

Using Brief Academic Assessments to Determine Generalization Strategies

Sara E. House Rich · Gary J. Duhon

Published online: 4 October 2014
© Springer Science+Business Media New York 2014

Abstract This study examined the utility of brief academic assessments to identify effective generalization procedures for individual students. Specifically, the study built on the proposal that brief assessments of antecedent and consequence manipulations can identify the most effective generalization strategy for individual students. The design was an alternating treatments design nested within a multiple baseline across six students. Students learned how to solve a set of multiplication facts using a common strategy while spontaneous generalization to other sets of facts was measured. Next, researchers determined whether an antecedent- or consequent-based generalization strategy would be more effective for increasing generalization across multiplication skills and conducted an extended analysis with an alternating treatment phase to confirm results of the brief assessment. Results indicated that the assessment correctly identified the most effective generalization strategy for five of the six students.

Keywords Generalization · Math interventions · Brief academic assessment · Multiple baseline · Alternating treatments design

Introduction

Over 30 years ago, Stokes and Baer (1977) wrote a seminal article on generalization in which they conceptualized it as an active process that does not always occur

S. E. House Rich (✉)
Union Public Schools, 8506 E. 61st Street, Tulsa, OK 74133, USA
e-mail: sara.e.house@okstate.edu

G. J. Duhon
School of Applied Health and Educational Psychology, Oklahoma State University,
Stillwater, OK, USA
e-mail: gary.duhon@okstate.edu

spontaneously and therefore should be programmed into interventions. Furthermore, Stokes and Osnes (1989) emphasized the importance of investigating the variables that produce and control generalized responses. Although the research on generalization has proven useful in identifying potentially effective procedures, less attention has been given to the underlying principles involved in producing it (e.g., Alessi 1987; Kirby and Bickel 1988; Mesmer et al. 2007; Stokes 1992; Stokes and Osnes 1989). This research is critical because individuals display differential responding to generalization procedures (e.g., Freeland and Noell 2002; Lloyd et al. 1981; Noell et al. 2000). For example, when taught an academic strategy, such as skip-counting to solve math facts, some students may immediately generalize the strategy to novel problems without any apparent programming, whereas other students may require a cue for generalization, and yet, other students may respond to the cue for generalization but at a slower and more variable rate than their peers (Lloyd et al. 1981). More research is needed to understand the influence of generalization principles on individual response patterns.

Broadly, generalization is a term used to explain the appearance of a relevant skill or behavior in the presence of a non-training condition (i.e., subject, settings, or time; Stokes and Baer 1977). More specifically, the current study focused on stimulus generalization, which describes the phenomenon of being able to produce the same generalized response in the presence of different stimuli (Alessi 1987; Cooper et al. 2007). For example, the ability of a child to describe three different stimuli such as the sky, a shirt and a car as blue could be stimulus generalization (Alessi 1987). Stokes and Baer posited that when generalization does occur, it is likely due to programming within the interventions, although the exact variables associated with the generalization may not be recognized. Kirby and Bickel (1988) extended this position by arguing that programming of generalization is due to the underlying principles of “stimulus control and reinforcement” (p. 116). Whereas generalization is often viewed as part of the behavioral learning process (Haring and Eaton 1978), Kirby and Bickel’s theory conceptualizes it as a separate behavior that occurs because of antecedent or consequent stimuli. For example, research has shown that antecedent cues, such as providing a common stimulus, can promote generalization of academic responding from training settings to non-training settings (Ayllon et al. 1983). If a student is taught how to read “_at” words (i.e., hat, rat, cat) with the “at” highlighted in each word and then shown a novel “_at” word with highlights, the highlighting procedure would be an example of an antecedent cue, or programmed common stimulus, to promote generalization. Research has also shown that reinforcing occurrences of generalization can increase the probability that the generalized behavior will occur (Weinstein and Cooke 1992). For example, providing a student with reinforcement for reading the words “at,” “sat,” and “cat” may increase the likelihood of generalization to reading “hat” or “mat.” Consideration of these potential underlying principles of generalization may help practitioners in selection of effective generalization procedures for individuals.

Brief academic assessments, such as brief experimental analyses (BEA) of reading interventions and skill versus performance assessments, are emerging tools for selection of effective individualized interventions in the school (e.g., Daly et al. 2006; Noell et al. 2001). Brief assessments utilize single-subject designs to compare

efficiently the effectiveness of two or more intervention strategies for increasing or decreasing a behavior (Noell et al. 2001). The interventions are typically selected a priori using a theoretical framework, such as the instructional hierarchy or skill versus performance deficits (Burns and Wagner 2008).

Research is needed to evaluate efficient procedures that account for differences in generalization across individuals to ensure appropriate selection of the most effective generalization procedures. Strategies that promote effective generalization with math are especially important, as few studies have addressed this topic. The current study utilized a brief assessment procedure to develop hypotheses about the most effective generalization strategies for individual students learning to employ a multiplication strategy. This proposed method has the potential of facilitating accurate and efficient delivery of treatments by providing a framework to identify what generalization strategy will be most effective for an individual.

The current study applied generalization procedures to a skip-counting math strategy that is presented commonly in math textbooks to teach multiplication and has been demonstrated to be effective for increasing accuracy and fluency (DuVall et al. 2003; Fennell et al. 1998). Based on the a priori theory that generalization is a behavior that can be affected by consequent and antecedent manipulations (Kirby and Bickel 1988), the authors utilized the results of a brief assessment to determine whether an antecedent- or consequent-based generalization strategy would be more effective to increase stimulus generalization across different multiplication facts. The antecedent-based procedure programmed the common stimulus of highlights on a number chart to facilitate generalization, whereas the consequent-based strategy rewarded occurrences of the generalized behavior. Given that the purpose of the study was to determine whether the brief academic assessment could accurately identify the most efficient and effective generalization procedure for each student, the procedures were applied separately instead of in combination. The authors hypothesized that the students would require an explicit procedure to demonstrate generalization and that the assessment would accurately identify the most effective strategy for each student.

Methods

Participants and Settings

Participants were from an elementary school in the Midwest United States with approximately 450 students (grades prekindergarten through fifth), 31 % of who qualified for free and reduced lunch and 14 % of whom were from ethnic minority backgrounds. Three second-grade teachers were each asked to identify four potential students for the study who demonstrated average performance in math. The 12 second-grade students referred by their teachers participated in pre-assessments, and six were included in the study. Five of the referred students were not included due to levels of accuracy over 25 % on the initial pre-assessment of multiplication skills, and one student was excused due to emotional frustration expressed during the pre-assessments. This student was frustrated by being

presented with probes of skills he had not been taught. Because multiplication skills are not typically taught in second grade, the researchers determined for this student that the risks of participating in the study outweighed the benefits of learning an advanced skill. The researchers consulted with the student's teacher and parent following the pre-assessment procedure regarding his emotional frustration. All six participants who were included in the study were white, and four of the participants were male. At the beginning of the study, Ryan, Matt, Kristin, and Beth were 7 years old and John and Travis were 8 years old. The six students were classified as general education students and had not been retained for any grades. Trained school psychology doctoral-level graduate students implemented all phases of the treatment, which occurred in a quiet hallway with tables and chairs outside the students' classrooms.

Materials

Throughout the phases of the study, students were given multiplication probe sets each having a single common factor of 6, 7, or 8, with the facts of 6×8 , 8×7 , and 6×7 removed from all sets to eliminate duplicate facts. The probes were created using an Excel[®] spreadsheet, which was specifically configured to generate random numbers for the factors of 6, 7, or 8 for a given problem so that multiple equivalent probes could be created. A number chart was included on the top right corner of each probe. The number chart consisted of a 10×10 grid with numbers 1–100 on it. During the treatment phase and the cueing procedures, the corresponding skip-counting multiples were highlighted with a yellow highlighter pen on the number chart prior to the session. For example, students receiving the 6s facts probes during the treatment phase had the multiples of 6 (e.g., 6, 12, 18) highlighted on the 10×10 grid.

During the pre-assessment and assessment phases, researchers utilized a grab bag of pencils and stickers as student rewards for beating individual goals along with explicit goal setting. During the reward generalization sessions, researchers presented a blank graph for students to record their daily progress toward their overall goal. On the *X*-axis of the graph, the dates of sessions were recorded and the *Y*-axis contained intervals of five for numbers 0–40. The students' individual goals were also marked on the graph with a dotted line. Researchers also used a sticker goal sheet during the reward generalization sessions, in which students were able to put a sticker next to their daily goal if they beat it.

Response Measurement

The dependent variables in this study were the students' accuracy and fluency on single-skill multiplication probes given as part of the intervention process. For each probe administered, researchers scored the number of digits correct (DC) and the number of correct problems (Shinn 1989). Accuracy was determined by dividing the number of correct problems by the total number of problems attempted to obtain a percentage of accurate problems. Fluency was determined by calculating the number of DC and dividing this number by 2 for the DC per min (DCPM).

Inter-scorer Reliability and Procedural Integrity

A trained independent scorer re-scored 39 % of the math probes across the six participants to determine inter-scorer reliability. The inter-scorer agreement was based on either the accuracy of problems attempted (for accuracy students) or the DC on problems attempted (for fluency students). The inter-scorer agreement was calculated by dividing the total number of probes re-scored by the number of re-scored probes that matched the original score. The average inter-scorer agreement was 88 % (range 77–100 %).

An independent observer collected procedural integrity data for 32 % of all intervention sessions. For each of the phases, an independent observer conducted a live observation of the intervention session and checked off steps on an intervention protocol as the research assistant completed each step. The number of steps for each protocol varied by phase, with a minimum of six steps for baseline and a maximum of 16 steps for the treatment and generalization phases (copies of the protocols may be obtained by contacting the first author). Based on step-by-step agreement, the procedural integrity was 98 % (range 91–100 %) for the intervention sessions.

Research Design

The design utilized in this study was an alternating treatments design nested within a multiple baseline across students. The multiple baseline demonstrated experimental control of the treatment condition and replication of the treatment effect for all students. The alternating treatments design provided an extended analysis for verification of the two conditions presented during the brief assessment (Mong and Mong 2012).

Establishing experimental control within generalization research can be difficult due to the possibility of spontaneous generalization of the skill between the treatment and generalization conditions. Generalization is spontaneous when it occurs without explicit programming (Stokes and Baer 1977). The spontaneous occurrence of generalization is likely due to the presence of unknown or unspecified variables in the training situation (Kirby and Bickel 1988; Stokes and Baer 1977). In the current study, a potential variable that may have affected spontaneous generalization is the students' prior learning experiences with the target skills and generalization procedures (Carnine 1980). For example, if a student had previously learned to skip-count on a 100s chart for addition, this may affect their performance of using the 100s chart for multiplication. The first two phases of the design attempted to control for these variables by establishing a stable baseline of both the treatment and generalization conditions and a baseline of the student's pre-treatment performance in the presence of the generalization strategies. Furthermore, to determine whether spontaneous generalization occurred during this study, the researchers intermittently administered probes of generalization facts during the treatment phase. If the students showed significant gains on the generalization probes during the treatment phase, it was determined that spontaneous generalization had occurred due to unknown variables. This study examined generalization for

both accuracy and fluency to determine whether students generalized one or both dimensions.

Procedure

Pre-assessment Procedures

All twelve students whose teachers referred them as potential participants in the study received one single-fact probe of each probe set (6, 7, and 8). An un-highlighted number chart was present at the top of each probe. Researchers instructed students to attempt all the problems and mark X's through any problems they could not complete. Researchers did not provide an explanation for the number chart; students were given 2 min to complete each probe. Students who were more than 25 % accurate or obtained more than 10 DCPM on the probe were excluded from the study because they had beginning multiplication skills.

To rule out performance deficits and prior exposure to the cueing strategy, researchers administered a baseline measure of both the reward and cueing procedures. First, researchers told students that if they could beat their score from the previous day's single-fact multiplication probes (the initial assessment probes), they could choose a novelty pencil from a pencil bag. During this procedure, the non-highlighted number chart was present at the top of the paper. During initial interviews with teachers regarding participation, they indicated that they often used pencils as prizes in the classroom and all participants indicated a desire to earn the pencils. If students increased their scores with the reward above 25 % accuracy or 10 DCPM, then it would have been determined that they had a performance deficit of the single-digit multiplication skills and they would have been removed from the study. Given that none of the students displayed this increase with the reward, researchers administered probes that included highlighted number charts at the top, but did not provide instructions on how to use the number chart. If the students increased their performance with the highlighted cue above 25 % accuracy or 10 DCPM, then it would have been determined that they had prior learning experience with the highlighted number chart and would have been removed from the study. None of the students increased their performance with the highlighting of the number chart; therefore, they proceeded to baseline phase 1.

Baseline (Phase 1)

Researchers assigned the six participants to a different combination of the multiplication facts for the treatment and generalization probes using a counterbalancing procedure. For example, Ryan and John received treatment on 6-fact probes, 7-fact probes were used for their cue strategy, and 8-fact probes were used as the reward strategy, whereas Kristen and Matt received treatment on 7- and 8-fact probes for cue strategy, and 6-fact probes for reward strategy.

During the baseline phase, researchers administered the single-fact multiplication probes of the treatment and generalization skills with instructions to the students to do their best on the problems. They had 2 min to complete each probe. The non-

highlighted number chart was present at the top of each probe; however, no explanation was given of the number chart; goals and feedback were not utilized during this phase. The decision to move to the treatment phase was based on at least three data points of stable performance with either no improvement or declining improvement for all three probes. Initiation of the treatment phase was staggered across participants, to demonstrate experimental control.

Skip-Counting Treatment (Phase 2)

Skip-counting treatment consisted of using a highlighted 100s chart to skip-count, goal setting, and incentive. On the first day of treatment, students received a probe of their treatment skill with a highlighted number chart at the top. The researcher gave students instructions on how to use the highlighted number chart to solve the problems with skip-counting. The researcher modeled the process on the first problem by showing the students how to count on the number chart to obtain the correct answer. For example, if the problem was 6×5 , multiples of six were highlighted on the number chart; the researcher counted five highlighted numbers, which was the number 30. The researcher circled this number and wrote it down as the answer to the problem. The researcher then used guided practice on the next two problems by reading the problems and asking the students how they could use the number chart to solve it and counted with the students on the number chart. If the students stated that they did not know how to use the number chart to solve the problem, the researcher gave the student-specific instructions to count the numbers on the number chart. After the students correctly completed two problems with guided practice, the researchers instructed them to complete the next three problems independently. If the students missed any of these three problems, they received corrective feedback. This process continued until the students correctly answered three problems without corrective feedback. After the students accurately completed three consecutive problems, they received a new probe with the highlighted chart; this probe was used to assess their skill for the first day of treatment. Each successive day of treatment, researchers reminded the students about how to use the number chart with the prompt: “Remember you can use the number chart to help you figure out the answers to these problems by counting the highlighted boxes.” For the treatment phase, the researchers administered 2-min probes of the target skills with a goal written at the top of each page, which were used as the dependent variable for the treatment phase. Goals were established using a percentile shaping procedure in which the daily goal was based on the students’ median performance from the previous 3 days’ probes (Galbicka 1994). The researchers instructed students that if they beat their goal, they would get to put a sticker on the sticker chart and mark their progress on the graph. The students received reminders that after they reached their overall goal on the graph, they would get their choice of either lunch in the classroom or extra computer time. During the initial interviews, the teachers identified a choice of computer and lunch as common rewards utilized in the classroom for which students expressed a desire to earn. During the treatment phase, researchers measured the students’ level of generalization by administering probes of the generalization facts using the baseline procedures at the end of every

third treatment session. Scores from these intermittent probes of the generalization facts were used to determine whether fluency or accuracy was the most appropriate measure for the students during the next phase

To group students based on either accurate or fluent response measures, the multiple baselines were re-arranged at this point in the study. Table 1 shows the accuracy and fluency performance for all six participants. For Travis, John, Ryan, and Matt, accuracy was identified as the primary dependent variable because their accuracy remained at zero during cue and reward baseline probes during the first treatment phase (skip-counting). For Beth and Kristin, fluency was identified as the primary dependent variable because their accuracy was elevated but their fluency did not show elevation during cue and reward baseline probes during the first treatment phase (skip-counting).

Brief Assessment (Phase 3)

The assessment procedures led to a hypothesis about which generalization procedure would produce the greatest degree of generalization. First, students received a probe of the reward generalization facts with the non-highlighted number chart present. Researchers informed students of a goal for performance, based on their average baseline performance of the reward probes. For example, if students had an average of 3 DC on the baseline probes, their goal for the hypothesis probe would be 4 DC. If students beat their goal, they received a choice of novelty pencil from the pencil bag. Then, the students received another probe of the cue generalization facts with a highlighted number chart. Researchers asked the students to complete the worksheet without any explanation of the number chart or the highlights. If a student's performance increased with the presentation of the goal setting and reward, then it was hypothesized that the consequent-based procedure would produce greater effects in the alternating treatment phase. If a student's performance increased with the presence of the highlighted number chart, it was hypothesized that the antecedent-based procedure would produce the greatest gains in the alternating treatment phase.

If students did not show clear differentiation of performance between the skills, the generalization hypothesis probe conditions were each repeated two more times, and the strategy that produced the higher level of performance on at least two of the trials was hypothesized to be the most appropriate generalization treatment for that student. If a student's performance did not increase on either probe for three trials during this phase, researchers added an additional verbal prompt of "use the number chart to help you solve these problems" to the cueing strategy instructions.

Extended Analysis of Intervention Strategies (Phase 4)

The alternating treatment design included implementation of the consequent-based reward procedure and the antecedent-based cueing strategy, with each session on a different school day. The order of the cue and reward days was counterbalanced across students. On the reward day, students received probes of the reward facts with a non-highlighted number chart. The goal was written at the top of each page,

and students received the same sticker chart and self-graphing procedures from the treatment phase. Researchers scored the probes in front of the students after completion and provided the students feedback on their performance, and they marked their score on the graph. On the antecedent days, students received a probe of the facts assigned to the cueing strategy with a highlighted number chart. The researcher instructed the students that they did not have a goal to beat on these days, but they needed to do their best. Researchers did not score the cueing probes in front of the students or provide the students feedback about their performance.

During the extended analysis, the original treatment probes were administered intermittently with a highlighted number chart present to ensure maintenance of these facts. The goal for the extended analysis phase was for each student to display stable, upward trending data with at least two scores above the goal for the intervention strategy indicated as most effective during the assessment.

Verification of Effective Generalization Strategy (Phase 5)

The final phase of the study replicated the effectiveness of the generalization strategies. In this phase, the most effective generalization strategy was applied to both sets of generalization probes. For example, if the cue strategy was more effective than the reward strategy, then the cue strategy was applied to the reward probes and the opportunity for reward was removed.

Data Analyses

This study used visual analysis and calculation of accurate or fluent responding to determine whether the skip-counting procedure was effective and which intervention was most effective during the brief assessment and extended analysis. Furthermore, researchers used a generalization ratio to determine whether spontaneous generalization had occurred during the initial treatment phase and to calculate the proportion of generalization that occurred during the extended analysis phases. The generalization ratio is the average performance of the last three data points for the generalization skill probes over the average performance of the last three data points for the treatment probes in each phase (House et al. 2009).

Results

Prior to the start of the study, the students were randomly assigned to one of two multiple baseline sets for the initial skip-counting treatment phase. Based on the student's level of spontaneous generalization for accurate and fluent responding during the initial skip-counting treatment phase, researchers regrouped the multiple baselines (Figs. 1, 2) and calculated the proportion of generalization for each student in phases 2, 4, and 5 (Tables 1, 2).

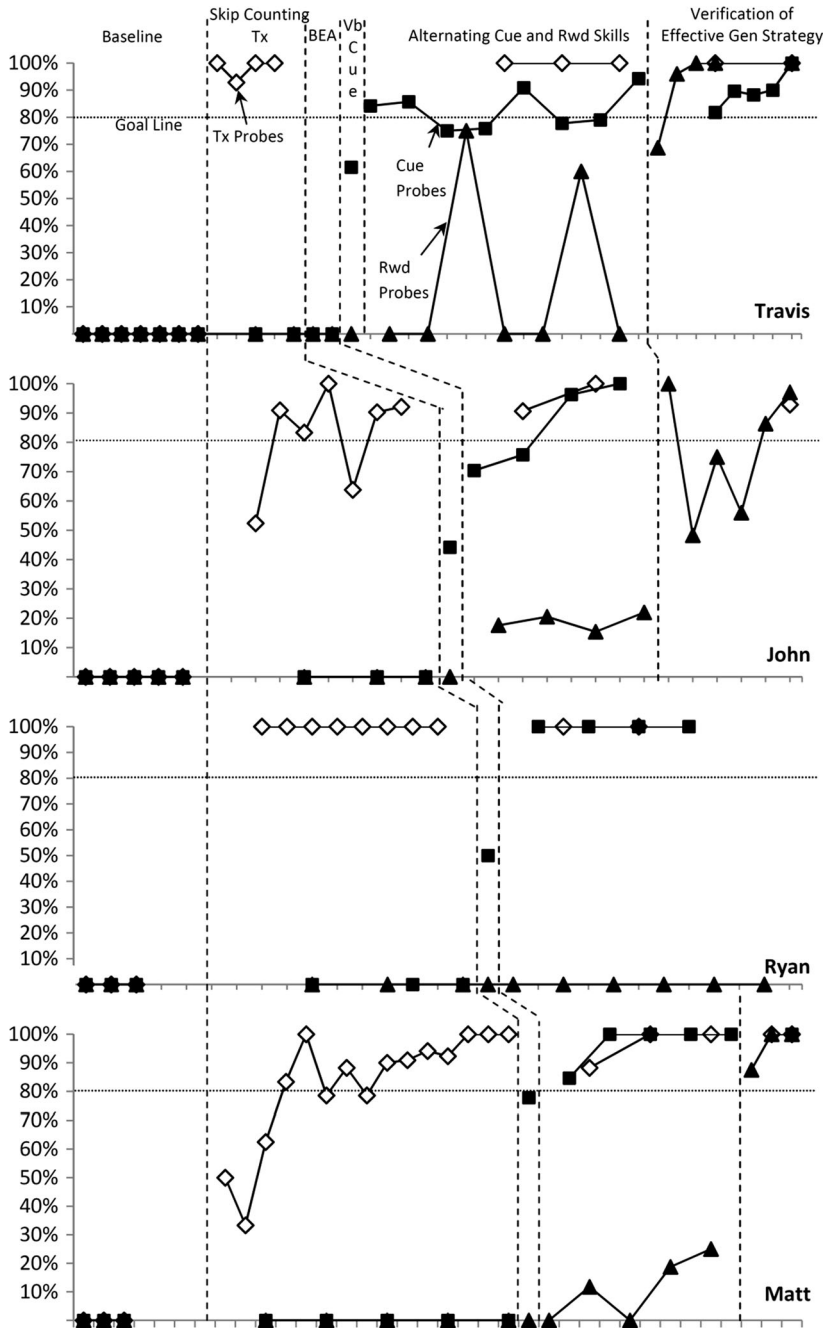


Fig. 1 Multiple baseline for participants with accuracy as the dependent variables. *Open diamonds* depict the treatment probes given throughout the phases. *Closed squares* depict the cue probes given throughout the phases. *Closed triangles* depict reward probes given throughout the phases of the study

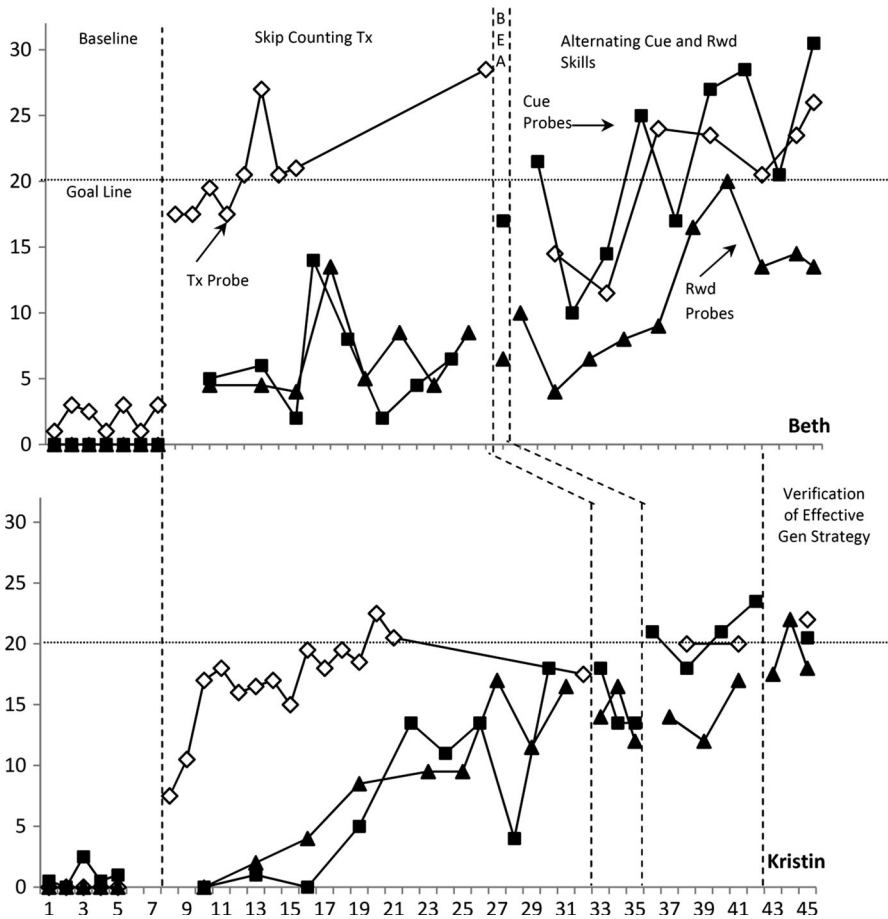


Fig. 2 Multiple baseline for participants with fluency as the dependent variables. *Open diamonds* depict the treatment probes given throughout the phases. *Closed squares* depict the cue probes given throughout the phases. *Closed triangles* depict reward probes given throughout the phases of the study

Baseline and Treatment phases

The results of the skip-counting treatment effectiveness were consistent for all six participants. All participants had levels of 0 % accuracy during the baseline phase, and all participants made immediate gains in accuracy after the initial training of the skip-counting procedure (see first row of data for all participants in Table 1). Ryan, Kristin, Beth, and Travis immediately increased their accurate level of performance above the goal criteria of 80 %. Matt had a more gradual growth in accuracy with three upward trending data points below the goal criteria before reaching the criteria in the fourth treatment session. John’s first score during the treatment phase was below the goal criteria at 52 %, but he increased to 91 % with the second session of treatment.

Table 1 Mean percentage of accurate and fluent scores during treatment phase

	Treatment phase accuracy (%)	Treatment phase fluency
Travis		
Treatment probes	98	20.5
Cue probes	0	0
Reward probes	0	0
John		
Treatment probes	82	20.36
Cue probes	0	1.33
Reward probes	0	1.68
Ryan		
Treatment probes	100	18.06
Cue probes	0	0
Reward probes	0	1.83
Matt		
Treatment probes	83	11
Cue probes	0	2.6
Reward probes	0	
Beth		
Treatment probes	100	21
Cue probes	71	6
Reward probes	88	7
Kristin		
Treatment probes	99	18
Cue probes	46	7
Reward probes	84	8

Figure 1 depicts the results from the four students who had low performance on both accurate and fluent responding of the generalization probes during the treatment phase; therefore, accuracy was selected as the primary dependent variable of investigation for these students. The antecedent cue was most effective for all four students with significant proportions of generalizations observed (see Table 2). During the skip-counting treatment, Travis, John, Ryan, and Matt all displayed a level of 0 % accuracy on the intermittent probes of the generalization skills, and their proportion of spontaneous generalization was 0 for both generalization probes.

Brief Assessment and Extended Analysis

As a result of the stable performance on both the treatment and generalization skill probes, the assessment and extended analysis started first for Travis. Initially, in response to the cue and reward procedures, Travis continued to display 0 % accuracy on both the cue and reward probes. After two trials of 0 % accuracy, a verbal cue to use the highlighted number chart was added to Travis's cueing

Table 2 Generalization ratios of generalization skills/treatment skill

	Treatment phase 2	Alternating treatments phase 4	Verification of effective gen strategy phase 5
Travis			
Cue probes	0	0.84	0.93
Rwd probes	0	0.20	1.00
John			
Cue probes	0	0.95	–
Rwd probes	0	0.20	0.85
Ryan			
Cue probes	0	1.00	–
Rwd probes	0	0	–
Matt			
Cue probes	0	1.04	1.00
Rwd probes	0	0.12	0.96
Beth			
Cue probes	0.26	1.14	–
Rwd probes	0.28	0.66	–
Kristin			
Cue probes	0.36	1.04	0.93
Rwd probes	0.43	0.78	0.91

Cue data were not obtained for John during the final phase. Ryan and Beth did not participate in the final phase

procedure. Upon initiation of the verbal cue, Travis displayed 62 % accuracy on the cue probes and 0 % accuracy on the reward probes. As a result of the assessment, researchers hypothesized the cueing procedure to be the most effective procedure for Travis, which was confirmed in the extended analysis phase. Travis consistently displayed higher levels of accurate performance for the verbal cue condition, with a mean of 83 % accuracy. For the reward probes, Travis had a mean of 19 % accuracy. Furthermore, Travis maintained a level of 100 % accuracy on intermittent probes of the treatment skill during the generalization phase. In the last phase of the study, the verbal cue was added to the reward probes and the opportunity for reward was removed. During this phase, Travis' performance increased on the reward probes to a mean of 93 % accuracy and maintained 90 % accuracy for the cue probes and 100 % accuracy for the treatment probes. The results of the last phase provided further evidence that the cueing strategy was effective for both generalization skill probes. Furthermore, the generalization ratio indicated that Travis displayed 0.84 generalization of the cue probe and 0.20 of the reward skill during the alternating treatment phase. In the last phase of the study, Travis displayed 0.93 generalization of the cue strategy and 1.00 generalization of the skills previously assigned to the reward condition.

John was the next student to begin the assessment and alternating treatment phases. During the brief assessment, John's performance on the cue skill probe

immediately increased to 44 % accuracy with the visual cue of the highlighted number chart, and his performance on the reward skill probe was 0 % accurate, which resulted in a hypothesis that the cue would result in consistently higher levels of performance in the extended analysis. During the extended analysis phase, John maintained a high level of performance on the cue probes with a mean level of 86 % accuracy and a low level of performance on the reward probes with a mean level of 19 % accuracy. Furthermore, John maintained a level of 95 % accuracy on intermittent probes of the treatment skill during the alternating treatment phase. In the last phase of the study, when the cue was applied to the reward probes, John's performance increased to 77 % accuracy and maintained 93 % accuracy on a treatment skill probe. The generalization ratio indicated that John displayed 0.95 generalization of the cue skill and 0.20 generalization of the reward skill during the alternating treatment phase. In the last phase, John displayed 0.85 generalization of the reward skill. John did not receive the cue probes during this phase.

Ryan began the assessment and alternating treatment phase after John. During the brief assessment, Ryan's performance on the cue probe increased to 50 % accuracy, and his performance on the reward probe remained 0 % accurate. During the alternating treatment extended analysis phase, Ryan maintained a high level of performance on the cue probes with 100 % accuracy and a low level of performance on the reward probes with 0 % accurate, which confirmed the results of the assessment. Furthermore, Ryan maintained a level of 100 % accuracy on intermittent probes of the treatment skill during the alternating treatment phase. Ryan displayed 1.00 generalization ratio of the cue skill and 0 generalization of the reward skill. During the alternating treatment phase, Ryan became frustrated with the intervention procedures and refused to complete the math probes. Therefore, Ryan did not complete the final phase.

Matt was the last accuracy student to enter the assessment and alternating treatment phases. During the brief assessment, Matt's performance on the cue skill probe increased to 78 % accuracy, and his performance on the reward skill probe remained 0 % accurate. The effectiveness of the cueing strategy was confirmed in the alternating treatment phase; Matt maintained a high level of performance on the cue probes with a mean level of 97 % accuracy and a low level of performance on the reward probes with 0 % accurate. His performance on the intermittent treatment skill probes was a mean of 96 % accuracy. In the last phase of the study when the cue was applied to the reward probes, Matt's performance increased on the reward probes to a mean of 96 % accuracy and maintained 100 % accuracy for the cue and treatment probes, which confirmed the effectiveness of the cueing strategy for both the generalization skills. Matt displayed 1.04 generalization of the cue skill and 0.12 generalization of the reward skill during the alternating treatment phase. In the verification phase, Matt displayed 1.00 generalization of the cue skill and 0.96 generalization of the reward skill.

Figure 2 depicts the results for the two students, Beth and Kristin, who displayed accurate but not fluent generalized responses during the initial skip-counting treatment phase, and Table 2 includes the proportion of generalization for these two students. During this treatment phase, Beth's proportion of spontaneous generalization for *accuracy* was 0.89 for the cue probes and 0.93 for the reward probes, and

Kristin's was 0.69 for the cue probes and 0.82 for the reward probes. Therefore, both students displayed a significant level of spontaneous generalization on accurate responses for both generalization probes during the treatment phase. Due to this level of spontaneous generalization, researchers utilized the fluency scores, which did not have the same level of generalization, to interpret the results of the generalization procedures.

During the baseline phase, Beth had a level of 0 DCPM on the cue probes and reward probes, and on the treatment probes, she had a mean score of 2 DCPM. Kristin's mean scores during baseline were 0 DCPM for the treatment and reward probes and 1 DCPM for the cue probes.

On the first treatment probe after teaching the skip-counting strategy, Beth's performance increased on the treatment probes to 17.5 DCPM and maintained a mean level of 21.06 DCMP during the treatment sessions. On the intermittent probes of the generalization skills during the treatment phase, Beth's performance was a mean of 6 DCPM for the cue probes and 6.63 DCPM for the reward probes. Beth's proportion of spontaneous generalization was 0.26 for the cue skill and 0.28 for reward skill during the treatment phase.

Kristin's performance on the treatment skill increased to 7.5 DCPM on the first treatment probe after teaching the skip-counting strategy, and she continued to increase her performance and maintained a mean level of 18.23 DCPM during the treatment phase. Kristin's performance on the intermittent probes of the generalization skills was a mean level of 7.33 DCPM for the cue probes and 8.72 DCPM for the reward probes. Kristin's proportion of spontaneous generalization was 0.36 for the cue skill and 0.43 for reward skill during the treatment phase.

As a result of the stable performance on both the treatment and generalization skills, Beth started the assessment and extended analysis first. During the brief assessment, Beth's performance on the cue probe immediately increased to 17 DCPM with the visual cue of the highlighted number chart and her performance on the reward probe was 6.5 DCPM. During the alternating treatment phase, Beth maintained a high level of performance for the cue probes with a mean level of 21.61 DCPM and a lower level of performance for the reward probes with a mean level of 11.5 DCPM, which confirmed the results of the assessment that the cue was a more effective procedure for Beth. The generalization ratio provided the proportion of generalization for each skill during the alternating treatment phase. Beth displayed 1.14 generalization of the cue skill and 0.66 of the reward skill. Due to the start of winter break toward end of the study, Beth did not participate in the verification phase.

Although Kristin's performance on the generalization skills continued to increase during the treatment phase, the assessment and alternating treatment phase began to determine whether the added cue or reward would result in differentiated effects and produce greater increases in her performance. On the first administration of the assessment, Kristin's performance was 18 DCPM for the cue probe and 14 DCPM for the reward probe. Due to the small difference in performance on these two skills, researchers implemented the brief assessment two more times. On the second administration of the cue and reward strategy, Kristin scored 13.5 DCPM for the cue probe and 16.5 DCPM for the reward probe; on the final implementation of the

assessment, Kristin scored 13.5 DCPM for the cue probe and 12 DCPM for the reward probe. On two of the three administrations of the assessment, Kristin performed slightly higher with the cueing strategy than the reward strategy; therefore, researchers hypothesized that the cueing strategy would be most effective for Kristin during the extended analysis. During the alternating treatment phase, Kristin's performance on the cue probes was a mean of 20.88 DCPM, and on the reward probes, her performance was a mean of 15.13 DCPM. Kristin maintained a high level of performance on the intermittent treatment skill probes with a mean of 20 DCPM. During the final phase of the study when the cue strategy was applied to the reward probes, Kristin's performance on the reward probes increased to a mean of 20 DCPM, and she maintained a mean of 20.5 DCPM on the cue probes and 22 DCPM on the treatment probes. The generalization ratio indicated that Kristin displayed 1.04 generalization of the cue skill and 0.78 of the reward skill during the alternating treatment phase. In the last phase of the study, Kristin displayed 0.93 generalization of the cue skill and 0.91 generalization of the reward skill.

Discussion

The primary purpose of this study was to examine the utility of a brief academic assessment to identify effective generalization procedures for individual students who learned a math strategy. Specifically, the basis of this study was the hypothesis that generalization should be viewed as a separate behavior that is under stimulus control and is affected by antecedent and consequent manipulations (Kirby and Bickel 1988). The results of the current study extend the generalization literature with the added notion that different types of generalization strategies can be tested with brief academic assessments to determine the most effective procedure for an individual student.

The results of the study confirm skip-counting as an effective strategy for solving specific multiplication facts. However, the effectiveness of the skip-counting is not of primary interest to the current study; of more importance is the amount of stimulus generalization that occurred with the multiplication facts not directly taught. The four accuracy students did not demonstrate spontaneous generalization to the untaught facts during the treatment phase and maintained a level of 0 % accuracy on all generalization facts during this phase. The other two students spontaneously generalized accurate responses and demonstrated minimal amounts of generalization for fluent responses, using the pre-established generalization criteria of 0.5. Examination of generalization is an important practical question because skip-counting is a common textbook strategy used to teach elementary students to solve multiplication problems (Fennell et al. 1998). Based on the current results, it appears that teachers cannot assume that students will demonstrate stimulus generalization without explicit programming.

Based on the previous literature, it should not be surprising that four of the students in the study did not generalize the simple skill of using a number chart to solve unfamiliar facts. The literature on generalization posits that spontaneous generalization is unlikely to occur without explicit programming (Stokes and Baer

1977; Stokes and Osnes 1989). In addition, previous research on generalization of skip-counting procedures demonstrated that students did not generalize the strategy to untaught skills (Lloyd et al. 1981). Research has also shown that prior training in basic skip-counting (e.g., counting by 2s, 3s, 5s) may facilitate generalization to untaught facts (Carnine 1980; McIntyre et al. 1991). Although the current study attempted to control for prior exposure to the skill with pre-assessments of the students' experience with a 100s number chart for multiplication, it is possible that the two students who did generalize had prior experience with the basic skip-counting procedure.

The primary goal of this study was to examine the utility of using a brief academic assessment to select effective generalization procedures for each individual. The assessment was an effective procedure for identifying which of the two strategies would be most beneficial for the individual student. For Kristin, the data indicated potential effectiveness of both strategies during the brief assessment.

An interesting and potentially important procedure was the use of a ratio to determine the proportion of generalization that has occurred. Generalization is often presented as a discrete phenomenon that either occurs or does not occur (e.g., Ayllon et al. 1983; Lloyd et al. 1981; Noell et al. 2000; Rhode et al. 1983; Weinstein and Cooke 1992). However, generalization is not an all or nothing phenomenon; therefore, it may be more beneficial to measure it as a proportion of a meaningful reference point. In stimulus generalization, the reference point would be the performance in the training condition. When comparing these two levels of performance, the ratio provides a metric to assist in determining if and when generalization programming is necessary and successful. For example, if a student completes addition problems in the training setting with 100 % accuracy and in the generalization setting completes the problems with 40 % accuracy, a generalization strategy is warranted. Additionally, if the generalization procedure used to improve performance in the generalization setting increases performance to 95 % accuracy, then it could be said that the generalization procedure was successful even though performance was not identical in both settings.

In the current study, the use of the generalization ratio allowed for estimates about the amount of generalization during each phase to identify the need for a generalization procedure. In addition, the ratio calculations provided estimates about the effectiveness of the generalization procedures. Visual analysis of the paths and levels for each skill in the extended analysis clearly showed that the antecedent-based procedure was most effective for the four students who did not generalize during the treatment phase. However, Kristin and Beth displayed some level of generalization during the treatment phase with both procedures. By using the ratio, researchers were able to compare the amount of generalization that occurred for each procedure to the student's performance on the initial treatment probes and measure it against the pre-established criteria for significant amounts of generalization.

Several limitations should be considered with this study. The first limitation is the difference in task difficulty between the two generalization strategies. Specifically, in the cueing procedure, the students simply had to count the number of highlights

to find the correct answer. The reward condition involved a more complex process, in which the students typically circled or marked the skip-count numbers on the chart first and then counted numbers to solve the problem.

The differences in response effort between the antecedent and consequent strategies may have had the greatest impact for Beth and Kristin, who demonstrated generalization of both probes. The antecedent procedure required less time to complete because the numbers were already highlighted, which directly affects fluency scores. If the two procedures were equal for effort and time required, differences in the extended analysis might not have been present, specifically for Kristin who displayed mixed results during the brief assessment.

Another limitation is the external validity of the procedures used to make decisions regarding the students' response to the interventions. Experimental control was maintained through the complexity of this research design; however, it is unlikely that school personnel would be able to replicate these procedures in a classroom environment. More research is needed on this topic to understand the most efficient process for implementing a brief academic assessment of generalization within the school environment.

Further, there was a limitation that Ryan and Beth did not complete the final verification phase of the study. Ryan did not complete due to work refusal he began to display on the days when he was given the reward probes and was not able to make progress. Because the multiplication skills were not part of the regular second-grade curriculum, the researchers determined it was in Ryan's best interest to end his participation in the study without the verification phase. Beth did not complete the verification phase due to time constraints of winter break beginning right after she met her goal for the cue strategy.

Finally, the use of the generalization ratio should be examined with caution. Although this formula has potential as a generalization measurement, its use has not been established in the literature. Furthermore, it is not a standardized measure, and the amounts of generalization cannot be directly compared across studies. This study used the arbitrary number of 0.50 as the criterion for considering a significant amount of generalization has occurred. For example, if a student had a 0.5 ratio on a generalization probe, their generalized performance would be about half of what their performance would be on a similar skill that they were explicitly taught. However, future researchers should develop criteria that are more rigorous based on larger samples of responses to generalization procedures.

Future research on generalization should focus on examining the variables involved in the production of generalization and strategies for identifying effective generalization procedures. More research is needed regarding generalization strategies that teachers and practitioners can easily integrate when students do not display generalization. This study provided one method for potentially identifying effective stimulus generalization techniques; however, the lack of differentiation across subjects indicates a need for more information about this assessment procedure. It is possible that other antecedent- and consequent-based procedures would produce greater differentiation between subjects. Furthermore, this strategy needs to be examined with other behaviors and across response generalization and maintenance.

Previous literature has proposed that generalization should be conceptualized as a behavior that can be changed with antecedent and consequent manipulations (e.g., Kirby and Bickel 1988; Stokes and Osnes 1989). In addition, research has shown both antecedent and consequent manipulations to be effective procedures for producing generalization (e.g., Ayllon et al. 1983; Mesmer et al. 2007; Noell et al. 2000; Rhode et al. 1983). However, researchers have not examined the differences in the individual responses to these procedures and a structure for selecting the most effective generalization procedure. The results of this study provide practical implications for practitioners developing academic interventions with the consideration of generalization and a potential structure for developing a procedure based on assessment data to remediate the generalization deficits.

References

- Alessi, G. (1987). Generative strategies and teaching for generalization. *The Analysis of Verbal Behavior*, 5, 15–27.
- Ayllon, T., Kuhlman, C., & Warzak, W. J. (1983). Programming resource room generalization using lucky charms. *Child & Family Behavior Therapy*, 4(2), 61–67. doi:10.1300/J019v04n02_06.
- Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review*, 37(1), 126–136.
- Carnine, D. (1980). Preteaching versus concurrent teaching of the component skills of a multiplication algorithm. *Journal for Research in Mathematics Educations*, 11(5), 375–379. Retrieved from <http://www.jstor.org/stable/748628>.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Education.
- Daly, E. J., I. I. I., Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foreman-Ya, K. (2006). Refining the experimental analysis of academic skills deficits: Part II. Use of brief experimental analysis to evaluate reading fluency treatments. *Journal of Applied Behavior Analysis*, 39(3), 323–331. doi:10.1901/jaba.2006.13-05.
- DuVall, T. D., McLaughlin, T. F., & Sederstrom, G. C. (2003). The differential effects of skip counting and previewing on accuracy and fluency of math facts with middle school children with learning disabilities. *International Journal of Special Education*, 18(1), 1–6. Retrieved from <http://www.internationalsped.com/documents/ACFFE.doc>.
- Fennell, F. M., Altieri, M. B., & Silver Burdett Ginn Firm. (1998). *In mathematics the path to math success!*. Parsippany, NJ: Silver Burdett Ginn.
- Freeland, J. T., & Noell, G. H. (2002). Programming of maintenance: An investigation of delayed intermittent reinforcement and common stimuli to create indiscriminable contingencies. *Journal of Behavioral Education*, 11(1), 5–18. doi:10.1023/A:1014329104102.
- Galbicka, G. (1994). Shaping in the 21st century: Moving percentile schedules into applied settings. *Journal of Applied Behavior Analysis*, 27, 739–760. doi:10.1901/jaba.1994.27-739.
- Haring, N., & Eaton, M. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23–40). Columbus, OH: Merrill.
- House, S., Duhon, G., Hastings, K., & Linden, S. (2009). *Using generalization ratio to determine the amount of generalization between early literacy skills*. Poster presented at American Psychological Association conference, Toronto, Canada.
- Kirby, K., & Bickel, W. (1988). Toward an explicit analysis of generalization: A stimulus control interpretation. *The Behavior Analyst*, 11, 115–129.
- Lloyd, J., Saltzman, N. J., & Kauffman, J. M. (1981). Predictable generalization in academic learning as a result of preskills and strategy training. *Learning Disability Quarterly*, 4(2), 203–216. Retrieved from <http://www.jstor.org/stable/1511005>.

- McIntyre, S., Test, D., Cooke, N., & Beattie, J. (1991). Using count-bys to increase multiplication facts fluency. *Learning Disability Quarterly*, *14*(2), 82–88. Retrieved from <http://www.jstor.org/stable/1510516>.
- Mesmer, E., Duhon, G., & Dodson, K. (2007). The effects of programming common stimuli for enhancing stimulus generalization of academic behavior. *Journal of Applied Behavior Analysis*, *40*(3), 553. doi:10.1901/jaba.2007.40-553.
- Mong, M. D., & Mong, K. W. (2012). The utility of brief experimental analysis and extended intervention analysis in selecting effective mathematics interventions. *Journal of Behavioral Education*, *21*(2), 99–118. doi:10.1007/s10864-011-9143-8.
- Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology*, *39*, 335–355.
- Noell, G., Roane, H., VanDerHeyden, A., Whitmarsh, E., & Gatti, S. (2000). Programming for communication in the classroom following an assessment of skill and performance deficits. *School Psychology Review*, *29*(3), 429–442.
- Rhode, G., Morgan, D., & Young, K. (1983). Generalization and maintenance of treatment gains of behaviorally handicapped students from resource rooms to regular classrooms using self-evaluation procedures. *Journal of Applied Behavior Analysis*, *16*(2), 171. doi:10.1901/jaba.1983.16-171.
- Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: The Guilford Press.
- Stokes, T. (1992). Discrimination and generalization. *Journal of Applied Behavior Analysis*, *25*(2), 429–432. doi:10.1901/jaba.1992.25-429.
- Stokes, T., & Baer, D. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*(2), 349. doi:10.1901/jaba.1977.10-349.
- Stokes, T., & Osnes, P. (1989). An operant pursuit of generalization. *Behavior Therapy*, *20*(3), 337–355. doi:10.1016/S0005-7894(89)80054-1.
- Weinstein, G., & Cooke, N. (1992). The effects of two repeated reading interventions on generalization of fluency. *Learning Disability Quarterly*, *15*(1), 21–28. Retrieved from <http://www.jstor.org/stable/1510562>.