GENERALIZATION AND MAINTENANCE OF PRESCHOOL CHILDREN'S SOCIAL SKILLS: A CRITICAL REVIEW AND ANALYSIS

Lynette K. Chandler and Roger C. Lubeck

SOUTHERN ILLINOIS UNIVERSITY

AND

Susan A. Fowler

UNIVERSITY OF ILLINOIS

This paper summarizes the results of a retrospective review of generalization in the context of social skills research with preschool children. A review of studies from 22 journals (1976 to 1990) that assessed generalization as part of social interaction research provided information concerning the prevalence of studies that have assessed generalization, common practices concerning the production and assessment of generalization, and the overall success of obtaining generalization and maintenance of social behaviors. A comparison of the most and least successful studies, with respect to generalization, revealed some differences concerning the practices employed by studies within each group. Differences differentially related to the production of generalization are discussed and recommendations are provided to guide and support future research efforts.
DESCRIPTORS: social skills, generalization, maintenance, preschool children

In 1968, Baer, Wolf, and Risley set forth the goals of applied behavior analysis. One of these goals was that applied research produce changes in behavior that generalize to a variety of environments, spread to a variety of relevant behaviors, and are maintained after an intervention has terminated. In their 20-year review of the goals of applied behavior analysis, Baer, Wolf, and Risley (1987) again pointed to the importance of generalization as "crucial . . . to the maximal effectiveness . . . of the discipline" (p. 321).

Since 1968, several types of generalization have been identified, including setting, time, and response generalization (Drabman, Hammer, & Rosenbaum, 1979), and strategies to promote generalization have been described (Stokes & Baer, 1977), elaborated, and examined (Haring, 1987;

Kirby & Bickel, 1988; Stokes & Osnes, 1986, 1988). In addition, designs to assess generalization have been described (Barlow & Hersen, 1984; Kazdin, 1982).

One area of applied behavior analysis in which generalization has been addressed is the social competence of preschool children. Social competence has been identified as a complex set of skills that includes effective peer interactions (McConnell & Odom, 1986). The development of desirable peer interactions and peer relationships during the first years of a child's life is considered important for several reasons. Peer interaction has been related to general developmental progress, communicative competence, and academic success (Curl, Rowbury, & Baer, 1985; Hendrickson, Strain, Tremblay, & Shores, 1981; Ichinose & Clark, 1990; Strain & Odom, 1986). Peer relations that develop early in life also have been related to adjustment in later years. Retrospective and longitudinal studies have described correlations between childhood social deficits and adjustment difficulties, mental health problems, alcoholism, and a variety of interpersonal problems experienced as adolescents and adults (Hartup, 1978; Strayhorn & Strain, 1986).

Researchers have developed a number of inter-

Table 1

List of Journals Reviewed

Analysis and Intervention in Developmental Disabilities
American Journal of Mental Deficiency
Behavior Assessment
Behavior Modification
Behavior Research and Therapy
Behavior Therapy
Child Behavior Therapy or *Child and Family Behavior Therapy*
Child Development
Child Study Journal
Education and Training of Mentally Retarded
Education and Treatment of Children
Exceptional Children
Exceptional Education Quarterly
Journal of Abnormal Child Psychology
Journal of Applied Behavior Analysis
Journal of the Division for Early Childhood and *Journal of Early Intervention*
Journal of Experimental Child Psychology
Journal of School Psychology
Journal of Special Education
Psychology in the Schools
Topics in Early Childhood Special Education
Volta Review

ventions to facilitate effective peer interactions and to decrease ineffective or aversive peer interactions of young children (e.g., McConnell, Peterson, Odom, & Fox, 1990; McEvoy et al., 1988; Sainato, Maheady, & Shook, 1986). Treatment programs have been employed with a variety of preschool children, including those considered aggressive, autistic, developmentally delayed, isolated, or withdrawn (e.g., Chandler, Ostrosky, Odom, & Rainey, 1990; Hodgens & McCoy, 1990; Strain, Hoyson, & Jamieson, 1985; Zahavi & Asher, 1978).

A majority of the social skills interventions applied to preschool children have produced favorable outcomes during training and treatment conditions. Unfortunately, a gap exists between our ability to effect desirable peer interactions and our ability to produce generalization and maintenance of these interactions. Generalization and maintenance have been particularly difficult to obtain in applied research with peer interactions and young children (Haring, 1987). Often the effects of social skills training with preschool children have been restricted to the treatment setting or to training conditions

and contingencies, or behavior has failed to endure when an intervention is terminated (e.g., Combs & Lahey, 1981; Odom, Hoyson, Jamieson, & Strain, 1985). The amount of generalization obtained in many peer interaction studies with young children limits the utility of social skills interventions for application in nontreatment settings (Baer, 1982; Kazdin, 1975; Stokes & Osnes, 1986).

A first step in addressing this problem is to conduct a retrospective analysis of social skills research with preschool children in order to (a) identify studies that have assessed generalization, (b) summarize common practices concerning generalization, and (c) identify practices associated with the production of generalization (Haring, 1987). In this paper, we summarize common practices among studies that have assessed generalization of peer interactions with preschool children. We also present a review and analysis of a subset of studies to identify differences between studies that produced generalization and studies that did not (i.e., studies that produced unintended stimulus control). As result, our analysis identifies practices related to the successful production of generalization and may serve as a framework to guide and support future intervention practices and research efforts.

METHOD

Analysis of the 51 Studies

The senior author reviewed 22 behavioral and nonbehavioral journals from the years 1976 to 1990. The journals selected for review are those that often publish research concerning young children. The list of the journals is presented in Table 1. This 15-year assessment period generally follows Stokes and Baer's (1977) seminal article, in which they described techniques that might promote generalization and urged investigators to plan for and assess generalization.

The senior author reviewed each journal to identify studies that (a) provided training or consequences for behaviors during intervention, (b) were successful at producing behavior change during intervention, (c) included preschool children as subjects, and (d) focused on peer interaction (i.e., social

Table 2

Type of Category and Variables within Categories That Were Coded for Each Study

Category	Variables within categories
Generalization dimension	Setting, time, setting/time, responses, persons or change agents, subjects
Generalization assessment design	Reversal and/or follow-up, probe, withdrawal, multiple baseline, alternating treatments
Behavior-change strategy	Discussion, feedback, instructions, modeling, prompting, rehearsal, positive reinforcement
Generalization-promotion strategy	Evidence of planning for generalization: (a) identify naturally occurring salient stimuli, (b) identify natural community of reinforcement, (c) identify functional target behaviors (with respect to natural or generalization setting) that are members of a large response class, (d) identify generalization-promotion strategies and integrate with behavior-change strategies as a treatment package;
	address functional target behaviors;
	specify a fluency criterion for termination of treatment or receipt of reinforcement;
	program common stimuli across treatment and generalization setting;
	train loosely (e.g., across stimuli, settings, responses, and persons);
	use indiscriminable contingencies (e.g., variable reinforcement, fade reinforcement, intermittent reinforcement);
	train sufficient exemplars (e.g., across settings, stimuli, persons, responses, cues);
	teach mediation strategies (e.g., problem solving, correspondence training);
	reinforce generalization: (a) unprompted generalization, (b) instruct to generalize, (c) target behavior emitted by peers;
	continue training; adjust fluency criteria;
	recruit natural communities of reinforcement: (a) train to solicit reinforcement, train/reinforce natural change agents to use contingencies, (c) reduce support for maladaptive or incompatible behavior;
	use sequential modification.

Note. See Drabman et al. (1979) and Stokes and Osnes (1986) for a review of the dimensions of generalization. See Barlow and Hersen (1984), Kazdin (1982), and Rusch and Kazdin (1981) for a review of assessment designs. See Haring (1987), Stokes and Baer (1977), Stokes and Osnes (1986, 1989), and Strain (1981) for a review of generalization-promotion strategies.

behavior directed to a peer). Seventy-three studies met these criteria (a list of these studies can be obtained from the senior author). Of these, 51 studies (70%) also assessed generalization of social behavior and therefore were used in the present analysis.

Each of the 51 articles was coded in four categories: (a) generalization dimension, (b) generalization assessment design, (c) behavior-change strategies, and (d) generalization-promotion strategies. These categories represent technological and methodological factors that often vary across studies and that may influence the success of producing generalization. For example, although behavior-change strategies are not usually discussed as variables that influence generalization, they were included in this review because effective behavior change may be a prerequisite for generalization. As Baer (1981) pointed out, "there are ways to make

behavior changes that encourage a more generalized change" (p. 7). Thus, the type of behavior-change strategy employed should be of interest to researchers.

The types of variables coded within the four categories are presented in Table 2. It was possible to code more than one variable within categories for individual studies. For example, in the generalization dimension category, response maintenance and setting generalization might both be coded for one study. Within each category, coding was based on the explicit textual identification of variables and strategies in each article, rather than on the coder's inference.

In addition to the categories presented in Table 2, we coded generalization comparison standards and the success of producing generalization. In judging the success of generalization, we coded (a) the standards of comparison that investigators used
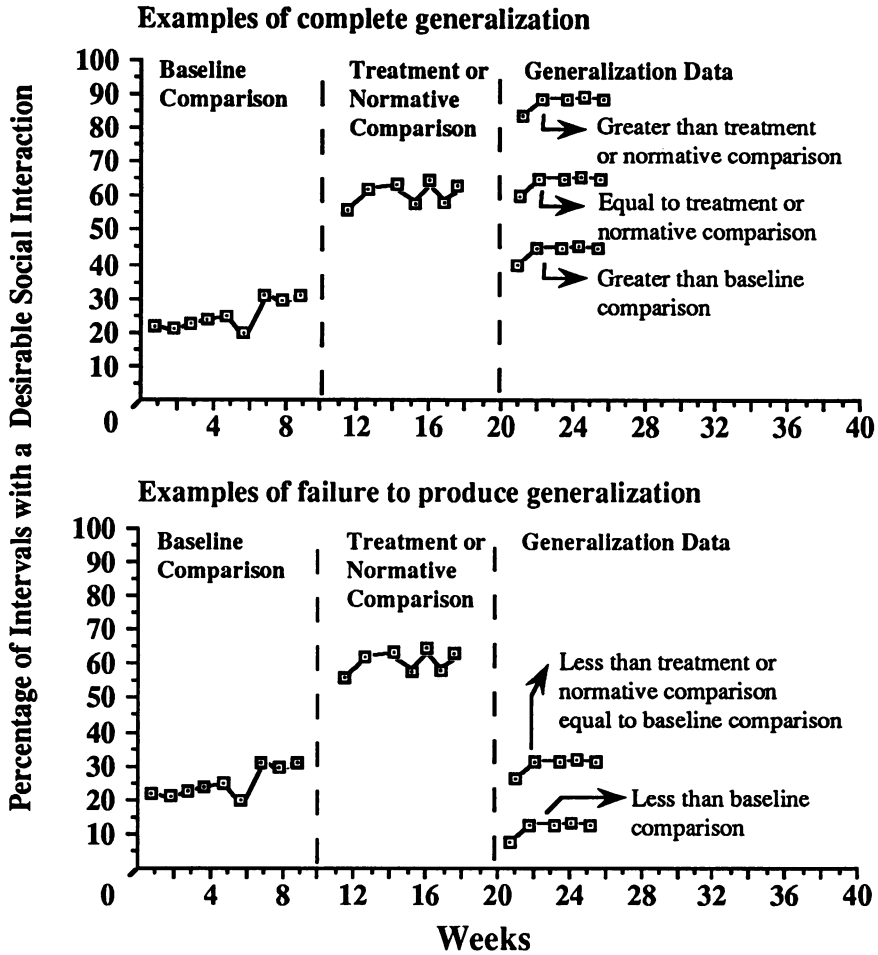
Figure 1. Examples of complete generalization and failure to produce generalization compared to baseline, treatment, and normative standards of comparison.

to assess generalization and (b) the success of generalization obtained relative to a standard for comparison. Comparison standards included baseline, treatment, and normative data (e.g., normative sample, control group, socially skilled peer). The standards for comparison consisted of only those specifically identified and discussed within each article. Coders did not impose comparisons that were not discussed in the articles.

The success of obtaining generalization is a judgment made with respect to a comparative measure. Often that judgment is subjective, based on graphic display of the data and the individual's research and clinical experience, rather than on predetermined criteria or statistical analysis. For the current analysis, primary and reliability coders (the first and second authors) relied upon subjective judgment to

code the success of producing generalization against a standard comparative measure. In making these judgments, the coders employed idiosyncratic visual inspection of graphic displays or tabled data. Judgments concerning the success of generalization were not based on predetermined criteria or statistical analysis.

Each comparison made within studies (e.g., each comparison against treatment data) was judged as complete generalization, partial generalization, or failure to generalize. Raters coded complete generalization when the amount of behavior obtained in generalization assessment was greater than or equal to the amount of behavior obtained in treatment or in normative data comparisons, or exceeded the amount of behavior obtained in baseline observations (Figure 1). Failure to generalize was cod-

ed when the level of behavior obtained in generalization assessment was equal to or less than a baseline standard for comparison, or less than treatment effects or normative data. Partial generalization was coded when complete generalization occurred for some subjects, settings, or responses, but not for others.

Analysis of the Most and Least Successful Studies

The 51 studies were examined to identify those most successful at producing generalization and those that failed to produce generalization (i.e., maintained stimulus control). Studies identified for inclusion in the successful generalization group produced complete or complete plus partial generalization across individual or multiple comparative measures (e.g., baseline and normative data) and across one or more dimensions of generalization (e.g., setting and response maintenance). For example, a study might obtain complete generalization using baseline data as a standard for comparison and partial generalization using treatment data as a comparative measure.

Studies selected for inclusion in the least successful generalization group failed to produce generalization across all comparative measures or dimensions of generalization. It should be noted that the analysis of the most and least successful generalization studies excluded studies that produced only partial generalization and studies that produced combinations of complete, partial, and failed generalization. In addition, all of the studies in this review were successful at producing behavior change during intervention. In other words, all studies demonstrated stimulus control. Studies included in the most and least successful groups were distinguished solely on the basis of generalization production.

In addition to the categories previously described, four additional categories were coded for studies in the most and least successful generalization groups: number of subjects employed, subject characteristics, target behaviors, and length of treatment. Children who served as subjects were coded as (a) having a handicapping condition or developmental delay plus social deficit, (b) normally

developing plus social deficit (e.g., a child described as withdrawn or aggressive), (c) at risk for developmental and social delay (e.g., low socioeconomic status), and (d) nonhandicapped subjects (the purpose of intervention was to increase the interaction of these subjects with children who exhibited social delays). The types of social behaviors trained during intervention were coded for each study (e.g., initiations, affectionate behavior, sharing). In addition, the number of days of treatment and length of treatment sessions were coded for all studies. This additional information was obtained to explore the possibility that the success of producing generalization was differentially related to these factors.

Interrater Agreement

Each article was coded by the senior author. Twenty-two percent of the articles were randomly selected and independently coded by the second author. The number of agreements and disagreements was assessed for each coded category and for all categories combined. An agreement was scored if both authors coded the same descriptors within categories, such as amount and type of generalization obtained and type of generalization-promotion strategy employed. A disagreement was scored if raters coded different descriptors within categories or one rater coded the occurrence of a descriptor and another did not. Reliability was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying that number by 100. Reliability across categories ranged from 81% to 90%, with a mean of 84%. The highest agreement occurred in the category of type of generalization-promotion strategy (90%), and the lowest agreement occurred in the category of generalization assessment design (81%).

RESULTS AND DISCUSSION

Review of the 51 studies that assessed generalization of preschool children's peer-directed social skills revealed some recurrent practices across studies and provided information concerning the overall success of producing generalization of social behavior. Analysis of the most and least successful generalization studies provided information about

variables that appeared to be associated with the production of generalization.

One measure of the robustness or scientific interest in an area is the number of studies conducted each year that address a specific topic. The 51 studies, including those in the most and the least successful generalization groups, occurred throughout the 15-year review period, with an average of 3.5 studies published per year. Fourteen of the 51 studies (27%) produced complete or complete plus partial generalization across all comparisons; these studies comprised the most successful group. Eight studies (16%) failed to produce complete or partial generalization across all comparisons; these studies comprised the least successful group. The remaining 29 studies produced a combination of complete, partial, and failed generalization or only partial generalization across all comparisons.

Encouragingly, the success of producing complete generalization has increased across the 15-year period. In the first 5 years of this period, 31% of the comparisons (i.e., comparison of generalization data against baseline, treatment, or normative data) produced complete generalization, whereas 37% of the comparisons failed to produce generalization. During the final 5 years, 59% of the comparisons produced complete generalization and only 13% failed to produce generalization. As encouraging as the more recent success of producing generalization may be, it is not clear if this increase reflects improved efforts to produce generalization or a more stringent criterion for publication (i.e., papers that produce generalization are more likely to be published than those that maintain stimulus control). In the absence of an analysis of publication criteria and a lack of knowledge of the number of studies that failed to produce generalization that were not published, this issue cannot be resolved. Nonetheless, examination of published studies may yield information regarding variables differentially associated with the production of generalization.

In any study, two factors may influence the success of producing generalization: technological features of the intervention and methodological features of the study. The remaining data are descried in terms of these two factors. Technological features

of interventions included (a) behaviors targeted for intervention, (b) behavior-change strategies, and (c) generalization-promotion strategies. Methodological features of the studies included (a) number of subjects and subject characteristics, (b) generalization dimensions, (c) length and duration of intervention, (d) success of producing generalization against standards for comparison, and (e) generalization assessment design.

Technical Features of Interventions

Target behaviors. The behaviors targeted for intervention varied between the most and least successful generalization groups, although the four most frequently selected behaviors were the same for both groups. Studies in the most successful group targeted initiations to peers most frequently (57%), followed by conversation or reciprocal interaction (43%), and responding to and sharing with peers (each selected in 38% of the studies). Studies in the least successful group also targeted initiations to peers most frequently (62%). However, the second most common target behavior was responding to peers (50%), followed by sharing with peers (37%) and conversation or reciprocal interaction (25%). The remaining behaviors (e.g., inappropriate social behavior, toy play, proximity) occurred in less than 25% of the studies. The types of behaviors targeted for intervention did not appear to be differentially associated with studies that produced generalization.

Behavior-change strategies. A majority of the 51 studies used multiple behavior-change strategies during intervention. Many of these strategies were combined as a treatment package (e.g., instructions, modeling, rehearsal, and prompting within a peer-mediation training package); however, only the components of intervention packages were coded due to variations among studies in the components that comprised similarly named packages. The order of the most often used strategies and percentage of behavior-change strategies employed across the 51 studies were as follows: positive reinforcement (67%), instructions (65%), prompting (57%), rehearsal (46%), modeling (41%), feedback (35%), and discussion (35%).

Table 3 depicts the percentage of behavior-change strategies employed in the most and least successful generalization groups. Differences between groups are most evident in the percentage of studies using prompting, positive reinforcement, and instructions; these were most common strategies employed in the most successful generalization group. A higher percentage of studies in the most successful group used prompting and positive reinforcement than studies in the least successful group. Interestingly, a higher percentage of studies in the least successful generalization group used instructions as a behavior-change strategy, although this was the third most common strategy employed by studies in the successful group. The most common single behavior-change strategies used by studies in the least successful group were instructions, modeling, and rehearsal.

Although there were differences in the percentage of studies using each behavior-change strategy, there are two reasons why it may be premature to recommend that one or more of these strategies be employed in studies seeking to produce generalization. First, all of the behavior-change strategies were used by studies in both groups. Therefore, it did not appear that any one strategy distinguished success from failure to produce generalization. Second, a majority of the studies in each group (13 of the 14 most successful and 7 of the 8 unsuccessful studies) used a combination of strategies. It may be that the combination of behavior-change strategies, rather than the application of a single strategy, influenced the success of producing generalization.

The two most common strategies combined by studies in the most successful group were prompting plus reinforcement (57%) and reinforcement plus feedback (50%). In contrast, only 25% of the studies in the least successful group employed prompting plus reinforcement and only 13% combined reinforcement and feedback. The most common sets of two strategies in the least successful group included combinations of instructions plus modeling, prompting, or rehearsal or combinations of rehearsal plus modeling or prompting.

The most common combination of three strat-

Table 3

Percentage of Behavior-Change Strategies Across Studies in the Most Successful and Least Successful Generalization Groups

Behavior-change strategies	Percentage occurrence	
	Most successful ($n = 14$)	Least successful ($n = 8$)
Prompting	79	50
Positive reinforcement	64	37
Instructions	57	75
Feedback	50	37
Rehearsal	50	62
Modeling	50	62
Discussion	36	50

Note. More than one behavior-change strategy could be coded for individual studies; thus, the percentage in either column may exceed 100%.

egies in the successful group was prompting, positive reinforcement, and feedback (43%). Only 13% of the studies in the least successful group employed this combination of strategies. The most common set of three strategies for studies in the least successful group included prompting, instructions, and rehearsal.

Baer (1981) advises investigators to consider the importance of behavior-change strategies in the production of generalization. Although no single strategy was differentially associated with successful generalization, examination of combinations of strategies suggests differences between the most and least successful generalization groups. Studies able to produce generalization appeared to use a combination of antecedent and consequence strategies (e.g., prompting and positive reinforcement). Studies that failed to produce generalization primarily employed combinations of antecedent strategies (e.g., modeling and rehearsal). Combinations of antecedent and consequence strategies may be more likely to produce generalization because they address both ends of the three-term contingency (antecedents, behaviors, and consequences). These may present a stronger form of control with respect to generalization than situations in which antecedent-only or consequence-only strategies are employed.

Generalization-promotion strategies. Thirty-seven of the 51 studies (73%) used a generalization-

Table 4

Order (Most Commonly Occurring) and Percentage of
Generalization-Promotion Strategies across Studies

1. Address functional target behaviors	65%
2. Train loosely	45%
3. Use indiscriminable contingencies	42%
4. Program common stimuli	37%
5. Reinforce generalization	34%
6. Specify a fluency criterion	29%
7. Teach mediation strategies	20%
8. Continue training	17%
9. Teach sufficient exemplars	17%
10. Recruit natural community of reinforcement	11%
11. Sequential modification	9%

Note. Multiple generalization-promotion strategies could be coded for individual studies; thus, the percentage may exceed 100%.

Table 5

Percentage of Generalization-Promotion Strategies across
Studies in the Most Successful and Least Successful
Generalization Groups

	Percentage occurrence	
Generalization strategies	Most success-ful (*n* = 14)	Least success-ful (*n* = 8)
Address functional target behaviors	50	25
Use indiscriminable contingencies	43	12
Specify a fluency criterion	36	0
Teach mediation strategies	29	12
Train loosely	29	37
Program common stimuli	29	25
Continue training	21	12
Teach sufficient exemplars	14	12
Reinforce generalization	7	12
Recruit natural community of reinforcement	7	12
Sequential modification	0	25

Note. Multiple generalization-promotion strategies could be coded for individual studies; thus, the percentage in either column may exceed 100%.

promotion technique; 27 of these used multiple strategies. The remaining studies did not use a promotion technique; that is, they adopted a "train and hope" strategy (cf. Stokes & Baer, 1977). Table 4 presents the most common strategies and the percentage of studies employing each strategy. The three most frequently used promotion strategies were (a) address functional target behaviors, (b) train loosely, and (c) use indiscriminable contingencies. The least commonly used strategies were (a) sequential modification and (b) recruit a natural community of reinforcement.

Only 51% of the articles provided textual evidence of planning for generalization by discussing procedures to identify common stimuli, natural communities of reinforcement, and functional target behaviors, and to integrate generalization-promotion strategies and behavior-change strategies as a treatment package. It is possible that more investigators actually planned for generalization; however, textual evidence was not provided in the articles.

Seventy-one percent of the studies in the most successful group and 62% of the studies in the least successful group used a generalization-promotion strategy. There were differences between groups in the percentage of studies that employed each strategy (Table 5). The most common strategies used in the successful group were (a) address functional

target behaviors, (b) use indiscriminable contingencies, and (c) specify a fluency criterion. These strategies were used in only 25% or fewer of the studies in the least successful generalization group; none of the studies in this group specified a fluency criterion.

Nine of the 11 studies (82%) in the most successful group that used a promotion strategy combined three or more strategies. Four of the five studies in the least successful group that used a promotion strategy combined two strategies; only one used a combination of three or more strategies. The most frequent combinations of two promotion strategies in the successful group were (a) address functional target behaviors and use indiscriminable contingencies (45%), (b) use indiscriminable contingencies and teach mediation strategies (36%), and (c) address functional target behaviors and teach mediation strategies (36%). The combination of addressing functional target behaviors and using indiscriminable contingencies was used in only 20% of the studies in the least successful group, and the

remaining combinations were not used at all in this group. The most frequent combination of two strategies for the least successful generalization group was addressing functional target behaviors and employing common stimuli (40%).

For studies in the most successful group, combinations of three and four generalization-promotion strategies usually contained a mix of four strategies: (a) address functional target behaviors, (b) specify a fluency criterion, (c) use indiscriminable contingencies, and (d) teach mediation strategies.

The different percentages of single and multiple combinations of strategies employed in the most and least successful generalization groups suggested that generalization-promotion strategies are not equally effective at producing generalization and should not be arbitrarily selected and implemented. As indicated by Baer (1981), generalization-promotion strategies can be effective at producing generalization; however, simply using one or more strategies does not guarantee the occurrence of generalization. Currently, there is little direction in the social skills literature concerning preschool children to guide our selection of one strategy or combination of strategies over another. Investigators who seek to produce generalization should select promotion strategies as part of a functional analysis of behavior in the context of desired generalization settings.

In the area of generalization, we may profit from research regarding the influence of single strategies and multiple strategies that typically are combined in interventions (e.g., the variables combined by studies in the successful group). It also will be important to examine interactions between combinations of variables across categories in order to develop systems for matching generalization-promotion strategies and behavior-change methods (Baer et al., 1987).

Methodological Features

Number of subjects and subject characteristics. The mean number of subjects used in the most successful generalization group was 6, ranging from 1 to 22. The mean number of subjects in the least successful group was slightly higher, at 9 subjects per study, and ranged from 3 to 32. There was little difference in the characteristics of the preschool children who participated in the most and least successful generalization studies. The majority of studies in both groups targeted children with handicaps and social delays (60% and 50% of the studies in the most and least successful groups, respectively), followed by typical children with social delays (20% of studies in the most successful group, 13% of studies in the least successful group), typical children (13% of studies in the most successful group, 25% of studies in the least successful group), and children at risk for school failure and social delay (7% and 12% of studies in the most and least successful groups, respectively).

It is often tempting to speculate that failure to obtain generalization is related to characteristics of the children selected for intervention (e.g., it may be more difficult to obtain generalization with children who exhibit severe rather than mild disabilities). Although subject characteristics may certainly influence generalization, they did not appear to be a factor that distinguished successful generalization and failure to produce generalization among the limited number of studies reviewed.

Generalization dimension. The most common types of generalization dimension assessed across the 51 studies were ordered as follows: (a) response maintenance (55%), (b) setting (51%), (c) setting/time (24%), (d) subjects (24%), (e) response (20%), and (f) persons or change agents (8%). Thirty of the 51 studies assessed generalization across two or more dimensions, typically combining response maintenance with setting/time, subject, and setting generalization.

The percentage of studies assessing each generalization dimension from the most and least successful generalization groups is presented in Table 6. In both groups, response maintenance and setting generalization were the most common dimensions assessed, although there were differences in the percentage of studies assessing each generalization dimension. Studies in the successful group assessed generalization along the dimensions of responses, subjects, and response maintenance more often than did studies in the least successful generalization group. In addition, 10 of the 14 studies

Table 6

Percentage of the Generalization Dimensions Assessed across Studies in the Most Successful and Least Successful Generalization Groups

Dimensions of generalization	Percentage occurrence	
	Most successful ($n = 14$)	Least successful ($n = 8$)
Response maintenance	57	37
Setting	43	37
Responses	43	12
Setting/time	21	12
Subjects	21	0
Persons or change agents	7	12

Note. Multiple dimensions could be coded for individual studies; thus, the percentage in either column may exceed 100%.

from the most successful group assessed generalization across two or more dimensions, whereas only one of the eight studies from the least successful group assessed multiple types of generalization.

It might be argued that assessment of multiple dimensions increases the probability of producing successful generalization; that is, the probability of finding a complete generalization increases as the number of dimensions assessed increases. If this were true, the increase in generalization obtained for studies in the successful group could simply be a statistical artifact of the increased number of dimensions assessed. However, if generalization were an artifact of the number of dimensions assessed, we might also expect to see more failures in studies assessing multiple dimensions. An analysis of the 51 studies appeared to support this possibility. Of the 31 studies that assessed multiple dimensions, 90% produced combinations of comparisons including complete generalization (e.g., comparisons may include complete plus partial generalization or complete, partial, and failed generalization); 64% of these combinations included comparisons with failed generalization. In the studies that assessed generalization for only one dimension, 30% included comparisons with complete generalization and 50% of the comparisons included failed generalization.

Although the failure rate across single and multiple comparisons appeared to be similar, the large

difference in the rates of complete generalization suggests two alternatives to a statistical sampling explanation. First, the strategies used to promote generalization are not equal; one strategy may be more effective at producing generalization across a particular dimension than another. For example, training sufficient exemplars may be more likely to produce response generalization than does applying indiscriminable contingencies. Therefore, assessing multiple dimensions may increase the likelihood of finding a match between generalization-promotion strategies and a dimension of generalization that results in complete generalization. A second explanation for the increased success rate across multiple dimensions might be that successful production of generalization in one dimension will likely produce generalization in multiple dimensions, due to the behavior change and/or promotion strategies employed or other factors that influence generalization. These disparate explanations merit further research.

Length of treatment. There was a small difference across groups in the duration of training sessions. Studies in the most successful group trained for an average of 21 min (range, 10 to 60 min). The average length of training sessions for studies in the least successful group was 27 min (range, 15 to 60 min).

There also were differences in the number of treatment sessions employed. Studies in the most successful group implemented intervention for an average of 33 sessions (range, 4 to 89 sessions). Studies failing to produce generalization implemented approximately half the number of sessions used by the successful generalization studies. The mean number of training sessions for studies in the least successful generalization group was 16 (range, 2 to 51 sessions).

The higher number of training sessions may have provided subjects with a sufficient history of performance to ensure that behavior was emitted and reinforced in natural or generalization situations. Although the determination of sufficient history is likely to vary across subjects, researchers might refer to normative rates of behavior or indicators of fluency (i.e., level of consistency and proficiency) before terminating treatment. For example, Strain

(1981) suggested that investigators overtrain behavior by selecting performance criteria that exceed normative rates of behavior, thus increasing resistance to extinction.

Success of generalization across standards for generalization comparison. For all 51 studies, baseline was the most common standard for comparison, followed by treatment and normative data. All studies presented baseline data as a comparison measure, 43 studies made additional comparisons against treatment data, and 17 studies used normative data as standards for comparison. Studies in the most successful generalization group conducted more comparisons against the three measures (baseline, treatment, and normative data) than studies in the least successful group. Ninety-three percent of the studies in the most successful group used treatment data and 43% used normative data, in addition to baseline data, as comparative measures. In contrast, treatment data were used as a standard for comparison by 62% of the studies in the least successful group; only one study in this group used normative data as a comparative measure.

In all 51 studies, the success of producing generalization varied across the standards for comparison. Complete generalization was most often obtained when baseline and normative data were used as comparative measures. Fifty-two percent of the comparisons against baseline data and 58% of the comparisons against normative produced complete generalization. Only 19% and 13% of the comparisons against baseline and normative data, respectively, failed to produce generalization. Failure to obtain generalization occurred most often when treatment data were used as a standard for comparison; 46% of the treatment comparisons failed to produce generalization. Partial generalization occurred at approximately 30% across each type of standard for comparison.

It is difficult to judge the seriousness of the failure to produce complete generalization using treatment data as a standard of comparison because it is difficult to judge the appropriateness of the amount of behavior produced in treatment sessions. For example, the rate of behavior produced during treatment may exceed normative rates in order to provide sufficient practice for children with social deficits (Strain, 1981). On the other hand, treatment data may reflect prompted and unprompted occurrences of behavior, artificially inflating the rate of naturally occurring behavior obtained during intervention. In both of these examples, it may not be appropriate to use treatment data as a comparative measure. This argues for the increased use of baseline and normative comparisons. Consideration should be given, however, to the defining characteristics of an appropriate normative group. For example, the rate of vocal interaction emitted by typical children would not constitute an appropriate normative comparison for children who are hearing impaired, use sign language, and present social deficits.

Although the failure to produce generalization using treatment data may be expected, the failure to obtain generalization using baseline and normative comparisons is not. Approximately half the comparisons failed to produce generalization with these measures. Unfortunately, we do not know why generalization failed to occur in these comparisons. We can speculate that generalization failed to occur for a number of reasons, such as (a) the intervention and generalization strategies were not implemented well, (b) the combinations of behavior-change and promotion strategies were not conducive to the production of generalization, (c) the behavior was not trained for a sufficient length of time or to fluency, (d) the intervention methods were not clearly related to conditions existing in the natural environment, or (e) a functional target behavior was not selected.

Empirical research is needed to explore the conditions controlling appropriate generalization. For example, studies that seek to systematically convert partial generalization into complete generalization, and investigations that report practices or conditions that failed to produce generalization and then (in the same study) report additional procedures resulting in generalization, would provide valuable information concerning functional variables that account for generalization (Stokes & Osnes, 1989).

Generalization assessment design. More than

Table 7

Practices Differentially Related to Studies in the Successful Generalization Group

1. Four generalization-promotion strategies were most frequently combined:
   A. Address functional target behaviors
   B. Specify a fluency criterion
   C. Use indiscriminable contingencies
   D. Use mediation techniques
2. Three behavior-change strategies were most frequently combined:
   A. Prompting
   B. Positive reinforcement
   C. Feedback
3. More than one dimension of generalization was assessed (e.g., setting, response maintenance).
4. Generalization was assessed with withdrawal designs.
5. Multiple comparative measures for generalization assessment were used (baseline, normative, and treatment data).
6. Complete generalization was most often obtained when baseline and normative data were used as comparative measures.
7. Intervention occurred over a longer period of time.

half of the 51 studies used multiple designs to assess generalization across various dimensions. For example, a study might assess setting generalization with a probe design and response generalization with a multiple baseline design. Probe designs were used most frequently (63%), followed by reversal and/or follow-up (47%), withdrawal (24%), multiple baseline (19%), and alternating treatments (6%) designs.

The preferred type of generalization assessment design varied between the most and least successful generalization groups. Probe and withdrawal designs were most commonly used in the most successful generalization group (93% of studies), followed by reversal and/or follow-up (43%) and multiple baseline (28%). Probe designs were the most common type of design used in the least successful group (62%), followed by reversal and/or follow-up (50%), multiple baseline (12%), and alternating treatments (12%).

The distinguishing feature between groups in terms of generalization assessment design was in the use of withdrawal designs. Fifty percent of the studies in the most successful group used with-

drawal designs, but only one study in the least successful group used this design. Withdrawal designs may have facilitated generalization because they interacted with or functioned as generalization-promotion strategies. More specifically, in the withdrawal design, components of the treatment program are gradually withdrawn from the training setting or are withdrawn across multiple baselines (Barlow & Hersen, 1984; Kazdin, 1982) exposing subjects to varying contingencies, stimuli, schedules, and so forth. This design is similar to using indiscriminable contingencies in that it may allow natural stimuli or consequences to gain control over behavior and may increase the length of training as components are gradually removed. Examination of the relationship between generalization assessment designs and the production of generalization requires empirical research to extend this descriptive analysis of existing literature.

SUMMARY

As a first step in improving our understanding of a possible set of best practices associated with generalization, we examined 51 studies that addressed generalization in the context of social interaction research with preschool children. We also reviewed a subset of studies that produced generalization and failed to produce generalization. There are some limitations to this type of analysis. This review is descriptive, rather than experimental or empirical. Clearly, the points and recommendations derived from this review must be supported through future research. In addition, studies within the most and least successful groups were very diverse, and the number of studies in these groups was small. In spite of these limitations, differences in the practices employed in both groups were noted, and many of these practices were frequently associated with successful generalization. These differences are summarized in Table 7.

It is tempting to argue that generalization will be more likely if these practices are incorporated in future research. However, this prediction may be premature. Not every successful generalization study used the same set of strategies, and there were

examples of failed generalization studies that used one or more of the practices identified in Table 7 (e.g., indiscriminable contingencies). The existence of a set of best practices must be experimentally verified; however, the strategies associated with successful generalization in this review should not be ignored. Therefore, the list of practices associated with successful generalization is presented as a possible starting point for further systematic investigations on the generalization of social skills.

In recommending these practices as a starting point for research, we want to emphasize that the area of social skills with preschool children is ready for studies that specifically address questions of generalization. Much of past research necessarily focused on the development and validation of behavior-change techniques, rather than on generalization. Generalization was used as a measure of training in many studies, but was not necessarily the focus of, or a goal of, research.

In the next decade of preschool social skills research, our focus must be on questions of generalization. The analysis of variables that produce and control generalization, as with behavior-change research, can only come from controlled studies in which subject and training variables are held constant as generalization techniques are systematically manipulated. For example, an investigator might assess the effectiveness of a single promotion tactic or a combination of promotion techniques across 3 subjects with similar characteristics, target behaviors, behavior-change strategies, and fluency criteria. Studies also are needed to identify empirically a set of best practices to promote generalization of preschool children's social behavior and to determine if these practices are consistent across children, target behaviors, environments, agents, and behavior-change strategies. The critical factor, however, is that the focus of future research be on generalization and maintenance of social behavior.

REFERENCES

Baer, D. M. (1981). *How to plan for generalization.* Lawrence, KS: H & H Enterprises.

Baer, D. M. (1982). The role of current pragmatics in the future analysis of generalization technology. In R. B. Stuart (Ed.), *Adherence, compliance, and generalization in behavioral medicine* (pp. 192–212). New York: Brunner/Mazel.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20,* 313–327.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.

Chandler, L. K., Ostrosky, M., Odom, S. L., & Rainey, S. (1990). *Removing the teacher from peer-mediated interventions: Cross site effects.* Paper presented at the annual meeting of the Association for Behavior Analysis, Nashville, TN.

Combs, M. L., & Lahey, B. B. (1981). A cognitive social skills training program: Evaluation with young children. *Behavior Modification, 5,* 39–60.

Curl, R. M., Rowbury, T. G., & Baer, D. M. (1985). The facilitation of children's social interaction by a picture-cue training program. *Child and Family Behavior Therapy, 7*(2), 11–39.

Drabman, R. S., Hammer, D., & Rosenbaum, M. S. (1979). Assessing generalization in behavior modification with children: The generalization map. *Behavioral Assessment, 1,* 203–219.

Haring, N. (1987). *Investigating the problem of skill generalization: Literature review III.* Seattle: Washington Research Organization.

Hartup, W. W. (1978). Peer interaction and the process of socialization. In M. J. Guralnick (Ed.), *Early intervention and the integration of handicapped and nonhandicapped children* (pp. 27–51). Baltimore: University Park Press.

Hendrickson, J. M., Strain, P. S., Tremblay, A., & Shores, R. (1981). Relationship between toy and material use and the occurrence of social interactive behaviors by normally developing preschool children. *Psychology in the Schools, 18,* 500–505.

Hodgens, J. B., & McCoy, J. F. (1990). Effects of coaching and peer utilization procedures on the withdrawn behavior of preschoolers. *Child and Family Behavior Therapy, 12*(2), 25–48.

Ichinose, C. K., & Clark, H. B. (1990). A review of ecological factors that influence the play and activity engagement of handicapped children. *Child and Family Behavior Therapy, 12*(3), 49–76.

Kazdin, A. E. (1975). *Behavior modification in applied settings.* Homewood, IL: Dorsey Press.

Kazdin, A. E. (1982). *Single case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Kirby, K. C., & Bickel, W. K. (1988). Toward an explicit analysis of generalization: A stimulus control interpretation. *The Behavior Analyst, 11,* 115–129.

McConnell, S., & Odom, S. L. (1986). Sociometrics: Peer-referenced measures and the assessment of social com-

petence. In P. S. Strain, M. J. Guralnick, & H. M. Walker (Eds.), *Children's social behavior: Development, assessment, and modification* (pp. 215–284). Orlando, FL: Academic Press.

McConnell, S., Peterson, C., Odom, S. L., & Fox, J. J. (1990). *Effects of child-specific interventions on social interaction rates of young children with disabilities: Selection of treatment components.* Paper presented at the annual meeting of the Association for Behavior Analysis, Nashville, TN.

McEvoy, M., Nordquist, V., Twardosz, S., Heckaman, K. A., Wehby, J., & Denny, K. (1988). Promoting autistic children's peer interaction in an integrated early childhood setting using affection activities. *Journal of Applied Behavior Analysis,* **21,** 193–200.

Odom, S. L., Hoyson, M., Jamieson, B., & Strain, P. S. (1985). Increasing handicapped preschoolers' social interactions: Cross setting and component analysis. *Journal of Applied Behavior Analysis,* **18,** 3–16.

Rusch, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis,* **14,** 131–140.

Sainato, D. M., Maheady, L., & Shook, G. L. (1986). The effects of a classroom manager role on the social interaction patterns and social status of withdrawn kindergarten students. *Journal of Applied Behavior Analysis,* **19,** 187–195.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis,* **10,** 349–367.

Stokes, T. F., & Osnes, P. G. (1986). Generalizing children's social behavior. In P. S. Strain, M. J. Guralnick, & H. M. Walker (Eds.), *Children's social behavior: Development, assessment, and modification* (pp. 407–443). Orlando, FL: Academic Press.

Stokes, T. F., & Osnes, P. G. (1988). The developing applied technology of generalization and maintenance. In R. H. Horner, G. Dunlap, & R. L. Koegel (Eds.), *Generalization and maintenance* (pp. 5–19). Baltimore: Paul H. Brookes.

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy,* **20,** 337–355.

Strain, P. S. (1981). Peer-mediated treatment of exceptional children's social withdrawal. *Topics in Early Childhood Education,* **1,** 94–105.

Strain, P. S., Hoyson, M., & Jamieson, B. (1985). Normally developing preschoolers as intervention agents for autistic-like children: Effects on class deportment and social interaction. *Journal of the Division for Early Childhood,* **9,** 105–115.

Strain, P. S., & Odom, S. L. (1986). Peer social initiations: Effective intervention for social skills development. *Exceptional Children,* **52,** 543–551.

Strayhorn, J. M., & Strain, P. S. (1986). Social and language skills for preventative mental health: What, how, who, and when. In P. S. Strain, M. J. Guralnick, & H. M. Walker (Eds.), *Children's social behavior: Development, assessment, and modification* (pp. 287–330). Orlando, FL: Academic Press.

Zahavi, S., & Asher, S. R. (1978). The effect of verbal instructions on preschool children's aggressive behavior. *Journal of School Psychology,* **16,** 146–153.