accuracy and fluency of basic math facts. In the case of students who did not respond to this intervention, perhaps a more intense peer tutoring intervention (more frequent and/or longer duration) would be more likely to result in remediation.

**Limitations and Future Research Directions**

There are several limitations in this study that warrant discussion. First, although classrooms were randomly assigned to groups, this did not result in random assignment of students to groups. As previously described, all students within a classroom participated in the same condition to control for preference effects. Although this may have prevented attrition or lack of cooperation, as many students in the nonreciprocal tutoring condition asked to switch roles multiple times and may have been very disappointed to know there was a reciprocal tutoring condition, it came at a cost in that assignment was only semirandom because of classrooms being preexisting groups. It is important to note, however, that multiple classrooms were randomly assigned to each group to try to account for any individual classroom differences. Follow-up studies can address this issue by randomly assigning students to conditions and informing them that they can participate in another condition once the experimental intervention is completed.

The use of only one measure of math skills is a limitation of the current study. Although significant differences were found, it is possible that other measures of academic performance may reflect more differences between groups or on the other hand, may indicate differences are not as large as those found in the current experiment. Another limitation is the dissimilarity between the tutoring material (flashcards) and the academic dependent measure (CBM probes). This difference assumes that students could generalize the individual facts on the tutoring flashcards to completing multiple problems on a probe worksheet; however, some students could not answer the same problems on the probe that they had mastered with the flashcards. Furthermore, after mastering multiple sets of flashcards, some students still performed below average on the post and follow-up CBM probes. Hence, the generalization of peer-tutored facts requires further investigation.
Another limitation concerns elements of the procedure that were arbitrarily chosen. Selecting a session length of 3 minutes, the number of flashcards in a set as 10, and the criteria for achieving mastery as two consecutive scores of 10 had no scientific basis. The possibility of manipulating these parameters and measuring any effects produced by variations is open to future research. Lastly, generalization of this study’s results is limited; the findings can only be applied to basic math facts taught to its efficiency of constant time delay. Similar comparisons of RPT and NPT need to be conducted to determine whether there are differences within other academic domains and with other instructional procedures.

Conclusion

The current study demonstrated that RPT and NPT were comparable in their efficacy in increasing low-performing students’ accuracy and fluency of basic math facts. However, the larger amount of students reciprocal tutoring is able to remediate simultaneously speaks to its efficiency compared to nonreciprocal tutoring. Furthermore, below-average achievers were able to accurately tutor each other to the extent that they both achieved significant gains. Hence, there are situations in which at-risk students do not need to be paired with average or above-average students to produce increases in performance. Moreover, the outcomes of this study demonstrate there are feasible peer-mediated interventions that may prevent future academic difficulties. These findings advocate for more widespread use of students, even those identified as at-risk, as primary delivery agents of academic interventions.

References


