The Effects of Peer Tutoring on Academic Performance of Students With Disabilities in Grades 6 Through 12: A Synthesis of the Literature

Nathern S. A. Okilwa and Liz Shelby

Remedial and Special Education 2010 31: 450 originally published online 5 January 2010
DOI: 10.1177/0741932509355991

The online version of this article can be found at:
http://rse.sagepub.com/content/31/6/450

Published by:
Hammill Institute on Disabilities

and

http://www.sagepublications.com

Additional services and information for Remedial and Special Education can be found at:

Email Alerts: http://rse.sagepub.com/cgi/alerts
Subscriptions: http://rse.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://rse.sagepub.com/content/31/6/450.refs.html

>> Version of Record - Nov 5, 2010
OnlineFirst Version of Record - Jan 5, 2010

What is This?
The Effects of Peer Tutoring on Academic Performance of Students With Disabilities in Grades 6 Through 12: A Synthesis of the Literature

Nathern S.A. Okilwa¹ and Liz Shelby¹

Abstract
This synthesis examined the effects of peer tutoring on academic performance of students with disabilities in Grades 6 through 12. Twelve studies met all the criteria for this synthesis: (a) original studies, (b) published in peer-reviewed journals between 1997 and 2007, (c) investigated peer tutoring in special education students in Grades 6 through 12, and (d) implemented peer tutoring as an intervention and measured the effect on the academic outcomes of students with disabilities. Findings revealed that peer tutoring has a positive academic effect on students with disabilities in Grades 6 through 12, regardless of disability type. Peer tutoring was reported as effective for special education students in both general education and special education settings. Peer tutoring implemented across subject areas also showed positive academic effects. Each of the 12 studies implemented peer tutoring in at least one content area (e.g., language arts, math, science, and social studies).

Keywords
peer tutoring, learning, support, disability, special education, academic achievement

During the political environment of general education school reform, the reauthorization of the Individuals with Disabilities Education Act (IDEA) of 1997 established Congress’s intent that most students in special education should be receiving grade-level instruction in the statewide curriculum and be included in state assessments at their chronological grade level. One of the main tenets of the No Child Left Behind (NCLB) legislation passed in 2001 was the need for scientifically based instructional practices to address the needs of all students, including those not functioning on grade level. In addition, the President’s Commission on Excellence in Special Education (2002) published the following finding:

Children placed in special education are general education children first. Despite this basic fact, educators and policy-makers think about the two systems as separate and tally the cost of special education as a separate program, not as additional services with resultant add-on expense. In such a system, children with disabilities are often treated not as children who are general education students and whose special instructional needs can be met with scientifically based approaches; they are considered separately with unique costs—creating incentives for misidentification and academic isolation—preventing the pooling of all available resources to aid learning. General education and special education share responsibilities for children with disabilities. They are not separable at any level—cost, instruction or even identification. (p. 7)

The President’s Commission on Excellence in Special Education noted that the current system “waits for a child to fail” (p. 21). Therefore, the commission recommended the implementation of prevention and early intervention models that are “based on response to intervention and progress monitoring” (p. 21).

Subsequently, the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004 established and strongly encouraged response to intervention (RTI) as an early intervention for all struggling learners. In the Department of Education’s training module on IDEIA regulations,

¹The University of Texas at Austin

Corresponding Author:
Nathern S.A. Okilwa, University of Texas at Austin, 3352 Lake Austin Blvd. #C, Austin, TX 78703
Email: nokilwa@mail.utexas.edu
Hozella (n.d.) explained that RTI consists of three tiers of intervention that begin with documented response to research-based instruction in the general education classroom and that may progress based on close monitoring of progress to more intensive intervention in the general education classroom prior to referral for special education.

Classroom Setting in Middle and High School

The legislative mandates placed greater expectations and demands on the general education teachers (Kunsch, Jitendra, & Sood, 2007). Middle and high school teachers are charged to utilize scientifically based practices to improve student achievement within an environment in which the school day consists of short class periods, movement between classes, and multiple teachers (Veerkamp, Kamps, & Cooper, 2007). Stenhoff and Lignugaris/Kraft (2007) noted that the middle and high school curriculum demands a shift from basic skills acquisition to content knowledge. The authors described middle and high school classrooms as utilizing lecture and independent, individual seatwork teaching styles. In addition, McIntosh, Vaughn, Schumm, Haagar, and Lee (1993) reported that the primary mode of instruction in general education was whole-class activities, and the instruction was not differentiated to address students with disabilities. Furthermore, the authors reported that the teachers were observed to interact less frequently with the students in the class who have learning disabilities.

Students With Disabilities in Middle and High School

Maheady, Sacca, and Harper (1988) reported that many secondary students with mild disabilities do not exhibit the academic or study skills necessary to compensate for their low academic achievement. In addition, Stenhoff and Lignugaris/Kraft (2007) noted that students with disabilities usually have “poor reading skills, poor notetaking skills, and poor organizational skills” (p. 9) when they enter middle and high school; students with disabilities enter middle and high school with limited academic skills and widening achievement gaps that impede their access to general education content-area knowledge. In addition, McIntosh and colleagues (1993) found that the students with learning disabilities in the general education class volunteered for participation at a very low rate. Utley and Mortweet (1997) reported that the academic achievement of students with disabilities in general classrooms is compromised without the necessary instructional support. The IDEA reauthorization of 1997, NCLB of 2001, and the IDEIA reauthorization of 2004 with the RTI initiative have echoed the need for evidence-based instructional practices. Gersten and colleagues (2005) suggested various indicators of evidence-based practice. Research has supported peer tutoring as an evidence-based strategy in improving the learning of students with and without disabilities (e.g., Calhoon & Fuchs, 2003; L. S. Fuchs, Fuchs, & Kazdan, 1999; Mastropieri et al., 2001; Mastropieri et al., 2006; Mastropieri, Scruggs, Spencer, & Fontana, 2003; Spencer, Scruggs, & Mastropieri, 2003; Stenhoff & Lignugaris/Kraft, 2007). Therefore, this synthesis is intended to examine the effectiveness of peer tutoring strategies on the academic performance of students with disabilities enrolled in Grades 6 through 12.

Peer Tutoring

Definition of peer tutoring. According to Slavin (1995), peer tutoring is a component of cooperative learning. Utley and Mortweet (1997) defined peer tutoring as “a class of practices and strategies that employ peers as one-on-one teachers to provide individualized instruction, practice, repetition, and clarification of concepts” (p. 9).

Models of peer tutoring. Different configurations of peer tutoring include classwide pairings, reciprocal relationship, and cross-age matching (Brady, 1997; Hughes & Fredrick, 2006; Mastropieri et al., 2001). Classwide peer tutoring (CWPT) was developed at the Juniper Gardens Children’s project in Kansas City, Kansas, in an effort to improve basic academic achievement of low socioeconomic, culturally diverse students in Chapter I schools in general education (Greenwood & Delquadri, 1995; Maheady, Mallette, & Harper, 1991). Because of the CWPT success, L. S. Fuchs and colleagues (1999) developed a variation of CWPT, which became known as peer-assisted learning strategies (PALS), to address reading skills. CWPT and PALS as well as other models are described in Table 1.

Eight of the studies in this synthesis implemented CWPT; four of the eight also utilized reciprocal relationship peer tutoring. The remaining four studies used reciprocal relationship peer tutoring but not on a classwide basis.

Benefits for students. The bulk of the research on the efficacy of peer tutoring has been conducted at the elementary level (i.e., K–5 or K–6) as opposed to middle and high school level (Bowman-Perrott, Greenwood, & Tapia, 2007; Stenhoff & Lignugaris/Kraft, 2007; Veerkamp et al., 2007). The benefits reported in the literature include improved academics (e.g., reading, comprehension, math computation), desirable behaviors (e.g., on-task, motivation), and improved social interactions or relationships such as making friends (Brady, 1997; Mastropieri et al., 2006). Extensive peer tutoring research at the elementary level has predominantly shown improved outcomes in language arts areas such as reading fluency and comprehension (e.g., Allor, Fuchs, & Mathes, 2001; D. Fuchs, Fuchs, & Burish, 2000; Mathes, Grek, Howard, Babayak, & Allen, 1999), sight word reading (e.g., Heward, Heron, Ellis, & Cooke, 1986; Kouy & Browder, 1986), spelling (e.g., Greenwood, Delquadri, &
Table 1. Models of Peer Tutoring

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classwide peer tutoring</td>
<td>All students in the class are paired into tutor–tutee dyads who work together on tasks structured by the teacher with a standard error-correction procedure. Each dyad is assigned to one of two competing teams to earn points, with the winning team determined daily or weekly (Greenwood, Delquadri, &amp; Hall, 1989).</td>
</tr>
<tr>
<td>Reciprocal relationship</td>
<td>The peer-assisted learning strategies (PALS) is a variation of classwide peer tutoring and includes partner reading, paragraph summarization, and prediction (Fuchs, Fuchs, &amp; Kazdan, 1999). The unique aspect of PALS is that the tutor–tutee dyads switch roles (Hughes &amp; Frederick, 2006).</td>
</tr>
<tr>
<td>Cross-age matching</td>
<td>The tutor–tutee dyads are students of different ages and grades. Older students in higher grades are matched with younger students in lower grades (Heron, Welsch, &amp; Goddard, 2003).</td>
</tr>
<tr>
<td>Other</td>
<td>Reverse-role tutoring in which students with mild disabilities tutor younger students with or without disabilities and group-oriented work to accomplish an outcome (Utley &amp; Mortweet, 1997).</td>
</tr>
</tbody>
</table>

Benefits for teachers. Teachers may become overwhelmed with the diversity of students’ academic and behavioral needs and, when compounded with limited instructional time, become eager to implement efficient, effective strategies that result in positive outcomes for all students (Hughes & Fredrick, 2006; Vaughn et al., 2001). Peer tutoring allows the teacher to share the responsibility of instruction with students, hence causing his or her role to change from primary deliverer of instruction to facilitator (Maheady, Harper, & Mallette, 2001). Peer tutoring also provides the teacher an excellent strategy for class management (Slavin, 1995). As students engage each other in the process of learning, there is limited opportunity to exhibit undesirable behaviors (McDonnell, Mathot-Buckner, Thorson, & Fister, 2001). Furthermore, peer tutoring allows every student to receive individual attention and immediate feedback, which the teacher cannot possibly provide during whole-class instruction sessions (Stenhoff & Lignugaris/Kraft, 2007). With RTI and increased inclusion of students with disabilities in general education, peer tutoring provides teachers a strategy to meet the diverse needs of students (Carter & Kennedy, 2006). Peer tutoring can help make diversity a resource rather than a challenge (Slavin, 1995).

Rationale and Purpose

A few research syntheses and reviews have been conducted recently on the use of peer tutoring with older learners in Grades 6 through 12 (Heron, Welsch, & Goddard, 2003; Kunsch et al., 2007; Ryan et al., 2004; Stenhoff & Lignugaris/Kraft, 2007). Kunsch and colleagues’ (2007) review addressed the effectiveness of peer-mediated instruction on math performance of both students with disabilities and those at risk for math disabilities. The authors found that elementary-level students demonstrated a higher effect size on mathematics performance using peer tutoring compared to secondary students. In general, Kunsch and colleagues reported, “peer-mediated interventions in mathematics are moderately successful in improving the performance of students with learning problems” (p. 10). Ryan and colleagues’ (2004) review focused on the effectiveness of peer-mediated intervention on students with emotional or behavior disorders (EBD) ranging from ages 5 to 18 (i.e., grades K–12). The authors reported a number of findings from their review:

(a) positive academic outcomes across all types of peer-mediated interventions; (b) inconsistencies in reporting participant characteristics; (c) study settings not reflective of actual placement for students with EBD; (d) a need for additional group studies; and (e) overall high levels of consumer satisfaction. (p. 336)

Heron and colleagues (2003) reviewed 15 studies that applied peer tutoring in specialized subject areas (i.e., art, music, horticulture, driver education, physical education, health, and social skills). Their findings showed a positive
effect of peer tutoring procedures in specialized (or noncore) subject areas, resulting in the authors’ concluding that “by systematically extending the application of tutoring systems into specialized areas, teachers can ensure that the positive benefits reported in academic areas will likewise be realized within this context” (p. 298). Stenhoff and Lignugaris/Kraft reviewed 20 studies and found that “all weighted effect sizes for primary dependent variables (e.g., words read, comprehension questions, chapter tests, unit tests, post/pretest scores) were positive” (p. 24).

The purpose of this research synthesis study is to examine the effectiveness of peer tutoring strategies in the academic performance particularly of students with disabilities enrolled in Grades 6 through 12. The synthesis of the data from the studies was organized to address the following: (a) the setting (general education and special education classrooms) in which peer tutoring was employed, (b) the disability type of students involved in peer tutoring procedures, (c) the use of peer tutoring in core content areas (e.g., math, language arts, science, social studies), and (d) the various types of measurement used to document academic outcomes.

Method

Search Procedures

Screening criteria. Predetermined criteria were established to refine the search items. The criteria included studies that (a) were original research, (b) were conducted with special education students in Grades 6 through 12, (c) were published between the years of 1997 and 2007, (d) included manipulation of a form of peer tutoring as an independent variable, (e) included at least one academic measure as a dependent variable, (f) were published in scholarly (peer-reviewed) journals, and (g) were conducted in the United States and in English. The authors chose to review publications after 1997 to coincide with the reauthorization of IDEA, which introduced requirements for grade-level instruction and inclusion of most special education students in statewide assessments at the chronological grade. The authors contended that IDEA of 1997 set the stage for subsequent mandates for scientifically based instructional practices in NCLB of 2001 and IDEA of 2004.

Phase 1. An electronic search for studies was conducted on ERIC (Educational Resources Information Center), Academic Search Premier, and PsycINFO databases. The following descriptors were used: peer tutoring, peer help, peer mediated learning, peer assisted learning, peer support, peer buddy, peer involvement, peer collaboration, and reciprocal tutoring, in combination with the terms disability* or special education paired with academic achievement, educational attainment, school success, or student success. The initial electronic search produced 26 articles with a repeat of 2 articles (i.e., appearing on two databases), thus resulting in 24 articles.

Phase 2. A second-tier search for more studies was carefully performed by reviewing the reference lists of research syntheses, meta-analyses, and reviews identified in the electronic search. In addition, the reference lists of previously identified articles (as long as they addressed the peer tutoring topic) were reviewed.

Phase 3. To locate any missed studies during electronic search and reference checks, a hand search was performed in the following major special education peer-reviewed journals: Education and Treatment of Children, Exceptional Children, Journal of Learning Disabilities, Journal of Special Education, Learning Disability Quarterly, Learning Disability Research & Practice, and Remedial and Special Education.

Final selection. After comparing studies to the predetermined criteria, the three search phases provided a total of 12 studies, which were included in this synthesis (see Table 2). Studies that were syntheses, reviews, or meta-analyses (e.g., Heron et al., 2003; Kunsch et al., 2007; Ryan et al., 2004; Stenoff & Lignugaris/Kraft, 2007) were excluded, but the lists of studies included in the articles as well as the reference lists were reviewed. In addition, studies that were descriptive in nature (i.e., not peer tutoring original studies) were excluded (e.g., Carter & Kennedy, 2006; DuPaul & Eckert, 1998; DuPaul & Weyandt, 2006; Harper & Maheady, 2007; Heron, Villareal, Yao, Christianson, & Heron, 2006; Maheady et al., 2001; Utley & Mortweet, 1997; Vaughn et al., 2001). Studies were excluded because they measured peer tutoring outcomes for general education students only (e.g., Cushing & Kennedy, 1997; Harris, Marchand-Martella, & Martella, 2000; Veerkamp et al., 2007) or did not report outcomes of students with disabilities separately from the rest of the participants (e.g., Allsopp, 1997) or reported only social outcomes (Carter, Cushing, Clark, & Kennedy, 2005). Finally studies were excluded because of implementation at the preschool level (Brady, 1997), and some were unrelated to peer tutoring (O’Rourke & Houghton, 2006; Stecker, Fuchs, & Fuchs, 2005; Woolsey, Harrison, & Gardner, 2004).

Coding Procedure

A predesigned code sheet (Kim, Vaughn, Wanzek, & Wei, 2004) provided the framework for organizing pertinent information from the studies, including data regarding participants (e.g., number, gender, age or grade, disability type, socioeconomic status, etc.), study design (e.g., research design, fidelity, outcome measures, measurement tools, and sample selection criteria), conditions (e.g., treatment or control, setting, implementation procedures, including person implementing intervention, length, frequency, total sessions, duration), clarity of causal inference (e.g., attrition, equating...
Table 2. Overview of Studies Included in the Synthesis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Participants</th>
<th>Setting</th>
<th>Procedure and Design</th>
<th>Intervention</th>
<th>Sessions and Duration</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuchs, Fuchs, and Kazdan (1999)</td>
<td>n = 102; T = 35 LD, 2 mild MR, 1 other, 14 remedial; C = 39 LD, 2 mild MR, 1 LD; 8 remedial</td>
<td>18 classes; Grades 8–10</td>
<td>Teacher-selected low-performing students</td>
<td>CWPT: T = reciprocal PALS w/ teacher matching high readers w/ low to read and summarize; C = no PALS; reported fidelity = 86.29%</td>
<td>5× every 2 wks for 16 wks</td>
<td>0.34 moderate ES reported on reading comprehension; no significance on reading words correctly based on standardized measurement</td>
</tr>
<tr>
<td>Hughes and Fredrick (2006)</td>
<td>n = 18 (3 w/ LD and 16 w/o disabilities)</td>
<td>Inclusive Grade 6 language arts class</td>
<td>Teacher pairs 9 dyads in 2 teams</td>
<td>Reciprocal PT w/ time delays of 0 and 5 s on vocabulary; reported fidelity = 99%</td>
<td>7 wks: around 2 hr for 15 words daily in 3–7 sessions</td>
<td>Improved mastery of words and second content area generalization</td>
</tr>
<tr>
<td>Mastropieri et al. (2006)</td>
<td>n = 213 (37 LD, 7 E/BD, 169 w/o disabilities)</td>
<td>13 general education science classes—Grade 8</td>
<td>Classes randomly assigned; classwide model design</td>
<td>T = CWPT w/ differentiated hands-on activities; C = lecture, notes, lab, and textbook; reported fidelity = 65% average</td>
<td>12 wks</td>
<td>Pre- to posttest: F(1, 10) = 10.89, p = .008; state high-stakes test: F(1, 10) = 4.71, p = .055</td>
</tr>
<tr>
<td>McDonnell, Mathot-Buckner, Thorson, and Fister (2001)</td>
<td>n = 6 (1 MR moderate, 2 severe MR, 3 w/o disabilities)</td>
<td>3 classes (history, prealgebra, and PE) in Grades 7 and 9</td>
<td>Multiple probe across-subjects design</td>
<td>CWPT: Students w/ disabilities paired w/ students w/o disabilities for multielement curriculum; reported fidelity = 92.4% average</td>
<td>15 min 2×/wk</td>
<td>Improved academic responding and reduced competing behaviors for both students w/ and w/o disabilities</td>
</tr>
<tr>
<td>Sideridis et al. (1997)</td>
<td>n = 6 (2 LD, 1 mild MR; 3 w/o disabilities)</td>
<td>Grade 6 general education spelling (inclusive)</td>
<td>A/BAB: 4 wks: baseline 5 wks: CWPT 3 wks: baseline 4 wks: CWPT</td>
<td>31 wks intervention: T: PALS: 2×/wk for 34 hr instruction; LST: 3×/wk for 51 hr instruction C: Saxon: 3×/wk for 51 hr instruction; SRA: 2×/wk for 34 hr instruction</td>
<td>All CWPT produced positive results on posttests; only CWPT w/ CSM had greater gains than control group; greater on-task behavior noted</td>
<td></td>
</tr>
<tr>
<td>Bowman-Perrott, Greenwood, and Tapia (2007)</td>
<td>n = 19 (all E/BD; some also w/ LD)</td>
<td>2 high school biology classes; 1 middle school spelling class</td>
<td>ABAB single-participant design</td>
<td>CWPT: T = reciprocal CWPT w/ reinforcement T2: reciprocal CWPT w/ CSM; C: teacher-led instruction; reported fidelity = 97% w/ observation checklist</td>
<td>T1: 3×/wk for 30 min; T2: 3×/wk for 30 min and CSM</td>
<td>Pre- and posttests: High ES for peer tutoring in word attack (ES = 1.10), word identification (ES = 0.99), and reading comprehension (ES = 0.94) but not for reading fluency (ES = −0.19).</td>
</tr>
<tr>
<td>Calhoon (2005)</td>
<td>n = 38 (38 w/ LD in reading)</td>
<td>2 middle schools: 4 self-contained language arts, Grades 6 (n = 32), 7 (n = 5), and 8 (n = 1)</td>
<td>Random assignment of teachers to treatment and control</td>
<td>CWPT: T = reciprocal PALS w/ LST; C = whole-class format using remedial reading programs (Saxon Phonics Intervention and SRA Skill Acquisition); reported fidelity: PALS = 90.7%, LST = 98.7%, Saxon = 96.6%, SRA = 91.7%</td>
<td>31 wks intervention: T: PALS: 2×/wk for 34 hr instruction; LST: 3×/wk for 51 hr instruction C: Saxon: 3×/wk for 51 hr instruction; SRA: 2×/wk for 34 hr instruction</td>
<td>Reported 0.40 moderate ES for peer tutoring w/ computational skills but not for concept/application skills (−0.02) or statewide test (−0.29).</td>
</tr>
<tr>
<td>Calhoon and Fuchs (2003)</td>
<td>n = 92; T = 35 LD, 2 BD, 6 MR, 2 other; C = 33 LD, 2 BD, 6 MR, 6 other</td>
<td>10 self-contained math classes, Grades 9–12</td>
<td>Random assignment of classes</td>
<td>CWPT: T = reciprocal PALS w/ CBM; C = workbook sheets; reported fidelity = PALS 96.2%; CBM 90.3%</td>
<td>30-min sessions 2×/wk for 15 wks (30 sessions total)</td>
<td>(continued)</td>
</tr>
<tr>
<td>Reference</td>
<td>Participants</td>
<td>Setting</td>
<td>Procedure and Design</td>
<td>Intervention</td>
<td>Sessions and Duration</td>
<td>Results</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------</td>
<td>----------------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mastropieri et al. (2001)</td>
<td>n = 24 (20 LD, 4 mild MR)</td>
<td>2 special education resource Grade 7 English classes</td>
<td>12 dyads of a high student w/ low student; assigned randomly</td>
<td>T = reciprocal PT w/ reading comprehension using summarization; C = class instruction; reported fidelity = observation w/ inconsistencies resolved through discussion</td>
<td>50 min once daily for 5 wks</td>
<td>High ES of 1.18 for reading comprehension (Cohen's d computation by reviewers)</td>
</tr>
<tr>
<td>Mastropieri, Scruggs, Spencer, and Fontana (2003)</td>
<td>n = 16 (14 LD, 1 MR, 1 LD/MR)</td>
<td>2 special education history (Grade 10)</td>
<td>Classes randomly assigned; dyads paired</td>
<td>T = reciprocal tutoring w/ reading and summarization; C = guided notes for class; reported fidelity = 98.3%</td>
<td>90-min sessions; (2×/wk; 3×/wk; 9 wks and 18 sessions total</td>
<td>Reported ES of 0.63 = pretest; 1.54 = chapter tests' average; 1.4 = open-ended tests; 2.2 = multiple choice; 0.67 = year-end average</td>
</tr>
<tr>
<td>Schloss, Kobza, and Alper (1997)</td>
<td>n = 6 (6 moderate MR, 1 ADD also)</td>
<td>Grades 8 and 10 math summer school program</td>
<td>Students chosen; comparison to baseline</td>
<td>Version of reciprocal PT: 3 dyads made from 6 peers. Peer gives 5 s for answer; dyads switched; reported fidelity = 100%</td>
<td>18 sessions in 16 wks</td>
<td>All students mastered next-dollar strategy and generalized skill to make actual purchases in consumer settings</td>
</tr>
<tr>
<td>Spencer, Scruggs, and Mastropieri (2003)</td>
<td>n = 30 (30 E/BD)</td>
<td>6 social studies special education Grades 7 and 8</td>
<td>Cross over design, classes randomly assigned</td>
<td>CWPT: 3 classes began w/ peer tutoring w/ summary strategy; 3 classes began in traditional instruction; both switched in 2 wks; reported fidelity = 97.5% average</td>
<td>4 wks (2 wks tutoring; 2 wks traditional teaching)</td>
<td>Reported ESs: 0.68 weekly quizzes, 0.88 multiple choice, 0.40 open-ended test, 0.89 increased on task</td>
</tr>
</tbody>
</table>

Note: T = treatment group; LD = learning disability; MR = mental retardation; C = control group; CWPT = classwide peer tutoring; PALS = peer-assisted learning; w/ = with; x = times; ES = effect size; w/o = without; PT = peer tutoring; E/BD = emotional or behavioral disorder; ABAB = design alternating intervention with baseline; CSM = classwide self management; LST = linguistic skill training; BD = behavioral disorder; CBM = curriculum-based measurement; ADD = attention-deficit disorder.
of groups, quasi-experimental, contamination elements), data collection, and results (e.g., mean, standard deviation, effect size, p values, F values, etc.). Two raters, yielding 93% interrater reliability, coded all 12 articles. Disagreements were resolved through discussion. The reviewers used Cohen’s d to calculate effect sizes for the Mastropieri et al. (2001) and Mastropieri et al. (2003) studies. Cohen’s d is determined by dividing the mean difference by the standard deviation (Gravetter & Wallnau, 2007). An effect size of 0 through 0.2 is considered small, 0.2 through 0.8 is moderate, and greater than 0.8 is high (Gravetter & Wallnau, 2007). In interpreting the findings of the studies, greater effect sizes indicate a larger impact that peer tutoring had on academic outcomes. Conversely, low effect sizes indicate that peer tutoring had little impact on academic outcomes.

Results

Table 2 presents an overview of pertinent data such as authors, number of participants with disability type, setting, research procedures and design with reported fidelity, intervention, intervention sessions and duration, and results extracted from the studies in this synthesis. The 12 studies included in this synthesis represented a variety of content areas in public school (i.e., Grades 6–12) curriculum. Subject areas included language arts (six studies), science (two studies), mathematics (three studies), and social studies (three studies). One study constituted three different settings, that is, math, physical education, and history. Another study focused on two content areas, that is, science and language arts. The 12 studies ranged from 3 very small studies (n = 6) to 3 large studies (n > 90) and were conducted with a total of 571 participants, 59.48% (n = 350) of whom were male. With the exception of one study that did not report ethnicity, participants were 47% Caucasian (n = 255), 35% African American (n = 190), 14% Hispanic (n = 77), 2% Asian (n = 13), and 2% multiracial or Eastern Indian (n = 12). Although two studies did not report the geographic location of the study, eight studies were in urban areas and one was conducted in a midsize community (i.e., 50,000 population). Four studies were conducted in the Southeast of the United States, three in the Midwest, two in the Mid-Atlantic, and one in the Southwest. Ten studies were conducted on traditional public school campuses. One study was conducted on an alternative school campus as well as in the educational component of a residential treatment facility. Another study was implemented on a public school campus that served only students with emotional or behavioral disorders.

Setting

The studies in the synthesis reflected two classroom settings (i.e., special education classroom or inclusion of special education students in general education classroom). Table 2 is organized by the two settings. Five studies reported on general education classes, and seven studies yielded results from special education classrooms. One study reported outcome data to compute effect sizes (i.e., reporting ES = 0.34) for students with disabilities in a general education classroom setting, and another study reported an F statistic indicating large differences on pre- and posttest mean scores of the treatment and control groups, that is, F(1, 10) = 10.89, p = .008. A majority of the studies that met research criteria for the development of effect sizes were conducted in special education classes. Responses as a result of peer tutoring ranged from effect sizes of 0.40 for abstract skills (math and open-ended test) to 2.2 for a multiple-choice test in history.

Students with disabilities in general education classes. Students with disabilities in four out of five studies conducted in general education settings were paired with students without disabilities during the peer tutoring sessions in all the content areas (i.e., language arts, science, mathematics, social studies, and PE). The five studies reported that students with disabilities implemented peer tutoring procedures with high accuracy, and the peer tutoring strategy resulted in improved reading comprehension, vocabulary mastery, spelling accuracy, science performance, and academic responding in history, prealgebra, and PE classes. L. S. Fuchs and colleagues (1999) reported a moderate effect utilizing reciprocal peer tutoring for reading comprehension. Mastropieri and colleagues (2006) noted an impressive difference between the mean science test scores of the treatment and control groups.

Students with disabilities in special education classes. Seven studies reflected instruction in the content areas of math, science, social studies, and English in special education classrooms. Students with disabilities who participated in these studies were paired with fellow students with disabilities during peer tutoring sessions. Across the content areas, the impact of peer tutoring was impressive. The highest effect size (i.e., 2.2) was observed on unit tests of the world history class (Mastropieri et al., 2003). Three studies reported moderate effect sizes from 0.40 to 0.68 (Calhoon & Fuchs, 2003; Mastropieri et al., 2003; Spencer et al., 2003). Spencer and colleagues (2003) explained that the low statistical power was associated with a low number of students completing an optional portion of the test. The lowest effect sizes were reported for a statewide math test with an effect size of –0.29 (Calhoon & Fuchs, 2003) and reading fluency with an effect size of –0.19 (Calhoon, 2005).

Disability Type

The studies included in this synthesis reported participants with disabilities as being diagnosed with “learning disabilities,” “learning disabilities/emotional disabilities,” “emotional/behavioral disorders,” and “mental retardation” with varying degrees. Table 2 displays the number of...
participants in each study by disability type. Data were not generally disaggregated or reported according to disability type. The most frequent disability reported was learning disabilities, with 256 study participants. In all, 60 participants with emotional or behavioral disorders were reported, and 32 were reported as being diagnosed with mental retardation. The studies indicated 22 of the participants were remedial students, and 10 participants were reported as “other” special education students. The studies included 191 students without disabilities. The vast majority of the studies included participants from various disability types; however, four studies included homogeneous disability groups (e.g., emotional or behavioral disorders in Bowman-Perrott et al., 2007; learning disabilities in Calhoon, 2005; mental retardation in Schloss, Kobza, & Alper, 1997; emotional or behavioral disorders in Spencer et al., 2003). Two studies exclusively focused on students with emotional or behavioral disorders, with one reporting an effect size (0.88) for multiple-choice tests in social studies (Spencer et al., 2003). Another study targeted students with learning disabilities and reported a high effect size (1.10) for peer tutoring to develop word attack skills (Calhoon, 2005). Schloss and colleagues (1997) studied students with mental retardation and reported mastery and generalization of a next-dollar math strategy.

**Content Areas**

Studies in this synthesis represented four different content areas (i.e., language arts, math, science, and social studies). Table 2 indicates peer tutoring being implemented in combination with a specific subject component such as vocabulary mastery, reading comprehension, summarization, computation, and concepts or application. The effect sizes across content areas ranged from −0.29 on statewide math tests (Calhoon & Fuchs, 2003) to 2.2 for world history (Mastropieri et al., 2003) for studies that reported effect sizes or provided sufficient data to calculate effect sizes.

**Language Arts**

*Reading comprehension.* Three studies that used peer tutoring to work on reading comprehension reported improved student reading comprehension skills. L. S. Fuchs and colleagues’ (1999) study reported the treatment group producing a moderate effect size (0.34) in reading comprehension, but the group failed to show substantial outcomes in reading fluency. The authors reported effect sizes ranging from 0.04 (indicating comparable growth in the number of words read correctly) to 0.34 (for the number of questions answered correctly). Mastropieri and colleagues (2001) also studied a reading comprehension component in English classes. The data reported in the article were used to compute the effect size (1.18) using Cohen’s $d$ calculated by the reviewers. The posttest (criterion-referenced tests) scores in reading comprehension indicated a higher performance by the peer tutoring group. Mastropieri and colleagues (2003) used a summarization strategy to promote reading comprehension in world history. Content tests included world history pre- and posttests (end-of-chapter tests), a cumulative unit test, and a delayed final exam covering the entire year’s content. A large effect was reported on content chapter tests (average $ES = 1.54$) and unit tests (average $ES = 1.85$), both in favor of peer tutoring. A medium effect (average $ES = 0.67$) was observed on the end-of-year exam. The medium effect on the end-of-year exam was explained by the fact that part of the content tested was not covered during the duration of the study. Items from the unit covered during the study reflected an even higher effect size ($ES = 1.28$), whereas those items that were not from the unit taught during the study had a very low effect size ($ES = 0.06$). Calhoon (2005) reported large effect sizes for use of peer tutoring in word attack ($ES = 1.10$), word identification ($ES = .99$), and reading comprehension ($ES = .94$). However, like L. S. Fuchs and colleagues (1999), there was no noteworthy improvement in reading fluency.

*Vocabulary.* Hughes and Fredrick (2006) reported on peer tutoring with a vocabulary mastery component in a language arts class. Students with disabilities mastered three vocabulary word sets in 4.6 sessions (median) to meet the criterion (i.e., 100% in two consecutive sessions). On rate of learning measures, students with learning disabilities needed 4.7 sessions (median) to meet the criterion (i.e., seven words). On average, students with learning disabilities needed one extra session to reach the criterion on accuracy and rate measures. Sideridis et al. (1997) reported improved spelling accuracy. The students with disabilities reported a mean spelling performance of 87%. Bowman-Perrott and colleagues (2007) reported positive results on posttests using peer tutoring with classwide self-management for spelling. Sufficient data to calculate effect sizes were not reported.

**Mathematics.** Calhoon and Fuchs (2003) reported a moderate effect size ($ES = 0.40$) for peer tutoring with computational skills but not for concept or application skills ($ES = −0.02$) or on the statewide test ($ES = −0.29$). The authors note that both the treatment and control groups “significantly increased in math concepts/applications and TCAP [Tennessee Comprehensive Achievement Test] skills. No significant differences were found between the groups for growth on concepts/applications skills [−0.01] and TCAP scores [−0.29]” (p. 241). Furthermore, the authors noted a moderate effect size (0.40) for computational skills. Schloss and colleagues (1997) studied currency skills and noted that all students mastered the next-dollar strategy and generalized the skill to make actual purchases in the community. Finally, McDonnell and colleagues (2001) reported improved academic responding and improvement.
in the three classes studied: prealgebra, physical education, and history.

**Social Studies.** In addition to McDonnell and colleagues’ (2001) study, two more studies addressed social studies. Mastropieri and colleagues (2003) compared two treatments (guided notes and peer tutoring) in 10th-grade history using various assessment components (pretest, chapter tests, unit tests, end-of-year exam), which produced effect sizes ranging from 0.63 to 2.2. In addition, Spencer and colleagues (2003) noted effect sizes ranging from 0.40 for open-ended tests to 0.68 on weekly quizzes in social studies classes.

**Science.** Mastropieri and colleagues (2006) conducted the only study in a general education science classroom. The authors reported improved test scores in the pre- and posttests for science content, $F(1, 10) = 10.89, p = .008$, interpreted as a large difference between the peer-tutoring treatment and control groups. In addition, the treatment groups’ scores on the state high-stakes test indicated improvement that approached significance, $F(1, 10) = 4.71, p = .055$. Bowman-Perrott and colleagues (2007) reported positive results on posttests using CWPT with classwide self-management for biology. Sufficient data to calculate effect sizes were not reported.

**Measurement Type**

**Standardized assessments.** Four studies used standardized measurements. L. S. Fuchs and colleagues (1999) used the Comprehensive Reading Assessment Battery, which includes four 400-word traditional folktales at a 2.5 reading level as well as weekly quizzes to measure outcomes. Calhoon (2005) administered the Woodcock–Johnson Test of Achievement–III reading subtests for Letter-Word Identification, Word Attack, Reading Fluency, and Passage Comprehension. Calhoon and Fuchs (2003) conducted pre- and posttests using the Math Operating Test–Revised for computation, the Math Concepts and Applications Test, and the Tennessee Comprehensive Achievement Test to document student progress. In the science content area, Mastropieri and colleagues (2006) used the state’s high-stakes test and unit tests. Moderate to high effect sizes were reported (i.e., ranging from $ES = 0.34$ to $ES = 1.10$) in the four studies that utilized standardized assessments. Calhoon and Fuchs’s (2003) study reported negative effect sizes on two of the mathematics assessments administered in the study: concept or application skills ($ES = -0.02$) and statewide ($ES = -0.29$) standardized assessments. In addition, Calhoon (2005) also elicited a negative effect size on one set of tests, that is, reading fluency ($ES = -0.19$).

**Teacher- or researcher-developed assessments.** The eight studies that did not employ standardized assessments utilized assessments created by the teachers and/or researchers. In the language arts content area, Hughes and Fredrick (2006) measured mastery of the meaning and spelling of 15 vocabulary words by comparing the baseline data to multiple probes conducted by the teacher. Mastropieri and colleagues (2001) used criterion-referenced tests for reading comprehension. Sideridis et al. (1997) used investigator-created assessment tools in combination with observation utilizing the New Code for Instructional Structure and Student Academic Response as well as the Multiple Optional Observation system for Experimental Studies software. In the content area of mathematics, Schloss and colleagues (1997) chose investigator-created criterion-referenced assessment tools for establishing a baseline, performance during peer tutoring, and subsequent community-based assessments, whereas McDonnell and colleagues (2001) used weekly tests in pre-algebra, physical education, and history. In the content area of social studies, Mastropieri and colleagues (2003) used pre- and posttests to measure improvement on social studies chapter tests, cumulative unit tests, end-of-year final exams, and oral reading fluency of social studies material. Spencer and colleagues (2003) used weekly quizzes, a multiple-choice test, and an open-ended test to measure student outcomes. Bowman-Perrott and colleagues (2007) used weekly and unit pre- and posttests for biology and spelling.

Of the eight studies that did not use standardized tests, only three (Mastropieri et al., 2001; Mastropieri et al., 2003; Spencer et al., 2003) reported sufficient statistical values to compute effect sizes using the Cohen’s $d$ statistical process. The effect sizes of the measures in the three studies ranged from $-0.29$ for a statewide math test to 2.2 for multiple-choice tests in history. The small number of participants may have influenced the large effect sizes.

**Reported Fidelity**

The fidelity reported in the studies indicated to what extent the implementation of peer tutoring was accomplished according to prescribed criteria. Fidelity is measured in percentages, with perfect fidelity being 100%. Eleven studies conducted observations to establish the fidelity of the implementation, as reported in Table 2.

**Discussion**

This synthesis of selected studies provides practitioners and researchers with pertinent information regarding the viability of peer tutoring as a technique for eliciting positive academic outcomes for students in Grades 6 through 12. This review examined the setting (general education and special education classrooms) in which peer tutoring was utilized, the disability type of participants, the content areas in which peer tutoring was employed, and the type of measurement used to report outcomes.
Setting

First, the findings of this review suggest that peer tutoring is effective regardless of the setting (i.e., general education or special education classes). The greatest peer tutoring effect size across studies was recorded in a special education setting, as supported by an effect size of 2.2 (for history), and the lowest effect size of –0.29 (for reading fluency) was also recorded in a special education classroom. In both settings, students with disabilities showed improved academic achievement using peer tutoring. Although some of the studies had a small number of participants, the consistent improvement based on reported effect sizes suggests that peer tutoring should be considered a promising practice in special education classes. The overall findings suggest that students with disabilities have the potential to utilize peer tutoring to teach and learn academic skills from each other. Furthermore, the academic achievement of the students is not dependent on the setting. However, Bowman-Perrott and colleagues (2007) suggested that lack of choice of peers to work with in the small classes might be considered a factor in the academic achievement of older students.

Disability Type

The effectiveness and benefits of peer tutoring transcend disability type. However, the degree of disability (i.e., mild, moderate, severe) may be a variable for further investigation. Studies included in this synthesis report primarily participants with high-incidence disabilities, that is, learning disabilities, emotional or behavioral disorders, and mild or moderate mental retardation. Because most studies in this review did not disaggregate outcome data according to disability, it is difficult to assess the outcome as a function of disability type. However, the effect sizes that were reported suggest that all three populations (learning disabilities, emotional or behavioral disorders, and mental retardation) experience substantial academic improvement utilizing peer tutoring.

Content Areas

Peer tutoring implemented across content areas shows positive effects regardless of subject area. The four subject areas (i.e., language arts, math, science, and social studies) represented in this synthesis all report positive academic outcomes. Mastropieri and colleagues’ (2003) 10th-grade history study reported the highest effect size on the multiple choice test component (ES = 2.2), which could be attributed, in part, to the small number of participants (n = 16). The lowest effect size (ES = –0.29) was recorded on a math statewide standardized test (Calhoon & Fuchs, 2003). The range of effect sizes could be a result of the fact that academic skills acquired through peer tutoring were very basic, such as computation skills, spelling, reading, vocabulary, and so on. It is unknown how students with disabilities will effectively utilize peer tutoring strategies to acquire complex academic skills that require critical thinking skills (Stenhoff & Lignugaris/Kraft, 2007).

Type of Measurement

Whether the researchers used standardized or teacher or researcher assessments to measure outcomes, the findings of this review reveal mostly positive growth in academic achievement using peer tutoring. The negative effect sizes that were reported could be attributed to the fact that these assessments tested complex knowledge skills. This concurs with the assertion that peer tutoring strategies are yet to prove efficacious in the ability of students to acquire complex skills (Kunsch et al., 2007; Stenhoff & Lignugaris/Kraft, 2007). The variation in reported effect sizes could have been a function of the type of measurement used to indicate academic improvement. For example, some standardized assessments may not be sensitive enough to measure changes over the short implementation duration. Statewide tests require knowledge of basic as well as complex thinking skills because all types of questions are included, for example, factual, inference, summarization, prediction, and application (Harper & Maheady, 2007).

Implications for Practice

In the wake of students with disabilities being served in general education as driven by NCLB and the IDEA reauthorizations of 1997 and 2004, peer tutoring is a promising strategy to enable students with disabilities to experience success in the general education curriculum. Carter and Kennedy (2006) posited that “although instructional goals for students with disabilities must be individually determined, the general education curriculum now assumes a more prominent role as the context for addressing those goals” (p. 284). Peer-mediated learning, such as CWPT, was developed to facilitate the inclusion of students with disabilities in general education settings (Greenwood, Maheady, & Delquadri, 2002, as cited in Kunsch et al., 2007). In addition, Carter and Kennedy (2006) stated that “peer support interventions were developed to offer an effective, practical approach for assisting students with severe disabilities to access the general curriculum and develop meaningful peer relationships” (p. 285). Collaboration between special and general education teachers is crucial as more and more students with disabilities are taught in the general education classroom (Murawski & Dieker, 2004). Special education teachers are an integral part of the collaboration efforts because of their expertise in curricular accommodations,
behavior management techniques, instructional design, and support services.

Heron and colleagues (2003) call for teacher training programs to teach peer-mediated interventions to equip preservice teachers with such evidence-based instructional strategies. Students with various disabilities, particularly high-incidence disabilities, can learn and also teach each other basic content area skills. Educators can entrust these students with their own learning. Heron and colleagues (2003) state that peer-mediated strategies employ “one of the most valuable—and most overlooked—resources in the classroom: the student” (p. 298). It is time for educators to utilize these resources that “cost little in terms of time and materials” (p. 298). Yet to be explored is the degree of complexity of the content (Kunsch et al., 2007; Stenhoff & Lignugaris/Kraft, 2007). Overall, the assessment of basic skills such as computation, vocabulary, spelling, and reading can be equally accomplished by the use of standardized or nonstandardized assessments. In the current standardized test era, it may just be that standardized measures are the norm for gauging the efficacy of peer-mediated strategies. Bowman-Perrott and colleagues (2007) offer one limitation to the use of peer tutoring; that is, the amount of time required to develop materials and establish and maintain the program.

Limitations

The reader must employ caution in the interpretation and application of the findings of this synthesis based on the following limitations. First, some of the disability types, particularly “learning disabilities,” may be determined differently among the studies, which may affect the interpretation of the findings of the synthesis. Second, some of the studies (especially in general education settings) did not disaggregate data to reflect scores or performance specific to students with disabilities by type of disability, gender, socioeconomic status, or ethnicity. Third, some studies did not meet the standards for rigorous research of random selection of participants and provision of statistical data to determine effect sizes regarding the impact of the intervention. Fourth, the synthesis included authors who conducted multiple studies, which may be interpreted as an overrepresentation by the same teams of researchers. Finally, an outside observer was used to validate the implementation of peer tutoring in 11 of the 12 studies. Having an outside observer in the classroom adds a dimension to the class that is normally not present.

Future Research

The limited literature in the use of peer tutoring in secondary settings suggests that more research is needed to establish the extent to which peer tutoring strategies can assist students with disabilities to access the general education curriculum because the students’ academic success is now measured using general education standards. The articles in this synthesis focused on the implementation of peer tutoring with the learning of basic skills. Further research is definitely needed to establish the effectiveness of peer tutoring in introducing more complex cognitive strategies (e.g., comprehension strategies, critical thinking, math concepts or application). In addition, rigorous research is needed to determine the effectiveness of using peer tutoring with students who have moderate to severe disabilities as well as students who have visual or auditory impairments who were not represented in the literature. Well-constructed studies are also needed to establish the effectiveness of peer tutoring interventions in a variety of other general education settings (i.e., specialized subject areas such as art, music, horticulture, driver education, etc.). In addition, studies are needed that focus on the impact of class size and choice of peer on academic outcomes, particularly by older students. Future studies on peer tutoring must address the ever-increasing diversity of the student population in schools, especially cultural diversity and English language learners. In addition, future research needs to disaggregate findings to inform researchers of the effects of peer tutoring in terms of gender, socioeconomic status, and ethnicity. Finally, teachers can profit from research conducted on the long-term effects of and generalization to the older students with academic skills learned using the peer tutoring strategy with older students.

Summary

This synthesis provides an analysis of 12 peer tutoring research studies conducted with older learners, Grades 6 through 12. The findings are organized into four themes: setting, disability type, academic content area, and type of measurement used to determine academic outcomes. The studies indicate that, first, peer tutoring is an effective strategy that improved academic outcomes of students with disabilities in both general education and special education settings. Second, regardless of the disability type of participants, peer tutoring resulted in positive academic outcomes for the students. Third, students with disabilities recorded academic growth in all the core subjects represented in the studies. Finally, regardless of the measurement type (i.e., standardized or teacher or researcher assessments), moderate to high effect sizes were reported for the learning of basic skills; however, complex knowledge skills (i.e., concept or application skills) elicited negative effect sizes. In summary, peer tutoring, as suggested by the findings of this synthesis, is a promising, effective instructional strategy for students in Grades 6 through 12.
Acknowledgment
The authors would like to thank Dr. Sharon Vaughn, professor, Department of Special Education, University of Texas at Austin, for encouragement and support during the writing of this article.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding
The authors received no financial support for the research and/or authorship of this article.

References
References marked with an asterisk indicate studies included in the research synthesis.
Heron, T. E., Villareal, D. M., Yao, M., Christianson, R. J., & Heron, K. M. (2006). Peer tutoring systems: Applications in...


### About the Authors

**Nathern S. A. Okilwa**, MA (SpEd), is a doctoral student in educational policy and planning at The University of Texas at Austin in Austin, TX.

**Liz Shelby**, MEd, is a doctoral student in special education administration at The University of Texas at Austin in Austin, TX.