

Social Skills Training for Students With Emotional and Behavioral Disorders: A Review of Reviews

John W. Maag
University of Nebraska–Lincoln

ABSTRACT: Teaching social skills to students with emotional and behavioral disorders (EBD) has become an accepted practice. Literally hundreds of social skills training (SST) efficacy studies for students with EBD appear in the literature. As a result, many authors have published both narrative and meta-analytic reviews of the literature. Reviews have highlighted various problem areas as targets for future research. Nevertheless, SST has subsequently often resulted in only modest changes in the social competence of students with EBD. The purpose of this article is to review the reviews on SST with students with EBD, discuss issues based on conclusions reached, and present implications for practice.

■ The acquisition and performance of prosocial behaviors is part of normal human development (Alexander & Entwisle, 1988). Youths who lack social competence have been at risk for many difficulties, including, but not limited to, aggression, rejection by peers, academic failure, loneliness, social dissatisfaction, difficulty maintaining employment and relationships with others, mental illnesses, and contact with the legal system (Parker & Asher, 1987). Consequently, many social skill training (SST) studies have appeared in the literature during the past 25 years as a way to lessen the impact of these deleterious outcomes.

SST has been conducted with a variety of youngsters throughout the years, including those with mental illnesses; adolescents who have been adjudicated; students who are unpopular, rejected, or neglected; and students with disabilities (e.g., Forness, Kavale, Blum, & Lloyd, 1997; Nanyang & Hughes, 2002; Parker & Asher, 1987). Lack of social competence has been considered characteristic of students with emotional and behavioral disorders (EBD) (e.g., Kauffman, 2005). Many authors have discussed the relevance, effectiveness, and caveats of SST for students with EBD (e.g., Gresham, 1997, 1998; Mathur & Rutherford, 1996; Smith & Travis, 2001; Strain, 2001).

Reviewing the research on SST has become a daunting task simply because of the sheer number of published studies in existence. Therefore, it may be more judicious to review the reviews. In addition, most reviewers have pointed out methodological flaws and made recommendations for future research. A review

of reviews may provide clues about why some mythological flaws and recommendations have been addressed and others have been ignored. Reviews of SST began appearing in the mid-1980s (e.g., Ladd, 1984) and continue (e.g., Kavale, Mathur, Forness, Rutherford, & Quinn, 1997; Mathur, Kavale, Quinn, Forness, & Rutherford, 1998). Their foci have been vast and—besides different populations—included methods of skill selection, assessment techniques, training approaches, outcome evaluation, and lack of generalization (e.g., Landrum & Lloyd, 1992; McIntosh, Vaughn, & Zaragoza, 1991; Zaragoza, Vaughn, & McIntosh, 1991).

A particularly troubling issue has been the heterogeneity of participants who received SST under the umbrella terms “emotional and behavioral disorder” or “seriously emotionally disturbed.” These terms, for better or worse, are used to classify students as being eligible to receive special education services under federal and state guidelines (Forness & Kavale, 2000). Participants in many studies, however, did not meet any federal or state eligibility criteria and, instead, were those who had been diagnosed with psychiatric disorders (e.g., conduct disorder, oppositional-defiant disorder), adjudicated (e.g., juvenile delinquents), “at-risk,” or simply nominated by their teachers as having behavior problems (Maag, 2005). Consequently, all conclusions about SST for students with EBD are, to a certain extent, problematic.

Although heterogeneity of participants makes it difficult for interpretation, there are some constants. For example, Gresham

(1998) reached several conclusions from his analysis of nine reviews of the SST literature (six narrative, three meta-analytic). First, little correspondence existed between behaviors that were assessed and those targeted for intervention. Second, researchers often failed to match SST interventions to identified deficits. Third, there has been a long-standing failure of researchers to program generalization into SST studies. These conclusions were reached, however, from analyzing reviews conducted on various populations. To date, no literature has focused exclusively on conclusions garnered from reviews specifically targeting students with EBD. Therefore, this article presents the results from reviewing the reviews on students with EBD. It also discusses issues based on conclusions reached and presents implications for practice.

Before proceeding, however, it is important to set the groundwork by differentiating between *social skills* and *social competence*. Social skills are the specific behaviors targeted in SST. Social competence is a general idiom referring to the adequacy of a person's social functioning; it is typically inferred when the targeted social skills result in increased ratings of acceptance from peers and positive judgments from important others (i.e., teachers, parents, community leaders) in a youth's life (Gresham, 1998).

Results From SST Reviews

Researchers have been investigating the effects of SST since Bandura (1977) conceptualized the skill deficit model—much of it resulting from his early investigations of modeling in the 1960s. A plethora of SST research articles with students with EBD began appearing in the late 1970s and early 1980s. Ladd (1984) was one of the first SST review articles. Since that time, there have been so many reviews that researchers have begun reviewing the reviews. For example, Forness et al. (1997) conducted what they called a "mega-analysis"—analyzing the results of meta-analyses on various interventions used in special education, including SST. Their conclusion was that SST had largely been ineffective. Besides Gresham's (1998) three conclusions described previously, he also found that the effect sizes (ESs) for the three meta-analytic reviews averaged only 0.35—a level associated with weak to moderate outcomes. The findings from these two reviews that SST, at

best, has been only marginally successful may be partially attributed to conceptual flaws in its implementation, which have been described in great detail (Maag, 2005).

It may be premature to abandon SST—especially for students with EBD. Myriad methodological nuances affect the outcome of a study, including, but not limited to, subject selection criteria, fidelity of SST implementation, type of dependent and outcome measures (specifying vs. impact), and experimental design (single subject vs. group). Therefore, this section gives a descriptive analysis of SST reviews focusing solely on students with EBD.

The following parameters were used for selecting reviews: (a) published in peer reviewed journals, (b) focused on youngsters with EBD, and (c) used school-based participants. Reviews were located using four procedures. First, a search of the social skills training data base was conducted using multiple combinations of the following descriptors: reviews/critiques of social skills training, emotional or behavioral disorder, and seriously emotionally disturbed. This search began in 1980, several years before Ladd's (1984) review appeared, and continued to October 2005. Second, an ancestor search was conducted of all reviews that met inclusion criteria. Third, references from these reviews were examined. Fourth, key journals in the area of EBD were scrutinized for titles related to SST: *Behavioral Disorders*, *Journal of Emotional and Behavioral Disorders*, *Journal of Clinical Child Psychology*, *Journal of Abnormal Child Psychology*, *Education and Treatment of Children*, *Journal of Child Behavior Therapy*, and *Journal of Behavioral Education*.

A total of 13 reviews were obtained. *Table 1* presents a summary of the reviews obtained: The reviewers, type of review, number and type of study, and major conclusions. This section presents results from the reviews.

Type of Review

Three types of review formats were evident. Nine reviews were narrative (Ager & Cole, 1991; Coleman, Wheeler, & Webber, 1993; Holinger, 1987; Landrum & Lloyd, 1992; Mathur & Rutherford, 1991; Olmeda & Kauffman, 2003; Schloss, Schloss, Wood, & Kiehl, 1986; Templeton, 1990; Zaragoza et al., 1991), three were meta-analytic (Beelmann, Pflingsten, & Losel, 1994; Mathur et al., 1998; Quinn, Kavale, Mathur, Rutherford, & Forness, 1999), and one was "quantitative" (Singh, Deitz, Epstein, & Singh, 1991). This last review

TABLE 1
Summary of Reviews of SST With Students With EBD

<i>Review</i>	<i>Type of Review</i>	<i>Number and Type of Design</i>	<i>Major Conclusions</i>
Ager & Cole (1991)	Narrative	22 studies 16 group design 6 single-subject design	<ul style="list-style-type: none"> • Inadequate treatment specification • Lack of investigation on different treatment outcomes • Lack of generalization/maintenance • Need to determine optimum ranges of training time necessary
Beelmann et. al. (1994)	Meta-analytic	49 studies All group design	<ul style="list-style-type: none"> • Lower ESs than those obtained for general treatment evaluation for children • Lack of generalization • Lack of process analysis to clarify how children change and breadth of modifications • Greatest effects for at-risk groups, lowest effects for externalizing disorders
Coleman et. al. (1993)	Narrative	9 studies All group design	<ul style="list-style-type: none"> • Social problem-solving can be learned by a variety of students • Little impact on observed behaviors • Lack of generalization • Need to individualize training
Hollinger (1987)	Narrative	15 studies Design type not specified	<ul style="list-style-type: none"> • Need to consider influence of peers and their perceptions • Failure to produce consistent treatment effects • Social perception bias among peers • Need to conduct training in peer group
Landrum & Lloyd (1992)	Narrative	12 studies Design type not specified	<ul style="list-style-type: none"> • Researchers rely on least analytic approaches to generalization • Generalized responding inadequately explored • Studies provide no analysis of presence or absence of generalization • Lack of assessing transfer effects of self-management techniques
Mathur et. al. (1998)	Meta-analytic	64 studies All single-subject design	<ul style="list-style-type: none"> • Effects modest with a PND of 62% • Social interaction skills more amenable to treatment than social communication skills • Delinquents benefited more from training than students with autism or EBD • SST not adapted to preschoolers
Mathur & Rutherford (1991)	Narrative	21 studies 4 group design 17 single-subject design	<ul style="list-style-type: none"> • Increase in use of generalization techniques but lack of examining efficacy • Need for systematic peer training • Lack of ways to determine cost-effectiveness of training outcomes • Need to examine social significance of behavioral change
Olmeda & Kauffman (2003)	Narrative	7 studies 6 group design 1 single-subject design	<ul style="list-style-type: none"> • Limited reporting of cultural, linguistic, and community characteristics • Need description of trainer characteristics • Lack of family involvement • No differences in efficacy based on ethnicity
Quinn et. al. (1999)	Meta-analytic	35 studies All group design	<ul style="list-style-type: none"> • Obtained small ESs • Students with EBD require specially designed, individualized instruction • Type of social skill deficit not considered • Interventions not implemented long enough to be effective • Need for formative evaluation

TABLE 1 (Continued)
Summary of Reviews of SST With Students With EBD

<i>Review</i>	<i>Type of Review</i>	<i>Number and Type of Design</i>	<i>Major Conclusions</i>
Schloss et. al. (1986)	Narrative	25 studies Design type not specified	<ul style="list-style-type: none"> • Lack of accepted conceptualization of social competence • Lack of attention to specific subject characteristics • Social validity ignored • Need for demonstrating integrity of training procedures • Lack of generalization
Singh et. al. (1991)	Quantitative	28 studies 2 group design 26 single-subject design	<ul style="list-style-type: none"> • Lack of subject classification criteria • Lack of functional analysis • Over-reliance on behavior-reduction procedures • Unsystematic use of training procedures • Lack of impact measures • Lack of individual-specific assessments
Templeton (1990)	Narrative	5 studies Design type not specified	<ul style="list-style-type: none"> • Evidence for efficacy inconclusive • Group format beneficial • Participants need cognitive ability to apply skills • Lack of instilling values during training • Lack of generalization/maintenance
Zaragoza et. al. (1991)	Narrative	27 studies Design type not specified	<ul style="list-style-type: none"> • Studies reported improvements • Lack of assessing subjects' type of social difficulty and matching training • Lack of positive changes in peer status • Subjects feel better about themselves after training

used a method of developing coded categories across studies that could be numerically summarized.

One review focused solely on single-subject design studies (Mathur et. al., 1998) and four focused solely on group design studies (Beelmann et. al., 1994; Coleman et. al., 1993; Quinn et. al., 1999; Templeton, 1990). Four others explicitly stated that both single-subject and group designs were reviewed (Ager & Cole, 1991; Mathur & Rutherford, 1991; Olmeda & Kauffman, 2003; Singh et. al., 1991). A final set of four reviews did not indicate the design type in either narrative or tables (Hollinger, 1987; Landrum & Lloyd, 1992; Schloss et. al., 1986; Zaragoza et. al., 1991). It is likely, however, that these last four reviews included both single-subject and group design studies. This conclusion was reached by examining tables for number of subjects (ranging from 1 to more than 100) and study reference (e.g., *Journal of Applied Behavior Analysis* would indicate a single-subject design whereas the *Journal of Consulting and Clinical Psychology* would reflect group designs). There is, however, a

possibility of error when extrapolating design type when it was not specifically stated.

Subject Characteristics

All but two reviews had in their title the following descriptors: students with emotional and behavioral problems, behavior-problem adolescents, students/youths with emotional or behavioral disorders, children with behavior problems, behaviorally disordered students/children, students who are seriously emotionally disturbed, or children/youth/adolescents with behavioral disorders. Two reviews were obtained from PsycINFO using the search terms described previously (Beelmann et. al., 1994; Coleman et. al., 1993). Nevertheless, both reviews had tables indicating subjects were youngsters with some type of emotional or behavioral disorder.

Titles, however, can be misleading. Upon examination of subject inclusionary criteria, only two reviews focused solely on students who were labeled seriously emotionally disturbed (SED) according to federal or state

special education eligibility criteria (Olmeda & Kauffman, 2003; Quinn et. al., 1999). Another review indicated that studies were included where participants were either formally (i.e., met federal or state SED criteria) or informally labeled as EBD (Singh et. al., 1991); however, nowhere in the review was "informally" ever defined. An additional review used the term EBD to describe subjects but never defined what was meant (Landrum & Lloyd, 1992). The remaining nine reviews included studies in which subjects were identified as displaying disruptive behavior, conduct disorders, autism, delinquency, or externalizing or internalizing syndromes; being rated by teachers or peers as aggressive, maladjusted, rejected, or neglected; having low social interaction; and being hospitalized psychiatric patients.

It is difficult to generalize from conclusions reviewers made based on the extreme variability of participants of studies described as having an "emotional or behavioral" disorder. On the other hand, excluding these samples who have often been identified as SED at some point in their school years would ignore a large body of pertinent literature. Therefore, it is important for researchers to carefully examine subject characteristics. Researchers often conduct studies to duplicate and extend the data base on SST efficacy. Science progresses, and conclusions are more robust on a construct, when a large body of research has accumulated. Currently, the large body of research on SST with students with EBD is so variable that conclusions should be reached cautiously. This limitation presents a major area of future research: conducting SST studies with students who meet federal or state eligibility criteria. This recommendation does not advocate any "trait-treatment" approach to SST. Students identified as EBD are a heterogeneous group and interventions should be tailored to their individual-specific deficits. Rather, research with students identified as EBD and receiving special education services in public schools would help identify the range and limitations of SST in this setting. Otherwise, educators may assume a particular SST procedure would be effective when, in fact, resources, personnel, or the logistics may not exist to implement them.

Types of Interventions

Social skills training is not actually an intervention but rather an outcome: improved social skills and social competence. There is no one intervention technique to train social

skills. The techniques will vary based on a youngster's pattern of deficits. If a student knows how to perform certain skills but his negative self-talk interferes with performing them, then self-instruction training may be indicated. Conversely, if the youngster does not possess the requisite social skills, then direct instruction would be warranted. Consequently, reviews contained a variety of interventions. Seven of the reviews included studies using behavioral, cognitive, or cognitive-behavioral intervention techniques (Ager & Cole, 1991; Beelmann et. al., 1994; Olmeda & Kauffman, 2003; Quinn et. al., 1999; Schloss et. al., 1986; Singh et. al., 1991; Zaragoza et. al., 1991). Four reviews focused on specific intervention approaches or outcomes. Coleman et. al., (1993) included only social problem-solving interventions whereas Mathur et. al., (1998) focused strictly on behavioral approaches. Mathur and Rutherford (1991) reviewed peer-mediated interventions. Landrum and Lloyd (1992) examined generalization outcomes. Two reviews did not specify exact training techniques (Hollinger, 1987; Templeton, 1990). In some instances it was easy to extrapolate; for example, both unspecified reviews made reference to instructions and modeling, which are behavioral techniques. Most of the narrative reviews, however, described studies as "social skills training resulted in . . ."

Taken together, the following intervention techniques were used in studies reviewed: coaching, modeling, rehearsal, feedback, reinforcement, goal setting, instructions, discussions, peer training, problem solving training, self-instruction, self-monitoring, self-evaluation, and self-reinforcement. Consequently, a reasonably sound conclusion is that SST studies incorporate both behavioral and cognitive techniques both independently and jointly—in the latter case reflecting a cognitive-behavioral orientation. Cognitive-behavioral interventions (CBI) focus on targeting a youngster's private speech self-instruction training, problem-solving training, attribution retraining, and cognitive restructuring approaches (Maag & Swearer, 2005). In addition, virtually all effective CBI techniques with youngsters include behavioral components such as modeling, role playing, and positive reinforcement (Braswell & Kendall, 1988). These techniques may work well for students who display cognitive deficits or distortions that interfere with accessing and performing social skills in their repertoire.

Results

Results from the three meta-analytic reviews are straightforward because effect sizes are calculated. The techniques vary, however, depending on whether group design or single-subject design studies were reviewed and analyzed.

Group designs use the *ES* statistic to depict the influence of the independent variable (i.e., SST) on the dependent variable (i.e., changes in measures of social skills). The level at which effect sizes approach significance is typically about 0.04, with effect sizes of 0.60 reflecting credible proof of a treatment being effective (Forness et al., 1997). Quinn et al. (1999) obtained a mean *ES* of 0.199, which was considered small. To place this in perspective, they stated that an *ES* of that magnitude indicates that only 58% of students with EBD will obtain any benefits from receiving SST. Meta-analytic results for Beelmann et al. (1994) fared somewhat better for one of the three groups: at-risk (*ES* = 0.85), externalizing syndromes (*ES* = 0.48), and internalizing syndromes (*ES* = 0.50). However, the effect of SST for two groups associated with EBD was small.

Meta-analyses of single-subject studies typically use the nonparametric approach developed by Scruggs, Mastropieri, and Castro (1987), which is based on the percentage of nonoverlapping data (*PND*) between baseline and successive intervention phases in graphic representations of data. Scruggs, Mastropieri, and their colleagues have provided guidelines for interpreting *PND* scores: > 50% = ineffective, 50%–70% = mildly effective or questionable, 70%–90% = moderately effective, and < 90% = highly effective (Mastropieri, Scruggs, Bakken, & Whedon, 1996; Scruggs & Mastropieri, 1994). Mathur et al. (1998) obtained a *PND* score of 64% for their sample of EBD studies, indicating that SST was mildly effective for this group.

Singh et al. (1991) indicated that they used quantitative analysis, although it was not meta-analytic. Instead, they developed coded categories across studies that could be numerically summarized on a 3 point scale: 0 = less than or equal to a 50% mean reduction from baseline; 1 = 51%–74% reduction, and 2 = 75%–100% reduction. They did not state how these criteria were generated. Nor did they submit the three group design/pre-posttest studies to this or any "quantitative" analysis. However, the majority (89%) of the single

subject studies reviewed obtained the following ratings: 16 received a rating of 2, seven were rated 1, and five were rated 0.

It is much more difficult to summarize the results from narrative reviews. The obvious reason is because the authors used their own internal professional judgment, which carries with it some subjectivity to appraise treatment efficacy. This assumption may account for the fact that four of the nine narrative reviews did not provide any statements regarding the efficacy of SST (Hollinger, 1987; Olmeda & Kauffman, 2003; Schloss et al., 1986; Templeton, 1990). They instead focused on issues for future research and implications for practice. The remaining five reviews reached conclusions regarding efficacy of SST using such words and phrases as "promising results," "cautious optimism," "lack of evidence," "marginal attention," "strong support," and "less positive" (Ager & Cole, 1991; Coleman et al., 1993; Landrum & Lloyd, 1992; Mathur & Rutherford, 1991; Zaragoza et al., 1991). Regardless of the subjectivity of the results provided in narrative reviews or the objective results of the quantitative reviews, they all raised numerous issues for future research and implications for practice.

Issues Raised From SST Reviews

The issues and recommendations raised varied depending on which aspects of SST they chose to focus. Some reviewers emphasized methodological issues whereas others addressed conceptual flaws. Nevertheless, four general areas of consensus deserve attention. In some instances they confirm methodological flaws (e.g., lack of generalization and assessing individual-specific deficits) but in other instances affirm positive changes (e.g., peer-mediated techniques).

Lack of Generalization

That lack of generalization was the prominent finding should not come as a surprise because it has been addressed and lamented in behavioral research since the seminal article by Stokes and Baer (1977). A decade later, Rutherford and Nelson (1988) reviewed 5,300 behavioral treatment studies and found that less than 2% addressed generalization and maintenance of educational treatment effects and less than 1% programmed for stimulus and response generalization. Landrum and

Lloyd (1992) reached similar conclusions when reviewing studies addressing the social behavior of students with EBD. Clearly, promoting generalization has been a thorny issue for basic behavioral treatment studies, let alone interventions as potentially complicated as SST. Ironically, several tactics for programming generalization could be infused in SST studies if researchers addressed three conceptual issues: selecting socially valid behaviors; focusing intervention on the peer group; and promoting entrapment.

Selecting socially valid behaviors. One of the most important issues when teaching youths social skills is determining whether targeted skills will enhance the quality of their lives—a concern often referred to as social validity (Wolf, 1978). Generalization will be enhanced when students with EBD see the relevance of using targeted skills in every day life. Unfortunately, the authors of some SST programs do not provide information on the social validity of the skills taught (e.g., Goldstein & McGinnis, 1997; Hazel, Schumaker, Sherman, & Sheldon-Wildgen, 1982; Walker et. al., 1988). In some instances, authors described more systematic approaches to selecting socially valid behaviors (e.g., Sheldon, Sherman, Schumaker, & Hazel, 1984).

Programming generalization also requires conducting functional assessment—specifically identifying and teaching students with EBD replacement behaviors (Maag & Katsiyannis, 1998). A replacement behavior is an appropriate behavior that serves the same function as the socially inappropriate behavior (Maag, 2005). Replacement behavior training (RBT) enhances generalization because if a student with EBD finds a replacement behavior to be an acceptable alternative that “gets him or her what he or she wants” (i.e., serves the same function), he or she is more likely to use that behavior outside the training setting. The very nature of replacement behaviors ensures that they will have greater social acceptability, which, in turn, may lead to two desirable outcomes: entrapment and improved peer status. The process of identifying replacement behaviors has been described previously (Maag & Kemp, 2003). Although it is not easy or frequently incorporated into SST, it nevertheless is an important component to promote generalization.

Including the peer group and promoting entrapment. Typically, a student who is socially incompetent is targeted for SST. Generalization is unlikely to occur, however, unless the peer group is included in intervention to promote

entrapment. Entrapment involves recruiting natural communities of reinforcement (McConnell, 1987). It occurs when peers reinforce a target student for performing a socially appropriate behavior. Strain, Odom, and McConnell (1984) coined the term *social reciprocity* to describe the mutually reinforcing interactions between individuals. Furthermore, certain categories of social initiations were found to have a high probability of gaining reciprocal responses from peers (Tremblay, Strain, Hendrickson, & Shores, 1981).

It is not easy to incorporate peers into training because of the many variables and dynamics of social networks that exist in schools. It is becoming increasingly evident that peer affiliations and classroom social positions influence the development of antisocial behaviors. For example, students who are highly aggressive tend to be members of peer groups characterized by high levels of antisocial behavior and low levels of prosocial behavior (Farmer & Hollowell, 1994). Similarly, a student's social prominence within a classroom or school social structure (i.e., social network centrality) has implications for the display of antisocial behavior. There are four prominence hierarchies: (1) highly prominent students in highly prominent peer groups are considered nuclear in the social structure; (2) average students in the social structure are secondary; (3) low prominence students are peripheral to the social structure; and (4) students who are not members are socially isolated (Farmer, Van Acker, Pearl, & Rodkin, 1999). Several researchers have found that aggressive and acting out behaviors are ways for some students to attain high status and that these behaviors are associated with high social positions (e.g., Adler, Kless, & Adler, 1992; Farmer & Farmer, 1996; Farmer & Rodkin, 1996; Xie, Cairns, & Cairns, 1999).

Promoting entrapment is not easy when social network researchers have found a respectable amount of aggressive and disruptive behaviors in general education classrooms (Farmer et. al., 1999). This social dynamic may encourage, or at least condone, antisocial behaviors that are characteristic of students with EBD. Surprisingly, Farmer et. al. (1999) found that in their study, the majority of students nominated as highly aggressive were students without disabilities, although there was a high level of variability in the nominations of students with disabilities. They also found that students with disabilities were part of a peer group within their classroom social

