Students With Emotional–Behavioral Disorders as Cross-Age Tutors: A Synthesis of the Literature

Gavin W. Watts
Diane Pedrotty Bryant
Megan L. Carroll

Follow this and additional works at: https://digitalcommons.tamusa.edu/sped_faculty

Part of the Special Education and Teaching Commons
Students with Emotional-Behavioral Disorders as Cross-Age Tutors: A Synthesis of the Literature

Abstract: The purpose of this synthesis was to evaluate the effectiveness and related outcomes of the cross-age tutoring model when students with or at-risk for emotional-behavioral disorders (EBD) serve as tutors. Research questions were posed to identify the shared and unique components (e.g., dosage, tutor training) of the cross-age tutoring model; the extent to which students with EBD can effectively serve as cross-age tutors (i.e., fidelity of implementation and tutees’ improvement); the extent to which the model was effective in promoting desired academic and/or social-emotional-behavioral outcomes for tutees and tutors with EBD; the generalization, maintenance, and social validity of the effects; and the overall methodological quality and rigor of the included studies. Findings showed common training and instructional components across interventions, and that tutors with EBD can implement cross-age tutoring procedures with fidelity. The cross-age model was shown to be effective in promoting academic and social-behavioral skills for the tutees as well as the tutors. Evidence for effectiveness in improving self-concept and attitude of the tutor with EBD was inconsistent. Implications and future research considerations are discussed.
Abstract

The purpose of this quantitative synthesis was to evaluate the effectiveness and related outcomes of the cross-age tutoring model when students with or at-risk for emotional-behavioral disorders (EBD) serve as tutors. Research questions were posed to identify the shared and unique components (e.g., dosage, tutor training) of the cross-age tutoring model; the extent to which students with EBD can effectively serve as cross-age tutors (i.e., fidelity of implementation and tutees’ improvement); the extent to which the model was effective in promoting desired academic and/or social-emotional-behavioral outcomes for tutees and tutors with EBD; the generalization, maintenance, and social validity of the effects; and the overall methodological quality and rigor of the included studies. Findings showed common training and instructional components across interventions, and that tutors with EBD can implement cross-age tutoring procedures with fidelity. The cross-age model was shown to be effective in promoting academic and social-behavioral skills for the tutees as well as the tutors. Evidence for effectiveness in improving self-concept and attitude of the tutor with EBD was inconsistent. Implications and future research considerations are discussed.

**Keywords:** emotional-behavioral disorders, cross-age tutoring, peer-mediated, disabilities
Students with Emotional-Behavioral Disorders as Cross-Age Tutors: A Synthesis of the Literature

Academic and behavioral needs of students with emotional-behavioral disorders (EBD) have been identified as some of the most difficult to address (Kern, 2015). However, these needs can be met through the use of effective academic planning and thoughtful selection of instructional techniques (Hughes & Fredrick, 2006). One such instructional technique, known as cross-age tutoring (i.e., an older student tutoring a younger student), shows evidence of being an effective model for teaching academic and social skills to students with disabilities, including students with EBD (Okilwa & Shelby, 2010; Spencer 2006; Spencer, Simpson, & Oatis, 2009). Additionally, this instructional technique requires minimal costs (i.e., time and materials) and can be implemented without substantial training time (Heron, Welsch, & Goddard, 2003). Given the demands placed upon special education classrooms for instructional techniques that are practical, low- to no-cost, and above all, provide effective individualized instruction, utilizing cross-age tutoring may provide a model for addressing the intensive needs of students with disabilities while also providing tutors with EBD opportunities to practice and develop social, behavioral, and academic skills in an instructional context.

Cross-Age Tutoring and Students with EBD

Research focusing on students with challenging behaviors in the role of cross-age tutor has been limited in recent years but has shown positive outcomes for the tutor, as well as for the tutee (i.e., the student receiving instruction from the tutor) (Blake, Wang, Cartledge, & Gardner, 2000; Gumpel & Frank, 1999). Improvements in the areas of reading (Cochran, Feng, Cartledge, & Hamilton, 1993; Top & Osguthorpe, 1987), mathematics (Robinson, Schofield, & Steers-Wentzell, 2005), spelling (Stowitschek, Hecimovic, Stowitschek & Shores, 1982), general test
STUDENTS WITH EBD AS CROSS-AGE TUTORS

scores, and grades (Maher, 1982; 1984) have been found for tutors with EBD. In addition to academic achievement, research on cross-age tutoring models also suggest positive outcomes in social, emotional, and behavioral skills, including discipline within the classroom setting and the reinforcement of peer relationships (Greenwood, Carta, & Hall, 1988; Maher 1982; 1984), social skills (Blake et al., 2000; Gumpel & Frank, 1999), on-task behavior (Greenwood, Delquadri, & Hall, 1989; Hogan & Prater, 1993), self-esteem and self-worth (Lazerson, 2005; Miller, Topping, & Thurston, 2010), and attendance rates (Maher, 1982). Given that social-behavioral and academic skills are frequently characterized as deficit areas for individuals with EBD (Landrum, Tankersely, & Kauffman, 2003; Trout, Nordness, Pierce, & Epstein, 2003), utilizing cross-age tutoring shows promise as a possible intervention for addressing these needs.

Existing Reviews

A number of systematic reviews completed within the last few years have focused on both academic outcomes, and less frequently, social-emotional and behavioral outcomes in regards to students with disabilities and peer-mediated interventions. Most recently, Bowman-Perrott, Burke, Zhang, and Zaini (2014) conducted a meta-analysis focusing on direct and collateral effects of peer tutoring on social and behavioral outcomes for students with disabilities. Findings showed peer tutoring had a greater effect on promoting social skills and reducing disruptive behaviors than increasing academic engagement for students with disabilities. Also, cross-age tutoring was found to be more effective than same-age or reciprocal tutoring for students with EBD. Similar findings were obtained through a meta-analysis of tutoring models for literacy instruction, where cross-age tutoring was found to be more effective than adult tutoring and computer-based tutoring, especially when students with disabilities served as tutors (Jun, Ramirez, & Cumming, 2010).
STUDENTS WITH EBD AS CROSS-AGE TUTORS

Bowman-Perrott and colleagues (2013) also examined peer tutoring effects on academic skills in a meta-analysis. Findings of this review showed the model to be highly effective and that students with EBD obtained greater benefit from the model than other disability types. Ryan, Reid, and Epstein (2004) also focused their review on the academic achievement of peer-mediated interventions for students with EBD. Overall findings of the synthesis suggest that peer-mediated interventions appear to be effective across content areas for students with EBD.

A review focused on students with EBD within cross-age and same-age peer tutoring models found that the cross-age tutoring model to be more effective than both the same-age and reciprocal tutoring in reading but less effective than the same-age tutoring model for mathematics (Spencer, 2006). It should be noted that only 13 studies provided sufficient data to calculate effect sizes. Spencer and colleagues (2009) continued the previous review by identifying nine additional studies that included students with EBD in tutor and tutee roles within peer tutoring models. The authors noted that although peer tutoring continues to show promise as an effective intervention for students with EBD as tutors or tutees, additional research is required for these students in secondary and generalized settings.

The number of studies containing students with EBD as cross-age tutors were limited in previous reviews (Spencer, 2006; Spencer, Simpson, & Oatis, 2009). Therefore, an expanded search is required to more comprehensively evaluate the effectiveness of this tutoring model for students with challenging behaviors. Considerations for research in this area, proposed by Greenwood, Carta, and Hall (1988), include identifying strategies used by students with disabilities as tutors that are sufficiently developed and validated, evaluating the fidelity of tutoring interventions, comparing procedures or materials with other conditions, and identifying potential areas for standardizing tutoring models for students with disabilities. Overall, the
existing reviews routinely focused on the general outcomes of peer tutoring with students with disabilities and rarely addressed the underlying, functional components of a given peer-tutoring model such as cross-age tutoring.

Therefore, the purpose of this synthesis was to examine the cross-age tutoring model components with students with EBD serving as tutors. Additionally, this review identified the shared, key components of the model (i.e., tutor training and implementation of tutoring sessions), and the extent to which fidelity, maintenance, generalization, and social validity were measured across included studies. Thus, this synthesis of the literature addressed the following research questions: (a) What were the components (e.g., dosage, tutor training, instructional components) of cross-age tutoring models with students with EBD as tutors? (b) To what extent was the cross-age tutoring model effective in promoting desired outcomes for tutees, and/or tutors with EBD? (c) To what extent were students with EBD effective in the role of tutor (i.e., fidelity of implementation and improvement of tutees)? (d) To what extent were generalization, maintenance, and social validity measured within cross-age tutoring studies with tutors with EBD? (e) What was the overall methodological quality and rigor of the included studies?

Method

Search Procedures and Inclusion/Exclusion Criteria

The procedures for this synthesis were designed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (PRISMA; www.prisma-statement.org). PRISMA is a process of directing and reporting systematic reviews and meta-analyses that has been mutually agreed upon by an international group of healthcare researchers. For the purpose of this synthesis, cross-age tutoring was defined as a school-age tutor (i.e., non-
STUDENTS WITH EBD AS CROSS-AGE TUTORS

adult) who is one or more years older and/or in a higher grade than the student (i.e., tutee) to whom they are providing instruction.

To be included in this systematic review, the study had to utilize single-case, experimental, or quasi-experimental designs and meet the following inclusion criteria: (a) cross-age peer tutoring model was used to deliver instruction/intervention, (b) the tutor was at least one year older or enrolled in a higher grade than the tutee, (c) the tutor was identified with, or at-risk for developing an emotional-behavioral disorder, and (d) at least one outcome (e.g., academic, social, behavioral) was measured for the tutee(s) and/or tutor(s), including fidelity of implementation of tutoring procedures. Studies that took place outside of school or school-like settings (e.g., after-school programs), involved adult tutors (Miller, 1995), or implemented a tutoring model that was not face-to-face (e.g., online; Smet, Keer, Wever, & Valcke, 2010) were excluded. Additionally, studies utilizing solely qualitative, anecdotal, or descriptive methods for identifying peer tutor outcomes were excluded, along with reviews and position papers. To account for publication bias, where studies yielding more favorable (statistically significant) results tend to be published, it was decided that there would be no restriction on the year of publication, and dissertations and unpublished manuscripts would also be considered when full-text was available (Dwan, Gamble, Williamson, & Kirkham, 2013; Pigott, Valentine, Polanin, Williams, & Canada, 2013). The search was limited to papers produced in the English language.

A search of the literature was conducted to identify the relevant studies utilizing five electronic databases: Academic Search Complete, Education Source, Education Resources Information Clearing House (ERIC), Professional Development Collection, PsycINFO, and ProQuest Dissertations & Theses Global. The following terms were entered into all six databases: “cross-age* OR mixed-age* OR coach*”, “emotion* OR behavi* OR emotional-
behavioral OR EBD OR social*”, “disab* OR disorder*”, and “tutor* OR support* OR mediat*”. Additionally, a hand search of the past four years of publications (i.e., 2013-2016) was conducted for five related journals (i.e., Behavioral Disorders, Behavior Modification, Beyond Behavior, Emotional and Behavioural Difficulties, and Journal of Emotional and Behavioral Disorders). Details pertaining to the method of literature search and inclusion/exclusion of studies are shown in Figure 1 (adapted from Moher, Liberati, Tetzlaff, & Altman; The PRISMA Group, 2009).

Screening and Coding Procedures

Prior to screening, two graduate research assistants with backgrounds in special education were trained on the screening and coding manual developed by the lead author in a 90-min training session and practice articles were coded until 100% consensus was met. At the end of the session, five practice articles (i.e., two single case, three group design) were screened and coded independently, according to the manual procedures. For screening and coding procedures, interrater reliability (IRR) was calculated by summing the number of agreements and then dividing by the total number of agreements plus disagreements. IRR on training articles was found to be 100% for screening and 93% for coding categories. Two researchers then independently screened the titles and abstracts of the identified papers within the initial pool to assess if the given paper utilized a peer-mediated intervention and contained participants with or at-risk for EBD (IRR = 98.3% for journal articles, 95.9% for dissertations). Disagreements were discussed until a consensus was reached. This step yielded 38 journal studies and seven dissertations for potential inclusion (n = 45 studies). The two researchers then screened the reference list titles of the included studies up to this point. They read and discussed the potential papers’ abstracts until a consensus was made to include or exclude. Finally, the full-text of the
papers were screened based on the inclusion criteria. Agreement during full-text screening was defined as both researchers approving that a paper should be included (i.e., all inclusion criteria were met) or excluded (i.e., one or more inclusion criteria were not met). If disagreements were found within a coding category, meetings were held to review the category information until a consensus was reached. The final pool included 15 papers that met the qualifications for inclusion in this synthesis (IRR = 94%; published = 11, dissertation/manuscript = 4).

Each of the included papers was then double-coded by the trained researchers according to the coding manual procedures. Descriptive information and page numbers were also noted for items that were not captured by the coding scheme (i.e., coded as ‘other’) and were later discussed and added to a mutually agreed upon coding category for added specificity. The following categories were used in coding the characteristics, quality, and outcomes of each study:

**Design.** Designs were coded as experimental (i.e., random assignment to conditions), quasi-experimental, or single-case, and descriptive information was recorded for specific design components (e.g., single group, number of control/comparison groups) as well as descriptions of how students, teachers, and/or classrooms were randomly assigned, if applicable.

**Setting.** Placement and treatment settings were both coded when reported. Placement setting refers to the school type and geographic location where the students attended. Geographic location was recorded as descriptive information as the authors reported it (e.g., Brooklyn, New York; central Texas). The description of the community where the school was located was also coded (urban, suburban, rural, not specified). The coding checklist of school types included: public, charter, lab/university, residential, clinic, hospital, other, and not reported. Treatment settings refer to the location within the school where the tutoring sessions took place and options
within this category included: general education classroom, special education classroom, hallway, office, observation room, hospital room, other, or not reported. Descriptive information was also recorded for added specificity. For example, if the setting was a special education classroom, and the authors reported it to be a resource or self-contained classroom, this description was recorded as the authors reported it.

**Implementer.** Implementer refers to the individual(s) providing tutor training and/or supervision of tutoring sessions. Codes included: researcher, research assistant, lead teacher (specify special or general educator), paraprofessional, related school staff (e.g., school counselor, school psychologist, social worker), other, or not reported. Descriptive information was recorded if provided (e.g., number of years/experience, educational background).

**Participants.** Demographics for both the tutor and tutee populations were accounted for when reported. Each of the following categories was coded as reported/not reported as well as quantity/descriptive data for each population (i.e., tutors and tutees): number of participants, grade level, age, gender, race/ethnicity, socioeconomic status, language status, and disability label. If more than one grade level/age/disability label was present, all were recorded. Disability categories included emotional disturbance/emotional-behavioral disorder, specific learning disability, intellectual disability (mental retardation in older studies), autism, visual impairment, deaf-blindness, deafness, hearing impairment, multiple disabilities, orthopedic impairment, traumatic brain injury, at-risk (difficulties/challenges as reported), no disability/general education/typically developing, other health impairment (OHI), other, or not specified. Additionally, to add further specificity, descriptive information related to qualification criteria utilized in determining disability risk status/categorization was recorded as reported by the authors.
Students with EBD as Cross-Age Tutors

Tutor training. Tutor training refers to any researcher or practitioner provided instruction or practice opportunities for tutors prior to the implementation of tutoring sessions with their tutee. Frequency and duration of individual tutor training sessions was recorded when reported. Additionally, the total duration (minutes) of tutor training was calculated when possible. Components of tutor training were defined as any instructional method or practice utilized to teach the tutors the procedures or strategies they would use within the tutoring sessions. The coding checklist contained the following categories and descriptive information was recorded to further specify type/features of each: curriculum name (scripted/unscripted), teacher developed lessons (scripted/unscripted), researcher-developed lessons (scripted/unscripted), purpose of training/introduction, greeting strategies, modeling, prompting/redirection, positive reinforcement, role-playing, lesson planning, evaluation/assessment/progress monitoring training, goal setting, problem solving/discussions, performance feedback from trainer, planning time, materials/manipulatives, review sessions, other; descriptive information for each, other (descriptive information recorded), or not reported.

Tutoring sessions. Tutoring sessions refer to the meetings where the tutor provided instruction to their tutee(s). Frequency and duration of individual tutoring sessions as well as total duration throughout the intervention was calculated when the necessary information was reported. When information was provided on how time was allocated within individual tutoring sessions, the disaggregated instructional (i.e., tutoring) time was calculated (i.e., does not include time for administration of measures). Components of tutor sessions were defined as any instructional method or practice utilized by the tutor to teach the tutee(s) the target skill or content. Descriptive information was recorded for future specification of coding categories for the following tutoring session checklist items: curriculum name (scripted/unscripted), teacher-
STUDENTS WITH EBD AS CROSS-AGE TUTORS

developed lessons (scripted/unscripted), researcher-developed lessons (scripted/unscripted),
purpose of tutoring/reviewing goals, introduction/greeting strategies, modeling,
prompting/redirection, corrective feedback, positive reinforcement, role-playing,
evaluation/assessment/progress monitoring, goal setting, problem solving/discussions,
performance feedback from tutor, planning time, instructional materials, manipulatives, review
sessions, reward/reinforcement system (type), tutor re-training/follow-up sessions, other
(descriptive information recorded), or not reported.

**Instructional focus or target skill(s).** The content area and/or
social/emotional/behavioral skill targeted for instruction or skill promotion within the tutoring
sessions was coded for both tutees and tutors, when applicable. Academic content area was
coded as basic reading skills (e.g., decoding), reading comprehension, written expression, early
numeracy skills (e.g., counting), mathematics (e.g., calculation, reasoning), history/social
studies, science, other academic area, or not specified. Non-academic skills were recorded based
on authors’ operational definitions of target behaviors.

**Dependent measures.** Names and descriptions were coded for dependent measures used
to evaluate outcomes for both tutor and tutee populations. Response categories included:
standardized, teacher-developed curriculum-based measures (CBMs), researcher-developed
CBMs, quiz/test grades, report card grades, other academic measure, attendance, observation of
operationally defined target behavior(s), researcher-developed social/behavioral rating scale,
self-concept/esteem scale, other social-emotional-behavioral measure, or not reported. Measures
used to evaluate the maintenance and/or generalization of effect of the intervention was assessed
using the same coding checklist.
Tutors’ fidelity of implementation. Fidelity was coded if information was reported regarding the extent to which the tutors implemented the instructional procedures specified by the researcher for use within the tutoring sessions. Evidence of the type of measures used, range, and mean fidelity scores were recorded when reported.

Social validity. Social validity was defined as any measure of consumer satisfaction from a tutor, tutee, teacher, and/or parent. Social validity was coded when measured for any of the participant populations. Participant codes included: tutor(s), tutee(s), supervisor(s) of tutoring program/sessions (e.g., teacher(s), paraprofessional(s), related staff), and/or parent(s). Descriptive information was recorded for all reported outcomes.

Initial IRR was found to be 93.4% for coding categories containing response options. Disagreements were found most frequently in identifying the primary, targeted skill/behavior for the tutors, as there were occasionally varying behaviors/social skills measured across tutors within the same study. Coding disagreements and descriptive data recorded under the ‘other’ code was discussed, agreed upon, and categorized by the coders prior to analysis. Each study was also assessed to determine the extent to which it met quality indicators for the design based upon categories outlined by Gersten et al. (2005) and Horner et al. (2005).

Quality indicators for single-case designs were examined for multiple components within the following categories: participant description/characteristics (e.g., gender, disability, diagnosis) and selection process/criteria described in sufficient detail to allow replication of process to obtain participants with similar characteristics; physical features/location of the setting described with enough detail for replication; dependent variable defined (operationally defined, countable index provided, evidence of validity/reliability, frequency of measurement, interobserver agreement measured/established); independent variable (operationally defined,
STUDENTS WITH EBD AS CROSS-AGE TUTORS

systematically manipulated, evidence of fidelity of implementation); baseline (conditions operationally defined, stable data trend); experimental control/internal validity (a minimum of three demonstrations of effect at three different times, controlled for threats to validity, demonstrated experimental control); external validity (effects replicated across participants, settings, or materials); and social validity (provided social importance of dependent variable, magnitude in change, practicality of intervention, cost effectiveness, and/or practitioner implementation).

Quality indicators for group designs were examined for components within the following categories: participants description/characteristics and selection process described with sufficient detail to allow replication of process to obtain a sample with similar characteristics (e.g., age, grade, disability/risk status); comparable population characteristics present across groups/conditions; differential attrition reported; setting described in enough detail for replication; dependent variable defined (operationally defined, aligned with intervention, evidence of validity/reliability, frequency of administration, interscorer agreement measured/established, multiple measures used and/or administered at different times, data collectors are blind/unfamiliar to conditions/participants); independent variable (operationally defined, comparison conditions described, fidelity of implementation measured/established); and data analysis (methods chosen are aligned to research questions, variability is accounted for, power analysis is provided). Quality indicator categories were scored as 1 (met standard without reservations; reported sufficient information for replication/outcome), 0.5 (met standard with reservations; met the minimum requirements for categories for replication/outcome), or 0 (did not meet standards; information provided was not adequate for replication/outcome; met less than half of quality indicators within a category) (Institute of Education Sciences, 2014). An
overall quality score was calculated for each study by dividing the score obtained by the total possible points and multiplying by 100%. A trained graduate research assistant in special education who had taken a course in quality indicators and experimental design assisted in assessing reliability. Reliability was calculated for 30% of the studies and interrater agreement was found to be above 90%.

Analysis of Outcomes

Percentage of data points exceeding the median of the baseline phase (PEM) approach was chosen to assess effectiveness in single-case design studies due to its assumed validity in assessing disruptive behaviors (Chen & Ma, 2007; Ma, 2006), a frequently targeted skill within the single-case studies included in this synthesis. PEM’s design does not rely on the most extreme datum point and therefore is recommended for use in instances where significant outliers may be identified within the baseline data or there is some variability over time. Furthermore, when floor or ceiling data points are present, PEM is still capable of reflecting effect size and has shown utility in meta-analysis of single-case research (Ma, 2009; Preston & Carter, 2009). PEM is calculated by identifying the median baseline point and drawing a median line from that point through intervention phases. The percentage of data points above or below the median line is calculated by summing all intervention data points above or below the line, depending on targeted skill or measure (e.g., increasing an academic skill, or decreasing a behavior), and dividing that sum by the total number of data points in the intervention phase. PEM results were interpreted using the following scale: 90-100% = large or highly effective, 70%-90% = moderately effective, and < 70% = small or questionable effectiveness (Ma, 2006).

Cohen’s $d$ was calculated for experimental and quasi-experimental, between group designs (Cohen, 1988). When raw data were not reported or were missing, $t$-scores were
provided in place. Cohen’s (1988) criterion was utilized for grading effect size values (i.e., .20 is small, .50 - .79 is moderate, and > .80 is large). For studies containing a single group, pre/post-test design, the mean of the pre-intervention assessment was subtracted from the mean of the post-intervention assessment, and the result was divided by the standard deviation of the pre-intervention.

Results

The search procedures and inclusion criteria resulted in 15 studies being identified for this synthesis. A summary of the study characteristics is shown in Table 1. All but four of the included papers were published in peer-reviewed journals, with three of the studies being doctoral dissertations (Hamelberg, 1987; Harrigan, 1994; Holecek, 2012), and one study unpublished at the time of this review (Watts & Bryant, 2017). The years of publication across the 15 studies ranged from 1972 to 2017. Group design was the most common study design (n = 9) and included treatment-comparison (n = 6) and pre-/post-test, single group methodologies (n = 3). Single-case designs were represented less frequently (n = 6). A multiple-baseline design was employed in all studies utilizing single-case methodology.

Setting, Implementer, and Participants

Of the studies that reported adequate information pertaining to the setting of the intervention, the most common location was in an urban environment (n = 6), with public (n = 5), suburban (n = 3), and rural school districts (n = 2) also being represented. Private, charter, lab, and self-contained special education schools were also used as settings for cross-age tutoring interventions (n = 5). Within these school settings, the most common location for tutoring sessions was in a special education classroom (n = 6).
A researcher was frequently the primary implementer of tutor training and supervision of tutoring sessions \((n = 6)\). It was also reported that research assistants or trained staff undertook these responsibilities as well \((n = 3)\). Practitioners, such as special educators or school psychologists \((n = 2)\), and paraprofessionals \((n = 1)\) were utilized as implementers less frequently. In three studies, this information was not reported (Harrigan, 1994; Holecek, 2012; Lazerson, 2005).

Across the included studies, the number of participants was 436, with the number of students serving in the role of cross-age tutor \((N = 126;\) comparison = 105) ranging from one to 39 per study, and the number students serving in the role of tutee \((N = 132;\) comparison = 73) ranging from one to 37. The grade level of the tutors ranged from second to high school, and the ages ranged from 9 through 18, with a majority of the studies including tutors in high school \((n = 7)\) and middle school \((n = 4)\). The tutees’ grade levels ranged from kindergarten to middle school, and ages ranged from five to 14 years old, with a majority of studies containing tutees in the elementary grades \((n = 11)\).

In all 15 studies, a student with or at-risk for EBD fulfilled the role of tutor. Students with EBD, or who had a comorbid label that contained EBD (e.g., learning and behavioral disorder), also commonly served in the role of the tutee in the included studies \((n = 8)\). Equally represented were tutees with or at-risk for specific learning disabilities \((n = 8)\); occasionally, this tutee role was filled by students with cognitive or intellectual disabilities \((n = 4)\). The tutors’ specific disability label, or area of challenge, was inconsistently defined: behavior disorder (Cochran et al., 1993; Lane et al., 1972), challenging behaviors and difficulties relating to peers (Blake et al., 2000), learning and behavior disorders (Scruggs & Osguthorpe, 1986; Top & Osguthorpe, 1987),
aggressive and withdrawn (Lazerson, 1980), and socially rejected and isolated (Gumpel & Frank, 1999).

Quality of Studies

For the nine studies utilizing group designs, the quality scores varied greatly, from 27.3 to 95.5, with the mean across studies being 59.1. Single-case designs proved more rigorous and quality scores were more consistent overall, although, it should be noted that the number of single-case studies included was less than the number of group design studies included. Single-case study scores ranged from 70.5 to 90.9, with a mean of 84.9 across studies.

Components of Cross-Age Tutoring

Table 2 summarizes the common and unique components of the cross-age tutoring model across the included studies.

**Tutor training.** The frequency of tutor training ranged from one to eight sessions, with the duration of individual training sessions ranging from 30 to 60 min. For studies that reported adequate information for determining the total number of minutes provided for tutor training, the durations ranged from 60 min (Harrigan, 1994) to 360 min (Gumpel & Frank, 1999) with the average length of training across studies being 177.5 min \((n = 6)\). The remaining studies reported solely the number of sessions/days of training \((n = 7)\) or no frequency/duration information at all (Holecek, 2012; Lane et al., 1972).

The components of tutor training were relatively consistent across studies. Mutual instructional features of the tutor training sessions (percentage across studies) included: instructions/procedures/objectives (60%), role-playing/practice opportunities (80%), performance/corrective feedback techniques (60%), providing positive reinforcement (60%), and modeling (47%). More unique features included training on scripted lessons (Blake et al., 2000),
greeting tutees (Holecek, 2012), individual interviews with tutors (Hogan & Prater, 1993), lesson planning (Maher, 1982; 1984), goal setting (Maher, 1982; 1984), self-monitoring procedures (Gumpel & Frank, 1999; Hogan & Prater, 1993; Scruggs & Osguthorpe, 1986), data collection procedures (Hamelberg, 1987; Top & Osguthorpe, 1987; Watts & Bryant, 2017), token reinforcement (Harrigan, 1994), proximity to tutee (Scruggs & Osguthorpe, 1986), review of previous skills (Hamelberg, 1987), and specific behavioral or academic instructional techniques (Gumpel & Frank, 1999; Harrigan, 1994; Lane, Pollack, & Sher, 1972).

**Tutoring sessions.** The frequency of tutoring sessions across all studies ranged from two (Holecek, 2012; Lane, Pollack, & Sher, 1972; Maher, 1982; 1984) to five sessions per week (Blake, Wang, Cartledge, & Gardner, 2000; Lazerson, 1980). The length of individual tutoring sessions varied from 15 min (Hogan & Prater, 1993; Top & Osguthorpe, 1987) to 60 min in length (Lazerson, 2005), with duration falling in the range of 20 to 30 min in all but five studies. The duration of the intervention was reported in all but two studies. Of those reported, the longest intervention phase was seven months (Lane et al., 1972) and the shortest was five weeks (Lazerson, 1980).

All but three studies selected academic skills as the target for tutoring instruction, while the remaining studies taught social skills or gave the tutors free manipulation of the content materials. When academics were the focus of instruction, the target skills most frequently fell within the domain of reading and/or literacy ($n = 10$). Mathematics instruction was represented in four of the studies (Lazerson, 2005; Holecek, 2012; Maher, 1984; Watts & Bryant, 2017). The instructional skills taught within tutor training sessions (e.g., modeling, corrective feedback, positive reinforcement) were frequently utilized within tutoring sessions as the primary instructional techniques. Supports for tutors included weekly planning sessions with special
educators (Maher, 1982; 1984), performance feedback or follow-up conferences (Maher, 1984), reteaching/retraining sessions (Lane, Pollack, & Sher, 1972; Watts & Bryant, 2017), and tangible reinforcers provided by researchers (Cochran, Feng, Cartledge, & Hamilton, 1993).

**Fidelity of implementation.** Fidelity of implementation of tutoring procedures by students with EBD was reported by four studies (Blake, Wang, Cartledge, & Gardner, 2000; Hamelberg, 1987; Maher, 1984; Watts & Bryant, 2017). Across the studies reporting fidelity, the rates of implementation ranged from 88% to 97% ($M = 94.2\%$). When fidelity was measured, the outcomes for all participants were moderate to large, with effects being maintained in each of the studies.

**Effectiveness of Cross-Age Tutoring**

Table 3 provides a summary of participant characteristics and common effects for targeted skill categories/content areas. Across studies, the measures used, in order of prevalence, were: curriculum-based ($n = 9$), standardized ($n = 7$), direct observation ($n = 6$), researcher-developed ($n = 2$), and school records ($n = 2$). Of the standardized measures, social-emotional scales, including attitude and self-concept assessments, were the most commonly used ($n = 5$), followed by behavior and social skill scales ($n = 3$), and academic tests ($n = 3$). Studies utilizing group designs contained effect sizes (Cohen’s $d$) for tutors and tutees ranging from null, non-significant effects to large, statistically significant outcomes. Single-case design studies reported effect sizes ($PEM$) for tutors ranging from 42.9 (Hogan & Prater, 1993) to 100 (Blake, Wang, Cartledge, & Gardner, 2000; Gumpel & Frank, 1999; Maher, 1984; Watts & Bryant, 2017), and for tutees, from 88.2 (Hamelberg, 1987) to $> 90$ (Blake et al., 2000; Gumpel & Frank, 1999; Maher, 1984; Watts & Bryant, 2017).
STUDENTS WITH EBD AS CROSS-AGE TUTORS

**Academic outcomes.** Tutees were most frequently assessed on academic skills and showed gains in three studies (Hogan & Prater, 1993; Lane, Pollack, & Sher, 1972; Maher, 1984), moderate gains in two studies (Hamelberg, 1987; Watts & Bryant, 2017), small gains in one study (Harrigan, 1994), null effects in one study (Cochran, Feng, Cartledge, & Hamilton, 1993), and mixed outcomes in three studies (Harrigan, 1994; Scruggs & Osguthorpe, 1986; Top & Osguthorpe, 1987). When tutors were assessed for academic outcomes, across studies, they showed large gains in five studies (Holeczek, 2012; Lane et al., 1972; Maher, 1984; 1982; Top & Osguthorpe, 1987), small gains in one study (Cochran et al., 1993), and mixed outcomes in one study (Scruggs & Osguthorpe, 1986).

Breaking down the effects by instructional content, the most frequently addressed skills fell in the domains of reading, spelling, and language arts. Effects on reading skills ranged greatly from study to study with null effects to significant increases being found for both tutees and tutors (Cochran et al.; Hamelberg, 1987; Hogan & Prater, 1993; Lane et al., 1972; Maher, 1982; Scruggs & Osguthorpe, 1986; Top & Osguthorpe, 1987). Spelling outcomes ranged from no effects to small effects in the group design study (d = -.31–.25; Harrigan, 1994) and large effects were found in the single-case study (PEM = 100; Hogan & Prater, 1993). Two studies that focused on basic mathematics and number sense during tutoring sessions showed moderate to large effects for tutees (d = .68; PEM = 91.7; Watts & Bryant, 2017), and also large effects for tutors (d = 1.0; Holeczek, 2012).

**Social/emotional/behavioral outcomes.** Tutors were more frequently assessed in the areas of social skills and behavioral outcomes, and showed significant gains in seven studies (Blake, Wang, Cartledge, & Gardner, 2000; Gumpel & Frank, 1999; Holeczek, 2012; Lazerson, 1980; Maher, 1984; 1982; Watts & Bryant, 2017), moderate gains in one study (Cochran, Feng,
STUDENTS WITH EBD AS CROSS-AGE TUTORS

Cartledge, & Hamilton, 1993), and small gains in one study (Hogan & Prater, 1993). Self-concept and attitude measures were regularly administered to this population as well and findings show significant, positive changes in two studies (Hamelburg, 1987; Lazerson, 1980), moderate gains in one study (Lazerson, 2005), null effects in two studies (Cochran et al., 1993; Scruggs & Osguthorpe, 1986), and mixed outcomes in one study (Top & Osguthorpe, 1987).

Tutees showed significant differences on an attitude measure compared to a control population (Scruggs & Osguthorpe, 1986) and moderate gains on a self-assessment of their own behavior (Cochran et al., 1993). Tutees were also assessed for behavioral outcomes and showed significant gains in four studies (Blake et al., 2000; Gumpel & Frank, 1999; Hogan & Prater, 1993; Lazerson, 1980), and mixed effects in one study (Cochran et al., 1993).

**Maintenance, generalization, and social validity.** Maintenance was measured inconsistently across the included studies \( n = 7 \) and generalization of targeted skills was measured even less frequently \( n = 2 \). Studies that administered follow-up and distal measures found that effects were readily maintained (Maher, 1982) and/or generalized to other settings/skills (Blake et al., 2000; Gumpel & Frank, 1999; Hamelberg, 1987; Hogan & Prater, 1993; Holecek, 2012; Maher, 1982; 1984; Watts & Bryant, 2017).

Nine of the 15 studies measured social validity, whether qualitatively or quantitatively. Measures included participant perceptions of tutee outcomes \( n = 2 \), tutor outcomes \( n = 8 \), tutor self-assessments \( n = 6 \), and overall tutoring program ratings \( n = 8 \). Across studies, almost all social validity measures reported overall positive responses from students, teachers, parents, and trainers regarding perceived outcomes for the participants and the tutoring program in general \( n = 8 \), with the exceptions being Hamelberg (1987) and Lazerson (1980).

**Discussion**

https://mc.manuscriptcentral.com/bd
The purpose of this synthesis was to evaluate the effectiveness and related outcomes of the cross-age tutoring model when students with or at-risk for EBD serve as tutors. The findings related to the first research question showed the prevalent use of modeling, role-playing, feedback, and positive reinforcement as common instructional components in both the tutor training sessions and the tutors’ instructional procedures. Considering that these components are also common instructional features within other peer-mediated instructional models (e.g., Peer Assisted Learning Strategies; Fuchs & Fuchs, 2005; classwide peer tutoring; Greenwood, Delquadri, & Hall, 1989), these findings could provide an opportunity for creating more standardized, systematic procedures for the cross-age model with tutors with EBD. Furthermore, although it was noted that the techniques of self-monitoring and self-assessment were used infrequently, the studies that included these components showed consistently large effects in promoting the academic and behavior skills targeted for tutoring instruction (Gumpel & Frank, 1999; Hogan & Prater, 1993). These self-management techniques have been shown to be effective strategies for students with EBD (Mooney, Ryan, Uhing, Reid, & Epstein, 2005) and may prove to be beneficial supports within the training and tutoring sessions of this model moving forward.

Results and findings from 11 studies reported intervention phases of 10 weeks or less. Findings from one study (Lane, Pollack, & Sher, 1972), which had the longest duration of the intervention phase (i.e., 7 months), yielded consistent, positive outcomes across both academic and behavioral measures. This duration finding aligns with previous research findings of positive effects for intensifying intervention dosage and duration to meet the intensive needs of students with disabilities (Bryant et al., 2011; Vaughn et al., 2012). It is suggested that future research is undertaken to determine effective intervention phase durations in regards to tutor training and
tutor instructional time, and additionally, utilize more rigorous reporting procedures in order to
determine accurate relations to participant outcomes (Conn & Chan, 2015; Conn & Groves,
2011).

In addressing research question two, results from this review are similar to previous
findings, which suggest the cross-age tutoring model can be effective in promoting academic and
behavioral outcomes for both the tutee and the tutor (Robinson, Schofield, & Steers-Wentzell,
2005). Reading and spelling skills, the most frequently assessed academic outcomes, showed
varying levels of effectiveness across studies, and also across targeted skills (e.g., fluency,
comprehension), while instruction in mathematics showed the most consistent, positive outcomes
for participants across the limited number of studies focusing on this content. The findings from
this synthesis related to content area differences are comparable with the results from a previous
review of peer-mediated instruction for students with EBD, which found larger effects for
mathematics skills than reading skills (Ryan, Reid, & Epstein, 2004). Perhaps the procedural
steps for some mathematics skills are conducive to the cross-age tutoring model; however,
instruction in conceptual understanding remain with the teacher. Furthermore, additional
empirical research on cross-age tutoring with students with EBD in the areas of mathematics,
science, writing, and social studies instruction is apparent.

The targeted skills for tutors with EBD were found to be more frequently socially or
behaviorally oriented, and typically addressing either targeted negative behaviors or self-concept.
In the social or behavioral domain, the cross-age tutoring model proved to be the most effective,
as a majority of the studies reported significant gains in tutors’ social and/or behavior skills. For
tutees, cross-age tutoring appears to be equally effective in developing desired behavioral and
social skills, with consistent decreases in negative behaviors and increases in pro-social
STUDENTS WITH EBD AS CROSS-AGE TUTORS

skills/behaviors. These findings support this model’s utility in providing opportunities to practice social and behavioral skills in natural, one-on-one interactions with other students in need of behavioral supports. Henceforth, findings related to fidelity of implementation show that students with EBD are able to implement tutoring procedures with a high level of fidelity, and in conjunction with the positive findings of tutee outcomes, the cross-age tutoring model shows promising effectiveness when students with EBD serve as cross-age tutors. This finding aligns with previous research results demonstrating that students with EBD can function effectively as tutors when provided with the appropriate training and supervision (Heron, Welsch, & Goddard, 2003), and with further research, this evidence may support practitioner use of the cross-age tutoring model for providing individualized instruction to younger students as well as practice opportunities for social-behavioral skills for students with EBD.

Findings from this synthesis for generalization and maintenance of target skills demonstrated consistent and lasting impact of tutoring instruction on target skills. The infrequent measurement of these distal outcomes across studies, especially for tutors, indicates the need for future research methodologies designed to directly assess the impact of tutor training and implementation on tutors’ academic, social, and behavioral skills in generalized settings (i.e., outside of the tutoring environment), as these are commonly identified areas of deficit for students with EBD. Furthermore, for tutors, the theory that providing the responsibility of tutoring another student in order to elicit gains in self-concept and/or self-esteem has mixed support in this review’s findings (Allen, 1976; Hogan & Tudge, 1999). These results could be related to the sensitivity of the emotional self-assessment measures used in the earlier studies, which contained overly general characteristics (e.g., anger, friendliness, studiousness), making it difficult to determine emotional change directly related to performing as a cross-age tutor.
(Hamelberg, 1987; Lazerson, 1980; Scruggs & Osguthorpe, 1986; Top & Osguthorpe, 1987). Interestingly, compared to tutors, tutees showed consistent gains in self-concept and positive attitude, which may be related to the uniqueness of the model, where attention, reinforcement, and consequently, motivation, are facilitated by an older student with a disability rather than the typical adult/teacher model (Jun, Ramirez, & Cumming, 2010; Topping & Ehly, 1998).

**Study Limitations**

Findings pertaining to the fifth research question (i.e., What is the overall methodological quality and rigor of the included studies?) must be considered in the interpretation of the results of this synthesis, as the overall quality of the studies included in a systematic review will affect the comprehensiveness of the findings. The majority of the studies reviewed for this synthesis were published before 2005 when recommendations for the indicators for quality in research studies were initially developed (Gersten et al., 2005; Horner et al., 2005). Included studies with low quality scores (i.e., less than 75%) frequently omitted descriptions of any specific training (e.g., amount of training, training to a criterion) or qualifications (e.g., professional credential) required to implement the intervention, reports of reliability (e.g., internal, interobserver, test-retest, parallel-form), and procedures for measuring implementation fidelity. Taking into account the variability in methodological rigor, this review shows promise for cross-age tutoring models utilizing students with EBD as tutors as an evidence-based practice. Henceforth, it is suggested that future research is undertaken in alignment with rigorous quality standards for design and reporting (i.e., Council for Exceptional Children, 2014). Moreover, considering the limited number of studies identified, and that half of these studies were published prior to 1990, it is possible that the findings of this review may not provide an accurate representation of the current population of students with disabilities, and thus may not attain external validity. Similar
finding were found in a related review showing a decline in the number of peer-mediated intervention studies for students with EBD in recent years, including the use of the cross-age model (Ryan, Reid, & Epstein, 2004). Although the number of studies devoted to the cross-age tutoring model containing tutors with EBD is limited, a majority of these studies showed moderate to large academic and/or behavioral effects for participants.

Another limitation is that researchers rather than school personnel were the most common implementers of the tutor training and supervision during the tutoring sessions. Overall social validity outcomes showed practitioners’ positive perceptions of the effectiveness and benefits of the model, and in the two studies where teachers reported mixed perceptions of student outcomes (Hamelberg, 1987; Lazerson, 1980), they still described the tutoring intervention as beneficial and stated that they would continue to utilize the model. These findings are promising, as when teachers and students perceive a tutoring model positively and see it as effective, they are more likely to continue using the practice (Zimmerman & Risemberg, 1997). To determine whether this model, with this unique population of students serving as tutors, is feasible and effective for practitioner implementation, further evaluations need to be undertaken.

By analyzing the findings of 15 cross-age tutoring interventions containing students with EBD as tutors, this synthesis makes a necessary contribution to the field of special education research and the literature base of peer-mediated interventions. Although findings show students with EBD to be able and effective cross-age tutors, and that the instructional model can facilitate positive outcomes for both tutee(s) and the tutor(s), there is still much research to be undertaken. The small number and varying rigor of the included studies provides useful insight into the need for more comprehensive research in determining if the cross-age instructional model provides unique advantages for tutors with EBD, such as promoting generalized improvements in deficit
social, emotional, and/or behavioral areas, in relation to more formalized, same-age peer support systems. Finally, in order to identify the extent of model’s utility, further empirical research is required to assess effectiveness under practitioner implementation and supervision, and also in providing instruction in content areas other than reading and spelling domains. These future areas of research will deepen our understanding of the potential of the cross-age tutoring model for tutors with EBD and their tutees.
References

*References marked with an asterisk indicate studies included in the synthesis.


STUDENTS WITH EBD AS CROSS-AGE TUTORS


For Peer Review

STUDENTS WITH EBD AS CROSS-AGE TUTORS


STUDENTS WITH EBD AS CROSS-AGE TUTORS


STUDENTS WITH EBD AS CROSS-AGE TUTORS


doi:10.1007/s10864-005-6298-1


doi:10.3102/0013189X13507104


STUDENTS WITH EBD AS CROSS-AGE TUTORS


*Watts, G. W., & Bryant, D. P. (2017). Promoting number sense through number line board games delivered by older students with EBD (Unpublished pilot study).


doi:10.1006/ceps.1997.0919
### Table 1

Study Characteristics

<table>
<thead>
<tr>
<th>Study &amp; research design</th>
<th>Setting, implementer (I), &amp; quality score (Q)</th>
<th>Tutors (T)/tutees (t)</th>
<th>Tutor training&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Tutoring sessions&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Dependent measure: T/t target skills</th>
<th>Maintenance (Ma), generalization (G), &amp; social validity (SV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blake et al. (2000)</td>
<td>Public, suburban self-contained special ed school I = researcher Q = 90.9%</td>
<td>T: n = 3, ages 11–13 yrs, Grades 6 &amp; 7, challenging behaviors</td>
<td>scripted social skills lessons &amp; materials, practice, feedback</td>
<td>5 sessions, 30 min each; 7 wks;</td>
<td>T &amp; t: Frequency count 93.5, large; (b) PEM = 100;</td>
<td>T: Ma (a) PEM = 97%</td>
</tr>
<tr>
<td>[Experiment 2], multiple baseline</td>
<td></td>
<td>t: n = 3, ages 9–10 yrs., Grades 3 &amp; 4, difficulties relating to peers</td>
<td></td>
<td></td>
<td>t: (a) PEM = large</td>
<td>t: Ma</td>
</tr>
</tbody>
</table>

<sup>a</sup> Tutor training

<sup>b</sup> Tutoring sessions

ES = Effect size; PEM = Percentage of meaningful change; SV = Social validity.
**STUDENTS WITH EBD AS CROSS-AGE TUTORS**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Interventions</th>
<th>Effect Sizes</th>
<th>Statistical Observations</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Cochran et al. (1993)             | Large, urban self-contained elementary school with/at-risk for EBD/struggling readers | T: $n^T = 4$, $n^C = 4$; 5 sessions, NR  
C: $n^C = 4$, Grade 5, 8 boys, 5 min, 5 days; 32 sessions, 28–30 min each, 8 wks; T & t: (a) CBM-sight words; (b) SSRS-S; (c) SSRS-T  
F = NR  
Ma/G = NR  
T: +/+ (self/program)  
T/t: +/+(T/t/program) | T: $T > C$, $d = .28$, small;  
C: $T = C$, $d = .05$, NE;  
T: $T > C$, $d = 1.1$, large;  
C: $T > C$, $d = .56$, moderate;  
T: $T > C$, $d = .54$, moderate | SV = Yes,  
t: $T < C$, $d = -.47$, NE;  
C: $d = -.19$,  
C: $d = .18$, NE;  
C: $d = .01$, NE | T: +/+ (self/program)  
T/t: +/+(T/t/program) |
STUDENTS WITH EBD AS CROSS-AGE TUTORS

Gumpel & Frank (1999), multiple baseline multiple baseline

Elementary school

T: n = 2, ages 11–12 yrs, Grade 6, 2 boys, at-risk for EBD, socially rejected & isolated. Grade K, at-risk for EBD, socially rejected & isolated.

Q = 84.1%

6–8 sessions, 45 min each, 1.5 wks; training on MSC procedures:

- Identify appropriate social stimuli,
- State possible behavioral options,
- Enter into that social interaction,
- Self-monitor, & observe

T: n = 2, age 5 yrs., training on MSC procedures:

- Modeling, role-playing, & discussion of tutee's self-monitoring sheet

4 times/wk; MSC procedures:

- Momentary time sampling

(a) No social interaction

(a) PEM = 97.5, large; (a) PEM = 100, large

(b) Positive social interaction

(b) PEM = 100, large; (b) PEM = 100, large

SV = Yes, Te: + (program)

T: (a) PEM = 100, large; (b) PEM = 100, large

T: Ma (a) PEM = 100, large; (b) PEM = 91.7, large

NE; (d) M change = +13.94

(e) M change = -1.4
## STUDENTS WITH EBD AS CROSS-AGE TUTORS

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Tutor Group</th>
<th>Student Group</th>
<th>Conditions</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamelberg (1987)</td>
<td>Public, suburban elementary school in central Ohio, resource &amp; computer room</td>
<td>4 years, males</td>
<td>3, ages 16–18 yrs, males, EBD</td>
<td>Multiple baseline</td>
<td>t: (a) Sight words read aloud, 29 possible sessions; t: (b) PHSC self-concept percentiles, 20 min each, 10 wks; goals/objectives, instructional steps, data collection, role-playing, review progress monitoring</td>
</tr>
<tr>
<td>Harrigan (1994)</td>
<td>Private, non-profit school for children w/ EBD &amp; neurological impairments in New Jersey</td>
<td>Treatment/comparison, RCT</td>
<td>10, ages 13–16 yrs, 16 boys, 4 girls, EBD &amp; neurological impairments in New Jersey</td>
<td>1 session, 30 min</td>
<td>t: CBM spelling word list</td>
</tr>
</tbody>
</table>

Note: I = researcher, Q = 86.4%
STUDENTS WITH EBD AS CROSS-AGE TUTORS

& neurological verbal F = NR
impairments reinforcement)
& token
reinforcement

T2: received 1
additional 30-
min session in
spelling method
based on
deCharms
(1972)

C: BAU

Hogan & Prater (1993),
Public, rural high
school in U.S.
multiple baseline

T: n = 1, age 15 yrs,
boy, EBD
3 days, min NR;
individual practice, error
t: n = 1, age 14 yrs,
boy, LD
interview,
correction,
provided modeling, positive
Q = 70.5%
procedures, reinforcement,
modeling, role-
playing, & record progress
behavior self-

4 days/wk, 15 min;
t: (a) Frequency count disruptive
t: (b) Time
PEM = 100, large; (c) PEM =
t: (a) PEM =
t: Ma/G
(b) PEM = 100,
large;
(c) PEM =
large;
(d) PEM =

4 days/wk, 15 min;
t: (b) Time
PEM = 100,

4 days/wk, 15 min;
t: (a) PEM =
t: Ma/G
(b) PEM = 100,
large;
(c) PEM =
large;
(d) PEM =

4 days/wk, 15 min;
t: (b) Time
PEM = 100,

4 days/wk, 15 min;
t: (a) PEM =
t: Ma/G
(b) PEM = 100,
large;
(c) PEM =
large;
(d) PEM =

4 days/wk, 15 min;
t: (b) Time
PEM = 100,

4 days/wk, 15 min;
t: (a) PEM =
t: Ma/G
(b) PEM = 100,
large;
(c) PEM =
large;
(d) PEM =

4 days/wk, 15 min;
t: (b) Time
PEM = 100,

4 days/wk, 15 min;
t: (a) PEM =
t: Ma/G
(b) PEM = 100,
large;
(c) PEM =
large;
(d) PEM =

4 days/wk, 15 min;
t: (b) Time
PEM = 100,
STUDENTS WITH EBD AS CROSS-AGE TUTORS

monitoring F = NR scores, (d)
training vocabulary test scores

Holecek (2012)\textsuperscript{a}, Middle/high school in the Midwest, pre/post test, single group
T: n = 3, Grade 1 or more above EBD min, 1 wk; 2 days/wk, 30 min, 6 wks; content from tutees’ math class T: (a) CBM basic math skills, (b) large T: G
F = NR
T: (a) d = 1.0, (b) d = 2.8, large
Q = 25.0%
T: (b) time engaged during math class

Lane et al. (1972), Private, urban middle school for students w/ EBD, pre/post test, single group
T: n = 8, Grades 8–9, EBD & struggling readers; IRM 2 sessions/wk, T & t: (a) MAT T: (a) M gain (in mos) = 19; SV = NR
T: (b) BBRS large
F = NR
T: (b) d = 3.4,
Q = 52.3%
T: (a) $M$ gain (in mos) = 14

F = NR
<table>
<thead>
<tr>
<th>Lazerson (1980), Lab school, university campus, Buffalo, NY</th>
<th>T: ( n^t = 20 ), ( n^c = 20 ), 2 sessions, NR</th>
<th>5 sessions/wk, 20–30</th>
<th>T &amp; t: (a) DESB behaviors, (b) LSSC self-concept</th>
<th>T &amp; t: (a) ( t &gt; \frac{13.7}{}, p &lt; .005, df = 38 ), SS; (b) ( t = 11.29, p &lt; .005, df = 38 ), SS</th>
<th>Ma/G = NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazerson (2005), Public school, western New York, self-contained classroom</td>
<td>T: ( n = 3 ), ages: 15–16 yrs, 2 boys, 1 girl, EBD &amp; LD</td>
<td>2 sessions, NR</td>
<td>2–5 sessions/wk, 45–60 min, 3 mos; instruction on basic reading comprehension/decoding &amp; math instruction on basic reading comprehension/decoding &amp; math</td>
<td>T: ( d = .23 ); moderate Ma/G = NR</td>
<td>SV = Yes, Te: +/+ (self/program); Te: + (tutors)</td>
</tr>
</tbody>
</table>

I = researcher Q = 47.7% I = NR Q = 27.3%
### STUDENTS WITH EBD AS CROSS-AGE TUTORS

**F = NR**

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Participants</th>
<th>Intervention</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maher (1982), treatment/comparison, RCT</td>
<td>Public urban high school, New Jersey</td>
<td>T: $n_T^1 = 6$, $n_C^1 = 6$, 18 yrs, 11 boys, 7 girls, ages 15–18 yrs; I = School psychologist &amp; SPED teacher; Q = 75.0%</td>
<td>T: Instruction &amp; practice on lesson planning/wk (15–20 min) with SPED teacher, planned &amp; provided social studies teaching, planned &amp; provided instruction on reading &amp; writing, planning filing for instructional goals; C1: Received peer tutoring</td>
<td>T: (a) grades $F(4, 30) = 6.41$ ($p &lt; .01$); SS; (b) $T &gt; C$, $F(4, 30) = 11.12$ ($p &lt; .01$), SS; (c) discipline $F(4, 30) = 8.90$ ($p &lt; .01$), SS; (d) % of days in attendance $&gt; C$, $F(4, 30) = 20.50$ ($p &lt; .01$), SS</td>
</tr>
<tr>
<td>Maher (1984), multiple baseline</td>
<td>Public urban school district, New Jersey</td>
<td>T: $n_T = 8$, ages 14–16 yrs, EBD; I = School psychologist &amp; SPED teacher; Q = 75.0%</td>
<td>T: Purpose/setting/evaluating instructional goals in reading &amp; writing content/skills, attendance; C1: Received peer tutoring; C2: Received group counseling</td>
<td>T: (a) PEM = 94.7, large; (b) PEM = 88.3, large; (c) PEM = 100, large</td>
</tr>
</tbody>
</table>

https://mc.manuscriptcentral.com/bd
STUDENTS WITH EBD AS CROSS-AGE TUTORS

Q = 86.4% about problems/concerns, role-playing of planning meetings w/ SPED teachers, role-playing of tutor sessions & performance feedback.

Scruggs & Osguthorpe (1986) [Experiment 1], Treatment/comparison, schools in the west

Public, rural elementary public schools in the west Grades 2–6, LD & EBD

T: Tutoring procedures, session 2–5 days/wk, 30 min, tutoring, session 1–2 sessions 10 wks; beginning reading curriculum (Harrison, 1979), directly supervised by project staff

T: n^T = 13, n^C = 20, Grades 1–5, LD & EBD

I = Project staff

Q = 65.9% EBD

Ma/G = NR

SV = NR

T & t: (a) T & t = C, NR, NS; (b) T & t > C, (c) T & t > C, t(33) = 2.46 (p < .05), SS

T: (c) T = C, t(13) < 1, NS

(b) PEM = 82.1, moderate

(c) PEM = 90.2, large; (b) PEM = 93.8, large; (a) PEM = 86.4%

T: +/+ (self/program)

Te: +/- (t/program)

SV = Yes,
<table>
<thead>
<tr>
<th>Top &amp; Osguthorpe</th>
<th>Large, suburban elementary school district</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment/comparison, RCT</td>
<td>Grades 4–6, LD &amp; EBD</td>
</tr>
<tr>
<td>I = teaching assistant</td>
<td>Grade 1, at risk for reading LD</td>
</tr>
<tr>
<td>Q = 72.7%</td>
<td>for reading LD</td>
</tr>
</tbody>
</table>

T: $n^T = 39$, $n^C = 39$, 3 sessions, 60 min each; 4 days/wk, 15–20 min, 14 wks; instruction on reading skills (phonics/ sight words); t: Reading skills, daily performance; C: BAU + additional instructional time to equal 4 days/wk, 15–20 min, 14 wks; instruction on reading skills (phonics/ sight words); paraprofessional recorded tutees’ positive reinforcement & progress monitoring methods.

C: BAU

T: (a) WJ T: (a) $T > C$, $d = .94$, large; $Ma/G = NR$

T: Procedures (Osguthorpe, 1984): modeling, prompting, positive reinforcement & progress monitoring methods

T: Self-concept (b) PHSC, (c) (b) $T = C$, $d = .15$, NE; (c) $T > C$, $d = .45$, small; (d) $T = C$, $d = .00$, NE

t: (a) t = C, $d = .16$, NE; (e) t $C$, $d = .86$, large

Ma/G = NR SV = NR
<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>School</th>
<th>Grades</th>
<th>Ages</th>
<th>Number of Sessions</th>
<th>Duration</th>
<th>Tutors</th>
<th>Procedural Training</th>
<th>Feedback</th>
<th>Reinforcement</th>
<th>Progress Monitoring</th>
<th>Fidelity</th>
<th>Effect Size</th>
<th>Overall Positive Responses</th>
<th>Overall Negative Responses</th>
<th>Teacher/Paraprofessional/Staff/Child-Care Worker</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple baseline</td>
<td>Charter elementary school, Texas, resource room</td>
<td>4–5, ages 9–11 yrs, at-risk EBD</td>
<td>3, ages 9–11 yrs</td>
<td>2 sessions, 45 min each</td>
<td>3 sessions/wk, 25 min</td>
<td>(a) partial</td>
<td>number line board game</td>
<td>corrective feedback, modeling, role-playing</td>
<td>positive interactions/behaviors</td>
<td>(b) TEMI, (c) TEMA-3</td>
<td>(a) PEM = 96.3, large</td>
<td>(b) PEM = 91.7, large; (c) d = .68, moderate</td>
<td>+ / + (self/program); + / + (tutors/program)</td>
<td>+ / + (tutors/program)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>I = Researcher</td>
<td>EBD</td>
<td>6, ages 5–7 yrs, at-risk math LD</td>
<td>3, ages 5–7 yrs</td>
<td>3 sessions/wk, 25 min</td>
<td>25 min each</td>
<td>(a) PEM = 96.3, large</td>
<td>number line board game</td>
<td>corrective feedback, modeling, role-playing</td>
<td>positive interactions/behaviors</td>
<td>(b) TEMI, (c) TEMA-3</td>
<td>(a) PEM = 91.7, large; (c) d = .68, moderate</td>
<td>+ / + (self/program); + / + (tutors/program)</td>
<td>+ / + (tutors/program)</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. WTSSC = Working Together Social Skills Curriculum; F = fidelity; PEM = percentage of data points exceeding the median; + = overall positive responses; – = overall negative responses; Te = teacher/paraprofessional/school staff/child-care worker; T/T = treatment/tutoring group;
STUDENTS WITH EBD AS CROSS-AGE TUTORS

C/C = comparison group; EBD = emotional-behavioral disorder; NR = not reported; CBM = curriculum-based measure; SSRS-S = Social Skills Rating System; SSRS-T = Social Skills Rating System-Teacher version; NE = no/negative effect; MSC = Model of Social Competence; LD = learning disability; PHSC = Piers-Harris Children’s Self-Concept Scale; RCT = ?; Tr1 = ?; Tr2 = ?; BAU = business as usual; ID = intellectual/cognitive disability; IRM = Intersensory Reading Method; MAT = Metropolitan Achievement Test; BBRS = Burk’s Behavior Rating Scale; DESB = Devereaux Elementary School Behavior Rating Scale; LSSC = Luszki & Shmuck Self-Concept Scale; SS = statistically significant; IEP = individualized education program; SPED = special education; WJ = Woodcock-Johnson Test of Achievement; BR-I = Beginning Reading Criterion Test; NS = non-statistically significant; SPAS = Students’ Perception of Ability Scale; ISC = Inferred Self-Concept Scale; TEMI = Texas Early Mathematics Inventories; TEMA-3 = Test of Early Mathematics Ability–Third Edition.

aFrequency, duration, components. bFrequency, duration, components, tutor supports, fidelity (F). cDissertation/unpublished.
### Table 2

**Summary of Common and Unique Components of the Cross-Age Tutoring Models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Sessions (n)</th>
<th>Avg. session length (range)</th>
<th>Avg. training/tutoring program duration (range)</th>
<th>Session frequency</th>
<th>Common features&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Unique features&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutor training</td>
<td>1–8</td>
<td>45 min (30–60 min)</td>
<td>177.5 min (30–360 min)</td>
<td></td>
<td>• Procedures/objectives</td>
<td>• Greeting tutees (7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(60%)</td>
<td>• Proximity to tutee</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Modeling (47%)</td>
<td>(7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Role-playing (80%)</td>
<td>• Review of skills (7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Corrective feedback (60%)</td>
<td>• Individual interviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Positive reinforcement (60%)</td>
<td>(7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Scripted lessons (7%)</td>
<td>• Lesson planning (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Goal setting (13%)</td>
<td>• Self-monitoring (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Data collection (20%)</td>
<td></td>
</tr>
</tbody>
</table>
STUDENTS WITH EBD AS CROSS-AGE TUTORS

- Token reinforcement (7%)
- Specific behavioral or academic strategies (20%)
- Modeling (54%)
- Corrective feedback (54%)
- Positive reinforcement (47%)
- Planning sessions w/ special educators (13%)
- Reteaching/retraining session (13%)
- Follow-up conferences (7%)
- Tangible reinforcers (7%)

<table>
<thead>
<tr>
<th>Tutoring session</th>
<th>20–30 min (15–60 min)</th>
<th>5 wks–7 mos</th>
<th>2–5/wk</th>
<th>Modeling (54%)</th>
</tr>
</thead>
</table>

*Percentage of included studies.*
## STUDENTS WITH EBD AS CROSS-AGE TUTORS

### Table 3

Summary of Participant Characteristics, Outcomes, and Common Effect Size Grades

<table>
<thead>
<tr>
<th>Group</th>
<th>Participants (n)</th>
<th>Disability/at-risk category</th>
<th>Academic overall&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reading/literacy&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Mathematics&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Social-behavioral&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Self-concept/attitude&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutees</td>
<td></td>
<td></td>
<td>SCD Group</td>
<td>SCD Group</td>
<td>SCD Group</td>
<td>SCD Group</td>
<td>SCD Group</td>
</tr>
<tr>
<td></td>
<td>Elementary = 163</td>
<td>LD = 90</td>
<td>Moderate/ Mixed</td>
<td>Moderate/ Mixed</td>
<td>Large/ Moderate</td>
<td>Large/ Mixed</td>
<td>N/A/ Moderate</td>
</tr>
<tr>
<td></td>
<td>Middle school = 41</td>
<td>Comorbid = 80</td>
<td>large (.01–.86)</td>
<td>large (-.3–.86)</td>
<td>(91.7) (.68)</td>
<td>(93.5–.19)</td>
<td>(.26–SS)</td>
</tr>
<tr>
<td></td>
<td>High school = 1</td>
<td>EBD = 25</td>
<td>(88.2–100)</td>
<td>(88.2–100)</td>
<td></td>
<td>100 (.26)</td>
<td>(.26–SS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ID = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutors</td>
<td>Elementary = 133</td>
<td>Comorbid</td>
<td>Large/ Large</td>
<td>N/A/ Large</td>
<td>Large/ Large</td>
<td>Large/ Mixed</td>
<td>Large/ Mixed</td>
</tr>
<tr>
<td></td>
<td>Middle school = 45</td>
<td>EBD+ = 150</td>
<td>(88.3–94.7) (.28–1.0)</td>
<td>(28–94)</td>
<td>(1.0)</td>
<td>(42.9–.56)</td>
<td>(.00–1.6)</td>
</tr>
<tr>
<td></td>
<td>High school = 53</td>
<td>EBD = 81</td>
<td></td>
<td></td>
<td></td>
<td>100 (.34)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* SCD = single-case design; Group = group design; LD = learning disability; EBD = emotional-behavioral disorder; ID = intellectual disability; SS = statistically significant effect.

<sup>a</sup>Most frequent/common effect size interpretation across included studies (range of effect sizes; SCD = PEM; Group = Cohen’s d).
Figure 1. Inclusion flow diagram illustrating the results of the literature search and inclusion process. Note. IRR = Interrater reliability.