December 2017

Generalization Programming and Performance Feedback: A Writing Intervention with Third-Grade Students

Rigby Dawn Malandrino
Syracuse University

Abstract

Despite the importance of explicitly programming generalization in the context of intervention studies (Stokes & Osnes, 1989), the research base is limited, especially with respect to academic interventions. Given that writing is a particular area of concern in the United States (National Center for Education Statistics, 2012), this is an important area to target. As such, the purpose of this study was to examine the benefit of incorporating explicit generalization programming tactics into a performance feedback intervention that has received support for increasing students' writing fluency (Hier & Eckert, 2014). Toward this aim, 52 third-grade students were randomly assigned to one of two conditions: (a) performance feedback, or (b) performance feedback with generalization programming. Four generalization assessments were administered during pre- and post-assessment phases. It was hypothesized that students receiving performance feedback with generalization programming tactics would outperform a condition receiving performance feedback alone across the generalization assessments. This hypothesis was not supported for any of the generalization assessments. Rather, students in both conditions demonstrated similar improvements in their post-assessment writing performance. As such, in the context of this study, there was not a significant benefit added to the performance feedback intervention by including generalization programming tactics.

Keywords: generalization programming, academic intervention, writing, randomized trial

GENERALIZATION PROGRAMMING AND PERFORMANCE FEEDBACK: A WRITING INTERVENTION WITH THIRD-GRADE STUDENTS

by

Rigby D. Malandrino, M.S.

B.A., West Virginia University, 2010
M.S., Syracuse University, 2015

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in *School Psychology*.

Syracuse University

December 2017

TABLE OF CONTENTS

Generalization Programming and Performance Feedback: A Writing Intervention with Third-Grade Students

Over the past several decades, researchers asserted the importance of explicitly programming and assessing generalization outcomes of intervention studies (Baer, Wolf, & Risley, 1968; Stokes & Osnes, 1989). Despite this recommendation, generalization research remains in a nascent state, as many intervention studies neither assess or report generalization outcomes, and to an even lesser extent, explicitly target generalization (Osnes & Lieblein, 2003). The underdevelopment of generalization literature is especially evident in academic intervention research, which is unfortunate, given that generalization is often an important outcome of academic interventions. It is pertinent that researchers target generalization as a dependent variable and incorporate tactics to promote generalization when engaging in intervention planning.

This introduction includes a review of the theoretical conceptualizations of generalization along with a critique of the research on generalization programming across academic content areas (i.e., mathematics, reading, and writing) focusing on academic intervention studies that either incorporated generalization programming or assessed generalization outcomes. The review concludes with a description of a study that examined the benefit of incorporating explicit generalization programming tactics into a performance feedback intervention targeting writing fluency with third-grade students.

Theoretical Conceptualizations of Generalization

‘Generalization’ refers to a behavior change that endures across time, settings, or is applied to other similar behaviors (Baer et al., 1968). Although generalization is an important outcome of most intervention studies, many researchers have taken a passive approach toward reaching these aims (Stokes & Baer, 1977). That is, researchers have commonly implemented interventions and hoped for positive generalization outcomes to occur, which has been conceptualized as a "train and hope" approach (Stokes & Baer, 1977).

Early proponents of generalization programming examined the literature base and categorized intervention elements employed toward the goal of generalization. Initially, Stokes and Baer (1977) described seven categories of generalization approaches, which included: (a) train and hope, (b) sequential modification, (c) introducing natural contingencies to maintain the behavior, (d) training sufficient exemplars, (e) training loosely, (f) using indiscriminable contingencies, and (g) programming common stimuli. At the time of publication, the authors reported that sequential modification and train and hope were the most commonly used tactics. They also reported that although positive generalization results were sometimes found in the absence of explicit programming, these results did not occur consistently. As such, the authors advocated the importance of explicit generalization programming.

Later, Stokes and Osnes (1989) expanded upon this work, by highlighting three categories of explicit generalization programming tactics, which included: (a) exploiting functional contingencies, which included using tactics involving contacting natural consequences, recruiting natural consequences, modifying maladaptive consequences, and reinforcing occurrences of generalization; (b) training diversely, which included using sufficient stimulus exemplars, using sufficient response exemplars, making antecedents less discriminable, and making consequences less discriminable; and (c) incorporating functional mediators, which included incorporating common salient physical stimuli, incorporating common salient social stimuli, incorporating self-mediated physical stimuli, and incorporating self-mediated verbal and covert stimuli. The authors argued that in order to continue to increase the sophistication of generalization programming, researchers needed to explicitly incorporate these tactics to promote generalization and to assess generalization outcomes. Further, they contended it was insufficient to solely provide information regarding favorable generalization outcomes without clarifying how those outcomes were achieved.

In addition to identifying tactics to actively promote generalization, generalization has been conceptualized within the context of skill development. The Instructional Hierarchy (Haring & Eaton, 1978) is a heuristic that specifies the order in which learning occurs. The model assumes a linear progression of skill development consisting of four main stages: (a) acquisition, in which a skill is learned, (b) fluency, in which the skill is able to be used quickly and accurately, (c) generalization, in which the skill is able to be used in response to novel stimuli (also referred to as stimulus generalization), and (d) adaptation, in which the skill can be modified in response to a novel demand (also referred to as response generalization). Although Haring and Eaton believed these skills are developed in a linear fashion, they noted some uncertainty regarding instructional timing, but speculated that perhaps generalization should be programmed during the fluency stage. Unlike the work of Stokes and Baer (1977) and Stokes and Osnes (1989), this heuristic conceptualized generalization programming within the context of academic performance, and as a result, it has the potential to inform generalization programming within the context of academic interventions.

In sum, the theoretical conceptualizations of generalization evolved from identifying a variety of methods that were used toward the goal of generalization (Stokes & Baer, 1977), to providing recommendations for explicit tactics that should be incorporated into interventions (Stokes & Osnes, 1989) as well as trying to pinpoint the timing of generalization programming within the context of academic skill development (Haring & Eaton, 1978). These conceptualizations provided a foundation for generalization programming in academic intervention studies.

Generalization Programming in Academic Interventions

Although the importance of generalization programming was established in the 1980s, relatively few existing studies have implemented explicit generalization programming tactics in the area of academic intervention research (Skinner & Daly, 2010). This is disheartening, considering that students are typically expected to learn and apply academic skills in ways that deviate from how they were originally taught (Poncy, Duhon, Lee, & Key, 2010). Additionally, when generalization is included in academic intervention research, generalization outcomes are often merely assessed rather than explicitly targeted or incorporated into the design of the academic intervention. As such, relatively few academic intervention studies have been conducive in providing information pertaining to the active programming and assessment of generalization outcomes. This area needs to be developed further considering that many of students in the United States are underperforming across academic content areas, with the most significant underperformance in the area of writing (National Center for Education Statistics, 2012; Persky, Daane & Jin, 2003).

Generalization Programming in Mathematics

Mathematics studies incorporating "Train and Hope" tactics. Wood, Rosenberg, and Carran (1993) examined the effectiveness of a tape-recorded, self-instruction intervention on the mathematics performance of nine 8- through 11-year-old students diagnosed with learning disabilities. The intervention targeted addition and subtraction skills with problems matched to each student's respective skill level. Following baseline, students were divided into one of three achievement levels (low, middle, and high) based on their performance on the Wide Range

Achievement Test – Revised (WRAT-R). Then they were randomly assigned to one of three conditions: (a) an experimental condition that received individualized self-instruction training, (b) an observer condition that permitted the students to observe the experimental condition during training, and (c) a control condition that was not exposed to any intervention elements. The authors used a multiple baseline design across student groups, which were each comprised of one student from each respective condition. The sequence varied by condition; students in the experimental condition received baseline, self-instruction training, a reversal, a second self-instruction training, a generalization assessment session in which students were provided tape-recorded instructions (without experimenter assistance), and a second generalization assessment session in which students did not receive access to the tape-recorded instructions. Students in the observer condition experienced an identical sequence, however, instead of receiving the self-instruction sessions directly, they observed the interventions being implemented with the experimental condition. The control condition did not receive or observe the self-instruction trainings; their sequence through the study included baseline followed by the two generalization sessions. Despite the unique nature of the study design, the study included sound elements to assist with the assessment of generalization outcomes, as the dependent variables (percentage of problems attempted, percentage of problems completed correctly, and duration of time spent completing the items) were examined concurrently over the course of the study.

Visual analysis of the results showed that students assigned to the experimental and observer conditions completed a higher percentage of problems following the second baseline phase in comparison to the control condition. In terms of percentage of problems completed correctly (accuracy), students assigned to the observer and control conditions demonstrated relatively low, stable performance throughout each phase of the study, whereas students assigned to the experimental condition substantially increased their accuracy following the second self-instructional phase. Despite these promising results, students in the experimental condition took more time completing mathematics problems in comparison to students in the other conditions. Results were similar during the final phase of the study in which generalization was assessed in the absence of tape-recorded instruction. That is, students in the self-instruction training and observer conditions completed more math problems, whereas students in the experimental condition increased their accuracy.

These results provide evidence of positive generalization outcomes for self-instruction training in the absence of specific generalization programing. However, the self-instruction condition included implicit generalization programming tactics (i.e., self-mediated verbal and covert stimuli) and this was not discussed or considered in relation to generalization outcomes. Additionally, intervention elements were included during the first generalization assessment phase. Therefore, performance during this phase was not truly indicative of generalized responding, as noted in Cuvo's (2003) work, which specifies that generalization should be assessed in the absence of intervention elements. Fortunately, the second generalization assessment phase did not include this limitation, and represented a sound assessment of generalization. Other limitations include the absence of procedural integrity results and the fact that only students with learning disabilities were included in the study.

In another study that did not incorporate explicit programming tactics, Codding, Eckert, Fanning, Shiyko, and Solomon (2007) implemented an alternating treatments design with three sixth-grade students to compare three variations of a cover-copy-compare intervention targeting mathematical fluency. The three variations of the intervention included: (a) cover-copy-compare alone, (b) cover-copy-compare with performance feedback regarding digits correct per minute, and (c) cover-copy-compare with performance feedback regarding digits incorrect per minute. Problem difficulty was individualized for each student and was identified by the classroom teacher. Generalization was assessed with pre- and post-intervention assessments. Follow-up assessments were administered four to twelve weeks following the conclusion of the intervention to assess stimulus generalization across time. Response generalization was also assessed at follow-up with a slightly more difficult mathematics problem.

The results of the study indicated no discernable differences between students' responding to the three interventions. Across students, an upward trend in performance was observed during the three conditions, with all students eventually demonstrating mastery-level performance. Additionally, students continued to demonstrate mastery-level performance during the follow-up assessment, which assessed stimulus generalization across time. Despite these improvements, students did not demonstrate evidence of response generalization on the slightly more difficult mathematics probe. As a result, this study provided evidence that the three variations of a cover-copy-compare intervention were effective with regard to increasing fluency and stimulus generalization across time, however, this performance was not amenable to a slightly more difficult measure of response generalization.

In another study that targeted computational fluency, Poncy et al. (2010) used a multi-component intervention with three fourth-grade students. A multiple baseline design across participants was used to examine the effectiveness of an intervention that included explicit timing, goal setting, performance feedback, and tangible reinforcement on students' addition skills. Response generalization was assessed concurrently with related and unrelated subtraction probes. If students did not demonstrate response generalization, additional treatment components were added after students reached a mastery level of 40 DCPM on their addition probes. The additional components included didactic instruction (i.e., assisting students in solving subtraction problems by using similar skills used to solve addition problems) and direction (instruction with a cloze procedure).

Overall, students demonstrated an increase in compositional fluency on the addition probes as a result of the intervention. However, these results did not generalize to either of the subtraction probes, even with the inclusion of the additional components. In fact, all of the students demonstrated relatively stable performance on the response generalization measures over the course of the study. The authors postulated that limited findings might have been a result of dissimilarity between the two types of mathematics skills. In terms of generalization programming, the additional components to increase generalization outcomes could be conceptualized as examples of multiple response exemplars, as these tasks required students to use their compositional skills in a different way; however, the authors did not refer to them as such. This is unfortunate, given that this study could provide useful information regarding the effectiveness of the tactics used. Perhaps the tactics used in this study were ineffective, or would have been more effective if incorporated concurrently with the interventions targeting fluency, as suggested by Haring and Eaton (1978). Despite these limitations, a particular strength of this study is that the authors used generalization assessments of graded difficulty levels (i.e., related and unrelated subtraction probes). This feature should be considered in the development of future studies examining generalization outcomes.

Codding, Archer, and Connell (2010) used a multiple baseline design across behaviors (i.e., multiplication problem sets) to examine immediate treatment effects (i.e., accuracy and fluency) in addition to stimulus generalization (i.e., generalization across multiplication probes and time) of mathematics computation skills with a 12-year-old girl. The intervention included incremental rehearsal targeting multiplication facts. Generalization was assessed with three similar multiplication problem sets administered concurrently with intervention sessions: (a) single skill mastery probes, (b) multiplying fractions, and (c) word problems. Outcome measures included measures of accuracy (i.e., percentage correct) and fluency (i.e., rate per minute). Generalization across time was assessed with the single skill mastery and multiplying fraction probes over the course of three sessions (roughly one session per week) directly following the conclusion of the intervention. The word problem sets were excluded due to scheduling difficulties.

In terms of immediate treatment effects, the participant demonstrated improved skill performance across the three probe types as demonstrated by sharp and stable improvements in accuracy, and gradual improvement in rate across the three multiplication problem sets. Additionally, gradual improvement was evidenced across the three stimulus generalization measures over the course of the intervention. The strongest improvement occurred during the final four intervention sessions, which the authors attributed to an increase in fluency of the computation facts. Thus, the authors argued that the findings supported the sequence outlined in the Instructional Hierarchy (Haring & Eaton, 1978), given that the participant's generalization skills were strengthened as fluency became more developed. Students also demonstrated stimulus generalization across time to the extent that their skills were maintained at intervention level across the three follow-up assessments. Although the study did not include explicit generalization programming, a strength of this study is that generalization was assessed with several types of multiplication problem sets. This study showed that in the absence of explicit generalization programming, an incremental rehearsal intervention was beneficial for increasing fluency and generalization outcomes of mathematics computation skills.

Mathematics studies incorporating explicit generalization programming tactics. In a study that used multiple generalization programming tactics, Roca and Gross (1996) examined stimulus generalization outcomes of a report-do-report intervention within the context of a multiple baseline design across subjects with a withdrawal phase. During the intervention, three third-grade students were instructed to prompt praise from their teachers following accurate mathematics problem completion. Reinforcements (i.e., praise and money) were delivered on a fixed ratio schedule by experimenters and intermittently by teachers during students' mathematics classes. As a result, this study incorporated two types of explicit generalization programming tactics; recruiting natural consequences (i.e., prompting praise from their teachers) and making consequences less discriminable (i.e., receiving intermittent reinforcement). Dependent variables included the number of students' praise prompts, the number of correctly completed math problems and correctly spelled words, and the ratio of correctly completed problems in students' regular mathematics and language arts classes. It should be noted that students were explicitly instructed to generalize praise prompting behaviors following correctly completed problems in their mathematics class, but not in their language arts classrooms.

Visual analysis of the results showed that all three students demonstrated increases in praise prompting behaviors in addition to accuracy, as evidenced by clear and immediate increases in the level of praise prompts and teacher praise in addition to accurately completed mathematics problems. These results were also demonstrated in the language arts classroom (i.e., stimulus generalization across settings). In addition to emitting more praise prompts and receiving more teacher praise, students increased their correctly spelled words. These results provided preliminary support for positive stimulus generalization outcomes as a result of the combined use of two explicit generalization programming tactics (i.e., recruit natural consequences and make consequences less discriminable) in the context of a single subject design.

Despite these promising findings, there were a few limitations associated with the study that inhibit one's confidence in the efficacy of the tactics employed to program generalization. First, instructional elements (e.g., modeling, feedback, rehearsal) were included as needed during baseline data collection when students demonstrated difficulty, which limits our ability to make comparisons between baseline and intervention performance. Also, it is unclear if the tactic involving the recruitment of naturally occurring consequences could really be considered as such, since reinforcement was provided on a fixed ratio schedule. This limits the extent to which we can surmise that the conclusions support the use of the tactic to recruit natural consequences.

Summary of Generalization Programming in Mathematics

In summary, of the five mathematics interventions reviewed, the vast majority incorporated "train and hope" procedures. Of these studies, two (Codding et al., 2010; Wood et al., 1993) reported evidence of stimulus generalization, one (Poncy et al., 2010) reported no evidence of response generalization, and the results were mixed for the remaining study (Codding et al., 2007), such that positive stimulus generalization results were found, but there was no evidence of response generalization outcomes. In this study, students demonstrated increases in fluent responding in addition to positive stimulus generalization outcomes.

It should be noted that two of the "train and hope" studies (Poncy et al., 2010; Wood et al., 1993) incorporated intervention elements that were similar to generalization programming tactics, self-mediated verbal and covert stimuli and multiple response exemplars; however, the authors did not explicitly report having used these tactics. This is unfortunate, given that without the explicit identification of tactics used, the results of these studies are easily overlooked in relation to the field of explicit generalization programming. In order to contribute to the literature base regarding tactics that are efficacious in increasing generalization outcomes, it is important that researchers employ consistent terminology and ensure that the presence or absence of explicit programming tactics are accurately detailed.

The Roca and Gross (1996) study was the only mathematics study to incorporate explicit generalization programming tactics. This study demonstrated successful generalization results, thus providing preliminary support for recruiting natural consequences and making consequences less discriminable (i.e., intermittent teacher praise) within the context of generalization programming in mathematics. However, even within this study, there were some limitations that preclude our ability to draw firm conclusions regarding the efficacy of the tactic of recruiting natural consequences. Specifically, it is unclear if the authors correctly applied the tactic of making consequences less discriminable, as reinforcement was administered on a fixed ratio schedule. However, this study provides some initial evidence that incorporating tactics to train diversely and exploit functional contingencies may increase the likelihood of positive generalization outcomes.

Although each of these studies had limitations associated with the programming of generalization (e.g., absence of explicit programming tactics, potential misuse of programming tactics), several of the studies included unique elements that are important to consider in the development of future research studies. For example, a couple of the studies administered assessments with graded difficulty levels to examine generalization outcomes (Codding et al., 2010; Codding et al., 2007; Poncy et al., 2010). Additionally, several studies examined generalization outcomes along multiple metrics (Codding et al., 2010; Roca & Gross, 1996;

Wood et al., 1993). Perhaps the inclusion of these elements would assist with the identification of potential generalization outcomes.

Generalization Programming in Reading

Reading studies incorporating "Train and Hope" tactics. Noell, Connell, and Duhon (2006) used a multi-element single subject design to examine students' accuracy of spelling and reading skills in addition to response generalization of whole word instruction with three first-grade students. During this study, students were trained on reading and spelling words. More specifically, students were trained with 10 words that they were directed to either spell or read, and error correction was provided as needed. If a student required assistance from the experimenter for unknown words, s/he was asked to repeat the correct answer and practice it five times. Generalization sessions were also administered concurrently throughout the study, which required students to provide the untrained response (i.e., spelling the words they were taught to read or reading the words they were taught to spell). Throughout the study, sessions were administered in a partially randomized order in four-session blocks, such that students first completed two training sessions (randomly ordered between reading and spelling) followed by the generalization sessions that occurred in the same skill order as the training sessions. Dependent variables included percentage of words read and spelled correctly. During control sessions, experimenters assessed students' reading performance on untrained words.

Students demonstrated evidence of response generalization, as observed through an increasing trend in both untrained responses (i.e., spelling and reading). As a result, these findings suggest that students were able to demonstrate response generalization in the absence of explicit generalization programming. Additionally, there were no significant differences between the two types of response generalization assessed (i.e., reading to spelling or spelling to reading); however, neither generalized skill improved to the same extent as the trained reading and spelling skills. Several limitations threated the internal validity of the study (i.e., word difficulty not assessed, lack of procedural integrity data). Outcomes of this study support that positive response generalization outcomes may be more likely to occur when the assessment material remains the same, but requires that the skill be modified.

Ardoin, Williams, Klubnik, and McCall (2009) used an alternating treatments design to compare outcomes related to fluency and generalization of two repeated readings interventions with four male students who attended a residential facility for students with emotional and behavioral disorders. Two of the participants were in the second grade, one participant was in fourth grade, and one participant was in fifth grade. The repeated readings interventions differed only by the number of times that the students read the passages (i.e., three versus six repeated readings). Components of the intervention included listening passage preview, followed by repeated readings (three or six), feedback related to number of words read correctly, and error correction. A separate design was not included for the generalization aspect of the study, however, stimulus generalization outcomes were assessed immediately following each intervention session with high word overlap passages, which were created by the authors. The same passages were administered a week later to examine stimulus generalization across time.

Results indicated that students demonstrated increases in reading fluency as a function of both repeated reading interventions, with increased level and trend observed across all students. Similar results were evidenced for students' stimulus generalization outcomes between the two interventions. These results suggest that the three repeated readings were a sufficient number of practice opportunities for students to demonstrate improvements in their oral reading fluency as well as demonstrate evidence of stimulus generalization. However, there were several limitations regarding the assessment of generalization. First, the design was not sufficient for analyzing generalization outcomes, as generalization was only assessed at the conclusion of each repeated readings intervention. This limits our ability to determine the presence of a functional relation between the treatment components and generalization outcomes. Additionally, students were provided reinforcement during generalization assessments, contingent upon improved performance, contrary to recommendations set forth by Cuvo (2003). Lastly, generalization passage difficulty was not controlled, so outcomes may have been confounded by variations in passage difficulty. As such, the findings from this study need to be replicated with studies using sound procedures to assess generalization.

In a similar study, Klubnik and Ardoin (2010) implemented an alternating treatments design with six second-grade students to compare the benefits of two repeated readings interventions, which differed by mode of administration (i.e., independent versus small group administration). The small group administration intervention was comprised of three students that engaged in sequential repeated readings (i.e., students took turns reading sentences appearing in text). In the independent administration intervention, students completed repeated readings individually. Both interventions included listening passage preview, error correction, and contingent reward for improved performance. A control condition was also implemented throughout the study, during which students read unrelated reading passages and no components of the intervention were provided. Following each intervention session, students were assessed individually with stimulus generalization passages and were provided tangible reinforcement (i.e., nickels) contingent upon improved performance.

Results showed that students demonstrated improved performance regardless of intervention type (i.e., individual or group administration). On average, students demonstrated the greatest generalization gains in words read correct per min following group intervention on the generalization passage (control, *M* = 3.33, *SD* = 8.33; group administration, *M* = 23.30, *SD* = 8.07; individual administration, *M* = 11.4, *SD* = 2.51). Follow-up assessments (i.e., stimulus generalization across time) administered a week following the conclusion of the intervention showed that students demonstrated even further improvement in relation to baseline. Similar to the Ardoin et al. (2009) study, a major limitation of this study is that intervention elements were incorporated during the assessment of generalization (i.e., reinforcement contingent upon improved performance). Therefore, it is unclear the extent to which positive generalization outcomes were due to the intervention or the reward. Additional limitations include that potential differences in passage difficulty between sessions were not controlled.

In a study that assessed both stimulus and response generalization, Cohen and Brady (2011) examined the effects of a reading intervention targeting acquisition of word decoding in the context of a multiple baseline design across behaviors. Although the intervention was implemented class-wide in a second-grade classroom, results were examined for only five students who met inclusionary criteria for the study (i.e., attended school regularly, full-scale IQ > 90, demonstrated adequate phonics abilities, and qualified for special education services for learning disability in reading). Components of the intervention included meaning-based activities (i.e., shared reading) and explicit phonics instruction with three different vowel patterns (i.e., magic e, double vowels, and closed vowels). Moreover, several additional elements were included in the intervention that required students to use the instructed skills in different ways (e.g., tracing vowels with a red marker, using American Sign Language to make the sign for vowels). Generalization was assessed concurrently throughout the study with word decoding tasks on untaught words and nonsense words (i.e., response generalization). Stimulus generalization across time was assessed with follow-up sessions, which commenced following instruction for each vowel pattern, and spanned over the course of seven weeks following the conclusion of the intervention.

Overall, study results showed that students demonstrated an increase in accurate decoding across the three types of vowel patterns, relative to baseline. Students also demonstrated evidence of response generalization with increased decoding accuracy on novel and nonsense generalization words. These results were maintained during the assessment of generalization across time. In addition to maintaining the skills across time, students demonstrated further improvement in the skill use. However, there were multiple threats to the internal validity of the study that limit our ability to draw firm conclusions regarding the efficacy of the intervention for increasing generalization outcomes. First, there were many instructional elements included in the intervention that were not sufficiently controlled experimentally, and procedural integrity data were not reported. As a result, it is difficult to discern which instructional elements resulted in positive generalization outcomes. It is also possible to conceptualize the intervention as having included the tactic of using multiple response exemplars. As such, positive generalization outcomes that were noted may be attributable to those tactics. Lastly, external validity was threatened by individualization of the intervention for students who demonstrated inadequate improvement over the course of the study, and the specific demographics of the students who were included as participants.

Reading studies incorporating explicit generalization programming tactics. In a study that targeted the development of basic reading skills, Mesmer et al. (2010) incorporated common salient physical stimuli to promote stimulus generalization of word decoding among four second-grade students. Using a multiple baseline design across three participants (with a replication of the procedures with a fourth participant), the authors examined the effects of highlighting common word-endings (i.e., end, en, et, ell) with color to improve the accuracy and generalization of students' word reading. During the training phase of the study, experimenters used flashcards with color-coded endings to direct students' attention to common stimulus features of targeted words as instructional reading decoding procedures were provided (i.e., experimenter model, practice, error correction), along with directions that emphasized the common features of the words (i.e., shared word endings and respective color code). Following each session, students' decoding accuracy was assessed on the same words but the flashcards did not contain the color-coding as a means of assessing stimulus generalization. Additionally, spontaneous generalization was assessed with 24 additional untaught words that did not include color-coded word endings. During the generalization phase, identical intervention procedures were followed, except all of the word endings presented on the flashcards (trained and untrained words) were color-coded as a means of prompting (i.e., common salient physical stimuli).

Results showed that three of the four participants demonstrated evidence of spontaneous stimulus generalization during the intervention; that is, they demonstrated improved performance on the 24 untrained words relative to baseline. However, all students demonstrated positive stimulus generalization outcomes during the generalization phase. These improvements were evidenced for both trained and untrained word sets. A limitation of the study was that insufficient baseline data were collected for two of the participating students, resulting in limited predictability of performance prior to intervention implementation. Additionally, the same set of flashcards was used over the course of the study, which may have resulted in improvements that were due to practice effects. This is especially problematic for the generalization words, which were administered throughout the intervention and generalization phases. Nonetheless, generalization assessments were also incorporated during baseline data collection, which is a strength in terms of the design incorporated to assess generalization outcomes. Similar to other studies (Cohen & Brady, 2011; Poncy et al., 2010), these findings provide promising implications for the use of graded generalization measures in the assessment of generalization outcomes. These results also provide preliminary support for using common salient physical stimuli to increase students' stimulus generalization across generalization measures of graded difficulty levels; however, these results need to be replicated with studies that address the limitations associated with inadequate baseline data and potential practice effects.

In another study that targeted word decoding, Peterson-Brown and Burns (2011) used incremental rehearsal with vocabulary (conceptualized as a 'train loosely' tactic) to promote the stimulus generalization of word decoding in 61 second- and third-grade students. Using a between-subjects design, an incremental rehearsal intervention (i.e., flashcards with eight known words and seven unknown words) was implemented with two groups: one that received the normal incremental rehearsal procedure, and the other, which included an added vocabulary component (i.e., students were asked to provide definitions for the words; if they could not, a short definition was provided for them). Stimulus generalization was assessed a week following the conclusion of the intervention session by having the students read the previously rehearsed word in the context of a novel sentence. Although generalization was assessed at the conclusion of the intervention, it was not included as a baseline assessment.

Results showed that incremental rehearsal with vocabulary was more effective than incremental rehearsal alone in increasing students' stimulus generalization across time ($d = 0.71$) and generalization of accurate word decoding ($d = 0.83$). These findings suggest that incremental rehearsal with vocabulary is effective at increasing two types of stimulus generalization (i.e., generalization across time and generalization across stimuli). However, a major limitation with regard to the design aspects of the study is that generalization was only assessed as a post measure. As such, it is impossible to rule out the possibility that there were pre-existing differences in students' performance on this measure. Additionally, the authors identified the programming tactic used as 'train loosely' (Stokes & Baer, 1977), a tactic which has since been described as a category which includes multiple programming tactics. The vocabulary tactic used in this study most closely matches the tactic of multiple response exemplars. As such, this study provided implicit support for using multiple response exemplars to increase stimulus generalization; however, the inadequate design to assess generalization precludes one's ability to draw firm conclusions. The study should be replicated with a design that includes both pre- and post-assessments of generalization.

Ardoin, McCall, and Klubnik (2007) implemented an alternating treatments design to compare the relative effectiveness of two variations of a repeated readings intervention targeting stimulus generalization of reading fluency. Participants included six third-grade students. Both interventions included listening passage preview, phrase drill error correction, and token reinforcement. The two interventions differed by the type of repeated readings materials; that is, the first intervention required students to read the same passage four times. The second intervention required students to read two high word overlap passages twice. As a result, this study incorporated the generalization tactic of multiple stimulus exemplars in an attempt to improve students' oral reading fluency during intervention and on generalization assessments. Generalization was assessed by comparing pre- and post-performance on a generalization assessment for each intervention session using the same probe for both pre- and post-assessment.

Contrary to the study hypotheses, the multiple stimulus exemplars intervention did not produce greater generalization outcomes. In fact, students evidenced greater stimulus generalization effects following the repeated readings intervention. As such, in the context of this study, the generalization tactic of incorporating multiple stimulus exemplars was not effective in increasing students' stimulus generalization. One possible reason for this finding was that there was insufficient exposure to the multiple exemplars. It is also important to consider that both repeated readings interventions consisted of several components (i.e., listening passage preview and phrase drill error correction), which could be considered multiple response exemplars. As such, it could be conceived that both intervention packages included generalization programming tactics.

Although this study attempted to experimentally examine the effectiveness of a generalization tactic to improve students' stimulus generalization, notable limitations were observed in the generalization assessment methods. First, the generalization assessment included components of the intervention (i.e., tangible reinforcement) and did not meet generalization assessment standards recommended by Cuvo (2003). Second, pre- and post-generalization measures were identical. As such, generalization gains may have been due to practice effects. Third, the nature of the experimental design does not protect from multiple treatment interference. As a result, some of the generalization effects observed during the repeated readings intervention may have been due to the multiple stimulus exemplar intervention. Future generalization studies should incorporate design components to reduce multiple treatment interference as well as develop isolated generalization assessments in addition to administering different assessments during pre- and post-generalization assessment to avoid the possibility of practice effects.

Ardoin, Eckert, and Cole (2008) conducted a within-subjects group design to evaluate the effects of two repeated readings interventions on students' initial reading fluency. In this study, second- (n = 25) and fourth-grade (n = 17) students read one passage three times and three different passages one time each (i.e., multiple stimulus exemplars). Both interventions also included listening passage preview and error correction (i.e., elements that could be regarded as multiple response exemplars). Stimulus generalization was assessed as an outcome measure with medium and high word overlap passages following each intervention session, however these assessments were not administered during baseline. The medium overlap passages were considered to be a more difficult assessment of generalization, given that they contained more novel words.

The results showed that although the repeated readings intervention led to significantly greater gains in students' oral reading fluency growth, greater stimulus generalization effects on the medium word overlap passages were observed following the multiple stimulus exemplars intervention t (42) = -2.54, p = .01. There were no discernable differences evidenced between the interventions on high word overlap passages t (42) = -.12, p = .90. It is important to note that although one type of generalization programming tactic was purportedly used in this study (i.e., multiple stimulus exemplars), the multiple elements included in both interventions (i.e., listening passage preview, error correction) could also be conceived as a generalization programming tactic, multiple response exemplars. As such, this study could be viewed as examining the added benefit to including multiple stimulus exemplars. A particular strength of this study is that generalization was assessed in a graded fashion. This type of generalization assessment may permit one to more accurately assess the strength of generalization outcomes. This might explain why positive stimulus generalization outcomes were not found in the Ardoin et al.

(2007) study. Despite these positive outcomes, it is a concern that generalization was only assessed as an outcome measure. Additionally, there may have been differences in the difficulty level of the medium overlap passages that were administered following the two interventions.

In another group design that compared repeated reading to a multiple stimulus exemplar intervention, Silber and Martens (2010) used a between-subjects group design to examine stimulus generalization of 111 first- and second-grade students' oral reading fluency. Students were randomly assigned to one of three conditions: (a) a multiple exemplar condition (i.e., listening passage preview, repeated readings with four sentences, which were representative of the intervention passage, and rewards), (b) a listening passage preview/repeated readings condition (i.e., listening passage preview, guided practice with 16 representative sentences, and rewards), and (c) a control condition (experimenter-administered mathematics probes, and rewards). Both interventions were conducted in a small group format. Stimulus generalization was assessed directly following the intervention session with a high word overlap passage, which was created by the authors. In addition to examining generalization outcomes, the authors also examined learning rates (i.e., gains in words read correct per minute following the intervention relative to baseline divided by intervention time in minutes).

Results indicated that students in both intervention conditions demonstrated significantly greater gains in their reading fluency on the intervention passage than students assigned to the control condition. Although results for the generalization measure indicated that students in the multiple exemplar condition outperformed students in the control condition ($p < .008$, $d = 0.70$), there were no statistically significant differences in performance on the generalization measure between the two intervention conditions ($p = .75$, $d = 0.15$). An analysis of learning rates (i.e., the change in reading fluency on both intervention and generalization passages from pre- to post-intervention) was highest for students in the multiple exemplar condition t (49.15) = 6.21, $p <$ .0005, d = 1.04, despite an increased number of practice opportunities for key words in the listening passage preview / repeated readings condition (i.e., 16 practice opportunities versus 4). Thus, outcomes examining learning rates demonstrated that the multiple exemplars intervention was more efficient at increasing fluency, as a faster learning rate was established than in the learning passage preview / repeated readings condition. Study limitations included the limited external validity of the materials (i.e., author-created to have high word overlap with intervention probes) and generalization assessment methods (i.e., repeated administration of probes during baseline and following the intervention). Additionally, similar to the previous studies examining repeated readings interventions (Ardoin et al., 2008; Ardoin et al., 2007), this study included listening passage previews along with repeated readings, which are similar to multiple response exemplars, given that students are required to read silently, then read aloud. Results of this study provide support for multiple exemplar training as a generalization programming tactic, however, future studies should include different assessments during pre- and post-assessment of generalization to rule out the threat of practice effects. Additionally, this study provided interesting implications for using different metrics in the assessment of generalization.

In a study that incorporated multiple generalization programming tactics (i.e., multiple response exemplars and common salient social stimuli) Duhon, House, Poncy, Hastings, and McClurg (2010) examined response generalization following a letter sound fluency intervention. This study used a multiple baseline design across participants with three first-grade students. Similar to Noell et al. (2006), this study examined whether response generalization would occur across related academic skills; in this case, letter sound fluency and letter sound blending. In this intervention, letter sound fluency was targeted first through direct intervention, which included repeated practice, goal-setting, and contingent reward (i.e., choice of prize). After meeting an established fluency criterion on letter sound fluency, the generalization phase began (with the exception of one participant, who never met the fluency criterion), wherein the authors implemented generalization tactics that varied as a function of students' responses. Generalization was assessed concurrently over the course of the study. Initially a cuing procedure was implemented, which involved the experimenters presenting instructions (i.e., "read the whole word") that were similar to those provided during the letter sound fluency intervention. If the students did not demonstrate increased letter sound blending per minute in response to this cue, the authors included a goal-setting component. If students still did not increase generalized responses, then the authors incorporated sufficient response exemplars (i.e., modeling, guided practice, and corrective feedback). The latter tactic was only implemented for one participant who did not meet the fluency criterion prior to the generalization phase.

Results showed that students demonstrated response generalization at different points in the study. One student demonstrated evidence of spontaneous generalization during the letter sound fluency intervention, however, demonstrated increased improvement after meeting the fluency criterion, and when the cuing procedure was implemented. A second student also demonstrated improved performance on letter sound blending following the inclusion of cuing and goal-setting procedures. Despite these promising outcomes, one student did not demonstrate response generalization despite having received both the cuing procedure and exposure to the sufficient response exemplars. It is possible that the lack of response was due to not having met the fluency criterion prior to generalization programming. These results highlighted inter-individual differences in response to an intervention targeting letter sound fluency and provided support for the use of cuing procedures to program response generalization. Results also suggest that the likelihood of response generalization increases once an initial performance criterion is met and in conjunction with generalization programming. As such, these results demonstrated that although some students may evidence some degree of spontaneous generalization, these outcomes are increased following an increase in fluency in basic skill development, and when generalization programming tactics are employed.

Summary of Generalization Programming in Reading

In summary, four of the ten reading intervention studies reviewed did not employ explicit generalization programming tactics; however, all of them reported positive generalization outcomes. Some authors related generalization outcomes to an increase in fluency (Ardoin et al., 2009; Klubnik & Ardoin, 2010). Despite these promising findings, the majority of these studies had serious limitations related to the design and assessment of generalization outcomes. For example, several studies only included generalization as a post measure following intervention sessions (Ardoin et al., 2008; Ardoin et al., 2009; Peterson-Brown & Burns, 2011). Additionally, in two of the studies, reinforcement was included during generalization assessments, contrary to recommendations by Cuvo (2003). Lastly, the Cohen and Brady (2011) included a tactic similar to multiple response exemplars. As such, there was some implicit support for including multiple response exemplars as a tactic to increase response generalization outcomes in reading.

Of the studies that included explicit generalization programming tactics, 83% showed some degree of positive stimulus (Ardoin et al., 2008; Mesmer et al., 2010; Peterson-Brown & Burns, 2011; Silber & Martens, 2010) and response generalization (Duhon et al., 2010) outcomes. Two of the studies administered generalization assessments prior to the commencement of explicit generalization programming in order to assess spontaneous generalization outcomes (Duhon et al., 2010; Mesmer et al., 2010). In both studies, there was evidence of slight spontaneous generalization (i.e., positive generalized responses varied among participants), but generalization outcomes were stronger following explicit programming. Effective programming tactics included common salient social (Duhon et al., 2010) and physical (Mesmer et al., 2010) stimuli, in addition to multiple exemplar training (Ardoin et al., 2008; Peterson-Brown & Burns, 2011; Silber & Martens, 2010). Limitations included only assessing generalization outcomes following the intervention (e.g., Peterson-Brown & Burns, 2011), including intervention elements (i.e., tangible reinforcement) in generalization assessments (Ardoin et al., 2007), and using the same probes during pre- and post-assessments (Ardoin et al., 2007; Silber & Martens, 2010). Additionally, several of the studies included intervention elements that could be considered as additional generalization programming tactics, but were not identified as such (Ardoin et al., 2008; Ardoin et al., 2007; Silber & Martens, 2010). Several of the reading studies reviewed included unique design elements that are important to consider for future generalization programming studies, such as the use of multiple assessments of graded difficulty (Ardoin et al., 2008; Mesmer et al., 2010) and separate metrics to examine generalization outcomes (Silber & Martens, 2010). Overall, multiple exemplar training received the most support for generalization programming in the area of reading.

Generalization Programming in Writing

Writing studies incorporating "Train and Hope" tactics. In one of the first intervention studies in the domain of writing to examine generalization outcomes, Van Houten (1979) used a multicomponent performance feedback intervention (i.e., explicit timing, self-scoring, and publicly posted performance feedback) targeting fluency of writing (i.e., total words written, different action words written). A multiple baseline design across settings (i.e., mixed second- and third-grade classroom; mixed third- and fourth-grade classroom) was used. Two follow-up sessions were administered one and three months following the conclusion of the intervention to assess stimulus generalization across time. Researchers first targeted the total number of words written followed by a second phase targeting different action words. For each skill targeted, students received 7 min to write, then they were instructed to count the number of words written or different action words, and their performance was posted on a classroom chart. Stimulus generalization assessments were conducted later in the day following each experimental session, but in the absence of treatment components (i.e., scoring, public posting).

All students demonstrated increases in their overall writing production in response to the intervention, as evidenced by an increase in their total words written. When different action words were explicitly targeted, there were differential outcomes such that students in the mixed third- and fourth-grade students evidenced some increase in performance, whereas the mixed second- though third-grade students did not. Similar stimulus generalization results were observed; all students evidenced generalization across settings for total words written. Additionally, the mixed third- and fourth-grade students evidenced generalization of number and percent of different action words; the mixed second- and third-grade students did not. Stimulus generalization across time at both follow-up sessions was evidenced for total words written; during this time, both classes demonstrated performance that approximated their performance during the intervention phase, which targeted total words written at both times. As such, this study provided some evidence of positive stimulus generalization outcomes when assessed at a later period in the day as a result of the multicomponent intervention; however, this was not consistent for writing skills that were not initially strengthened during the intervention. This suggests that fluent use of a skill might be important before stimulus generalization is targeted.

In an intervention targeting spelling, Pratt-Struthers, Struthers, and Williams (1983) implemented an add-a-word program with nine 10-12-year-old students in a multiple baseline design across behaviors (i.e., target spelling words). The intervention targeted the acquisition and generalization of spelling words with students who were classified as learning disabled and received instruction in a special education setting. The add-a-word intervention consisted of daily practice with individualized spelling lists. Students were required to copy each word from their respective spelling list, write it again from memory, and compare the word to a target. If a word was misspelled, the process was repeated until the student spelled the word correctly. Following each add-a-word intervention with the itemized words, spelling tests were administered; words were considered 'mastered' when the student responded correctly during two consecutive sessions. Following the spelling intervention, students were required to engage in creative writing activities. Response generalization was assessed by examining the number of target words that students correctly spelled in their creative writing products. Stimulus generalization was assessed by measuring the students' spelling accuracy on creative writing assessments.

All students demonstrated sharp increases in spelling accuracy on target words that were written in creative writing products immediately upon introduction of the intervention. At the conclusion of the study, all students demonstrated at least 80% accuracy on targeted spelling words referenced in their creative writing assessments, providing evidence of positive stimulus generalization outcomes as a result of the add-a-word intervention. A limitation of this study is that the authors did not provide information regarding specific words that were targeted or how often those words were used during creative writing periods. Given that students were not instructed to use the words targeted in the add-a-word intervention, it is unclear the extent to which they continued to use the words in their writing samples. Despite these limitations, it is plausible that positive stimulus generalization effects occurred because they were preceded by evidence of acquisition, which relates to the Instructional Hierarchy (Haring & Eaton, 1978).

Using a between-subjects group design, Schunk and Swartz (1993) examined the effectiveness of strategy instruction to increase 40 fourth-grade students' writing performance. Although this study did not specify the use of generalization programming tactics, across all conditions, four different strategies for writing paragraphs (i.e., descriptive, informative, narrative story, and narrative descriptive) were embedded in the instruction, which can be conceptualized as a form of multiple response exemplars. Additionally, two different topics for each of the instructed writing strategies were used, which could be conceptualized as a form of multiple stimulus exemplars. In this study, students were randomly assigned to one of four conditions: (a) product goal (i.e., learning to apply correct strategies for different types of writing), (b) process goal (i.e., composing paragraphs), (c) process goal plus feedback, or (d) instructional control (i.e., working productively). Students' writing performance was assessed along the following dimensions: (a) self-efficacy, (b) writing achievement, (c) self-reported use of writing strategy steps. In addition, response generalization was purportedly assessed by students' writing performance on two different types of writing tasks (i.e., compare and contrast, expressive). Stimulus generalization was assessed with a follow-up assessment administered six weeks following the intervention, in the absence of treatment elements.

Unfortunately, only the overall treatment effects were reported in this study. No comparison tests were conducted to determine which experimental conditions resulted in differences in students' responding. However, in terms of direct treatment outcomes, the authors noted a trend based on their descriptive findings, which suggested that students assigned to the process goal plus feedback condition outperformed students in the other conditions, followed by students assigned to process goal condition. Additionally, the authors did not report results for the response generalization measures (i.e., the compare and contrast and expressive writing assessments), therefore, the extent to which the students demonstrated response generalization was unclear. The authors reported that the immediate treatment effects were also demonstrated in the follow-up assessment to examine stimulus generalization across time. Limitations associated with the intervention include absence of procedural integrity data and incomplete analysis of treatment effects. Additionally, increases in stimulus generalization performance may have been due to implicit generalization programming tactics (i.e., multiple stimulus and response exemplars) that were included in the study. As such, these results provide some support for including multiple exemplars in a performance feedback and goal-setting intervention to increase stimulus generalization in the area of writing; however, more studies are needed.

In another study that incorporated implicit generalization programming tactics, Medcalf, Glynn, and Moore (2004) examined the effectiveness of peer tutoring in increasing students' writing performance (e.g., planning, editing) and stimulus generalization (i.e., generalization across individuals and time). A between-subjects group design was used that included two conditions: a peer tutoring condition (*n* = 7) and a control condition (*n* = 4). Students assigned to the peer tutoring condition received guidance from peer tutors during the writing process with regard to assistance with planning, text generation, and editing. Peer tutors also provided praise contingent upon utilization of proper writing skills. This intervention is similar to the generalization programming tactic of incorporating common salient social stimuli, however, it was not explicitly defined as such by the authors. Students assigned to the control group practiced writing without the help of peer-tutors. To analyze the results, five writing samples were taken at three periods (i.e., baseline, during intervention, and at follow-up) for each participating student. Stimulus generalization was assessed during the tutees' regularly scheduled class time (in the absence of tutor support) and was analyzed with measures of rate (i.e., total words written, total sentences written), accuracy (% correct punctuation, % words spelled correct), and quality (i.e., teacher ratings of enjoyment, clarity).

Although statistical analyses were reported, it was unclear if the data presented related to immediate intervention effects or stimulus generalization; therefore, outcomes will be reported in terms of descriptive findings. For immediate intervention effects, the students who received peer tutoring demonstrated an overall mean increase in total words written in comparison to the controls, not quite as strong for total sentences written. However, it appears that students in the control condition exhibited stronger performance at baseline. Similar findings were described for punctuation (i.e., an increase from 4.8% to 62.8%) and spelling accuracy (i.e., an increase from 53% to 91%) in addition to teacher ratings of enjoyment and clarity. These results were similar for stimulus generalization assessments that were conducted in the absence of peer-tutoring and during the follow-up (i.e., stimulus generalization across time) assessment. Overall, it is difficult to compare the two interventions with the results provided. Given the group differences at baseline, it might have been beneficial to control for baseline performance. Improvements that were noted by the authors may be have been related to the use of the implicit generalization programming tactic, common salient social stimuli; however, it is difficult to make strong claims regarding the outcomes of the study due to the lack of information pertaining to results of the intervention.

More recently, Hier and Eckert (2014) conducted a randomized controlled trial to examine the effectiveness of a performance feedback intervention on third-grade students' (N = 103) writing fluency. As part of the study, stimulus generalization was assessed across measures and time (i.e., two, four, and six weeks following the intervention). For this intervention, students received individualized performance feedback regarding their performance on a 3 min writing task. Writing probes used for the intervention included self-referenced story-starters (e.g., "One day when I went to school…"). Stimulus generalization was assessed with a writing probe that included a story-starter that was "others referenced" (e.g., "One day when he went to school…") and was administered only orally, rather than the standard oral and visual administration. Stimulus generalization effects were analyzed by comparing pre- and post-performance. An ANCOVA was used to examine between group differences on the stimulus generalization measure, to control for baseline performance. Stimulus generalization across time was measured during three periods that occurred two, four, and six weeks following the conclusion of the intervention.

Results showed that in comparison to a practice-only condition, students assigned to the performance feedback condition demonstrated significantly greater growth in their writing fluency over the course of the study. Additionally, students in the performance feedback condition significantly outperformed the control condition on a measure of stimulus generalization F (1, 87) = 5.62, p = .02, partial η^2 = .06. Despite these promising findings for the performance feedback condition, the practice only condition demonstrated stronger stimulus generalization effects in terms of stimulus generalization across time. Overall, these results provide support that in the absence of programming generalization tactics, students who improve their writing fluency during intervention are more likely to demonstrate stimulus generalization across generalization measures. A particular strength of this study is that a plan for generalization assessment was included in the design element of the study, such that pre-post performance could be examined; additionally, baseline performance was used as a covariate.

Writing studies incorporating explicit generalization programming tactics. In one of the first writing interventions to interrelate generalization programming tactics and the Instructional Hierarchy (Haring & Eaton, 1978), Jackson (1995) implemented a performance feedback intervention which incorporated the generalization tactic of self-mediated physical stimuli (i.e., self-recording of specific compositional variables) to examine response generalization of writing fluency using a multiple baseline design across behaviors (total words written, action verbs, describing words) with six students. During this study, three phases were examined: (a) baseline, (b) self-mediated performance feedback intervention of compositional variables, and (c) generalization programming. Generalization probes (i.e., teacher-generated writing assignments) were administered during the baseline and self-monitoring phases, which were completed during the students' regularly scheduled class time and were not followed by self-management. The generalization programming phase commenced once reinforcement (i.e., points) was applied to all three compositional variables.

Results showed that self-mediated feedback was at least moderately effective at increasing students' writing fluency (i.e., total words written) for all but one of the students. That is, the percentage of non-overlapping data points was 83% or higher for five of the participants between the highest baseline data point and performance during the intervention. These results were not as strong for different action verbs or different describing words. Descriptively, students who showed increases in rate of responding during the intervention were more likely to demonstrate generalized responding following intervention. Alternately, skills that were in the acquisition phase were less likely to generalize. However, conclusions are limited regarding the effectiveness of the intervention due to considerable variability in participants' responding during the baseline and intervention conditions. Ultimately, this study provided support for the use of self-mediated physical stimuli as a tactic to promote generalization of writing skills that are amenable to interventions targeting increases in fluency.

In a study that incorporated multiple stimulus exemplar training, Campbell, Brady, and Linehan (1991) examined the effects of peer-mediated instruction on the acquisition and generalization of capitalization skills with three nine-year-old students. For this intervention, peer tutors worked with tutees to review capitalization rules, provide feedback regarding capitalization use in tutees' written products, and provide guided practice for capitalization in sentence writing. This intervention included a combination of single subject designs, which were unique to each participant. For the first participant, a multiple baseline design across peer partners was used to target the acquisition of capitalization with peer partners who provided direct training, and to assess stimulus generalization with peer partners who did not implement the training. For the remaining two participants, an ABAB withdrawal design (A = baseline, B = peer teaching) was used to examine acquisition of capitalization. Stimulus generalization was assessed concurrently with capitalization probes that were administered by other peers. Additionally, maintenance probes (i.e., stimulus generalization across time) were administered once weekly for three weeks following the conclusion of the intervention.

All participants demonstrated variability in capitalization use during baseline data collection. However, there were clear improvements in performance at the onset of the peer teaching intervention for two of the participants. These participants also demonstrated slight improvement on measures of spontaneous generalization; however, this improvement was strengthened following the inclusion of generalization programming tactics. These improvements were also demonstrated at follow-up, which provides evidence of stimulus generalization across time. The third participant demonstrated more gradual improvements along the generalization measures throughout the intervention, but did evidence a degree of stimulus generalization throughout the study and during the measure that assessed stimulus generalization across time. In conclusion, this study provided support for the use of peer teaching (i.e., multiple stimulus exemplars) in increasing stimulus generalization across peers and time. This study included elements across the different designs to allow for the examination of generalization effects that occurred spontaneously and as a result of explicit programming. Limitations of the study include lack of procedural integrity data and due to the variability that the participants demonstrated during baseline assessments. Additionally, as with any single subject design, limitations to external validity due to the characteristics of the students who participated in the study. Despite these limitations, this study included sound elements in the assessment of generalization and provided promising results for multiple exemplar training.

Somewhat similarly, in a study that assessed response generalization, Graham, Harris, and Mason (2005) examined the effects of incorporating a peer support component (i.e., incorporating common salient social stimuli) into a Self-Regulated Strategy Development (SRSD) intervention with 73 general and special education students in the third grade. This study utilized a between-subjects group design with students randomly assigned to one of three conditions: (a) SRSD instruction only, (b) SRSD plus peer support, and (c) comparison group of students who received typical writing instruction (i.e., Writer's Workshop) from their teachers. Students in both SRSD conditions received instruction geared towards composing stories and persuasive essays, with a focus on planning. The students in the SRSD plus peer support group received additional guidance from a peer that focused on providing prompts for clarifying and expanding upon ideas. Although the authors only specified the use of one tactic to promote generalization, there are several other programming tactics that are inherent to the SRSD intervention. That is, students are taught self-regulatory skills during writing by asking themselves questions to facilitate text generation (e.g., “Who are the main characters?”) and expanding upon ideas (e.g., “What do the main characters want to do?”). Both of these skills are examples of self-mediated verbal and covert stimuli generalization strategies. Additionally, sufficient stimulus and response exemplars were used during the SRSD intervention given that students receive practice with different types of writing prompts, as well as instruction and practice with different writing strategies (e.g., planning, revising, and genre specific writing strategies). Response generalization was assessed with two novel writing tasks: informative writing and personal narratives. Dependent variables included compositional length (i.e., number of words written), compositional quality, and the use of basic story elements necessary for the given genre.

Overall, the results indicated that students in the two SRSD conditions consistently outperformed the students assigned to the comparison condition on all dependent variables (i.e., length, story elements, and quality). In terms of response generalization effects, the results of this study indicated that both SRSD groups showed statistically significant gains relative to the comparison group with respect to length of informative essays (ES = 1.57 for SRSD and 1.58 for SRSD plus peer support), number of story elements used in personal narratives (ES = 1.28 for SRSD plus peer support), and personal narratives quality (ES = 1.08 for SRSD and 1.15 for SRSD plus peer support). However, the only statistically significant difference between the two SRSD conditions on the generalization outcomes was for the number of elements used in personal narratives, with students assigned to the SRSD plus peer support condition writing more elements in their personal narratives than students assigned to the SRSD condition. This makes sense, given that students receiving peer support received assistance with identifying areas in their writing products that needed to be clarified or expanded upon.

It is important to note that although peer support (i.e., incorporating common salient social stimuli) was explicitly used as a tactic to enhance the students' generalization in writing, other generalization programming tactics are inherent in the SRSD intervention (i.e., self-mediated verbal and covert stimuli, and multiple stimulus and response exemplars). Thus, it is not necessarily surprising that there were few differences between the two SRSD groups on measures of generalization. Rather, the study provides some indication regarding the amount of growth that may occur in writing when common social stimuli are incorporated with the other generalization tactics. This study included sound elements to assess generalization outcomes (i.e., pre- and post- generalization assessments), and provided positive implications for the inclusion of common salient social stimuli is beneficial in increasing students' ability to include more elements in their written compositions among novel writing tasks. However, the authors did not employ statistical analyses to control for baseline performance in when examining generalization outcomes.

In another study that incorporated multiple generalization programming tactics into a performance feedback intervention, Malandrino (2015) conducted a randomized controlled trial with 116 third-grade students. Students were randomly assigned to one of three conditions: (a) a practice-only condition that received weekly writing practice on a 3 min writing task; (b) a performance feedback condition that received individualized performance feedback as described in Hier and Eckert (2014); and (c) a generalization programming condition that received individualized performance feedback in addition to explicit generalization programming. As part of the generalization programming, three tactics were implemented, which included common salient physical stimuli (i.e., a large, cardboard pencil) and multiple stimulus and response exemplars (i.e., writing probes that were administered only visually and an expository writing probes). A pre-post design was used to examine generalization effects, which were assessed with a stimulus generalization measure (i.e., 3 min writing task was administered visually only) and a response generalization measure (i.e., 3 min expository writing task).

Generalization results were analyzed with two ANCOVAs. The first ANCOVA examined group differences in performance on the stimulus generalization probe while controlling for baseline performance. Results showed that there were no significant group differences on the stimulus generalization measure F (2, 88) = 0.93, p = .39, but there were statistically significant differences on the response generalization measure, showing that students in the generalization programming condition significantly outperformed students in the practice only condition F (2, 96) = 3.82, p = .03, η^2 = .07. Limitations to the assessment of generalization include that perhaps the stimulus generalization probes were an inadequate method to assess stimulus generalization, and due to experimenter error, the common salient physical stimulus was not presented until the fourth intervention session. Despite these limitations, strengths of the study included that generalization was assessed at pre- and post-assessment in the absence of intervention elements, and statistical analyses were used to examine group differences on the outcome measures while controlling for baseline performance. Additionally, multiple intervention assessments were included to assess both stimulus and response generalization outcomes. This study provided support for the use of multiple generalization programming tactics (i.e., common salient physical stimuli, multiple stimulus and response exemplars) toward the goal of response generalization.

In a similar study, Hier and Eckert (2016) implemented generalization programming tactics (i.e., multiple stimulus and response exemplars) into a performance feedback intervention in writing with 118 third-grade students. Students were randomly assigned to one of three conditions: (a) a practice-only condition, that received practice with weekly WE-CBM probes, (b) a performance feedback intervention, that received weekly WE-CBM probes in addition to feedback regarding total number of words written in addition to feedback regarding how that number related to previous performance, and (c) a multiple exemplar training condition, that was identical to the performance feedback condition, however, they received diverse writing probes that was alternated on a weekly basis. The multiple stimulus exemplar probes were identical to the WE-CBM probes, however, the story-stems were administered only orally, rather than orally and visually. The multiple response exemplar probes required students to write in response to a prompt that required them to compare and contrast two items or ideas. All writing probes used during the intervention allotted students one minute to plan and three minutes to write.

Generalization outcomes were assessed with a stimulus generalization probe, a response generalization probe, and a maintenance probe (i.e., a probe that assessed stimulus generalization across time). Students were not provided performance feedback when generalization was assessed. Additionally, when response generalization was assessed, students were provided with 10 minutes to write their compare and contrast essays. Maintenance (i.e., stimulus generalization across time) was assessed four months following the final intervention session. Conversely to the study hypotheses, students assigned to the multiple exemplar training condition did not outperform the other conditions on post-assessment measures of stimulus or response generalization. However, these students did outperform the other conditions on a measure that examined maintenance (i.e., stimulus generalization across time). As such, this study demonstrated some support for multiple stimulus and response exemplars as tactics to improve students' generalization across time.

Summary of Generalization Programming in Writing

In the content area of writing, only a small number of studies used explicit generalization programming tactics to improve students' writing skills. Of the studies that did not incorporate explicit generalization programming tactics, three reported evidence of stimulus generalization (Hier & Eckert, 2014; Medcalf et al., 2004; Pratt-Struthers et al., 1983), and two reported weak evidence of stimulus generalization (Schunk & Swartz, 1993; Van Houten, 1979). It is important to note that although these studies did not identify having included explicit generalization programming tactics, two of them included tactics that were similar to common salient social stimuli (Medcalf et al., 2004) and multiple stimulus and response exemplars (Schunk & Swartz, 1993).

Of the studies that did include explicit generalization programming tactics (Campbell et al., 1991; Graham et al., 2005; Jackson, 1995; Hier & Eckert, 2016; Malandrino, 2015), all reported at least partial evidence of generalization outcomes. As with reading studies that explicitly targeted generalization, tactics that received the most support for increasing stimulus generalization included multiple stimulus exemplars (Campbell et al., 1991; Hier & Eckert, 2016). For response generalization, beneficial tactics included common salient social (Graham et al., 2005) and common salient physical stimuli (Malandrino, 2015) in addition to multiple stimulus and response exemplars (Graham et al., 2005; Malandrino, 2015), and self-mediated physical stimuli (Jackson, 1995). Similar to several studies reviewed in the area of reading, there were a few studies that examined spontaneous generalization (i.e., positive generalization performance occurring in the absence of explicit programming) in addition to programmed generalization. Campbell et al. (1991) found that generalization performance was improved following explicit programming. Additionally, two studies provided several generalization assessments to examine both stimulus and response generalization outcomes (Hier & Eckert, 2016; Malandrino, 2015).

Among studies that examined generalization in the area of writing, many studies indicated that generalization was more likely to occur for fluent skills (Hier & Eckert, 2014; Jackson, 1995; Van Houten, 1979). This was true among studies that did or did not include explicit generalization programming tactics. As such, this supports the Instructional Hierarchy, which states that fluency is typically developed prior to generalization (Haring & Eaton, 1978).

Similar to the content areas of mathematics and reading, there were several studies in which it was difficult to draw firm conclusions regarding the findings. Some of the issues that were common to writing studies included the lack of clarity in information related to statistical analyses (Medcalf et al., 2004; Schunk & Swartz, 1993). Additionally, many of the single subject subjects only presented results in terms of visual analyses. As a result, alternative approaches to examine the magnitude of treatment effects (e.g., effect size, percent of non-overlapping data points) were not reported.

Purpose of the Current Study

Generalization programming is an essential component of intervention studies, especially those pertaining to academics, where students are often expected to demonstrate their learning through versatile application. However, limited academic intervention studies have incorporated explicit generalization programming tactics. An even smaller number of studies have implemented generalization programming tactics within the context of class-wide interventions while assessing generalization outcomes using appropriate design elements (e.g., pre- and post-assessments).

Of the studies that did incorporate explicit generalization programming tactics, a few tactics that were found beneficial included incorporating multiple stimulus exemplars (Ardoin et al., 2007; Hier & Eckert, 2016; Malandrino, 2015; Silber & Martens, 2010), multiple response exemplars (Duhon et al., 2010; Hier & Eckert, 2016; Malandrino, 2015), and common salient physical stimuli (Malandrino, 2015; Mesmer et al., 2010). Some researchers have also included important elements such as assessing different forms of generalization (Campbell et al., 1991; Codding et al., 2010; Codding et al., 2007; Hier & Eckert, 2016; Malandrino, 2015), including multiple assessments of generalization (Codding et al., 2010; Hier & Eckert, 2016; Malandrino, 2015) and have thus provided information related to outcomes among graded difficulty levels (Ardoin et al., 2007; Mesmer et al., 2010).

The purpose of the current study was to incorporate explicit generalization programming tactics into a performance feedback intervention targeting third-grade students' writing fluency in an effort to improve their performance across generalization outcome measures of graded difficulty levels. To address this purpose, the primary research question examined whether students who received explicit generalization programming tactics (i.e., common salient physical stimuli, multiple stimulus and response exemplars) in addition to the performance feedback intervention demonstrated greater evidence of generalization in comparison to students who did not receive explicit generalization programming tactics. Due to previous research in other academic domains that suggests incorporating the generalization tactic of providing students with multiple response exemplars (Duhon et al., 2010; Hier & Eckert, 2016; Malandrino, 2015) and common salient physical stimuli (Malandrino, 2015; Mesmer et al., 2010) increases students' performance on generalization outcomes, as well as previous research in the content area of writing (Hier & Eckert, 2016; Malandrino, 2015), it was hypothesized that students who received explicit generalization programming tactics would outperform the performance feedback only condition on measures of generalization that are of graded difficulty levels in terms of number of correct writing sequences.

Method

Participants and Setting

Instructional Review Board approval was received from Syracuse University and the participating school district. Upon approval, third-grade students receiving general education programming from a local school were invited to participate in the study. First, participating students were required to have their parents sign and return permission forms, indicating consent for their child to participate in the study. Then students were required to sign an assent form, formalizing their agreement to participate in the study.

A total of 108 students were screened for eligibility. A subset of these students (n = 25) were assigned to an alternative intervention session, which was not included in the current study. The remaining students were deemed ineligible if they met any of the following criteria: (a) teachers reported the student was eligible for special education programming or limited English proficiency / English language learner status that would negatively impact their ability to participate in the project (n = 9), (b) scoring < 1^{st} percentile or less on AIMSweb Written Expression Measure for total words written Winter benchmark (i.e., less than 8 words) (n = 4), or (c) standard score < 50 on the WIAT Alphabet Writing Fluency subtest (i.e., 0 letters written) (n

= 1), or (d) scored below criterion for both the AIMSweb Written Expression Measure and the WIAT Alphabet Writing Fluency subtest (*n* = 1).

A total of 52 students met the eligibility requirements and participated in the study (see Table 1). Most of the students were female (51.9%) and their race was identified as White (57.7%) or Black or African American (26.9%). A smaller percentage of students were identified as being of two or more races (11.5%), American Indian or Alaska Native (1.9%), or Asian (1.9%). In terms of ethnicity, most students were not Hispanic or Latino (88.5%). The average age of the students was 8 years (range, 8 years, 0 months to 9 years, 1 month). A small percentage of students (3.8%) were eligible for special education services due to speech and language impairment. An additional 3.8% were eligible for special education services due to learning disability classifications, but still met inclusionary criteria.

Eligible students were randomly assigned to one of two conditions: (a) a performance feedback condition (*n* = 25) or (b) a performance feedback with generalization programming condition (*n* = 27). There were no statistically significant differences between the two conditions with regard to sex, x^2 $(1, N = 52) = .07, p = .79$, ethnicity, x^2 $(4, N = 52) = .00, p = 1.00$ special education status, x^2 $(2, N = 52) = .01, p = 0.99$, or age, $F(1, 50) = 0.59$, $p = .45$ (see Table 1). There were also no statistically significant differences between conditions with regard to race, when examining students identified as White and non-White (x^2 $(2, N = 52) = 2.09, p = 0.15$).

The students recruited for this study were enrolled in one elementary school located in a moderately sized city in the northeast. According to 2014-2015 enrollment data, the school population was 938 students, 50.4% were male. Most students were identified as White (43%) or Black or African American (33.8%). A smaller percentage of students were identified as

Hispanic or Latino (8.2%), Asian or Native Hawaiian or Pacific Islander (7.9%), Multiracial (5.7%), or American Indian or Alaska Native (1.5%). Additionally, over half (68%) of the student population received services from economic assistance programs (e.g., free or reduced-price lunch, social security insurance, food stamps). The schools were selected due to their proximity to the university, and represented a sample of convenience.

Experimenters

Doctoral students in school psychology served as primary experimenters with the assistance of advanced undergraduate research assistants. Research assistants were required to complete formal training in research ethics, as required by Syracuse University. Training consisted of completing the Social and Behavioral Focus and Responsible Conduct of Research courses through the Collaborative Institute Training Initiative (CITI) regarding the protection of human research subjects. Documentation of successful training completion was submitted to the Institutional Review Board. Research assistants also received training in the administration and scoring of dependent measures in addition to data entry and procedural integrity assessments. All research assistants were required to demonstrate 100% proficiency scoring dependent measures and conducting procedural checks prior to participation.

Materials

Several assessments were used over the course of the study to assess eligibility, pre-assessment performance, response to intervention, and to measure generalization outcomes. The measures include assessments of immediate treatment effects (i.e., WE-CBMs self-referenced narrative probes), and four generalization measures of graded, increasing difficulty levels (i.e., WE-CBMs others-referenced narrative probes, WE-CBM expository probes, and WE-CBM compare and contrast writing probes).

Wechsler Individual Achievement Test – Third Edition. The Wechsler Individual Achievement Test – Third Edition (WIAT – III; Pearson, 2009) is a standardized, norm-referenced writing measure that is used to measure the academic skills of children aged 4 to 19 years. For the purposes of this study, only the Alphabet Fluency subtest of the WIAT-III was used to determine students' eligibility to participate in the study. The Alphabet Fluency subtest requires students to write as many letters of the alphabet as they can over the span of a 30 sec interval.

This subtest has been shown to have moderate test-retest reliability ($r = .69$) among children eight to nine years of age, with a test-retest interval that averaged 13 days and ranged from 2 to 32 days (Pearson, 2009). It is also moderately correlated ($r = .41$) with the Written Language composite of the WIAT-II.

Curriculum-Based Measurement in Written Expression Self-Referenced Narrative Probes. Curriculum-Based Measurement in Written Expression (WE-CBM) self-referenced narrative probes were used to assess eligibility and were administered during the intervention (see Table 2). For these writing probes, students were provided with a two-page packet. At the top of the first page of the packet a self-referenced story stem was presented (e.g., "I was walking down the street, when all of a sudden…"). The rest of the page contained horizontal lines for the students to write their story. No other instructional material was presented to the students. McMaster, Wayman, Deno, Espin, & Yeo (2010) has examined the technical adequacy of these probes and reported strong alternate form reliability ($r = .73$ to $.90$) and low to moderate criterion-related validity ($r = .29$ to $.63$).

WE-CBM Others-Referenced Narrative Probes. WE-CBM narrative probes that were others-referenced were used to assess generalization (see Table 2). In addition, these probes were used as a specific generalization tactic for participants assigned to the performance feedback with generalization programming condition (see Procedures section). That is, students in the performance feedback with generalization programming condition received performance feedback in response to their performance on these and other writing probes over the course of the intervention. For these writing probes, students were provided with a two-page packet. At the top of the first page of the packet an others-referenced story stem was presented (e.g., "One night the boy found the most interesting thing and…"). The rest of the page contained horizontal lines for the students to write their story. No other instructional material was presented to the students.

These probes were modified from work of McMaster et al. (2010) to be others-referenced. Other than the changes to the pronoun in the story stem, the material is identical to the narrative WE-CBM probes that are self-referenced. These probes were conceptualized to be slightly more difficult than the self-referenced probe, because they required students to slightly alter their response. That is, students were required to write from another's perspective rather than their own. Because these probes were modified for the purposes of this study, there is currently no psychometric evidence available.

WE-CBM Expository Probes. WE-CBM expository probes were used to assess generalization (see Table 2). In addition, these probes were used as a specific generalization tactic for participants assigned to the performance feedback with generalization programming condition (see Procedures section). That is, students in the performance feedback with generalization programming condition received performance feedback in response to their performance on these and other writing probes over the course of the intervention.

For these writing probes, students were provided with a two-page packet. At the top of the first page of the packet a position statement was provided for the student to generate a written composition (e.g., "Write about your favorite time of year and why you like it."). The rest of the page contained horizontal lines for the students to write their story. No other instructional material was presented to the students.

McMaster et al. (2010) reported that these probes have strong alternate form reliability ($r = .75$ to .85) and low to moderate criterion-related validity ($r = .38$ to .64). These probes are conceptualized to be of a moderate difficulty level due to the increased demands of associated with created an expository composition rather than a narrative composition. That is, students were required to state their stance on a specific issue and explain their position. Additionally, these types of prompts are described in the Common Core Standards as one writing production genre that should be strengthened in the third-grade.

WE-CBM Compare and Contrast Probes. WE-CBM compare and contrast probes were used to assess generalization (see Table 2). In addition, these probes were used as a specific generalization tactic for participants assigned to the generalization programming condition (see Procedures section). For these writing probes, students were provided with a two-page packet. At the top of the first page of the packet a directional statement was presented that required students to compare and contrast two concepts (e.g., "Write about how gym is the same as and different from art class."). The rest of the page contains horizontal lines for the students to write their story. No other instructional material was presented to the students.

These probes were developed by the author and based on curricular expectations for third-grade students as evidenced by the New York State Common Core standards and grade-level curriculum (i.e., Treasures; Macmillan McGraw-Hill, 2006). These probes are conceptualized to be of highest difficulty level due to the increased task demands of describing similarities and differences between two items or concepts. Additionally, these writing probes are similar to what is listed in the Common Core standards as a type of writing genre that should be developed in the fifth-grade. However, the prompts used for this study were simplified, given that students were not be required to write a compare and contrast composition in response to literary or informative text. Because these probes were developed by the author, there is currently no psychometric evidence.

Kids Intervention Profile. The Kids Intervention Profile (KIP; Eckert, Hier, Malandrino, & Hamsho, in press) was administered to students receiving performance feedback to assess their perceptions of the intervention. The KIP contains 8 items that are rated on a 5-point Likert-scale, with items ranging from “not at all” to “very, very much”. Eckert et al. demonstrated that this measure has adequate internal consistency ($r = .76$) and test-retest reliability ($r = .69$) over the span of 3 weeks. Items load onto two factors, labeled “General Intervention Acceptability” (e.g., “How much do you like writing stories with us each week?”) and “Intervention Skill Improvement” (e.g., “Do you think your writing has improved?”). For the present study, Cronbach’s alpha was .79.

Behavior Intervention Rating Scale. Teachers were asked to complete a rating scale regarding their perception of acceptability and effectiveness of the intervention (BIRS; Elliott & Treuting, 1991). This scale contains 24 items that are rated on a 6-point Likert-scale with items ranging from “strongly disagree” to “strongly agree”, which load onto three factors: (a) acceptability, (b) effectiveness, and (c) time of effectiveness. For the purposes of this study, the scale was modified so that questions related to ‘problem behavior’ were reworded to reference difficulties in the area of writing. Because modifications were made to the scale, the internal consistency of the scale was examined and Cronbach's alpha was .98.

Procedures

The study was conducted in four phases that included sessions conducted in a group format in two general education classrooms, within one local elementary school, during regularly scheduled class time. Although the initial plan was to conduct weekly sessions, several disruptions occurred due to unanticipated rescheduling (see Figure 1). The five phases included: (a) eligibility; (b) pre-assessment; (c) intervention; (d) generalization programming, which was only instated for the generalization programming condition; and (e) post-assessment. These phases were implemented sequentially; the eligibility phase spanned over approximately two sessions, followed by the pre-assessment phase, which spanned over the course of four sessions. Following eligibility and pre-assessment, students were randomly assigned to one of two conditions: (a) performance feedback (n = 25) or (b) performance feedback with generalization programming (n = 27). The intervention phase spanned over the course of approximately eight sessions. Following the final intervention session, the post-assessment phase was conducted during four sessions.

Manipulation checks were conducted after every session to assess the writing performance of the students assigned to the performance feedback with generalization condition in comparison to their performance on the corresponding pre-assessment probe. These checks were used to determine whether additional generalization programming sessions needed to be provided in order to improve the students' writing performance before switching to the next generalization programming tactic. For the purposes of this study, the generalization programming tactic needed to be in effect for at least three sessions and students needed to demonstrate an average 25% increase in their writing performance, relative to their baseline performance on the respective measure. This criterion was met for both the WE-CBM self-referenced narrative probes and the WE-CBM others-referenced narrative probes. However, due to unexpected scheduling issues, it was not possible to implement the final generalization programming tactic (i.e., performance feedback in response to the WE-CBM expository probe) for more than two sessions (see Figure 1), although the students exceeded the criterion during the two sessions. Specifically, students' average increase in writing performance relative to their baseline performance was 18% during the seventh intervention session and 61% during the eighth intervention session, resulting in an overall average increase of 39.5% across the two sessions.

Eligibility determination. During the first phase of the study, students were administered a WE-CBM (self-referenced) narrative probe, as well as the Alphabet Writing Fluency subtest of the WIAT. Students were considered ineligible if they obtained a standard score below 50 on the WIAT Alphabet Writing Fluency subtest, or if they wrote less than eight words on the WE-CBM (self-referenced) narrative probe.

Pre-assessment. During the pre-assessment phase of the study, students completed a CBM-WE self-referenced narrative probe as well as generalization measures of graded difficulty levels, which were administered during separate sessions. Generalization measures included a WE-CBM others-referenced narrative probe (conceptualized to be of low difficulty level), a WE-CBM expository writing probe (conceptualized to be of moderate difficulty level), and a WE-CBM compare and contrast probe (conceptualized to be of high difficulty level). Across all WE-CBM probes, the same administration procedures were followed. Specifically, students were informed, "I am going to read a sentence to you first, and then I want you to write a story about what happens next. You will have some time to think about the story you will write and then you will have some time to write it. Please turn to the next page of your packet. For the next minute think about writing a story that begins with this sentence – [insert story stem]. Remember, take time to plan your story. A well-written story usually has a beginning, a middle, and end. It also has characters that have names and perform certain actions. Use paragraphs to help organize your story. Correct punctuation and capitalization will make your story easier to read. Please do not write the story yet. Just think of a story that begins with this sentence – [insert story starter]. Ready? Begin thinking and I'll tell you when a minute for thinking is up." Students were given 1 min to plan their story and then the students were directed, "When I tell you to start, please begin writing your story. Remember, if you don't know how to spell a word, you should try your best and sound it out. It is important that you do your best work. If you fill up the first page, please turn to the next page and keep writing. Do not stop writing until I tell you to. Do your best work. Okay, you can start writing." Students were given 3 min to write their stories.

Performance feedback condition. Students assigned to the performance feedback condition received individual writing packets, the first page was printed on colored paper that signified their respective group. It also included their identifying information (i.e., name, homeroom teacher). The second page presented a stop sign to prevent them from seeing their feedback prematurely. The third page included individualized performance feedback, which consisted of a box in the center of the page with a number inside, denoting how many words they wrote during the previous session. Beside the box was an upward or downward facing arrow or an equal sign to denote how the number in the box compared to the number of words that they wrote two sessions previous. The primary experimenter followed a procedural script to explain performance feedback to the students and to administer the WE-CBM probe. Specifically, the research assistant stated to the students, "The box in the middle of the page [The research assistant should point to the box] tells you how many words you wrote last week. Next to the box you will see an arrow. If the arrow is pointing up towards the sky, then that means you wrote more words since the last time I worked with you. If the arrow is pointing down towards the floor, then that means you wrote fewer words since the last time I worked with you. If you have an equal sign instead of an arrow, then that means you wrote the same number of words as you did the last time I worked with you. Every week when we work with you, we are going to tell you how you are doing with your writing." Following this review, the respective WE-CBM probe was administered.

Performance feedback plus generalization programming condition. Students assigned to the performance feedback plus generalization programming condition received explicit generalization programming tactics in addition to individualized performance feedback, as described in the previous section. First, students received the customary performance feedback in response to their performance on WE-CBM self-referenced narrative probes during the first three intervention sessions in addition to introducing one generalization programming tactic. That is, a large (approximately 42" in height) cardboard pencil was placed at the front of the classroom and was explicitly referred to during administration of the writing probes. This served as a common salient physical stimulus that was referenced during each session within the context of a prompt designed to explicitly state the importance of increasing the students' writing output, as well as providing the students with directions regarding what to write about. Specifically, the research assistant would state, "This pencil will be here to remind you to WRITE MORE. It is also here to help you to think of the Triple I's, which include "I think, I felt, I did". As you are writing your story, include information about what you think, what you felt, and what you did...".

During sessions four through six, a second generalization programming tactic was introduced. The writing task was changed so that WE-CBM others-referenced narrative probes were administered to program for stimulus generalization. Students continued to receive the performance feedback intervention as well as were exposed to the common salient physical stimuli. During these sessions, the instructions pertaining to the pencil changed slightly to reflect the difference in the story stem. Specifically, the researcher stated, "This pencil will be here to remind you to WRITE MORE. During our last session, we spoke about the Triple I's when you were writing your story, but today we are writing through the perspective of a young boy / girl. So when you write your story today, think of the Tripe S/He's. The Triple S/He's include "He thought, he felt, he did." That means as you are writing your story, include information about what s/he thought, what s/he felt, and what s/he did."

During the final two sessions of the study, the writing task was changed again so that the WE-CBM expository probe was administered to program response generalization. Students continued to receive the performance feedback intervention and were exposed to the common salient physical stimuli (see Figure 1). During these sessions, the instructions pertaining to the pencil changed again to reflect the difference in the story stem. Specifically, the researcher would state, "This pencil will be here to remind you to WRITE MORE. In previous sessions, the pencil was here to remind you of the "Triple I's" or the "Triple She's or He's". Today we are writing a composition which requires you to state and opinion and include information to support your opinion. So when you write your story today, the pencil is here to remind you of your two main goals, which are to 1) state your opinion, and 2) support your opinion with reasons. To assist with your writing, you should include words such as “because”, “therefore”, “since”, and “for example”.

Post-assessment. The post-assessment of generalization measures of graded difficulty levels was administered at the conclusion of the study. One WE-CBM self-referenced narrative probe, one WE-CBM others-referenced narrative probe, one WE-CBM expository probe, and one WE-CBM compare and contrast writing probe were administered on separate occasions. Standard administration procedures for WE-CBM probes were followed. For the generalization programming condition, the common salient physical stimuli (i.e., the large cardboard pencil) was present. In addition, the Kids Intervention Profile (KIP; Eckert et al., in press) was administered to all of the students to assess their perceptions of the intervention and all of the teachers were asked to complete a modified version of the Behavior Intervention Rating Scale (BIRS; Elliot & Treuting, 1991) to assess their perceived acceptability and effectiveness of the intervention.

Dependent Measures

Primary measures. Correct writing sequences were calculated for all WE-CBM measures administered over the course of the study, as a measure of writing fluency and quality. Procedures developed by Shapiro (2004) were used for scoring correct writing sequences. That is, each adjacent word in students’ writing products was examined for accuracy based on spelling, capitalization, punctuation, and syntax. Results of several comprehensive reviews were provided in the AIMSweb manual (Powell-Smith & Shinn, 2004) and in a meta-analysis by McMaster and Espin (2007). Reliability coefficients were moderate to high (r = .46 to .86). Studies examining validity of the measure indicated that correct writing sequences were moderately to strongly correlated with criterion measures (e.g., holistic ratings and with regard to the Test of Written Language (Hammil & Larsen, 1996).

Experimental Design

A randomized controlled trial was used to examine whether students who received explicit generalization programming tactics (i.e., common salient physical stimuli, multiple stimulus and response exemplars) in addition to the performance feedback intervention demonstrated greater evidence of generalization in comparison to students who did not receive explicit generalization programming tactics on generalization measures of graded difficulty levels (i.e., the WE-CBM self- and others-referenced narrative probes, WE-CBM expository probes, and WE-CBM compare and contrast probes). Previous randomized controlled trials (RCTs) (Hier & Eckert, 2014; Malandrino, 2015) included a practice-only control group, which received only weekly writing practice without individualized performance feedback. Because previous studies have shown that students in the performance feedback condition outperformed those who have received writing practice alone ($d = 0.89$) (Hier & Eckert, 2014), the current study did not include a no-treatment control condition.

Prior to the commencement of the study, 108 students were assessed for eligibility (see Figure 2). A total of 55 students were excluded due to the following reasons: (a) students were allocated to another experimental condition, not included in the described study ($n = 25$), (b) did not meet baseline criteria for total words written on the initial WE-CBM probe and / or the basal for the WIAT-III spelling measure ($n = 6$), (c) had an individualized education program under the classification of speech and language impairment, received special education programming, and were identified by their teachers as experiencing significant delays in the their language that would impede their performance in the study ($n = 3$), (d) had an individualized education program under the classification of Other Health Impairment (*n* = 2), (e) had an individualized education program under the classification of Autism Spectrum Disorder (*n* = 1), (f) students moved to another school (*n* = 4), (g) student arrived too late to participate in the study (*n* = 1), (h) students were identified as having limited English proficiency or as an English language learner (*n* = 9), or (i) parents declined to consent (*n* = 4). This resulted in a total sample size of 52 students.

A covariate adaptive randomization method was used to assign eligible students within their respective classroom to conditions based on their average performance on the four pre-assessment measures using a random number generator. That is, students in each classroom were ranked in terms of their average pre-assessment performance (i.e., number of CWS), which is an important baseline characteristic to equate across conditions. Within each classroom, students were randomly assigned in sequential order to one of two conditions: (a) performance feedback (*n* = 25), or (b) performance feedback plus generalization programming (*n* = 27). This method of randomization controls for the possible influence of students' initial writing performance while retaining equal sample sizes across conditions.

An a priori power analysis was conducted to determine an adequate sample size for testing group differences between the two conditions. Sample size was calculated by setting α equal to 0.05 and power equal to 0.80. The results indicated that 26 third-grade participants per condition should be included, resulting in a total sample size of 52 participants. Based on pilot work by Eckert et al. (2006), the sample size was calculated to detect a minimum meaningful difference in slopes of 0.60.

Procedural Integrity

Primary research assistants followed a procedural script throughout each session and tallied each step as it was completed. Secondary research assistants completed procedural integrity checks for 46% of the sessions. Specifically, procedural integrity checks were completed for 50% of the performance feedback condition sessions and 42% of performance feedback with generalization programming condition sessions. To do so, research assistants followed additional procedural scripts and tallied steps as they were completed by the primary experimenter. Procedural agreements were tallied as instances when the secondary research assistant indicated that the primary research assistant correctly implemented that portion of the procedure. Integrity was determined and reported as the lower total count of agreements divided by the total number of possible procedural steps and multiplied by 100%. Procedural integrity was exact across all sessions (M = 100%) and no deviations were reported (see Table 3).

Interscorer Agreement

A total of 37% of the WE-CBM probes (37% from the performance feedback condition, 38% from the performance feedback with generalization programming condition) were randomly selected across all writing measures and re-scored. Interscorer agreement and kappa coefficients were calculated to examine the extent to which the two scorers agreed. Instances of disagreement between scorers were re-examined by the primary experimenter to determine the final score. Interscorer agreement was calculated by dividing number of agreements by number of agreements plus disagreements times 100. Kappa was calculated by subtracting chance agreements from observed agreements and dividing that number by the product of one minus chance agreements. This product was multiplied by 100. The mean percentage of interscorer agreement was 96% (range, 40% to 100%, and the mean kappa values were .90 (range, .16 To

1.00). It should be noted that there were 17 instances in which kappa could not be calculated because there was a ‘0’ value in an ‘agreement’ cell. In these instances, the proportion of agreement value was used instead.

Results

Data Preparation

The primary researcher, along with trained research assistants, were in charge of entering data into a Microsoft Excel file. Another researcher double-checked all imputed data to ensure accuracy. Data were transferred from Microsoft Excel to SPSS 21.0 (SPSS Inc., 2012). SPSS was used to perform descriptive statistics in addition to statistical analyses.

There was no missing pre-assessment data. However, for the post-assessment sessions, there was a small number of missing data (see Figure 2). For the performance feedback condition, two post-assessment sessions were missing (i.e., one student missed two sessions). For the performance feedback with generalization programming condition, there were four post-assessment sessions missing (i.e., one student was absent for all four sessions). Cases with missing post-data were excluded from the analysis.

Descriptive Results

Students in both conditions increased their correct writing sequences from the pre- to post-assessment writing period across all measures (see Table 4). Moreover, the mean number of correct writing sequences for both conditions were slightly higher than the expected mean number of correct writing sequences based on national norms (i.e., 26 CWS) collected during the Winter assessment period (AIMSweb®, 2012) for the WE-CBM self-referenced narrative probe (*M* performance feedback condition = 28.11; *M* performance feedback with generalization programming condition = 28.17), the WE-CBM others-referenced narrative probe (*M* performance feedback condition = 29.55; M performance feedback with generalization programming condition = 29.32) and the WE-CBM compare and contrast writing probe (*M* performance feedback condition = 28.90; *M* performance feedback with generalization programming condition = 27.71). This pattern of writing performance was not observed for the WE-CBM Expository probe (*M* performance feedback condition = 24.26; *M* performance feedback with generalization programming condition = 25).

Major Analyses

To examine whether there were differences in students' responses on the post-assessment generalization measures (i.e., WE-CBM self-referenced narrative probes, WE-CBM others-referenced narrative probes, WE-CBM expository probes, and WE-CBM compare and contrast probes), four analyses of covariance (ANCOVAs) were conducted. Prior to conducting the ANCOVAs, the underlying statistical assumptions were tested, including normality, homogeneity of variance, correlations among covariates, linearity, and homogeneity of regression slopes. These assumptions were mostly upheld, aside from one instance in which the assumption of homogeneity of variances was met, which is described below. To reduce errors associated with conducting multiple comparisons, a Bonferroni correction was applied to the alpha values, such that each alpha value was increased to .012.

Generalization Outcomes. Prior to conducting the ANCOVAs, the homogeneity of regression assumption was analyzed for each of the four measures to ensure that there was not an interaction between the covariate and the conditions. Univariate analysis of variance results indicated that this assumption was met for the WE-CBM self-referenced narrative probe, F (1, 47) = 1.20, p = .28, the WE-CBM others-referenced narrative probe, F (1, 47) = 1.23, p = .27, the

WE-CBM expository probe, $F(1, 46) = 0.63$, $p = .43$, and the WE-CBM compare and contrast probe, $F(1, 46) = 0.53$, $p = .47$.

After adjusting for scores on the pre-assessment measure, there was not a statistically significant difference between the two conditions on the post-assessment CBM-WE self-referenced narrative probe, $F(1, 48) = 0.0001$, $p = .98$, the post-assessment CBM-WE others-reference narrative probe, $F(1, 48) = 0.01$, $p = .95$, the post-assessment CBM-WE expository probe, $F(1, 47) = 0.05$, $p = .83$, or the post-assessment compare and contrast probe, $F(1, 47) = 0.12$, $p = .73$. However, due to the violation of the assumption of homogeneity of variances for the compare and contrast probe, a Mann-Whitney U Test was also conducted to examine between-group differences on this measure. Similar to the results of the ANCOVA, these results were not statistically significant ($p = .95$). Students performed similarly across the two conditions in terms of mean CWS (see Tables 5 and 6).

Secondary Analyses

Intervention performance. Although the present study's primary aim was to examine post-assessment outcomes, the students' performance during the intervention was descriptively examined to ensure that students demonstrated improvements over time, with respect to their correct writing sequences. Figure 3 illustrates the students' performance during each intervention session. For students assigned to the performance feedback condition, an increasing trend in mean correct writing sequences was observed over the course of the intervention phase of the study. For students assigned to the performance feedback with generalization programming condition, increases were also observed over the course of the intervention phase of the study. However, following the introduction of each generalization tactic, which included a different writing probe, students in this condition demonstrated a slight decrease in their responding.

Student acceptability outcomes. Students' acceptability ratings on the Kids Intervention Profile (KIP; Eckert et al., 2015) showed that students in the performance feedback condition (*M* = 3.43, *SD* = 0.58) and performance feedback with generalization programming conditions (*M* = 3.25, *SD*, 0.53) rated the intervention as acceptable. There were no statistically significant differences in ratings for the two factors. Students assigned to the performance feedback intervention rated the overall acceptability of the intervention (*M* = 3.59, *SD* = 0.82) similarly to students assigned to the performance feedback with generalization programming condition (*M* = 3.78; *SD* = 0.63), *t* (47) 1.05, *p* = .30. In addition, students assigned to the performance feedback condition rated their skill improvement (performance feedback *M* = 2.93; *SD* = 0.68) similarly to students assigned to the performance feedback with generalization programming condition (*M* = 2.88, *SD* = 0.65) *t* (47) 0.26, *p* = .79.

A descriptive review of individual items on the KIP suggested that the items rated highest included, "Do you think your writing has improved?" (performance feedback condition *M* = 4.48, *SD* = 1.28; performance feedback with generalization programming condition *M* = 4.42, *SD* = 0.95), "How much do you like writing stories with us each week?" (performance feedback condition *M* = 4.43, *SD* = 0.95; performance feedback with generalization programming condition *M* = 3.88, *SD* = 1.21). The lowest rated item was, "How much do you think it helped when you were told how many words you wrote?" (performance feedback condition *M* = 1.65, *SD* = 1.11; performance feedback with generalization programming condition *M* = 2.04, *SD* = 1.37). There were no statistically significant differences in student acceptability ratings or individual items between the two conditions (see Table 6).

Teacher acceptability outcomes. Teachers' acceptability ratings on the Behavior Intervention Rating Scale (BIRS; Elliot & Treuting, 1991) suggested that the two teachers, whose classrooms were used for the intervention conditions, reported high levels of overall acceptability and effectiveness regarding the procedures used in the intervention (*M* = 4.92, *SD* = 1.12; see Table 7). These levels were high across the three factors (acceptability *M* = 5.00, *SD* = 1.13; effectiveness *M* = 4.71, *SD* = 1.01; time of effectiveness *M* = 5.00, *SD* = 1.13). Items rated highest by teachers included "this would be an acceptable intervention for students' writing difficulties" (*M* = 5.50, *SD* = 0.71), "this intervention should prove effective in changing students' writing difficulties" (*M* = 5.50, *SD* = 0.71), "students' writing difficulties are severe enough to warrant use of this intervention" (*M* = 5.50, *SD* = 0.71), "the intervention would not result in negative side-effects for the student" (*M* = 5.50, *SD* = 0.71), "this intervention would be appropriate for a variety of students" (*M* = 5.50, *SD* 0.71), and "I like the procedures used in this intervention" (*M* = 5.50, *SD* = 0.71). An item of the scale that was rated lower was "the intervention is consistent with those I have used in the classroom" (*M* = 3.00, *SD* = 0.00).

Discussion

Generalization is an important component of intervention studies (Baer et al., 1968; Stokes & Osnes, 1989), however, there are few academic intervention studies that have explicitly targeted generalization (Skinner & Daly, 2010). In addition, many of the existing academic intervention studies that have incorporated these tactics have methodological flaws that limit the conclusions that can be drawn regarding the effectiveness of the generalization tactics used. This is especially true in writing intervention research, which is disappointing, given that writing is a content area in which students are most severely underperforming (National Center for Education Statistics, 2012).

As a result, the purpose of the current study was to examine the effectiveness of implementing explicit generalization programming tactics into an academic intervention in the area of writing. Toward this goal, generalization tactics were included into a performance feedback intervention that has demonstrated effectiveness in increasing students writing fluency (Hier & Eckert, 2014). The results of this study showed that students receiving performance feedback in addition to generalization programming demonstrated improvements in their writing fluency on four post-assessment generalization measures following the performance feedback intervention. However, they did not outperform students assigned to a condition receiving performance feedback alone.

Effects of Explicitly Programming Generalization into a Performance Feedback Intervention

Given that previous researchers demonstrated successful generalization outcomes as an effect of using common salient physical stimuli (Malandrino, 2015; Mesmer et al., 2010), multiple stimulus (Ardoin et al., 2008; Campbell et al., 1991; Graham et al., 2005; Hier & Eckert, 2016; Malandrino, 2015; Silber & Martens, 2010), and response exemplars (Graham et al., 2005; Malandrino, 2015; Peterson-Brown & Burns, 2011), it was hypothesized that students receiving this package of explicit generalization programming tactics would outperform students assigned to a performance feedback condition that did not receive explicit generalization programming tactics. Similar to previous studies that examined generalization outcomes with measures of graded difficulty level (Ardoin et al., 2008; Codding et al., 2010; Codding et al., 2007; Poncy et al., 2010), this study used multiple measures to assess generalization outcomes. It was hypothesized that students receiving generalization programming tactics would outperform the performance feedback condition on these measures. This hypothesis was not supported for any of the four measures used to assess generalization. Students assigned to both conditions performed similarly across all generalization post-outcome measures assessed.

Although results of secondary analyses examining students' acceptability ratings of the intervention indicated that the intervention was rated as moderately acceptable and no statistically significant differences between the two conditions were observed, there was a pattern of results suggesting that students attributed improvements in their performance to the weekly writing practice embedded in the intervention (i.e., "How much do you think it helped to write with us each week?") versus components of the intervention (i.e., "How much do you think it helped when you were told how many words you wrote?"). Further, teachers' responses on the Behavior Intervention Rating Scale (BIRS; Elliott & Treuting, 1991) suggested that the teachers participating in this study did not provide their students with performance feedback in the classroom and did not prioritize writing fluency during their instructional practices. Although the teachers' responses are consistent with national classroom instructional practices in writing (Cutler & Graham, 2008), the students' perceptions are inconsistent with prior performance feedback studies (Malandrino, 2015).

There are a few considerations associated with the null findings. First, although there were not statistically significant differences between conditions, students in both conditions demonstrated improvements in their post-assessment performance on each of the generalization measures relative to their pre-assessment performance. Prior research demonstrated that students are more likely to generalize their skills following initial increases in fluency (Ardoin et al., 2009; Jackson, 1995; Klubnik & Ardoin, 2010; Silber & Martens, 2010; Van Houten, 1979). As such, it is possible that the lack of differences between the students' performance in the two conditions is related to intervention increases in fluency that occurred across both conditions.

For example, in the Malandrino (2015) study, students who received performance feedback and generalization programming demonstrated statistically significant differences in their generalization outcomes in comparison to students who were assigned to a practice-only condition. However, no statistically significant differences were observed among students assigned to a performance feedback condition. This suggests that the performance feedback intervention in the absence of specific generalization tactics may be sufficient for establishing generalized effects.

It is also possible that unintentional generalization programming tactics (i.e., multiple stimulus exemplars) were embedded in both conditions. For example, the story stem varied during each intervention session, which could be considered a form of multiple stimulus exemplars. Additionally, due to scheduling conflicts, the same experimenters were not always available to conduct each intervention session with their respective groups. This could be conceptualized as a generalization programming tactic, as the conditions of the intervention (i.e., the experimenters who administered the intervention) varied. This may have impacted responding, as students may have demonstrated generalized performance across experimenters. Finally, in an effort to avoid missing data, make-up sessions for students who were absent from the original pre- or post-assessment sessions were conducted outside of the classroom setting (e.g., in the school library or in a meeting room). As a result, it is possible that the process of completing intervention sessions in an alternate setting served as a multiple stimulus exemplar (across settings), and impacted student performance during the generalization assessments.

The possibility of intervention elements potentially functioning as unintentional multiple exemplars was noted in other studies specifically targeting generalization outcomes (Ardoin et al., 2007, 2008; Silber & Martens, 2010). Within these specific studies, students unintentionally received explicit generalization programming tactics, which included elements that could be regarded as potential multiple response exemplars (e.g., listening passage preview, phrase drill error correction. In these studies, group differences favoring the condition receiving explicit generalization programming tactics were either not observed at all (Ardoin et al., 2007) or potentially weakened. For example, results from the Ardoin et al. (2008) study demonstrated that a condition receiving explicit multiple exemplar training outperformed a condition not receiving the explicit package of multiple exemplar training on one generalization measure, but not another. Along those lines, the Silber and Martens (2010) study showed that while a multiple exemplar condition outperformed a control condition on a generalization measure, this group did not outperform an additional experimental condition receiving listening passage preview and repeated readings. As such, it is plausible that the current study included intervention elements that functioned as multiple exemplars for both conditions. Accordingly, the results of the current study suggest that there was no added benefit to including the additional generalization programming tactics (e.g., multiple response exemplars, common salient physical stimuli) in relation to the outcomes assessed.

It is also possible that the scheduling of generalization programming sessions may not have been sufficient for adequately achieving generalization outcomes. One way that the current study sought to control for this possibility was by conducting ongoing manipulation checks for the generalization programming condition throughout the intervention. Accordingly, manipulation checks were conducted following each intervention session for the performance feedback with generalization programming condition to compare to their performance with the corresponding pre-assessment probe. Results of these checks were used to determine whether to implement additional generalization programming sessions or to move on to the next generalization programming tactic. Prior to the commencement of the intervention, a criterion of three sessions with each tactic and an average increase of 25% in writing performance, relative to their baseline performance on the respective measure, was chosen. These criteria were met for the WE-CBM self-referenced and others-referenced narrative probe, but not for the WE-CBM expository probe. As such, it is possible that an additional session would have been beneficial in improving generalization outcomes. It is also possible that higher threshold should have been instated for the manipulation checks.

Additionally, perhaps the current study did not incorporate an adequate number of intervention sessions to program generalization outcomes. Other academic studies have implemented varying numbers of intervention sessions including generalization programming tactics. Some have conducted a single intervention session (Ardoin et al., 2008; Peterson-Brown & Burns, 2011; Silber & Martens, 2010), while others scheduled sessions within the context of single-subject designs that varied based on individual performance (Duhon et al., 2010). With regard to studies examining generalization outcomes specific to writing, Hier and Eckert (2016) conducted a total of six intervention sessions prior to finding partially successful generalization outcomes. In this study, students assigned to a multiple exemplar training condition outperformed students assigned to a condition receiving performance feedback alone and students assigned to a condition receiving only writing practice on a measure examining maintenance (i.e., stimulus generalization). The Malandrino (2015) study also found partially successful generalization outcomes following three sessions of response generalization programming. Based on the results of these studies, it is unclear the specific number of sessions that are beneficial toward the goal of generalization outcomes, however, there has been at least partial support for positive generalization outcomes following even one intervention session.

Despite these findings in prior research, it is plausible that the number of sessions of generalization programming implemented in the current study was insufficient toward the goal of generalization.

Finally, assessing students' generalization outcomes within the context of a RCT resulted in a delay between each generalization programming tactic and the assessment of the corresponding generalization outcome. All post-assessment generalization outcomes were assessed at the conclusion of the study. As a result, the latency between the final generalization programming session for the performance feedback with generalization programming condition and the generalization post-assessment with the WE-CBM self-referenced narrative probe was six weeks. For the WE-CBM others-referenced narrative probe, the latency was three weeks. For the WE-CBM expository probe, the latency was a little over one week. The two studies (Hier & Eckert, 2014; Malandrino, 2015) reporting positive generalization outcomes in elementary students' writing fluency conducted assessments immediately following the intervention sessions. Thus, the latency between generalization programming and assessment likely decreased the sensitivity of detecting generalization outcomes.

Generalization Programming: Developmental and Instructional Considerations

Within the existing literature base for generalization programming in academic intervention studies, several studies have targeted generalization outcomes in elementary-aged students (e.g., Graham et al., 2005; Hier & Eckert, 2014; Malandrino, 2015; Noell et al., 2006; Van Houten, 1979). Most of the studies demonstrated at least partial support for generalization outcomes (Graham et al., 2005; Hier & Eckert, 2014; Malandrino, 2015; Noell et al., 2006), beginning in the first grade. In the content area of writing, two studies (Hier & Eckert, 2016; Malandrino, 2015) have specifically targeted generalization of writing fluency with third-grade students following explicit generalization programming tactics in a performance feedback intervention and reported evidence for generalization. These findings suggest that it is developmentally appropriate to incorporate generalization programming tactics in academic interventions, including academic interventions that focus on writing performance, for children in the third grade.

Although developmental levels have rarely been stipulated in the literature with regard to appropriate timing and effectiveness of generalization programming, researchers have discussed generalization in the context of skill development. For example, Haring and Eaton (1978) identified generalization within the Instructional Hierarchy. Within this framework, generalization of skills occurs following the development of skill fluency and the authors postulated that generalization programming should be targeted in conjunction with fluency.

Applying the Instructional Hierarchy framework to the present study, it is possible that the generalization programming may have been negatively impacted by the students' instructional learning levels. Because this study was implemented in a class-wide, group setting, generalization programming tactics could not be tailored depending on students' individual levels of functioning within the Instructional Hierarchy. For example, of the 85.2% students in the generalization programming condition who were performing at a frustrational level (based on the total words written) during the pre-assessment, 50% continued to demonstrate frustrational level performance at post-assessment. As a result, a large number of students may not have foundational levels of fluency in order to generalize this skill to alternative writing tasks. This is supported by previous studies showing that students are more likely to generalize skills that have been developed to fluency (Hier & Eckert, 2014; Jackson, 1995; Van Houten, 1979), and are less likely to generalize skills that are not developed to fluency (Jackson, 1995; Van Houten, 1979).

As such, it may be that generalization programming tactics are more effective if matched to students' instructional learning needs.

Advances in Programming and Assessing Generalization

Although recommendations have long been in place toward the explicit pursuit of generalization outcomes (Baer et al., 1968), the field of generalization programming has been slow to advance in the area of academic intervention research (Poncy et al., 2010). As a result, there is not a strong literature base with regard to tactics that have been consistently shown to lead to generalization outcomes. In addition to the paucity of research studies, substantial weaknesses within many of the existing studies were identified. Of these weaknesses, some of the more basic limitations included failing to incorporate explicit tactics to program generalization, or properly assessed generalization outcomes. Other more nuanced limitations included incorporating explicit strategies to promote generalization without identifying them as such, or including intervention elements in the generalization assessment, thus conflating the results.

An examination of prior studies that specifically incorporated specific tactics suggests a variety of approaches were applied. Many studies incorporated packages of intervention tactics, that is, they have used more than one generalization programming tactic toward the aim of increased generalized outcomes (i.e., Duhon et al., 2010; Hier & Eckert, 2016; Malandrino, 2015; Roca & Gross, 1996). There have also been studies that have demonstrated successful generalization outcomes on the basis of incorporating individual generalization programming tactics. The tactics that have received the most support include incorporating multiple stimulus and response exemplars. Despite that several studies have demonstrated positive outcomes for programming tactics, often, generalization results have been described as partially effective, and in one case, not effective (Ardoin et al., 2007).

The present study attempted to address prior limitations observed in academic intervention generalization studies in an attempt to advance programming and assessing students' generalization. Specifically, the current study included explicit generalization programming tactics, which have received support in previous studies targeting reading (Ardoin et al., 2007; Duhon et al., 2010; Mesmer et al., 2010; Silber & Martens, 2010) as well as previous studies in the content area of writing (Hier & Eckert, 2016; Malandrino, 2015). Additionally, generalization assessment was incorporated into the design of the intervention, such that generalization results were assessed both prior to and following implementation of the intervention. This differs from prior academic intervention generalization studies in the area of reading that have only assessed generalization outcomes, and did not include pre-assessment generalization measures (Ardoin et al., 2008; 2009; Peterson-Brown & Burns, 2011). To further strengthen the current study, generalization assessments of graded difficulty levels were also included, similar to studies in the area of reading that have included multiple assessments of generalization (Ardoin et al., 2008; Mesmer et al., 2010).

Limitations

There are several limitations of the current study. First, given that the study took place in a school setting, where schedule conflicts were unavoidable, there were several instances in which the intervention sessions could not be administered during consecutive weeks. A second threat to the internal validity is the effect of testing (i.e., the effect of pre-testing on post-assessment outcomes). In the current study, participants received pre-assessments that were similar in nature to the assessments administered at post-assessment. As such, having received similar assessment probes may have impacted their performance on the post-assessments. It is possible that the effect of testing may have served as a generalization programming tactic for both conditions, making it difficult to distinguish between outcomes related to explicit generalization programming. Further, compensatory rivalry is also a potential threat to the internal validity of the study. Given that students came from the same school, and students from each of the classrooms were assigned to both conditions, it is possible that students between conditions discussed the strategies employed in the generalization programming conditions, thereby impacting their performance on the generalization assessments.

There were two threats to the external validity of the study. First, it is possible that there were reactive effects of the assessment, such that exposure to the pre-assessments interacted with the participant's responsiveness to the intervention. Given that several measures were administered during the eligibility and pre-assessment phases, it is unclear what the results of the study would have been in the absence of such testing. A second threat is that although the results of this study are meant to inform intervention practices for general education students, this study population was limited to third-grade students in an urban school district, which included a large percentage of general education students who qualified for free or reduced-price lunch. As such, the results may not be indicative of how other populations would respond to the intervention elements.

Lastly, given that Cuvo (2003) recommended that generalization should be tested under extinction conditions, the generalization assessments that were conducted in the present study did not meet this assertion. For students assigned to the performance feedback with generalization programming condition, a prompt was included that referenced the common salient physical stimuli (i.e., the cardboard pencil) during each generalization assessment, as well as explicit directions with regard to reminding students to increase their writing productivity. However, the nature of this strategy necessitates that the stimuli remain common between intervention and non-intervention sessions. As a result, this tactic was included during the generalization assessments and the assessment of generalization was not conducted in total absence of intervention elements.

Implications for Practice

Generalization is an important outcome of academic interventions that are implemented in the classroom setting, as students are typically expected to learn skills and apply them across relevant settings and situations. Although the existing literature base is broad with respect to academic content area, student grade level and setting (e.g., individual or group settings), a number of design features associated with the existing research base limits the generalization of findings and implications for practice. The current study, which was based on prior findings reported by Malandrino (2015), implemented a package of intervention elements (i.e., multiple stimulus and response exemplars, common salient physical stimuli) into a class-wide performance feedback intervention. Results of the current study showed that although there were not differences in students' performance between the two conditions, students in both conditions improved their post-assessment performance across measures of generalization.

Although these outcomes do not provide clear implications for teachers with respect to explicit tactics that should be incorporated into instruction to promote generalization, support from existing literature provides some guidelines for practice. First, generalization should be considered a standard component of instructional practice. That is, generalization should be considered with respect to instructional strategies, and should be measured before and after the commencement of instruction or intervention. The most heavily researched strategies (i.e., multiple stimulus and response exemplars) are often inherently included in instructional practices. However, given the importance of generalization outcomes, it is important that these strategies be utilized in an intentional manner.

Educators should also explicitly target generalization outcomes of instruction, and assess students to ensure that generalization outcomes are met. As such, educators should go beyond a "train and hope" scenario in which they implement instruction and assume that generalization outcomes will occur. Rather, they should incorporate diversified training (e.g., multiple exemplar training) into instructional practices. Assessment of generalization should occur in the absence of instructional assistance (Cuvo, 2003) and should be examined with multiple assessments.

Although generalization has been successfully targeted with students beginning in the elementary grades, consideration should be taken with regard to students' performance levels (i.e., the Instructional Hierarchy) when implementing instructional practices. That is, instructional tactics should vary when students are performing at an acquisition level as opposed to when students are preforming at a fluency level. For example, Haring and Eaton (1978) proposed that when students are performing at an acquisition level, instructional strategies should include modeling and corrective feedback. As such, it is important that students' skill level be assessed prior to the implementation of generalization programming tactics.

Directions for Future Research

The state of generalization programming is underdeveloped in intervention research (Skinner & Daly, 2010), especially in the area of writing where students demonstrate severe underperformance. Studies that have examined generalization outcomes as a result of explicit programming in the area of writing have demonstrated mixed results (Hier & Eckert, 2016;

Malandrino, 2015). Given the importance of improving students' writing skills, it is important that research in this area continue. Results of the current study showed that although students receiving performance feedback demonstrated improvements in writing fluency and accuracy over the course of eight weekly writing sessions, the addition of generalization programming tactics did not result in greater generalized outcomes in comparison to performance feedback alone.

Similar to the Malandrino (2015) study, results showed that students assigned to the performance feedback intervention, which did not receive explicit generalization programming tactics, demonstrated improvement in post-assessment generalization measures relative to pre-assessment scores, similar to the condition that received explicit programming. When considering how these results align with the Instructional Hierarchy (Haring & Eaton, 1978), it is possible that due to an increase in writing fluency, these students were inherently more likely to demonstrate generalization and adaptation of their writing skills. As such, future studies should further examine the notion of a "functional fluency" level, and how this relates to generalization outcomes.

In the current study, generalization outcomes were assessed at the conclusion of the intervention phase. As a result, there was a delay in administration between when the generalization programming probes and the corresponding post-assessment probes. Given this delay, it is possible that there were immediate generalization effects that went undetected. Future studies should consider examining these outcomes more proximal to their programming.

Conclusions

It is important that researchers examine generalization outcomes of intervention studies (Stokes & Osnes, 1989; Skinner & Daly, 2010). This is particularly true in the area of academic intervention research. Given that writing is a particular area of weakness, this is a prime area to target. Although there have been several studies establishing the effectiveness of a performance feedback intervention to increase students' writing fluency, there have been mixed results regarding the effectiveness of these interventions with regard to generalization outcomes (Hier & Eckert, 2014; 2016; Malandrino, 2015). The current study aimed to improve the existing research base by including explicit generalization programming tactics (i.e., multiple stimulus and response exemplars and common salient physical stimuli) into a performance feedback intervention in writing intervention. Generalization was included in the study design with pre- and post-assessment phases to ensure that results were not impacted by pre-existing differences in students' performance. Additionally, results were assessed with multiple generalization assessments that were conceptualized to be of graded difficulty level. Results of the study showed that students assigned to a performance feedback with generalization programming condition did not outperform students assigned to a performance feedback condition. However, students in both conditions demonstrated improvements in their post-assessment writing performance. Future studies should continue to examine the effectiveness of generalization programming.

Table 1

Student Demographic Information (N = 52)

	Total Sample		Condition: Performance Feedback		Condition: Performance Feedback with Generalization Programming			
Characteristics	%	(*n*)	%	(*n*)	%	(*n*)	X^2	*p*
Sex							.071	.789
Female	51.90	(27)	48.00	(12)	55.60	(15)		
Male	48.10	(25)	52.00	(13)	44.40	(12)		
Race							2.096	.148
American Indian or Alaska Native	1.90	(1)	4.00	(1)	.000	(0)		
Asian	1.90	(1)	0.00	(0)	3.70	(1)		
Black or African American	26.90	(14)	20.00	(5)	33.30	(9)		
White	57.70	(30)	68.00	(17)	48.10	(13)		
Two or More Races	11.50	(6)	8.00	(2)	14.80	(4)		
Ethnicity							.000	1.00
Hispanic or Latino	11.50	(6)	12.00	(3)	11.10	(3)		
Not Hispanic or Latino	88.50	(46)	88.00	(22)	88.90	(24)		
Special Education Eligibility							.006	.997
Specific Learning Disability	3.80	(2)	4.00	(1)	3.70	(1)		
Speech or Language Impairment	3.80	(2)	4.00	(1)	3.70	(1)		
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	F	*p*
Age	8.03	(.47)	8.49	(.49)	8.39	(.45)	.593	.445

Table 2

Curriculum-Based Measurement in Written Expression Probes (WE-CBM)

Probe	WE-CBM (self-referenced) narrative probes	WE-CBM (others-referenced) narrative probes	WE-CBM expository probes	WE-CBM compare and contrast probes
1	One night I had a strange dream about…	One day the boy found the most interesting thing and…	Write about your favorite day of the week and why you like it.	Write about how gym is the same as and different from art class.
2	It was the first day of school so I decided to…	One day the boy went to school but nobody was there except him so he…	Describe the friends you have and tell why they are your friends.	Write about how school is the same as and different from home.
3	One morning I found a note under my pillow that said…	The boy was watching TV when he heard a knock at the door and…	Describe your favorite time of the year and why you like it.	
4	I was walking down the street when I saw…	The girl was walking home when she found a $100 bill on the sidewalk and…	Describe a place you like to go and tell why you like to go there.	
5	One day I went to school but nobody was there except me, so I…	The girl was on her way home from school and…		
6	It was a dark and stormy night and I…			
7	I was talking to my friends when all of a sudden…			
8	One day I woke up and was invisible…			
9	One day my friend told me the strangest story and I…			
10	One day when I got home from school…			

Table 3

Results of Procedural Integrity Assessments

	Sessions Assessed			Procedural Integrity Results		
Condition	Percentage	(*n*)	(*N*)	Mean Percentage	(*SD*)	Range of Percentage
[a] Performance Feedback	50%	(6)	(12)	100%	N/A	100-100
[b]Performance Feedback with Generalization Programming	42%	(5)	(12)	100%	N/A	100-100
	13%	(1)	(8)	100%	N/A	100-100

Notes. [a]Contained between 21 and 22 steps. [b]Contained between 22 and 23 steps.

Table 4

Adjusted Means, Standard Deviations, and ANCOVA Results for Pre- and Post-Assessment Measures

	Performance Feedback Condition		Performance Feedback with Generalization Programming Condition		ANCOVA Outcomes		
	M	(*SD*)	*M*	(*SD*)	F	partial n^2	*p*
WE-CBM (Self-Referenced) Narrative Probe					.001	.000	.982
Pre-Assessment	20.92	(12.09)	19.30	(11.55)			
Post-Assessment	28.11	(1.79)	28.17	(1.76)			
WE-CBM (Others-Referenced) Narrative Probe					.005	.000	.945
Pre-Assessment	20.36	(10.62)	21.70	(11.70)			
Post-Assessment	29.55	(2.30)	29.32	(2.25)			
WE-CBM Expository Probe					.046	.001	.831
Pre-Assessment	21.20	(12.21)	18.63	(9.44)			
Post-Assessment	24.26	(2.46)	25.00	(2.36)			
WE-CBM Compare and Contrast Probe					.118	.002	.733
Pre-Assessment	18.64	(11.03)	21.26	(12.12)			
Post-Assessment	28.90	(2.50)	27.71	(2.40)			

Table 5

Raw Means and Standard Deviations for Pre- and Post-Assessment Measures

	Performance Feedback Condition		Performance Feedback with Generalization Programming Condition	
	M	(*SD*)	*M*	(*SD*)
WE-CBM (Self-Referenced) Narrative Probe				
Pre-Assessment	20.92	(12.09)	19.30	(11.55)
Post-Assessment	29.16	(14.78)	27.15	(15.84)
WE-CBM (Others-Referenced) Narrative Probe				
Pre-Assessment	20.36	(10.62)	21.70	(11.70)
Post-Assessment	29.20	(14.06)	29.65	(17.33)
WE-CBM Expository Probe				
Pre-Assessment	21.20	(12.21)	18.63	(9.44)
Post-Assessment	24.92	(13.10)	24.38	(16.06)
WE-CBM Compare and Contrast Probe				
Pre-Assessment	18.64	(11.03)	21.26	(12.12)
Post-Assessment	27.38	(13.43)	29.12	(18.81)

Table 6

Descriptive Results of the Kids Intervention Profile

	Performance Feedback Condition		Performance Feedback with Generalization Programming Condition			
	M	*(SD)*	*M*	*(SD)*	*t*	*p*
Factor 1: Overall Intervention Acceptability	3.59	(0.82)	3.78	(0.63)	1.049	.300
Factor 2: Skill Improvement	2.93	(0.68)	2.88	(0.65)	.263	.793
How much do you like writing stories with us each week?	4.43	(0.95)	3.88	(1.21)	1.756	.086
How much do you like being told what to write about?	2.83	(1.53)	2.81	(1.47)	.043	.966
Were there times when you didn't want to write with us?	2.04	(1.55)	2.54	(1.53)	-1.123	.267
Were there times when you wished you could work more on writing stories with us?	3.83	(1.59)	2.96	(1.62)	1.870	.069
Do you think your writing has improved?	4.48	(1.28)	4.42	(0.95)	.173	.863
Do you think your writing has gotten worse?	3.96	(1.40)	3.65	(1.47)	.737	.465
How much do you like being told how many words you wrote?	4.22	(1.24)	3.69	(1.32)	1.429	.160
How much do you think it helped when you were told how many words you wrote?	1.65	(1.11)	2.04	(1.37)	-1.072	.289

Table 7

Descriptive Results of Teachers' Behavior Intervention Rating Scale (BIRS)

Item	*M*	*(SD)*
Factor 1: Acceptability	**5.00**	**(1.13)**
Factor 2: Effectiveness	**4.71**	**(1.01)**
Factor 3: Time of Effectiveness	**5.00**	**(1.13)**
Overall Acceptability	**4.92**	**(1.11)**
This would be an acceptable intervention for students' writing difficulties.	5.50	(0.71)
Most teachers would find this intervention appropriate for writing difficulties in addition to the one described.	5.00	(1.41)
The intervention should prove effective in changing students' writing difficulties.	5.50	(0.71)
I would suggest the use of this intervention to other teachers.	5.00	(1.41)
Students' writing difficulties are severe enough to warrant use of this intervention.	5.50	(0.71)
Most teachers would find this intervention suitable for the writing difficulties described.	5.00	(1.41)
I would be willing to use this intervention in the classroom setting.	4.50	(2.12)

Item	*M*	*(SD)*
The intervention would not result in negative side-effects for the student.	5.50	(0.71)
The intervention would be appropriate for a variety of students.	5.50	(0.71)
The intervention is consistent with those I have used in the classroom.	3.00	(0.00)
The intervention is a fair way to handle students' writing difficulties.	4.50	(2.12)
The intervention is reasonable for the writing difficulties described.	5.00	(1.41)
I like the procedures used in the intervention.	5.50	(0.71)
The intervention is a good way to handle students' writing difficulties.	5.00	(1.41)
Overall, the intervention would be beneficial for students.	5.00	(1.41)
The intervention would quickly improve students' writing difficulties.	5.00	(1.41)
The intervention would produce lasting change on students' writing difficulties.	5.00	(1.41)
The intervention would improve students' writing difficulties to the point that they would not deviate from other classmates' skills.	5.00	(1.41)
Soon after using the intervention, the teacher would notice a positive change in writing skills.	5.00	(1.41)
Students' writing skills will remain at an improved level even after the intervention is discontinued.	4.50	(0.71)

Using the intervention should not only improve students' writing difficulties in the classroom, but in other settings (e.g., other classrooms, home.	4.50	(0.71)
When comparing students who were poor writers at the beginning of the study with a good writer before and after the use of the intervention, the students' and peer's behavior would be more alike after using the intervention.	5.00	(1.41)
The intervention should produce enough improvement in the student's skills So that writing difficulties are no longer a problem.	4.50	(0.71)
Other skills related to the writing difficulties also are likely to be improved by the intervention.	4.50	(0.71)

Note. $N = 3$. Answers were based on a Likert-type scale with 1 = strongly disagree, and 6 = strongly agree.

Figure 1. Timeline, List of Sessions by Condition, and Percent Increases for Generalization Programming Condition

Date	Performance Feedback Condition	Performance Feedback with Generalization Programming Condition	Percent Increase For Performance Feedback with Generalization Programming Condition
1/22/16	Pre-assessment session 1 WE-CBM (self-referenced) narrative probe 1	Pre-assessment session 1 WE-CBM (self-referenced) narrative probe 1	--
1/28/16	Pre-assessment session 2 WE-CBM (others-referenced) narrative probe 1	Pre-assessment session 2 WE-CBM narrative probe (others) 1	--
2/5/16	Pre-assessment session 3 WE-CBM expository probe 1	Pre-assessment session 3 WE-CBM expository probe 1	--
2/22/16	Pre-assessment session 4 WE-CBM compare and contrast probe 1	Pre-assessment session 4 WE-CBM Compare and contrast probe 1	--
3/4/16	Intervention session 1 WE-CBM (self-referenced) narrative probe 2	Intervention session 1 WE-CBM (self-referenced) narrative probe 2	11%
3/18/16	Intervention session 2 WE-CBM (self-referenced) narrative probe 3	Intervention session 2 WE-CBM (self-referenced) narrative probe 3	29%
3/28/16	Intervention session 3 WE-CBM (self-referenced) narrative probe 4	Intervention session 3 WE-CBM (self-referenced) narrative probe 4	67%
4/1/16	Intervention session 4 WE-CBM (self-referenced) narrative probe 5	Intervention session 4 (stimulus generalization programming) WE-CBM (others-referenced) narrative probe 2	43%
4/18/16	Intervention session 5 WE-CBM (self-referenced) narrative probe 6	Intervention session 5 (stimulus generalization programming) WE-CBM (others-referenced) narrative probe 3	37%
4/22/16	Intervention session 6 WE-CBM (self-referenced) narrative probe 7	Intervention session 6 (stimulus generalization programming) WE-CBM (others-referenced) narrative probe 4	39%
5/2/16	Intervention session 7 WE-CBM (self-referenced) narrative probe 8	Intervention session 7 (response generalization programming) WE-CBM expository probe 2	18%
5/6/16	Intervention session 8	Intervention session 8 (response generalization programming)	61%

	WE-CBM (self-referenced)narrative probe 9	WE-CBM expository probe 3	
5/9/16	Post assessment session 1 WE-CBM (self-referenced) narrative probe 10	Post assessment session 1 WE-CBM (self-referenced) narrative probe 10	--
5/13/16	Post assessment session 2 WE-CBM (others-referenced) narrative probe 5	Post assessment session 2 WE-CBM (others-referenced) narrative probe 5	--
5/16/16	Post assessment session 3 WE-CBM expository probe 4	Post assessment session 3 WE-CBM expository probe 4	--
5/20/16	Post assessment session 4 WE-CBM compare and contrast probe 2	Post assessment session 4 WE-CBM compare and contrast probe 2	--

Figure 2. Particpant Flow Chart

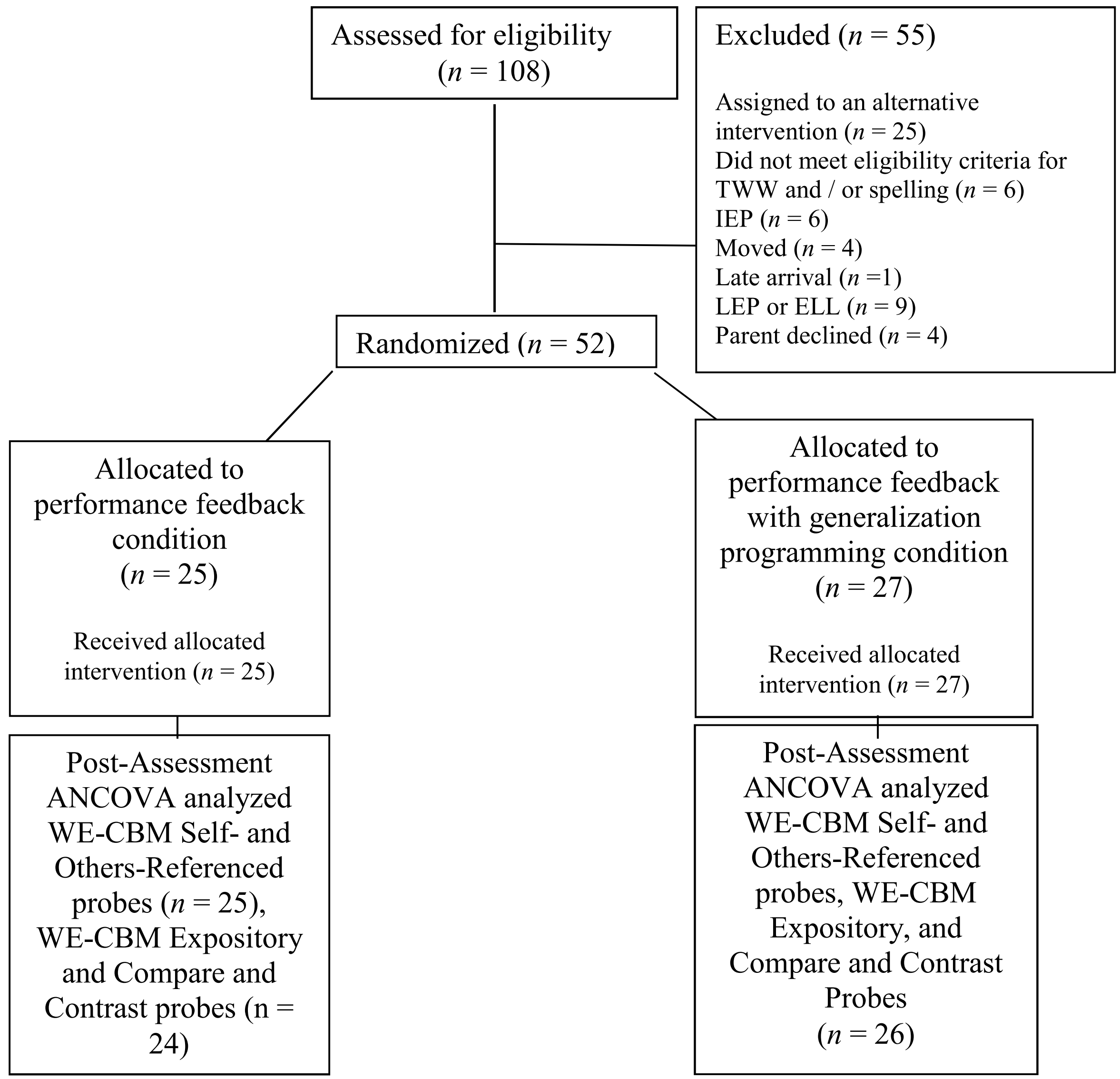
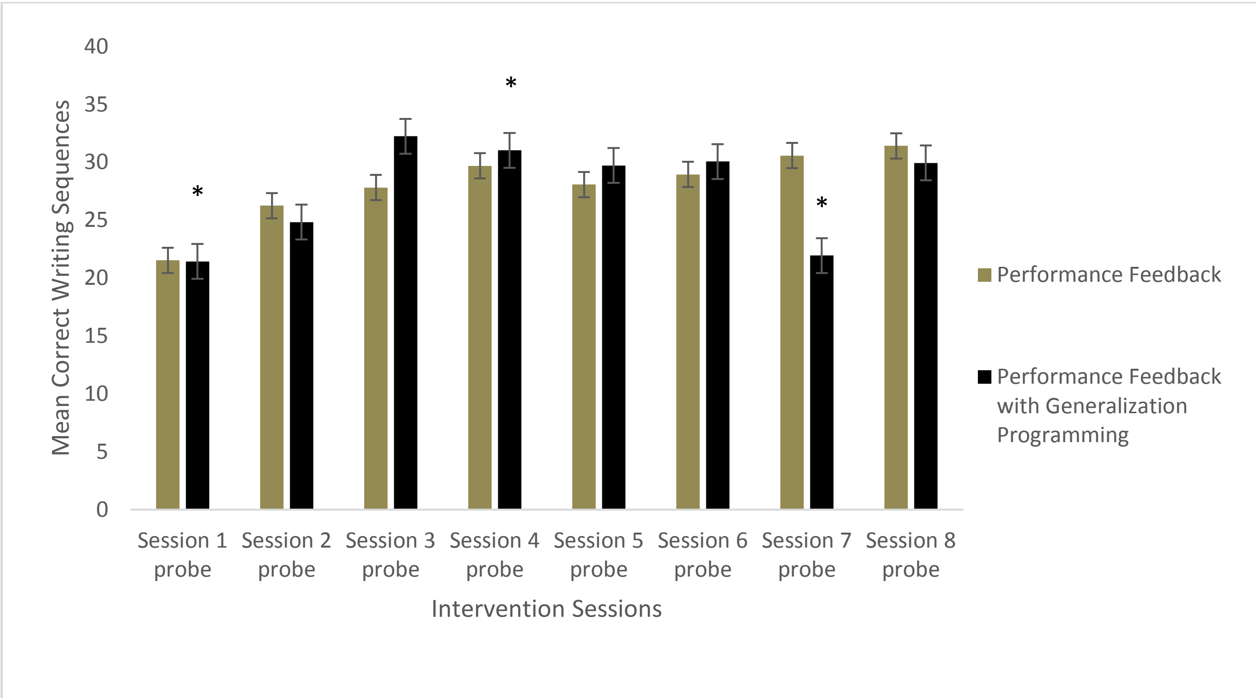
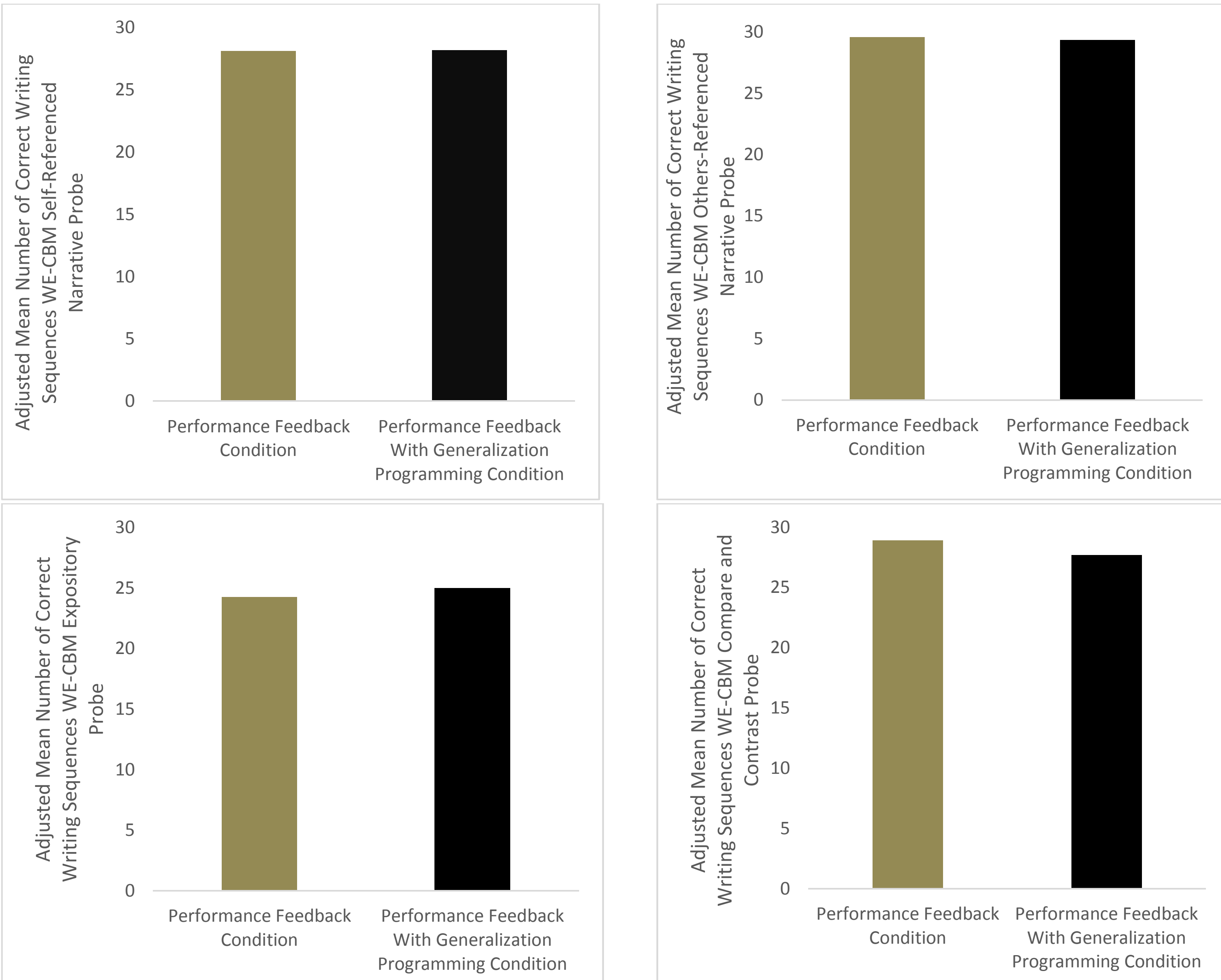


Figure 3. Students' Mean Correct Writing Sequences Across Sessions and Conditions.



Note. Sessions in which new generalization programming tactics were introduced for the performance feedback with generalization programming condition are marked with an asterisk.

Figure 4. Adjusted Mean Scores by Condition on the Post WE-CBM Self-Referenced Narrative Probe, Others-Referenced Narrative Probe, Expository Probe, and Compare and Contrast Probe.



References

AIMSweb®, Inc. (2012). AIMSweb. Retrieved July 27, 2017, from *http://www.aimsweb.com*.

Ardoin, S.P., Eckert, T.L., & Cole, A.S. (2008). Promoting generalization of reading: A comparison of two fluency-based interventions for improving general education student's oral reading rate. *Journal of Behavioral Education, 17,* 237-252.doi: 10.1007/s10864-008-9066-1

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16*, 55-70. doi: 10.1007/s10864-006-9020z

Ardoin, S. P., Williams, J. C., Klubnik, C., & McCall, M. (2009). Three versus six rereadings of practice passages. *Journal of Applied Behavior Analysis, 42*, 375-380. doi: 10.1901/jaba.2009.42-375

Baer, D., Wolf, M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*, 91-97. doi: 10.1901/jaba.1968.1-91

Campbell, B. J., Brady, M. P., & Linehan, S. (1991). Effects of peer-mediated instruction on the acquisition and generalization of written capitalization skills. *Journal of Learning Disabilities, 24*, 6-14. doi: 10.10.1177/002221949102400103

Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education*, *19*, 93-105. doi: 10.1007/s10864-010-9102-9

Codding, R. S., Eckert, T. L., Fanning, E., Shiyko, M., & Solomon, E. (2007). Comparing mathematics interventions: The effects of cover-copy-compare alone and combined with performance feedback on digits correct and incorrect. *Journal of Behavioral Education, 16*, 125-141. doi: 10.1007/s10864-006-9006-x

Cohen, E. J., & Brady, M. P. (2011). Acquisition and generalization of word decoding in students with reading disabilities by integrating vowel pattern analysis and children's literature. *Education and Treatment of Children, 34*, 81-113. doi: 10.1353/etc.2011.0006

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology, 100*, 907–919.

Cuvo, A. J. (2003). On stimulus generalization and stimulus classes. *Journal of Behavioral Education, 12*, 77-83.

Duhon, J. D., House, S. E., Poncy, B. C., Hastings, K. W., & McClurg, S. C. (2010). An examination of two techniques for promoting response generalization or early literacy skills. *Journal of Behavioral Education, 19,* 62-75. doi:10.1007/s10864-010-9097-2

Eckert, T. L., Hier, B. O., Malandrino, R. D., & Hamsho, N. (in press). Assessing children's perceptions of academic interventions: The Kids Intervention Profile. *School Psychology Quarterly.*

Elliot, S. N. & Treuting, M. V. B. (1991). The Behavior Intervention Rating Scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology, 29*, 43-51. doi: 10.1016/0022-4405(91)90014-i

Graham, S., Harris, K., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*, 207-241. doi:10.1016/jcedpsych.2004.08.001

Haring, N. G., & Eaton, M. D. (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: Charles E Merrill Publishing Co.

Hier, B. O., & Eckert, T. L. (2016). Programming generality into a performance feedback writing intervention: A randomized controlled trial. *Journal of School Psychology, 56*, 111-131. doi: 10.1016/j.jsp.2016.03.003

Hier, B.O., & Eckert, T.L. (2014). Evaluating elementary-aged students' abilities to generalize and maintain fluency gains of a performance feedback writing intervention. *School Psychology Quarterly, 29(4)*, 488-502. doi: 10.1037/spq0000040

IBM Corp. (2012). IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

Jackson, H. I. (1995). *Performance feedback and generalization programming in writing instruction.* Retrieved from ProQuest Information and Learning. (AAM9522528).

Klubnik, C. & Ardoin, S. P. (2010). Examining immediate and maintenance effects of a reading intervention package on generalization materials: Individual versus group implementation. *Journal of Behavioral Education*, 7-29. doi: 10.1007/s10864-009-9096-3

McMaster, K., Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education, 41*, 68-84. doi:10.1177/00224669070410020301

McMaster, K.L., Wayman, M.M., Deno, S.L., Espin, C.A., & Yeo, S. (2010). *Examining technical features of progress monitoring measures across grade levels in writing.* (Report No. 38). Minneapolis, MN: Research Institute on Progress Monitoring.

Macmillan McGraw-Hill (2006). *Treasures grade 3.* New York, NY: Macmillan/McGraw-Hill.

Malandrino, R. (2015). *Generalization programming and the instructional hierarchy: A performance feedback intervention in writing* (Order No. 1593898). Available from Dissertations & Theses @ Syracuse University. (1710514360). Retrieved from https://search-proquest-com.libezproxy2.syr.edu/docview/1710514360?accountid=14214

Medcalf, J., Glynn, T., & Moore, D. (2004). Peer Tutoring in Writing: A school systems approach. *Educational Psychology in Practice, 20,* 157-178.

Mesmer, E. M., Duhon, G. J., Hogan, K., Newry, B., Hommema, S., Flethcher, C., & Boso, M. (2010). Generalization of sight word accuracy using a common stimulus procedure: A preliminary investigation. *Journal of Behavioral Education, 19,* 47-61.doi: 10.1007/s1064-010-9103-8

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history / social studies, science, and technical subjects.* Washington, DC: Authors.

Noell, G.H., Connell, J.E., & Duhon, G.J. (2006). Spontaneous response generalization during whole word instruction: Reading to spell and spelling to read. *Journal of Behavioral Education, 15,* 121-130. doi:10.1007/s10864-006-9016-8

Osnes, P. G., & Lieblein, T. (2003). An explicit technology of generalization. *The Behavior Analyst Today, 3*, 364-374. doi: 10.1037/h0099994

Pratt-Struthers, J., Struthers, T. B., & Williams, R. L. (1983). The effects of the add-a-word spelling program on spelling accuracy during creative writing. *Education and Treatment of Children, 6*, 277-283.

Pearson. (2009). *Wechsler Individual Achievement Test – Third Edition*. San Antonia, TX: PsychCorp.

Persky, H. R., Daane, M. C., & Jin, Y. (2003). *The Nation's Report Card Writing 2002* (NCES 2003-529). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Peterson-Brown, S., & Burns, M.K. (2011). Adding a vocabulary component to incremental rehearsal to enhance retention and generalization. *School Psychology Quarterly, 26*, 245-255. doi:10.1037.a0024914

Poncy, B.C., Duhon, G.J., Lee, S.B., & Key, A. (2010). Evaluation of techniques to promote generalization of basic math fact skills. *Journal of Behavioral Education, 19,* 76-92. doi: 10.1007/s10864-010-9101-x

Powell-Smith, K., & Shinn, M. (2004). Administration and scoring of written expression curriculum-based measurement (WE-CBM) for use in general outcome measurement. Eden Prairie, MN: Edformation.

Roca, J. V. & Gross, A. M. (1996). Report-do-report: Promoting setting and setting-time generalization. *Education and Treatment of Children, 19*, 408-424.

Schunk, D., & Swartz, C. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18,* 337-354. doi:10.1080/02783199309553512

Shapiro, E. S. (2004). *Academic skill problems: Direct assessment and intervention*. New York, NY: Guilford.

Silber, J. M., & Martens, B. K. (2010). Programming for the generalization of oral reading fluency: Repeated readings of entire text versus multiple exemplars. *Journal of Behavioral Education, 19,* 30-46.doi: 10.1007/s10864-010-9099-0

Skinner, C. H., & Daly III, E. J. (2010). Improving generalization of academic skills: Commentary on the special issue. *Journal of Behavioral Education, 19*, 106-115. doi: 10.1007/s10864-010-9100-y

Stokes, T.F., & Baer, D.M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-367. doi: 10.1901/jaba.1977.10-349

Stokes, T.F., & Osnes, P.G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20,* 337-355. doi: 10.1016/s0005-7894(89)80054-1

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2012*. Washington, DC: Author.

Wood, D. A., Rosenberg, M. S., & Carran, D. T. (1993). The effects of tape-recorded self-instruction cues on the mathematics performance of students with learning disabilities. *Journal of Learning Disabilities, 26*, 250-258. doi: 10.1177/002221949302600405

Van Houten, R. (1979). The performance feedback system: Generalization of effects across time. *Child Behavior Therapy*, *1*, 219-236. doi: 10.1300/J473v01n03_0

Rigby Malandrino

Vita

Contact:
430 Huntington Hall
Syracuse, NY 13244

Education:

Syracuse University Department of Psychology, School Psychology Ph.D. Program	2012-2017
West Virginia University Department of Psychology, B.A.	2006-2010