A Comparison of Two Spelling Strategies with respect to Acquisition, Generalization, Maintenance, and Student Preference

By

Tracie B. Mann

Submitted to the graduate degree program in Applied Behavioral Sciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Chairperson Edward K. Morris, Ph.D.

____
Florence DiGennaro Reed, Ph.D.

____
Claudia Dozier, Ph.D.

____
Pamela Neidert, Ph.D.

____
Nancy Brady, Ph.D.

Date Defended: June 2, 2014

The Dissertation Committee for Tracie B. Mann

certifies that this is the approved version of the following dissertation:

A Comparison of Two Spelling Strategies with respect to Acquisition, Generalization, Maintenance, and Student Preference

Chairperson Edward K. Morris, Ph.D.

Date approved: June 2, 2014

Abstract

Correct spelling is a learned performance, but effective and preferred procedures to develop accurate spelling in young children have not been adequately described. We evaluated the effectiveness of two strategies for teaching spelling to 10 elementary students of typical development. In the traditional rehearse and test method commonly used in elementary classrooms, we gave students a list of ten words on Monday, they practiced spelling the words throughout the week, and then were tested on Friday. We also taught students to use the cover-copy-compare (CCC) method to practice their spelling words within a similar time frame. During CCC, we also taught students to say each phoneme of a word ("sound out") as they practiced each word. Interobserver agreement was collected for 33% of sessions; agreement was 100% for all measures. A reversal design showed that CCC was clearly more effective for promoting acquisition of spelling words for six students, and for promoting generalization and maintenance for two students. No difference between conditions was observed with the remaining students. Nine of the ten students preferred CCC to rehearse and test. Implications for the design of an effective spelling curriculum are discussed.

Keywords: CCC, concurrent chains arrangement, generalization, maintenance, preference, spelling

Table of Contents

Introduction 1
Method 21
Participants 21
Materials 22
Setting 22
Procedure 22
Pretest 23
Practice 24
Rehearse and test 24
Copy, cover, and compare plus sounding out 25
Acquisition test 27
Generalization test 28
Maintenance test 29
Preference Assessment 29
Response Measures 31
Experimental Design 32
Interobserver Agreement 33
Results 33
Generalization test words 33
Efficacy 33
Preference 34
Pretest error analysis 35
Practice error analysis 36
Duration of practice 36
Discussion 37
Efficacy 37
Preference 41
References 47
Figures 55

Introduction

Spelling is an important and ubiquitous part of every elementary student's formal curriculum and has been since school began (Heron, Okyere, & Miller, 1991; Scott, 2000). Learning to spell is important because it predicts the amount and quality of written composition in elementary students. Students who are effective spellers are more likely to be effective writers (e.g., Okyere, Heron, & Goddard, 1997; Scott, 2000). By contrast, students who spell poorly are more likely to forget their formulated writing ideas, and limit their writing by avoiding words they cannot spell (Alber & Walshe, 2004). In addition, poor spellers are often labeled uneducated or careless (Okyere, Heron, & Goddard, 1997; Scott, 2000.), and their problems with spelling often persist into later years and have lasting effects on their writing skills. This is especially true of children with learning disabilities who are far more likely to be poor spellers than their typically developing peers (Graham, 2000).

Despite its importance, spelling has received less attention in recent years and is often noted as a subject commanding only modest concerns in the elementary curriculum (Dagdag, McLaughlin, & Weber, 2002; Heron, Okyere, & Miller, 1991). This is perhaps due to the emergence of state-testing mandated by the No Child Left Behind Act, which focuses primarily on reading, math, writing, science, and social studies and not directly on spelling. Several researchers have noted that elementary teachers do not have sufficient knowledge about effective spelling instruction (e.g., Schermerhorn & McLaughlin, 1997) and, consequently, may be incorporating teaching strategies that are effective only for some students. Further, spelling is consistently rated by teachers and students as one of the least preferred subject areas in the curriculum (Nies & Belfiore, 2006).

Traditional “assign and test” methods of spelling instruction in which students are given a list of 10 – 20 words on Monday, engage in some type of practice throughout the week, and then are tested on Friday, are not effective with many students (e.g., Alber & Walshe, 2004; Cuvo, Ashley, Marso, Zhang, & Fry, 1995). Traditional approaches are implicit approaches; that is, they provide the words to be learned, but not specific strategies for how to learn them. Often, students are left to develop their own study tactics or their parents take on the teaching task during homework. They also vary with respect to the practice component from very little (i.e., assigning the word list on Monday and asking students to practice at home for the Friday test) to extensive (i.e., assigning daily spelling practice sessions that may consist of rehearsing the words several times, using the words in sentences, etc.). Either way, some spelling researchers argue that these methods do not teach at all. They simply focus on the memorization of a finite set of words, which cannot possibly prepare all students to become competent spellers, readers, and writers (e.g., Alber & Walshe, 2004; Scott, 2000). Further, all traditional approaches lack several components of effective instruction, including individualization and the content relevance of word lists (Scott, 2000), ample response and error correction opportunities, and immediate and frequent reinforcement.

Fortunately, researchers have demonstrated a better way to teach spelling. Some research on spelling has focused on specifically promoting spelling acquisition by manipulating antecedent or practice variables, such as varying the number of words to be practiced, the amount of practice, interspersing known words with unknown words, breaking up words into syllables, and implementing a constant time delay (CTD) procedure.

Bryant, Drabin, and Gettinger (1981) compared the effects of varying the number of words taught at one time on the spelling acquisition of 64 5th graders with disabilities. All groups learned the same number of words by the end of the week period; however, students committed many more errors when presented with more words to practice and there was more response variability in the more-words group. In a similar line of research, Cuvo et al. (1995) conducted a parametric study also examining, among other variables, the amount of practice on spelling acquisition and maintenance in four students with disabilities. The results showed minimal differences between the effects of small, moderate, and large amounts of practice on acquisition and maintenance. The data from these two studies may seem counterintuitive, but they suggest that the number of words to be practiced at one time and the amount of rehearsals are not significant variables in learning spelling words.

Van Houten and Van Houten (1991) and Mann, Bushell, and Morris (2010) examined the effects of breaking down words into smaller units on spelling acquisition. In the former study, Van Houten and Van Houten compared the effects of presenting words as a whole and words broken down into syllables in 5 elementary students. Results indicated that students learned words faster when the words were broken down into units. Mann, Bushell, and Morris also examined a method to "break-down" words by having students sound out words and write what they say. The sounding-out strategy increased spelling performance compared to no-sounding out for all five elementary students. These studies suggest that breaking down words into smaller units improves spelling acquisition; however, the effects on maintenance and generalization to a writing context were not tested.

Interspersal and high-probability (high-p) sequencing are procedures that have been effective in teaching new skills to children (Cates, et al., 2003). Interspersal techniques alter the academic lesson by adding mastered tasks among tasks that students are learning. High-p sequencing is a procedure in which researchers present several tasks that are likely to be completed accurately first and then present a task that is less likely to be completed accurately. Neef, Iwata, and Page (1980) compared the effects of interspersing known items during spelling instruction and a high-density reinforcement condition to a control condition functionally similar to a traditional procedure in 3 students with disabilities. Interspersal training was the most effective in improving acquisition and maintenance in all 3 students. Koegel and Koegel (1986) also showed evidence that interspersal training is superior to a more traditional approach. They compared the effects of interspersal training to a traditional baseline on academic skill acquisition in an 8 yr. old stroke victim. Interspersal training dramatically improved spelling acquisition relative to baseline performance.

In contrast, Cates et al. (2003) compared the effects of high-p sequencing and an interspersal procedure with traditional drill and practice on spelling acquisition in 5 typically developing 2nd graders. Results showed little difference between conditions with respect to number of words mastered; however, students required more time to learn the words in the high-p sequencing conditions than any other condition, and all but one student mastered target words fastest in the traditional drill and practice condition. This is the only known study to show a traditional procedure to be more effective than a specialized approach. It is possible that because these students were typically developing, a traditional approach was sufficient to improve spelling. Another possibility is that the words used in this study were all 3 letter consonant-vowel-consonant (CVC) words, relatively easy words to learn for many children. The undifferentiated results may have been partially a function of a ceiling effect.

Perhaps the words used were not challenging enough to detect a difference between practice conditions.

A final manipulation occurring during spelling practice that has received attention in the literature is the application of a constant time delay (CTD) to improve spelling performance. Typically in these studies, the experimenter delivers a prompt immediately following antecedent stimulus, which in this case, is a model of the correct spelling of the word (i.e., the experimenter dictates the spelling word and immediately holds up a card with the correctly spelled word for the student to copy). In subsequent sessions, prompts are delivered 5 s following the target word or immediately following an error. The desired effect is that control will transfer from the prompt to the target word, so that it will come to function as a discriminative stimulus evoking the correct spelling response. In a single-case study, Stevens and Schuster (1987) investigated the effects of CTD on spelling acquisition, generalization, and maintenance in a 6^{th} grade boy with a learning disability in a multiple baseline design across word sets. Using the procedure described above, spelling acquisition increased when and only when the CTD procedure was implemented. Further, the effects generalized across settings and tasks and maintained at high levels over a 2-week period. Stevens, Blackhurst, and Slaton (1991) extended these findings by including more participants and combining the CTD procedure with computer-assisted instruction. They used the same general procedure except that the experimenters designed a computer program to dictate words and deliver prompts and feedback. This procedure improved spelling for 4 of 5 students and the effects generalized to a written test and maintained on a two-week maintenance probe.

In more recent studies, Cates et al. (2007) and Coleman-Martin and Heller (2004) further examined the effects of CTD in students with disabilities. Consistent with other studies, Coleman-Martin and Heller demonstrated that CTD improved spelling in three children with physical disabilities. Cates extended the literature by comparing CTD to cover, copy, and compare (CCC), another commonly used spelling strategy for children with disabilities, in three typically developing 3rd grade boys. When using the CCC procedure, students first copy the model spelling word, then cover the word and write it from memory, and finally compare their written word with the model. CCC slightly improved spelling acquisition compared to CTD in all students, although it took more instructional time. The effects of the two strategies on maintenance and generalization were not as robust. Both CTD and CCC promoted maintenance over time and generalization to reading; however, CTD was slightly more effective than CCC for two students with respect to maintenance and for one student with respect to generalization. The authors concluded that because CCC took more time and only slightly improved spelling acquisition, either procedure could be used effectively.

In sum, the research from studies that examined only the effects of practice variables may inform educational practices in the following ways: amount of practice does not matter that much, breaking words down into smaller units may be beneficial to many learners, interspersing novel words with mastered words may facilitate acquisition, and using a constant time delay procedure will likely improve spelling in children with disabilities. Research investigating how antecedent variables can improve spelling is important, but is likely to be far more useful to teachers if the data from these studies can be evaluated concurrently with data from studies that address the consequences for spelling (i.e., error-correction).

Error-correction is thought to be one of the most functionally important components in any teaching strategy (e.g., Carnine, 1980; McGuffin, Martz, & Heron, 1997). In general, error-correction is a broad term that refers to a teacher response following a student error. Rodgers and Iwata (1991) delineated four general types of error-correction procedures: (a) the absence of programmed consequences, (b) programming a delay prior to the next learning trial, (c) presenting some kind of discrete event following an error in which the student remains passive, and (d) presenting a remedial trial contingent on errors. With respect to spelling, the most common form of error-correction is the last, a remedial trial (often trials) in the form of a teacher or student-delivered prompt to engage in the appropriate response (Worsdell et al., 2005). These error-correction procedures are variably labeled positive practice, directed rehearsal, or overcorrection and all involve repeated practice under appropriate stimulus conditions (Barbetta, Heron, & Heward, 1993; Cuvo, Ashley, Marso, Zhang, & Fry, 1995). Research on error-correction has generally focused on examining: (a) the effects of various procedural variations (i.e., timing, amount, and source), (b) the function of error-correction, and (c) how error-correction procedures compare to traditional spelling strategies.

With respect to procedural variations, researchers (e.g., Cuvo, et al., 1995; Foxx & Jones, 1978) have evaluated the effect of the number of remediation trials on spelling performance. Intuitively, it would seem that the larger number of remediation trials contingent on an error would produce better performance. Either more correct responding under appropriate stimulus conditions should improve spelling or repeated practice may be aversive and may be expected to improve performance via negative reinforcement. However, consistent with the results of the practice variable research described above, this is not always the case. For example, Cuvo et al. (1995) compared conditions in which individuals with disabilities engaged in 5, 10, and 15 practice trials contingent on errors and found no significant difference on rate of spelling acquisition. By contrast, Foxx and Jones (1978) found that increased amounts of repeated practice improved spelling performance. It is possible that the conflicting results are a function of the relative difference between amounts of error-correction. Foxx and Jones compared very large amounts of practice to very little. The most effective condition was the one in which students engaged in two remediation procedures, one following a pretest and then one following a posttest, for a total of 14 extra rehearsals for each misspelled word. This procedure improved spelling relative to a traditional approach, and a procedure in which students engaged in only one remediation procedure for each misspelled word. It is possible that Cuvo et al. would have observed a difference had they included a practice condition in which the students only engaged in one error-correction.

Another procedural variation that has been evaluated is the effect of the timing of error-correction on spelling performance. Morton, Heward, and Alber (1998) compared the effects of immediate error correction (i.e., evaluating each word) and delayed error-correction (i.e., evaluating all words upon completion of the list) on the acquisition and maintenance of spelling in 5 elementary students with disabilities. Results showed that immediate error-correction improved acquisition for all 5 students. In a systematic replication of Morton et al., Alber and Walshe (2004) also found that immediate error-correction produced higher scores on acquisition and maintenance tests. The efficacy data from these studies are consistent with research that has demonstrated the superiority of immediate error correction for improving other academic skills such as math (Bennett & Cavanaugh, 1998) and sight-word acquisition (Barbetta et al., 1994).

Research has also examined the effects of teacher-directed vs. student-directed evaluation and correction on spelling performance. Gettinger (1985) compared teacher-directed error correction and student-directed error correction with a traditional control condition and found that while both error-correction procedures improved spelling, all nine students performed better on posttests when they detected their own errors. Similarly, Viel-Ruma, Houchins, and Fredrick (2007) evaluated the effects of self-correction compared to a traditional rehearse and test condition on spelling performance in 3 high-school students. The purpose of this study was to add to the literature on self-correction, and not necessarily to compare teacher versus student-detected errors; however, the researchers designed the conditions in such a way that the data support Gettinger's findings. In most self-correction studies, self-correction is compared with a traditional procedure that involves delayed feedback or no feedback at all from the teacher. In this study, the teacher immediately corrected spelling errors and instructed students to write the misspelled words again. This procedure was compared with self-correction, in which students detected their own errors by comparing their spelled words with a model. For all three students, self-correction improved spelling performance compared to the traditional approach.

While the above research has focused on the effects of several important characteristics of error-correction, researchers have also attempted to isolate the function of repeated practice error-correction on spelling acquisition. This line of research is important because a better understanding of the behavioral mechanism(s) underlying the error-correction component of an instructional strategy would contribute to both the conceptual analysis of error-correction and, from a practical perspective, may help teachers design effective and preferred curricula. One way that researchers have attempted to isolate the function of repeated practice is to compare “relevant” and “irrelevant” practice conditions. The rationale for the comparison is that error-correction likely improves spelling acquisition in two ways: (a) by providing repeated practice under appropriate stimulus conditions, thereby enhancing stimulus control over correct responding, or (b) by improving performance via negative reinforcement. Thus, if irrelevant practice were to improve performance, it would suggest a negative reinforcement contingency, whereas if relevant practice were superior, it would suggest a more educative, positive reinforcement contingency.

Axelrod, Kramer, Appleton, Rockett, and Hamlet (1984) and Cuvo et al. (1995) compared the effects of relevant and irrelevant practice in 3 and 5 students with disabilities, respectively. In the former, the researchers compared relevant and irrelevant practice to a traditional procedure using an alternating treatments design. The results showed little difference in the effects of relevant and irrelevant practice; that is, both types of error-correction procedures improved spelling performance relative to the traditional procedure. These results are consistent with those demonstrated by Cuvo et al. in their comparison of relevant and irrelevant practice in 5 students with disabilities; however, Cuvo et al. extended their findings in that maintenance of spelling was measured. The authors did not observe a difference in conditions. Taken together, these results suggest that error-correction procedures may simply be sufficiently aversive to improve spelling performance via negative reinforcement.

The above research, particularly the research examining procedural variations of error-correction, led researchers to design and evaluate a self-correction procedure to increase spelling acquisition in students (e.g., McGuffin, Martz, & Heron, 1997; McNeish, 1992; Okyere, Heron, & Goddard, 1997; Vargas et al., 1997; Wirtz, 1996;). In these studies, the self-correction procedure is typically as follows: (a) spelling words are dictated either from a tape recorder or by a researcher, (b) the student writes the word on his paper, (c) the student stops the tape recorder and checks his work against a written model that is covered during practice, (e) if it is incorrect, the student corrects the error, letter by letter, using formal proofreading marks and writes the word again, and (f) if it is correct, the student turns on the tape recorder, listens to the next word and repeats the process. The critical components of this procedure are "best practice" as determined by previous research; that is, students: (a) self-correct (i.e., student-directed), (b) after each word, (c) engage in only one contingent repeated practice per error, and (d) the practice is relevant (i.e., they practice the same word).

In general, the research on self-correction is more sophisticated than what has been reviewed in the above sections in that many studies include multiple measures in addition to acquisition. For example, one of the earliest empirical evaluations of the effects of self-correction, McNeish (1992), compared the effects of self-correction to a traditional approach in five middle-school students diagnosed with a learning disability with respect to acquisition, maintenance, and generalization. Self-correction resulted in far greater improvements in spelling for all students. In the first replication of McNeish, Wirtz et al. (1996) extended the findings by comparing the same two procedures with six typically developing, but low achieving, 3rd grade students. Interestingly, although all six students were more successful in the self-correction condition, only one student spelled more than

90% of words correctly on posttests (i.e., earning an ‘A’ while other students earned ‘Bs’ and ‘Ds’). Wirtz et al. reported similar results with respect to maintenance and generalization; that is, students performed better during the self-correction condition, but only maintained an average of 75% of words on delayed posttests and on an oral recitation test. In a satisfaction survey, all six students reported that they preferred self-correction, but only 3 reported that it helped them spell better.

Okyere, Heron, and Goddard (1997) further replicated these findings in a clinical setting in 6 elementary children who scored below the 5th percentile in a standardized test of spelling achievement. McGuffin, Martz, and Heron (1997) extended the research by showing that self-correction was effective when students engaged in whole-word correction instead of letter-by-letter.

The self-correction procedure is clearly an effective strategy to improve spelling performance. However, the trial and error nature of student responding inherent in this procedure may be improved upon by adding a practice component in which the student is highly likely to emit a correct response on the first practice trial. Two such teaching strategies that combine effective practice variables with a self-correction component are Cover, Copy, Compare (CCC; McGuigan, 1975; Skinner, Ford, & Yunker, 1991) and Add-A-Word (McGuigan, 1975; Pratt-Struthers, Struthers, & Williams, 1983). The generic procedure used in both CCC and Add-A-Word is as follows: (a) the student looks at the spelling word and copies it, (b) the student covers the word and writes it again, (c) if it is correct, the student moves on to the next word, (d) if the response is incorrect, the student corrects the error and then engages in another rehearsal of the word. Both procedures are self-paced and self-managed, components that researchers have shown to improve academic skill acquisition with respect to spelling (e.g., Rafferty & Arroyo, 2011), mathematics (e.g., Farrell & McDougall, 2008), and reading (e.g., Edwards, Salant, Howard, Brougher, and McLaughlin, 1995). The only difference between CCC and Add-A-Word is that the latter is more of a complete spelling curriculum. That is, Add-A-Word uses the CCC procedure, but includes rules for classroom use such as incorporating mastery criterion and a review component (e.g., students must spell a word correctly on 3 consecutive trials before it is removed from the list, and words are programmed for review every so often). In sum, the practice and self-correction components in Add-A-Word are functionally identical to CCC; therefore, research on both strategies is discussed concurrently.

Several researchers have used CCC or Add-A-Word to increase spelling performance in young children. Larsen and McLaughlin (1997) used CCC to successfully teach a typically developing preschool girl to spell 20 words using an AB design. Hubbert, Weber, and McLaughlin (1997) extended these findings by comparing the effects of CCC and a traditional approach on the spelling acquisition in an adolescent girl with a conduct disorder and demonstrating that CCC was more effective. The generality of these single-case studies is limited, but they both showed that CCC can work.

Schermerhorn and McLaughlin (1997), Murphy et al. (1990), and Pratt-Struthers et al. (1983) compared the effects of Add-A-Word or CCC with traditional spelling approaches in well-controlled studies using more than one participant. Schermerhorn and McLaughlin taught 16 typically developing 5th and 6th graders to use the Add-A-Word strategy and found that Add-A-Word increased spelling acquisition for all students compared to a traditional approach. Pratt-Struthers et al., and Murphy et al. demonstrated similar results in students with disabilities. In a more recent study, Nies and Belfiore (2006) compared the effects of

CCC with a copy only approach (a common traditional spelling strategy) in two elementary students with disabilities. Results indicated that CCC increased the number of words learned and maintained compared to the copy only approach.

Jaspers et al. (2011) compared the effects of CCC with CCC plus sentence definition (CCC + SD), which consisted of the experimenter reading the word in a sentence and then providing the definition immediately before the student engaged in the CCC procedure. Both procedures improved spelling acquisition relative to a no-practice control condition; however, acquisition percentage was lower in this study than in other studies. Three participants acquired on average, only 57% of words in the CCC condition and 58% in the CCC + SD condition. It is possible that the mastery criterion (spelling the word correctly on two consecutive daily assessments) accounted for this discrepancy; also, the participants used in this study were younger than those in most other CCC studies and the authors note that their writing fluency may have interfered with their spelling performance.

Erion et al. (2009) extended the literature on CCC by evaluating the differential effects of one versus three error-correction rehearsals on acquisition and maintenance in three elementary students who displayed poor spelling skills. Consistent with results from other studies (i.e., Cuvo, 1995; Foxx & Jones, 1978), the researchers did not observe a significant difference in acquisition or maintenance, suggesting that number of rehearsals may not affect spelling performance.

In further analysis of CCC, Mann, Bushell, and Morris (2010) evaluated the effects of CCC by adding a “sounding out” component to the general procedure. Sounding out improved spelling performance for all five students compared to CCC alone. The data suggested that adding a sounding-out component further enhances the already beneficial effects of CCC and the combination of both may offer a simple, effective strategy to improve spelling in young children. However, similar to other studies that examined the effects of CCC, Mann et al. (2010) did not investigate the effects on the generalization and maintenance of spelling, nor was student preference evaluated. No study has empirically evaluated the effects of CCC with respect to all three measures.

Maintenance

Assessing maintenance of spelling over time is clearly an important measure of the overall efficacy of an instructional strategy. In fact, the lack of maintenance data is a common criticism of traditional spelling measures (e.g., McNeish et al., 1992). Students learn to spell words for their Friday test and then are unable to spell many of them on assessment probes later in the school year. Some of the above research included maintenance probes and may suggest variables that can improve the likelihood that students will continue to spell words correctly following acquisition (Alber & Walshe, 2004; Gettinger, 1985; Morton, Heward, & Alber, 1998)

Wirtz et al. (1996) found that the above-described self-correction strategy improved maintenance over a 10-day period relative to a traditional procedure; however, only 75% of the words maintained. In contrast, McGuffin, Martz, and Heron (1997) and Okyere, Heron, and Goddard (1997) found that the self-correction strategy significantly improved maintenance. Students in these studies maintained, on average, 86% and 94% of words, respectively. With respect to the research on CCC and Add-A-Word, Nies and Belfiore (2006) and Cuvo et al. (1995) included maintenance probes, but the data show that, on average, students retained a high percentage of words on delayed tests using the CCC method, 95% and 94%, respectively. In contrast, Jaspers, et al. (2011), reported that students maintained, on average, only 75% of words on delayed tests and Erion et al. (2009) also reported a retention percentage of less than 80%. These data are limited, however, in that they represent the performance of only 8 participants and the maintenance probes were scheduled within a week following acquisition.

In their analyses of CTD, Stevens, Blackhurst, and Slaton (1991), Stevens and Schuster (1987), and Cates et al. (2006) included maintenance probes. Stevens and Schuster reported that the participant retained 93% of mastered words on a 3-week retention test and Cates et al. reported that CTD promoted maintenance for 2 of 3 participants on a maintenance test delivered one day after mastery. Stevens, Blackhurst, and Slaton reported that maintenance was "often below the 100% correct responding exhibited at the end of training" (p. 158); however, visual inspection of the data revealed that CTD did improve maintenance significantly. It should be noted that in the Stevens, Blackhurst, and Slaton and Stevens and Schuster studies, the CTD procedure dictated that mastered words be added to a maintenance list and continue to be interspersed during practice. Presumably, repeated practice with mastered words likely promotes maintenance over time, and, while this limits the study in that the maintenance probes are atypical and not really comparable to those in other studies, the data suggest that interspersing mastered words every so often is beneficial. Maintenance data from Neef, Iwata, and Page (1980) also support this assertion. They demonstrated that interspersing known words with unknown words during acquisition trials resulted in maintenance of at least 85% of words on a retention test 10 days following mastery for all 3 participants.

Although relatively few studies included maintenance probes, the data may still inform educational practice. Interspersal, CTD, CCC, and self-correction procedures all, to a certain extent, promote maintenance over time compared to traditional procedures. When faced with students who continually fail to retain words, teachers may use any of these specialized approaches and will likely see improved performance. Further, systematically interspersing words from previous weeks into current weekly word lists may improve maintenance on a class-wide scale. Future research on maintenance should address methodological issues including how and when to administer retention tests and may also continue to evaluate the differential effects of instructional strategies on maintenance.

Generalization

Equally important is the extent to which the effects of a spelling strategy will generalize to novel contexts. Fewer studies included generalization measures and those that did represent the variety of ways in which it is possible to measure generalization of academic skill acquisition. While it may be important to know if the effects of an instructional strategy will generalize across settings, teachers and modes of responding, it is critical to know whether they will generalize to a writing context and to related words. With respect to the former, Wirtz et al. (1996) found that self-correction promoted generalization to an oral test for 3 of 6 students. Stevens, Blackhurst, and Slaton (1991) and Stevens and Schuster (1987) reported that the effects of computer assisted instruction (i.e., students practiced words by typing them on a computer) generalized to a written test for 5 students. And, Okyere, Heron, and Goddard (1997) showed that the effects of a self-correction procedure generalized, to a certain extent, to a home setting (all students spelled at least 11 of 15 words correctly per list). Jaspers et al. (2011) found that practicing spelling words facilitated reading the words for 3 participants, but did not test the effect on related words.

However, the main, if not only, reason to learn how to spell is to communicate effectively through writing; thus, the success of any spelling strategy should heavily depend on its ability to promote generalization to a writing context and to related words. Cuvo et al. (1995) found that dense amounts of practice (10 rehearsals per trial) promoted generalization to a sentence writing task for 4 students regardless of how they practiced the words (traditional practice, CCC, or oral practice). Similarly, Okyere, Heron, and Goddard (1997) demonstrated that self-correction promoted generalization to sentence writing and that students spelled an average of 96% of target words correctly. Further, students were able to spell variations of target words at least 73% of the time (e.g., achieve/achievement). However, in both of these studies, the discriminative stimulus controlling the spelling response was similar during acquisition and generalization sessions in that the experimenter dictated the word. Generalization tests only differed in that students had to spell the word concurrently with other words. Perhaps a better test of generalization would represent a more natural writing exercise in which students use target words in the absence of teacher dictation.

The extent to which these studies can inform educational practices is limited given the relatively sparse data available on a wider variety of important generalization measures. For example, using a CTD or CCC procedure may facilitate acquisition and maintenance, but it is unknown if it will produce generalization to students' written work. On the other hand, self-correction appears to promote generalization to writing, but the methodology for assessing generalization could be improved. Future research may answer some of these questions, as well as identifying variables of spelling strategies that will promote generalization to novel words.

Student Preference

Student preference for instructional strategies that compliments measures of the efficacy of a procedure is a final important measure that has not been directly measured. Some spelling studies have employed indirect methods (i.e., surveys and questionnaires) to assess student preference. For example, Nies and Belfiore (2006) included a written survey at the end of their study. Survey items included statements such as, “Learning spelling using the cover copy compare method is better than learning the usual way,” and student were asked if they agreed or not. Indirect methods of assessing preference, such as these surveys, are problematic for individuals with limited language skills (i.e., young children and individuals with disabilities) in that they may be unable to respond verbally or their verbal responses may not be reliable.

Exposing individuals to different interventions or instructional strategies (i.e., contexts) and directly observing which one(s) they repeatedly select may improve the questionable reliability and validity of indirect measures. For example, Hanley, Piazza, Fisher, Contrucci, and Maglieri (1997) described a method that directly measures preference by using a modified concurrent chains arrangement. In this arrangement, individuals are repeatedly exposed to different contexts that are correlated with salient stimuli (i.e., color, verbal cues) and then are asked to select which context they prefer. Procedurally, access to different contexts (the terminal link) is contingent on pressing a micro switch or selecting a color card (i.e., the initial link). By separating out responding in the initial link trials from the contingencies operating in the terminal link trials, researchers are able to isolate preference as a dependent variable. Hanley et al. used this procedure to evaluate preference for two common problem behavior reduction strategies, functional communication training (FCT) and non-contingent reinforcement (NCR). Results suggested that even though both interventions reduced problem behavior to near-zero levels, both individuals preferred FCT. This arrangement has also been used to evaluate the value of choice in preschool children (Tiger, Hanley, & Hernandez, 2006), motivation systems during instruction (Heal & Hanley, 2007), and different teaching strategies (Heal, Hanley, & Layer, 2009).

One potential limiting factor of this procedure is the amount of time it takes to implement, particularly for large participant groups. Layer, Hanley, Heal, and Tiger (2008) extended the concurrent-chains arrangement literature by evaluating the feasibility for use in a group context. In this study, the authors compared conditions in which preference for food items was determined individually through a typical multiple stimulus arrangement (DeLeon & Iwata, 1996) and in a group setting using a concurrent chains arrangement. The primary question was whether or not the delayed and probabilistic nature of receiving a selected food item in the group arrangement would impose variability on children's preference. The data showed that preference hierarchies were stable between assessment types for all children and that the group arrangement did not impose variability on food selection. In fact, the group arrangement decreased selection variability for 11 of the 14 children. Taken together, the results suggest that a group arrangement may be an efficient alternative to conducting individual preference assessments for preferred items. Further, the authors suggest that this arrangement may be appropriate for determining the preferences for more complex contexts in larger groups of children.

In sum, results from the research on improving spelling performance in children are promising. Researchers have identified several procedures that can improve spelling, and facilitate maintenance and generalization. Further, results from indirect preference assessments indicate that students prefer the more "specialized" approaches to traditional ones. Of these specialized approaches, CCC is a promising technology because it combines empirically validated practice and error-correction components. Moreover, CCC is advantageous because it can be implemented in large classrooms as it does not require a large amount of teacher supervision. However, data from CCC studies show that while it improves acquisition (i.e., students are earning As and Bs on their tests), it does not always improve maintenance or generalization to a significant level. Adding a sounding-out component enhances the effects of CCC on acquisition, but whether or not it would also improve maintenance and generalization is unknown. Further, researchers have not directly measured student preference for CCC relative to a traditional teaching approach.

The purpose of the current study is threefold: to evaluate the effects of CCC with respect to acquisition, maintenance, generalization and student preference in a class-wide setting, to extend the findings of Mann, Bushell, and Morris (2011) by comparing the effects of CCC plus sounding out with a traditional approach, and to extend the preference assessment literature by conducting a group preference assessment using a concurrent chains arrangement.

Method

Participants

Five male and five female elementary students who attended a small nonprofit private school participated in the study. Three students were selected for the study at the request of their parents (Nic, Ted, and Ali) and the remaining students were selected due to their regular morning attendance during the school's summer session. An independent tester administered the *Woodcock-Johnson Test of Achievement (Revised)* (1998) to each student within one month of the beginning of the study to determine reading and dictation ages (year, month). Table 1 summarizes key student information.

Materials

The words used in the study were taken from *Programmed Reading, Series 1* (Buchannan & Sullivan Associates, 1978), from which the students were reading at the time of the study. The pool of words, or the word bank, included all of the words except proper names in the reading series beginning with book 1A and ending with book 23. Words in each condition were roughly matched by number of letters. Words used in the first Rehearse and Test condition were, on average, 5.4 letters long and words used in the CCC and second Rehearse and Test condition were, on average, 5.8 letters long.

Setting

Sessions occurred four to five times per week and were generally between 5 and 30 minutes, depending on the session type and condition. All sessions occurred in the morning in the school's lunchroom. The lunchroom measured 15 x 30 feet and contained two large tables and ten chairs. Prior to each session, the two tables were pushed together to make one large table.

Procedure

There were two parts to this study: an efficacy analysis and a preference assessment. The efficacy analysis was conducted first and the preference assessment was run immediately after completion of the former. During the efficacy analysis, the students participated in three general sessions throughout the course of the study: pretests, practice, and three types of test sessions (acquisition, generalization, and maintenance). The first session for all students was a pretest and all pretests occurred on Monday. Pretests were followed by practice sessions on Tuesday, Wednesday, and Thursday, and by an acquisition test on Friday. Generalization tests were always administered the following Monday and maintenance tests were given two weeks following the initial acquisition test. This cycle of pretest, practice, and test was repeated for 12 weeks and for 12 word sets.

Pretest. On Monday of each week, the experimenter and a research assistant administered a pretest so that only words the students did not spell correctly were included in that week's word list. Due to the unique repertoires of each student, it was not possible to administer one pretest to the group. Due to the number of participants, it was also not practical to administer the pretests individually; thus pretests were conducted with groups of two to three students at a time. During pretest sessions, students sat at the table with a piece of lined notebook paper and a pencil. The experimenter began each pretest session with the following instruction: "This is a pretest. I am going to read a word to you and I want you to say it back to me and try to write it as best you can." The experimenter than dictated one word at a time to each student sitting at the table. Because the students were seated close together during the pretest it was possible for them to look to at one another's papers. However, after the first pretest, students were never dictated the same list of words at the same time. Thus, any possible effects from cheating were limited.

The first pretest for all students began with the first word in the word bank and continued until they misspelled ten words. Subsequent pretests began with the word in the word bank immediately following the tenth misspelled word from the preceding pretest. No programmed feedback was delivered for spelling accuracy during pretest sessions. However, students could gauge how well they were spelling words by how quickly their pretest was completed. At the end of each pretest session, the experimenter printed the ten misspelled words from each student onto a piece of construction paper to be used in the practice sessions.

Practice. Students participated in two practice conditions: Rehearse and Test (RT) and Copy, Cover, and Compare plus Sounding Out (CCC). All practice sessions began on the Tuesday following the pretest. The first practice condition for all students was RT.

Rehearse and test. During the RT condition, practice sessions occurred on Tuesday, Wednesday, and Thursday. The students participated in practice sessions as a group, all ten students sitting together at two tables that had been pushed together. At the beginning of each practice session, the experimenter placed a piece of lined notebook paper, a pencil, and a list of the ten words each student had misspelled on the pretest on the table in front of each student, and gave the following instructions:

> "This is a rehearse and test practice session. We are going to practice spelling your words three times today. Please point to the first word on your list and read it to yourself. Then, write the word three times on your sheet of paper. Do this for each word on your list. Make sure you check your spelling with the paper I have given you and make corrections if you make a mistake. You may begin."

The instructions were the same on Thursday and slightly different on Wednesday. On Wednesdays in the RT condition, the experimenter instructed students to practice spelling each word twice instead of three times. This was done to more closely approximate common RT procedures that introduce practice variability throughout the week. No programmed feedback was delivered for spelling accuracy, however students may have been able to produce automatic reinforcement for correct responding in comparing their correct spelling to the model (if "being right" was a reinforcer). Occasionally, the experimenter praised students for staying on-task; however, praise was not delivered systematically and treatment fidelity was not measured for this variable. Students were allowed to return to their classrooms when they finished practicing their words. The total number of rehearsals written for each word list throughout the week was 80 (8 rehearsals for each of 10 words) for this condition. A treatment fidelity analysis of the practice sheets showed that students engaged in the correct number of rehearsals 99% of the time across all Rehearse and Test practice sessions.

The experimenter collected the practice sheets at the end of the session and graded them privately. Scores of 90% or above produced a sticker and a word of praise written on the paper such as "Nice Job!" Errors were corrected and the appropriate spelling was modeled in the margin. The graded practice sheets were returned at the beginning of the subsequent practice session (Thursday practice sheets were returned on Friday before the acquisition test). The experimenter asked the students to look at their errors and the correct spellings, and to be sure not to make them again. In general, the students made very few errors in practice sessions.

Copy, cover, and compare plus sounding out (CCC). During the CCC condition, practice sessions occurred on Wednesday and Thursday. Tuesday was a "No Practice Day" during this condition. This was done to more closely approximate the procedural characteristics of other research on CCC in which the actual number of rehearsals ranged from 2 to 5. To start each practice session, the experimenter placed a piece of lined notebook paper and pencil in front of each student. The students folded the paper in half lengthwise to produce four columns. Next, the experimenter placed the list of the ten words the students had misspelled on the pretest in front of each student and delivered the following instructions:

(a) Read the first word on the list out loud.

(b) Say each sound of the word (i.e., each phoneme).

(c) Copy the word.

(d) Look at the word you have written and read it out loud.

(e) Turn the paper over onto the word list so that the other side of the paper is showing and the word list is covered.

(f) Say the word you just wrote, sound it out, write the word and then read the word you have just written.

(g) Move the paper off the word list and compare the word you have written with the word on the word list.

(h) If they are the same, put a check by the word, unfold the paper and refold the paper so that the columns on the backside of the paper are now showing. Arrange the paper so that one column is covering the word list. Say the word you just wrote, say each sound, write the word and then read the word you have written. Move the paper off the word list and compare the word you have written with the word on the word list. If they are the same, repeat the same steps with the next word on the word list (note that the students engaged in this procedure for one word at a time)

(i) If the word you wrote is not the same as the word on the list, correct the misspelled word and repeat the steps until you have accurately rehearsed the word 3 times with no errors.

As the students gained experience using the CCC strategy, the experimenter instructed less and less. Eventually, all students independently and accurately engaged in practice sessions. No teacher-delivered consequences were programmed for spelling during practice sessions. Again, however, automatic reinforcement could have been produced for “being correct.” No treatment fidelity data were taken on whether or not students sounded out. Anecdotally, all but one student routinely sounded out. Ike occasionally did not sound out and was prompted to do so when the experimenter observed him practicing without sounding out. A treatment fidelity analysis confirmed that students completed the correct number of rehearsals 100% of the time.

Acquisition test. The experimenter administered acquisition tests every Friday in both conditions. All acquisition tests were administered to the whole group, even though every student had a different list of ten words. All students sat at one of two tables with a piece of notebook paper and a pencil in front of them. During tests in the RT condition, the experimenter began each session with the following instruction: “This is your test. I am going to read you a spelling word and I want you to write it on your paper.” The experimenter then read the first student’s word quietly to him, moved to the next student and read his word, and so on until all ten words were dictated to the group of five students. A research assistant engaged in this same procedure with the other group of five students.

During acquisition tests in the CCC condition, the experimenter gave the same instructions and followed the same general procedure. Additionally, the experimenter requested that the students sound out each word quietly like they did in their practice sessions. All but one student (Ike) regularly engaged in some form of sounding out during acquisition tests.

Generalization test. Generalization tests were administered every Monday except for the first Monday of the study. They were always administered before the Monday pretests. Similar to the acquisition tests, generalization tests were administered to the whole group. Students sat at the table with a piece of lined notebook paper, a pencil, and a sheet of construction paper with five picture stimuli in front of them. The five pictures represented a sample of five of the student's spelling words from the previous week's word list. The experimenter selected words for the generalization test based on two criteria. First, the experimenter selected words for which it was relatively easy to find pictures (i.e., "drink" was chosen over "think"). Second, selected words were representative of student performance on the relative acquisition test. That is, the experimenter included both words a student misspelled and words spelled correctly, unless, of course, a student spelled all words correctly. The experimenter performed a treatment fidelity analysis on the number of misspelled words and words spelled correctly included on the generalization tests (see Figure 1) The experimenter obtained all pictures from Google Images™.

Generalization test sessions began with the following instructions: "Here are five pictures that represent five of the words you practiced spelling last week. Please look at the picture and write a good sentence about the picture using the appropriate word." For the first generalization test session, the experimenter provided a model of an appropriate written response. The target word was "digging," and the picture showed a dog digging in the ground with a bone in his with. The model sentence read, "The dog is digging in the ground and burying his bone."

Occasionally, the pictures did not evoke the targeted response; either the students wrote a sentence that did not contain the target spelling word or they did not know what to write at all. In both cases, the experimenter provided the verbal prompt, "Look at this picture and think about the words you practiced last week. Can you think of a word that tells about this picture?" This prompt was successful most of the time, but occasionally, students required additional prompts such as, "What is this person doing?" or "What is another word for that?" Rarely, neither of the prompts evoked the target spelling word. In this case, the experimenter dictated the first part of the word to the student (e.g., if the word was "kingdom," the experimenter said "king…").

Sometimes, students simply named the spelling word in sentence form. For example, when writing a sentence about a picture of a ship sailing in the ocean, they wrote, "It is a ship," or "A ship." In this case, the experimenter requested that the students be more descriptive; to try to think about where the ship was going, what kind of ship it was, etc.

Maintenance test. The experimenter administered maintenance tests on Fridays two weeks following the acquisition test. Thus, maintenance tests did not begin until week 3 of the study. On Friday of week 3 and all subsequent Fridays, the experimenter administered the maintenance test after that week's acquisition test. The procedures for the maintenance tests were identical to those for the acquisition tests.

Preference assessment. To enhance discrimination in the efficacy analysis, all RT sessions were correlated with the color blue and all CCC sessions were correlated with red. Specifically, during RT sessions (pretest, practice, and the 3 tests), the experimenter and research assistant wore blue shirts, covered the table with a blue table cloth, gave the students blue pencils, and printed the word lists and generalization picture stimuli on blue construction paper. During CCC, the same procedures were used, except that everything was red. To further improve discrimination, the experimenter displayed a poster board on the wall of the session room naming the condition and describing the particular spelling strategy for the week (the RT board was blue; the CCC board was red). The experimenter reviewed the board with the students as a group before the start of each session. For example, on Tuesdays in the RT condition, the experimenter pointed to the board and said, “This is a blue week. On Tuesdays in blue weeks, we practice spelling words three times.”

The students also practiced the selection response that was to be the initial-link response in the preference assessment during the efficacy analysis. The efficacy analysis was “forced choice;” that is; students could not choose what type of strategy they wanted to use. However, to practice choosing, the experimenter passed out a colored poker chip (blue during RT and red during CCC) on Mondays and instructed each student to place the chip in a brown paper sack. This was done primarily to improve the efficiency of the initial-link responses in the preference assessment (i.e., so the students clearly understood the procedures), but also to further enhance discrimination between RT and CCC.

A concurrent chains arrangement was used to conduct the group preference assessment immediately following the final session in the efficacy analysis. In this arrangement, two concurrently available response options are associated with identical but independent schedules of reinforcement during initial links; responding in the initial links results in different schedules of reinforcement in the terminal links. Thus, relative response rates in the initial links are a direct measure of preference for options in the terminal links (Hanley et al., 1997; Layer et al., 2008). In this study, the students were given a red and blue poker chip and were asked, in private, to select the spelling strategy they wanted to do by placing the corresponding colored chip into the brown paper sack (the initial link). After every student made their selection, the experimenter called all students into the session room, publicly and randomly selected a chip from the paper sack, and arranged the session room for the appropriate condition (i.e., if a blue chip was selected, the room was arranged for a RT pretest session; and everything was blue). In this way, responding in the initial link resulted in exposure to one of the two spelling strategies (the terminal link).

Due to time constraints (the summer session was ending and several student participants were matriculating on to public school), the week-long exposure format used in the efficacy analysis was shortened to three days in the preference assessment. The experimenter ran pretests on day 1, practice on day 2, and acquisition tests on day 3. Further, no generalization or maintenance tests were given during the preference assessment. A student met completion criterion in the preference assessment when they selected a color three consecutive times. Note that a student could possibly meet the completion criterion without ever having their choice of strategy selected. That is, a student could choose RT on three consecutive occasions, but a CCC chip could be pulled from the bag each time (assuming that at least one student selected CCC). To be clear, the student only needed to choose a strategy three times in a row, not experience it.

Response Measures and Data Analysis

The primary dependent variable for the efficacy measure was the percent of words spelled correctly on the acquisition, generalization, and maintenance tests. A word was scored correct if it matched the spelling in the book. Letter reversals were counted as correct throughout the study. Individual effect sizes (Cohen's *d*) were also calculated to measure the magnitude of the difference between performance on CCC and RT posttests. The primary dependent variable for the preference measure was selection; defined as placing a small colored chip in a bag.

Secondary dependent variables included the number of letters sequenced correctly on misspelled pretest words, the number of correctly spelled rehearsals during practice, and the duration of practice sessions. To measure the number of letters sequenced correctly on misspelled pretest words, the experimenter counted the number of letters that were in the correct, relative order. Extra and missing letters were subtracted from the total. However, missing letters, particularly those missing from the beginning of the word, did not result in all subsequent letters to be counted as out of sequence. For example, a common error was to omit the first vowel sound in a word like *mitten* and spell it as *mtten.* If only letters that were correctly sequenced in absolute order were counted, this spelling would score only 1 point out of a possible 6. In this study, this word only loses 1 point for omitting the -*i* and scores 5 points. Another common example was students spelling *dog* as *dig;* this spelling would have received 2 out of 3 points. Students also would have received 2 out of 3 points if they spelled *dog* as *dogg* (3 points for 3 correctly sequenced letters, minus 1 point for an extra letter). Less common examples include *crack* as *ckark*. The student would have received points for the *c, a,* and *k* for a total of 3 points out of a possible 5.

Experimental Design

A reversal design was used to compare the effects of a rehearse and test strategy with a copy, cover, and compare strategy on spelling acquisition, generalization, and maintenance. The order of conditions was four weeks of RT, four weeks of CCC, and then four weeks of RT. A concurrent chains arrangement was used to assess student preference for the CCC and RT strategies.

Interobserver Agreement

A second observer independently scored at least 67% of students' written posttest responses across RT and CCC conditions and independently recorded 80% of students' selections during the preference assessment. Interobserver agreement was calculated as the number of agreements divided by number of agreements and disagreements multiplied by 100. Interobserver agreement for both measures was 100%.

Results

Generalization test words.

Figure 1 displays the number of words included on generalization tests throughout the study that were correctly (black bars) and incorrectly (white bars) spelled on corresponding acquisition tests. For all participants, more words correctly spelled on acquisition tests were tested again for generalization to a writing context; however, this varied by participant and was influenced by how many errors were made on acquisition tests. Nic and Ted made the most errors on acquisition tests and were thus more likely to have more misspelled words included in their generalization tests. By contrast, Cat and Ali made the fewest errors on acquisition posttest, so they were likely to have fewer misspelled words included in their generalization posttest.

Efficacy.

Data for individual students are shown in rows in Figures 2 and 3. The data for acquisition, generalization, and maintenance are displayed from left to right in columns. CCC increased spelling acquisition relative to RT for Ann, Dan, Nic, Rai, Ted, and Ike. Jon, Cat, and Ali's spelling performance remained high in both conditions, suggesting the presence of a ceiling effect. Thus, experimental control was demonstrated over the acquisition of spelling in six out of ten students. CCC also improved the generalization of spelling for Ann, Dan, Nic, and Rai and maintenance for Ann, Dan, Nic, and Jon. In sum, CCC improved spelling with respect to all measures for three students (Ann, Dan, and Nic), with respect to acquisition and maintenance for one student (Rai), acquisition only for two students (Ted and Ike), and maintenance only for one student (Jon). Equally important, CCC never systematically decreased performance for any measure relative to RT. That is, for all measures, all students' performance was identical or better during CCC when compared with RT. Effect sizes were calculated for each student and are indicated in the right columns of Figures 1 and 2. From a statistical perspective, effect sizes for all but two students (Meg and Cat) were above .8, suggesting a "strong" effect in favor of CCC (Cohen, 1992).

Figure 4 displays class-wide, mean percent correct on acquisition, generalization, and maintenance tests. As a group, students spelled better on all tests during CCC compared to RT. These effects were slightly more robust for the acquisition and maintenance measure than for the generalization measure. Effect sizes were also calculated for the class-wide data and are displayed on each panel in Figure 4. The calculated effect size of 3.72, 3.38, and 4.86 for acquisition, generalization, and maintenance, respectively, also suggest a "strong" effect of CCC for all measures.

Figure 5 displays the mean percent correct across each measure for each student for each set of words (i.e., the mean composite score). CCC improved total, class-wide spelling performance relative to RT.

Preference.

Figure 6 displays preference data for each student. Six students (Ann, Dan, Ted, Jon, Meg, and Cat) demonstrated a clear preference for CCC, choosing it first and never sampling

RT. Three students (Rai, Ike, and Ali) sampled both strategies, but ultimately showed a preference for CCC. Only one student, Nic, preferred RT over CCC.

Pretest error analysis.

Figure 7 shows the degree to which students were misspelling words on the pretests across each condition. Individual effect sizes are also labeled next to each student's name. The black bars indicate the average percentage of correctly sequenced letters in misspelled pretest words and the white bars indicate what performance would have looked like if the participants had made only one error per word. The difference in height represents the degree of spelling errors. A large difference between the black and white bars suggests that students were making several errors per word and a small difference between bars suggests that the students were only slightly misspelling words during the pretest. Ann, Nic, Jon, and Ali made more letter sequencing errors in their misspelled pretest words than the other six students; however, they nevertheless correctly sequenced at least 60% of letters per word. Ann, Nic, and Ali were all students with strong effect sizes and relatively large degrees of spelling error, perhaps suggesting that the two variables are related.

Figure 8 show these data by Word Set (top panel) and aggregated by condition (bottom panel). The word set analysis indicates that, with the exception of Set 9, the degree of spelling error is relatively stable across sets. The larger degree of difference for Set 9 can be attributed primarily to Jon, Ann, and Nic's performance on set 9 pretest words. The bottom panel of Figure 8 shows that, on average, and of the 10 words misspelled on their pretests, students sequenced 72% of letters correctly in the first RT condition, 74% correctly in the CCC condition and 74% in the second RT condition. If students had only made one error per word, they would have, on average, sequenced 81%, 82 %, and 82% of letters correctly in the first RT, CCC, and second RT conditions, respectively. These data show minimal differences in the degree of misspellings on pretests across conditions, suggesting that the words in each condition were roughly the same level of difficulty.

Practice error analysis.

Figure 9 displays the mean percent of misspelled rehearsals during practice for each student across each condition. Note that the total number of rehearsals in the RT and CCC conditions was different, 80 and 50, respectively. Also note the scale only goes to 10% suggesting that overall, students made very few errors during spelling practice. Dan, Ann, Ali, Cat, Meg, Ted, and Rai made errors on fewer than 5% of rehearsals in each condition. Jon, Ike, and Nic made errors on more than 5% of rehearsals in at least one condition. Further, these data show that 7 students (Ann, Ali, Cat, Meg, Ted, Jon, and Nic) made more errors in RT practice sessions than in CCC practice sessions (shown by a relatively higher black or gray bar compared to the white bar). Only Dan, Rai, and Ike made more errors during CCC practice sessions than during RT sessions.

Duration of practice.

Figure 10 shows that mean duration of practice sessions for all students for each word set across conditions. Overall, students took longer to complete CCC practice sessions than RT practice sessions. There is a clear downward trend in the first RT condition and slight downward trend in the CCC and second RT condition, suggesting that students were becoming proficient with the practice procedures over time. Individual data (not shown) revealed that all participants, with the exception of the very first set of words, took longer to practice using CCC than RT.

Discussion

Efficacy.

CCC was more effective than RT in improving acquisition in six students and generalization and maintenance in four students. Equally as important, CCC never systematically decreased performance relative to RT. Effect-sizes calculated for each participant revealed that 8 out of 10 were larger than .8, suggesting a strong effect of CCC. This measure is important because even though experimental control over spelling was not demonstrated for Jon and Ali, the effect size suggests that they performed better using CCC. Class-wide, CCC was more effective than RT for all 3 measures, but particularly for acquisition and generalization. The composite data show that when the mean of all posttest scores for a word set was calculated, CCC was again superior. CCC was also preferred by 9 out of 10 students, 6 of whom never selected RT during the preference assessment.

The pretest and practice error analyses revealed important findings as well. First, as a group, the students were not misspelling words on the pretest by a large degree. Most of the time, students were able to sequence 3 to 4 letters correctly in 5 to 6 letter words. Four students made relatively more errors in pretests than their peers: Nic, Ann, Ali, and Jon. The effect sizes for Nic, Ann, and Ali were large at 2.28, 2.01, and 1.20, respectively. At least for these three, a relationship appears to exist between degree of error during pretest and posttest performance. Similar to Nic, Ann, and Ali, participants with learning or developmental disabilities (Nic and Ali) or young, naïve spellers (Ann) who may be unlikely to approximate correct spellings may be more likely to benefit from a specialized spelling strategy like CCC. The practice-error analysis also contributed an important finding. The individual data showed that students did not make many errors during practice throughout the study, but when they did, they were more likely to make errors during RT practice than CCC practice.

This study adds to the growing body of research demonstrating the benefits of CCC on spelling performance. CCC is likely more effective than traditional rehearse and test spelling procedures for several reasons. As many other researchers have suggested (e.g., Skinner, McLaughlin, & Logan, 1997) CCC includes several components of effective instruction missing from RT including the availability of immediate and frequent reinforcement, and ample response and error correction opportunities. Students also detect their own errors and do so immediately for each word. In addition, the “copy first” component of CCC reduces the likelihood of errors on the first (and perhaps subsequent) rehearsal, aligning it with other errorless learning procedures like the CTD procedure described in the introduction.

This study also extends previous research by comparing the effects of CCC and RT on generalization and maintenance. With the exception of Meg, Cat, and Ali, students' average scores on CCC generalization and maintenance tests were higher than in RT. However, for all students, performance deteriorated on generalization and maintenance tests relative to acquisition. Similar effects were found in the few studies that evaluated at least one of the measures and perhaps are related to the time between acquisition and later generalization and maintenance posttests. Both Nies and Belfiore (2006) and Cuvo et al. (1990) reported high retention rates (95% and 94%, respectively) on maintenance tests that were delivered within 3 days and 1 week, respectively. By contrast, Jaspers et al. (2011) found that only 75% of words maintained on posttests delayed between two and four weeks and Erion et al. (2009) reported that performance on retention tests two weeks following acquisition deteriorated significantly to below 80%. Visual inspection of the data from this study (Figure 4) show that students in the current study maintained fewer words than participants who were administered maintenance tests closer in time to acquisition tests, but more words than participants with comparable delayed posttests.

With respect to generalization, the only other study to evaluate the effects of CCC on generalization to a writing context was Cuvo et al. (1990). They found that a CCC procedure that included dense amounts of practice resulted in moderate generalization, between 63% and 64% on a sentence-writing task administered immediately after acquisition. In the current study, the generalization test was delayed by three days and produced roughly 80% generalization, significantly higher performance than in Cuvo et al. Interestingly, when the methods were comparable, the CCC procedure used in this study resulted in improved generalization and maintenance. The difference may be at least partially attributed to the sounding-out component.

Mann et al. (2010) found that CCC plus sounding out improved spelling acquisition compared to CCC alone for all 5 participants. Although conclusions may not be drawn about the relative contribution of the sounding out component in this study because it was only evaluated in conjunction with CCC, it likely contributed to the overall efficacy of the CCC procedure. Theoretically, sounding out may enhance the effects of CCC by strengthening the stimulus control of the auditory stimulus over spelling. The sounding out component required students to emit auditory stimuli before spelling the word. In this way, they were dictating the sounds of the words to themselves and writing what they heard themselves say, thereby strengthening the stimulus control of the auditory stimulus over spelling. By contrast, in RT, students practiced their words by copying a model. Thus, during practice, their spelling responses were likely under textual control rather auditory control. In both conditions the spelling tests were dictation tests in which the discriminative stimulus was auditory. If the sounding out procedure strengthened stimulus control of the auditory stimulus over spelling, then the auditory stimulus presented during posttests may have been more effective for words practiced in this way.

Sounding out may also be a type of self-instruction, typically defined as a verbal response that directs other responses of the speaker (Vintere, Hemmes, Brown, & Poulson, 2004) and conceptualized as a stimulus that functions to mediate behavior change (Guevremont, Osnes, & Baer, 1988). Self-instruction is beneficial for learning academic skills, including spelling, because it recruits and maintains attention, which enhances the salience of the cues to be discriminated; slows performance to improve accuracy; and turns the student into an active learner (Duarte & Baer, 1994). For spelling, sounding out recruits and maintains attention by requiring the student to read each word and say each sound. In turn, saying each sound enhances the salience of the letter sounds to which the student must respond. Further, sounding out slows student responding by increasing the response requirement associated with each written rehearsal. Both of these components require students to be active rather than passive learners and generally increase the overall amount of responding with respect to each word. Note that self-instruction is different than self-management. CCC is often referred to as a self-managed procedure in that it is student directed and does not require teacher intervention, but it is in and of itself, not a form of self-instruction.

Preference.

The students in this study overwhelmingly preferred CCC to RT. Preference was likely influenced by the density of reinforcement and response effort associated with each condition. If in comparing a written word to a model automatic reinforcement was produced for "being correct," then the schedule of reinforcement associated with CCC was relatively dense. The procedure required that students compare their word to the written model for each and every written rehearsal, resulting in at least 20 or 30 opportunities for reinforcement each practice session. By contrast, the RT procedure did not require that students compare their word to the model, but it was possible for them to do so. There are no data to show whether or not students were comparing their spelling to a model during RT practice, but the speed at which they completed these sessions suggests that they were not, at least not to the same degree as in CCC.

Response effort is another variable that can influence preference. In this study, both the number of errors committed during practice and duration of practice may be indicators of response effort. Errors may have increased effort by increasing the overall amount of responding required for each practice session (e.g., erasing, correcting, engaging in extra word rehearsals, etc.). Six out of seven participants that made fewer errors during CCC practice sessions preferred CCC, and five of these six never chose RT during the assessment. Interestingly, the two participants that committed more errors during CCC practice initially chose RT, but ultimately showed a preference for CCC. Nic made significantly fewer errors in CCC but showed a strong preference for RT.

For all participants, the CCC practice sessions took more time than the RT practice sessions (with the exception of the very first RT session), suggesting that, at least for most of these participants, some quality of CCC (perhaps the density of reinforcement associated with it) was more valuable than the effort associated with longer practice. In contrast, Nic's preference assessment suggests that response effort was the salient feature for him. He was the only student who preferred RT, despite being one of three students for whom CCC increased all measures of spelling (and arguably the student for whom it was most beneficial). It is likely that, for him, the CCC strategy was considerably more effortful due to his poor reading skills (see his WJ-R scores in Table 1). Sounding out requires a basic letter-sound correspondence repertoire, which was not fully present for Nic. Without an intact letter-sound correspondence repertoire, sounding-out may be relatively more effortful and perhaps aversive. For Nic, the effort associated with sounding-out was likely more salient than possible reinforcers for correct responding. His verbal behavior with respect to the CCC condition supports this hypothesis. During the preference assessment, when the experimenter asked why he chose RT, he responded, "Because I don't have to sound out…I hate sounding out."

The results from the preference assessment support Layer et als.' (2008) suggestion that a group-oriented arrangement can be a reliable and efficient alternative to conducting individual preference assessments for instructional contexts. Arranging for and measuring initial-link responding only took minutes at the beginning of each week. Exposure to the terminal-link was more time intensive; however, these sessions were not superfluous; they were still providing valuable instruction to students. Seven of the nine students who preferred CCC were more effective spellers using the CCC approach and the remaining two were equally effective in the different conditions. For these two, and others like them, allowing them to choose how they practice their words from time to time may be beneficial.

Unfortunately, Nic chose an instructional strategy that was ineffective. For Nic, and others like him, future research may identify the conditions under which individuals do not choose effective contexts and evaluate interventions to shift preference so they do.

There were several limitations to the current study. Perhaps the most significant limitation is the lack of treatment fidelity measures for several important variables. First, the experimenter did not conduct a fidelity measure on the sounding-out component of CCC. Anecdotally, most students engaged in some form of sounding out during CCC practice and test sessions. However, it is unknown how often they did not and whether or not they were making errors while sounding out. This study was designed to compare CCC plus the SO component in a class-wide setting and it would have been difficult for the experimenter to measure this. Nonetheless, a fidelity measure of SO is important, particularly because it was a component of the CCC strategy that is believed to result in improved performance. Future research should continue to evaluate the relative contributions of sounding-out on all measure of spelling *and* as a strategy in its own right.

Second, no treatment fidelity measures were conducted for the frequency and type of prompting that occurred during generalization tests. Because the generalization tests were designed to improve upon previous procedures by including a more naturally occurring discriminative stimulus for writing, it is important to know how often the experimenter dictated part of the spelling word as a prompt. Anecdotally, this did not occur often, but data were not taken. Relatedly, a second limitation is the lack of interobserver agreement (IOA) on fidelity and error-analysis measures. Fortunately, the pretest and practice error analyses were conducted using permanent products from the study. In the future, IOA should be conducted for these measures by having an independent observer record the number of letters correctly sequenced on misspelled pretest words, the number of correctly spelled rehearsals during practice, and the number of words included in generalization tests that were spelled correctly and incorrectly on corresponding acquisition tests. Unfortunately, no IOA is available for duration of spelling practice.

A third limitation is that the experimenter did not systematically deliver feedback during practice sessions. It is possible that this differentially influenced responding, but not likely. As described in the procedures, only general praise statements were delivered for staying on-task and no feedback was delivered for accuracy. Further, the contingencies associated with each practice condition were likely far more powerful than any feedback delivered by the experimenter. Future research should include procedures to systematically deliver feedback during all sessions to ensure the amount is equal across conditions.

In addition to the few listed above, the results from these studies have several implications and may inform several areas of future research. Copy, cover, compare is clearly an effective procedure for improving spelling in children with and without developmental and learning disabilities. This study showed that a CCC procedure plus sounding out is also more effective than a traditional rehearse and test strategy. However, the effects were more robust for acquisition than for generalization and maintenance. Future studies should continue to evaluate modifications to the CCC procedure that may facilitate generalization and maintenance. To this end, research may continue to evaluate the effects of sounding out alone to see if the procedure might facilitate maintenance and generalization. The data are limited, but adding an interspersal component by which mastered words from previous weeks are included on a rotating basis may also be a strategy to facilitate maintenance.

Future researchers should also compare the CCC plus sounding out procedure to more thorough traditional procedures that involve more than daily rehearsal. One possible comparison would be to a traditional approach in which students engage in a variety of exercises throughout the week including using words in sentences, alphabetizing words, reading stories with the spelling words, etc. This variation of traditional spelling was chosen because it was a common procedure in the literature, but it certainly falls on the “less thorough” end of the traditional spectrum.

Researchers may also continue to evaluate the relative contribution of sounding out. Mann et al. (2010) showed that adding a sounding out component improved the effects of CCC for all 5 participants. It is reasonable to believe that the sounding out component was helpful for at least some of the participants in this study. Sounding-out may reduce the amount of rehearsals needed to promote spelling acquisition; thus, future researchers could conduct a parametric analysis of CCC+SO with varying numbers of rehearsals. Researchers may also want to evaluate if SO would improve traditional rehearse and test approaches. If so, perhaps teachers would be more willing to add a component to their current spelling procedures than to replace them altogether.

Finally, future research should begin to evaluate assessment tools for identifying those students who are likely going to struggle with learning how to spell words, and those for whom specialized approaches would be beneficial. Certainly, not all students will need to use a specialized strategy like CCC and will acquire spelling with little instruction. For these students, exposing them to a variety of instructional strategies and allowing them to choose how they would like to learn may be ideal. For the others, particularly those with developmental or learning disabilities, identifying spelling deficits early is critical. One possible warning sign is the degree of spelling errors students make on pretests. In the current study, three of the four students that showed a large degree of error on pretests also performed much better on spelling tests in the CCC condition. Once at-risk or poor spellers are identified, the CCC procedure with the added sounding out component may be an attractive alternative to traditional rehearse and test strategies for promoting the acquisition, generalization, and maintenance of spelling.

References

Alber, S.R., & Walshe, S.E. (2004). When to self-correct spelling words: A systematic replication. *Journal of Behavioral Education, 13,* 51-66.

Axelrod, S., Kramer, A., Appleton, E., Rockett, T., & Hamlet, C.C. (1984). An analysis of relevance of topographical similarity on positive practice of spelling errors. *Child & Family Behavior Therapy,* 6, 19-31.

Barbetta, P.M., Heron, T.E., & Heward, W.L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis*, 26, 111-119.

Barbetta, P.M., Heward, W.L., Bradley, D.M., & Miller, A.D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis*, 27, 177-178.

Bennett, K.., & Cavanaugh, R.A. (1998). Effects of immediate self-correction, delayed self-correction, and no correction on the acquisition and maintenance of multiplication facts by a fourth-grade student with learning disabilities. *Journal of Applied Behavior Analysis,* 31, 303-306.

Bryant, N.D., Drabin, I.R., & Gettinger, M. (1981). Effects of varying unit size on spelling achievement in learning disabled children. *Journal of Learning Disabilities,* 14, 200-203.

Buchannan & Sullivan Associates (1978). *Programmed Reading, Series 1.* Phoenix Learning Resources: New York.

Carnine, D.W. (1980). Phonic versus whole-word correction procedures following phonic instruction. *Education and Treatment of Children,* 3, 323-330.

Cates, G.L., Dunne, M., Erkfritz, K.N., Kivisto, A., Lee, N., & Wierzbicki, J. (2007). Differential effects of two spelling procedures on acquisition, maintenance, and adaption to reading. *Journal of Behavioral Education,* 16, 70-81.

Cates, G.L., Skinner, C.H., Watson T.S., Meadows, T.J., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision-making: An evaluation of interspersing procedures. *School Psychology Review, 32,* 601-616.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.),* Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Coleman-Martin, M.B., & Heller, K.W. (2004). Using a modified constant prompt-delay procedure to teach spelling to students with physical disabilities. *Journal of Applied Behavior Analysis, 37,* 469-480.

Cuvo, A.J., Ashley, K.M., Marso, K.J., Zhang, B.L., & Fry, T.A. (1995). Effect of response practice variables on learning spelling and sight-vocabulary. *Journal of Applied Behavior Analysis, 28,* 155-173.

Dagdag, R., McLaughlin, T.F., & Weber, K.P. (2002). The use and evaluation of a sound out or error only sound out procedure on the spelling performance of a third grade student. *International Journal of Special Education, 17,* 92-97.

DeLeon, I.G., & Iwata, B.A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforce preferences. *Journal of Applied Behavior Analysis,* 29, 519-532.

Duarte, A.M.M., & Baer, D.M. (1994). The effects of self-instruction on preschool children's sorting of generalized in-common tasks. *Journal of Experimental Child Psychology,* 57, 1-25.

Edwards, L., Salant, V., Howard, V.F., Brougher, J., & McLaughlin, T.R. (1995). Effectiveness of self-management on attentional behavior and reading comprehension for children with attention deficit disorder. *Child and Family Therapy,* 17, 1-17.

Erion, J., Davenport, C., Rodax, N. Scholl, B., & Hardy, J. (2009). Cover-copy-compare and spelling: One versus three repetitions. *Journal of Behavioral Education,* 18, 319-330.

Farrell, A., & McDougall, D. (2008). Self-monitoring of pace to improve math fluency of high school students with disabilities. *Behavior Analysis in Practice,* 1, 26-35.

Foxx, R.M., & Jones, J.R. (1978). A remediation program for increasing the spelling achievement of elementary and junior high school students. *Behavior Modification,* 2, 211-230.

Gettinger, M. (1985). Effects of teacher-directed versus student-directed instruction and cues versus no cues for improving spelling performance. *Journal of Applied Behavior Analysis, 18,* 167-171.

Graham, S. (2000). Should the natural learning approach replace spelling instruction? *Journal of Educational Psychology,* 92, 235-247.

Guevremont, D. C., Osnes, P. G., & Stokes, T. F. (1988). The functional role of preschoolers' verbalizations in the generalization of self-instructional training. *Journal of Applied Behavior Analysis, 21,* 45-55.

Hanley, G.P., Piazza, C.C., Fisher, W.W., Contrucci, S.A., & Maglieri, K.A. (1997) Evaluation of client preference for function-based treatment packages. *Journal of Applied Behavior Analysis,* 30, 459-473.

Heal, N.A., Hanley, G.P., & Layer, S. (2009). An evaluation of the relative efficacy of and children's preferences for teaching strategies that differ in amount of teacher directedness. *Journal of Applied Behavior Analysis,* 42, 123-143.

Heal, N.A., & Hanley, G.P. (2007). Evaluating prechool children's preferences for motivational systems during instruction. *Journal of Applied Behavior Analysis,* 41, 249-261.

Heron, T. E., Okyere, B. A., & Miller, A. D. (1991). A taxonomy of approaches to teach spelling. *Journal of Behavioral Education, 1,* 117-130.

Hubbert, E. R., Weber, K. P., & McLaughlin, T. F. (2000). A comparison of copy, cover, and compare and a traditional spelling intervention for an adolescent with a conduct disorder. *Child & Family Behavior Therapy, 22,* 55-68.

Jaspers, K.E., Williams, R.L., Skinner, C.H., Cihak, D., McCallum, R.S., & Ciancio, D.J. (2011). How and to what extent to two cover, copy and compare spelling interventions contribute to spelling, word recognition, and vocabulary development? *Journal of Behavioral Education,*

Koegel, L.K., & Koegel, R.L. (1986). The effects of interspersed maintenance tasks on academic performance in a sever childhood stroke victim. *Journal of Applied Behavior Analysis,* 19, 425-430.

Larsen, D., & McLaughlin, T. F. (1997). The effects of cover, copy, and compare on spelling accuracy of a preschool child. *Reading Improvement, 34,* 189-192.

Layer, S.A., Hanley, G.P., Heal, N.A., & Tiger, J.H. (2008). Determining individual preschoolers preferences in a group arrangement. *Journal of Applied Behavior Analysis,* 41, 25-37.

Mann, T.B., Bushell, D., & Morris, E.K. (2010). Use of sounding out to improve spelling in young children. *Journal of Applied Behavior Analysis,* 43, 89-93.

McGuffin, M.E., Martz, S.A., & Herren, T.E. (1997). The effects of self-correction versus traditional spelling on the spelling performance and maintenance of third grade students.

McGuigan, C. A. (1975). The Add-A-Word spelling program. (Working paper No. 53). *Experimental Education Unit,* Seattle, WA: University of Washington.

McNeish, J., Heron, T., & Okyere, B. (1992). Effects of self-correction on the spelling performance of junior high students with learning disabilities. *Journal of Behavior Education,* 2, 17-37.

Morton, W.L., Heward, W.L., & Alber, S.R. (1998). When to self-correct?: A comparison of two procedures on spelling performance. *Journal of Behavioral Education, 8,*321-335.

Murphy, J. F., Hern, C. L., Williams, R. L. & McLaughlin, T. F. (1990). The effects of the copy, cover, compare approach in increasing spelling accuracy with learning disabled students. *Contemporary Educational Psychology, 15,* 378-386.

Neef, N.A., Iwata, B.A., & Page, T.J. (1980). The effects of interspersal training versus high-density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis, 13,* 153-158.

Nies, K. A., & Belfiore, P. J. (2006). Enhancing spelling performance in students with learning disabilities. *Journal of Behavioral Education, 15,* 163-170.

Okyere, B. A., Heron, T. E., & Goddard, Y. (1997). Effects of self-correction on the acquisition, maintenance, and generalization of the written spelling of elementary school children. *Journal of Behavioral Education, 7,* 51-69.

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.

Rafferty, L.A., Arroyo, J., Ginnane, S., & Wilczynski, K. (2011). Self-monitoring during spelling practice: effects on spelling accuracy and on-task behavior of three students diagnosed with attention deficit hyperactivity disorder. *Behavior Analysis in Practice,* 4, 37-45.

Rodgers, T.A., & Iwata, B.A. (1991). An anlaysis of error-correction procedures during discrimination training. *Journal of Applied Behavior Analysis,* 24, 775-781.

Schermerhorn, P.K., & McLaughlin, T.F. (1997). Effects of the add-a-word spelling program on test accuracy, grades, and retention of spelling words with fifth and sixth grade regular education students. *Child and Family Behavior Therapy,* 19, 23-35.

Scott, C. M. (2000). Principles and methods of spelling instruction: Applications for poor spellers. *Topics in Language Disorders, 20,* 66-82.

Skinner, C. H., McLaughlin, T. F., & Logan, P. (1997). Cover, copy, and compare: A self-managed academic intervention effective across skills, students, and settings. *Journal of Behavioral Education, 7,* 295-306.

Skinner, C.H., Ford, J.M., & Yunker, B.D. (1991). A comparison of instructional response requirements on the multiplication performance of behaviorally disordered students. *Behavior Disorders, 17*, 65-65.

Stevens, K.B., Blackhurst, A.E., & Slaton, D.B. (1991). Teaching memorized spelling with a microcomputer: Time delay and computer-assisted instruction. *Journal of Applied Behavior Analysis, 24,* 153-160.

Stevens, K.B., & Schuster, J.W. (1987). Effects of a constant time delay procedure on the written spelling performance of a learning disabled student. *Learning Disability Quarterly,* 10, 9-16.

Tiger, J.H., Hanley, G.P., & Hernandez, E. (2006). An evaluation of the value of choosing with preschool children. *Journal of Applied Behavior Analysis, 39,* 1-16.

Van Houten, R., & Van Houten, J. (1991). The effects of breaking new spelling words into small segments on the spelling performance of student with learning disabilities. *Journal of Behavioral Education, 1,* 399-411.

Vargas, A.U., Grskovic, J.A., Belfiore, P.J., & Halbert-Ayala, J. (1997). Improving migrant students' spelling of English and Spanish words with error correction. *Journal of Behavioral Education,* 7, 12-23.

Viel-Ruma, K., Houchins, D., & Fredrick, L. (2007). Error self-correction and spelling: improving the spelling accuracy of secondary students with disabilities in written expression. *Journal of Behavioral Education,* 16, 291-301.

Vintere, P., Hemmes, N. S., Brown, B. L., & Poulson, C. L. (2004). Gross-motor skill acquisition by preschool dance students under self-instruction procedures. *Journal of Applied Behavior Analysis, 37,* 305-322.

Wirtz, C.L., Gardner, III. R., Weber, R., & Bullara, D. (1996). Using self-correction to improve spelling performance of low achieving third graders. *Remedial and Special Education, 17,* 48-58.

Worsdell, A.S., Iwata, B.A., Dozier, C.L., Johnson, A.D., Neidert, P.L., & Thomason, J.L. (2005). Analysis of response repetition as an error-correction strategy during sight-word reading. *Journal of Applied Behavior Analysis, 38,* 511-527.

	Developmental Category	Chronological Age	Dictation Age (WJ-R)	Reading Age (WJ-R)
Cat	Typical	6,3	7,7	8,11
Ann	Typical	6,6	7,4	7,11
Rai	Typical	6,10	7,5	8,5
Jon	Typical	7,6	7,10	8,5
Dan	Typical	7,9	7,2	8,2
Meg	Typical	8,5	9,3	14,10
Nic	Moderate Language Delay	10,3	7,4	7,5
Ike	Typical	10,3	9,0	10,10
Ted	Dyslexia	10,9	8,3	8,11
Ali	Autism, mild mental retardation	11,4	7,7	8,11

Table 1. Participant characteristics.

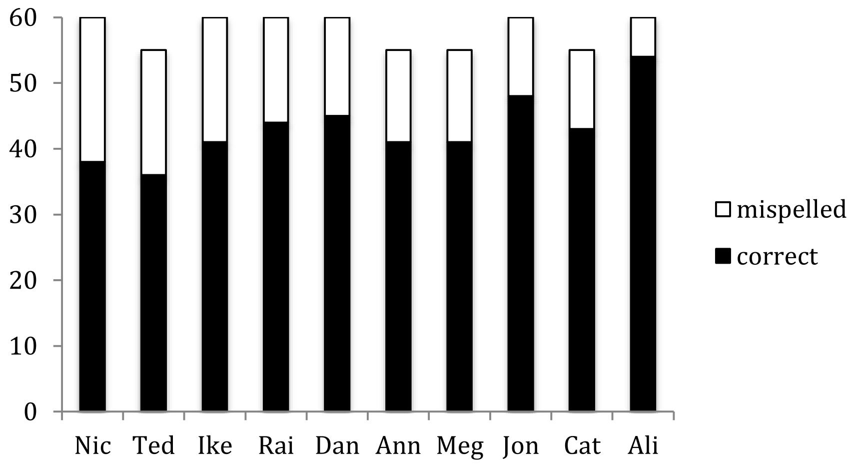

Figure 1. Number of words included on generalization tests that were spelled correctly (black bars) and incorrectly (white bars) on acquisition tests.

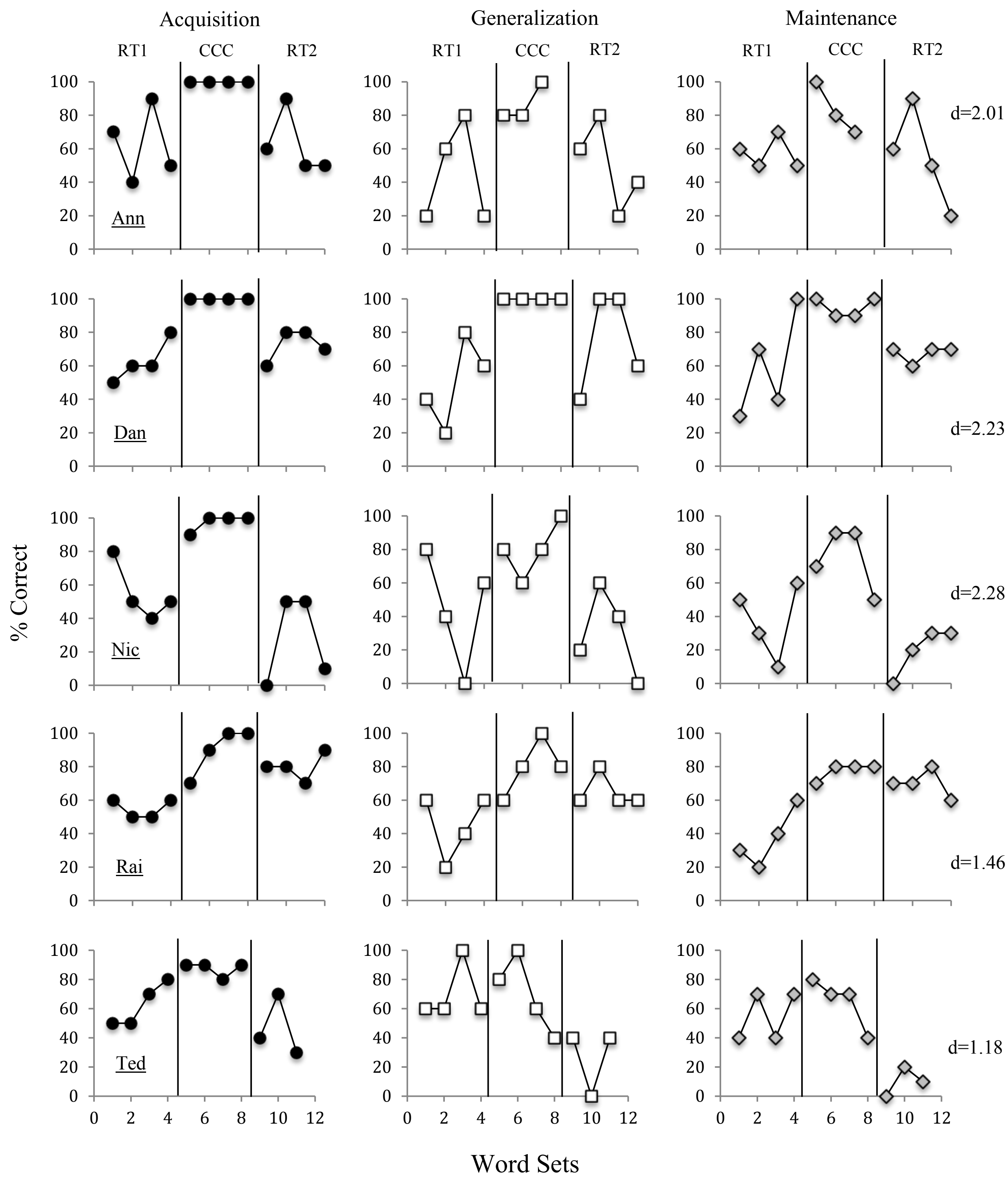

Figure 2. Percent correct on acquisition tests (left column), generalization tests (middle column), and maintenance tests (right column) for five students. Effect sizes for each participant (Cohen's *d*) are shown in the far right column.

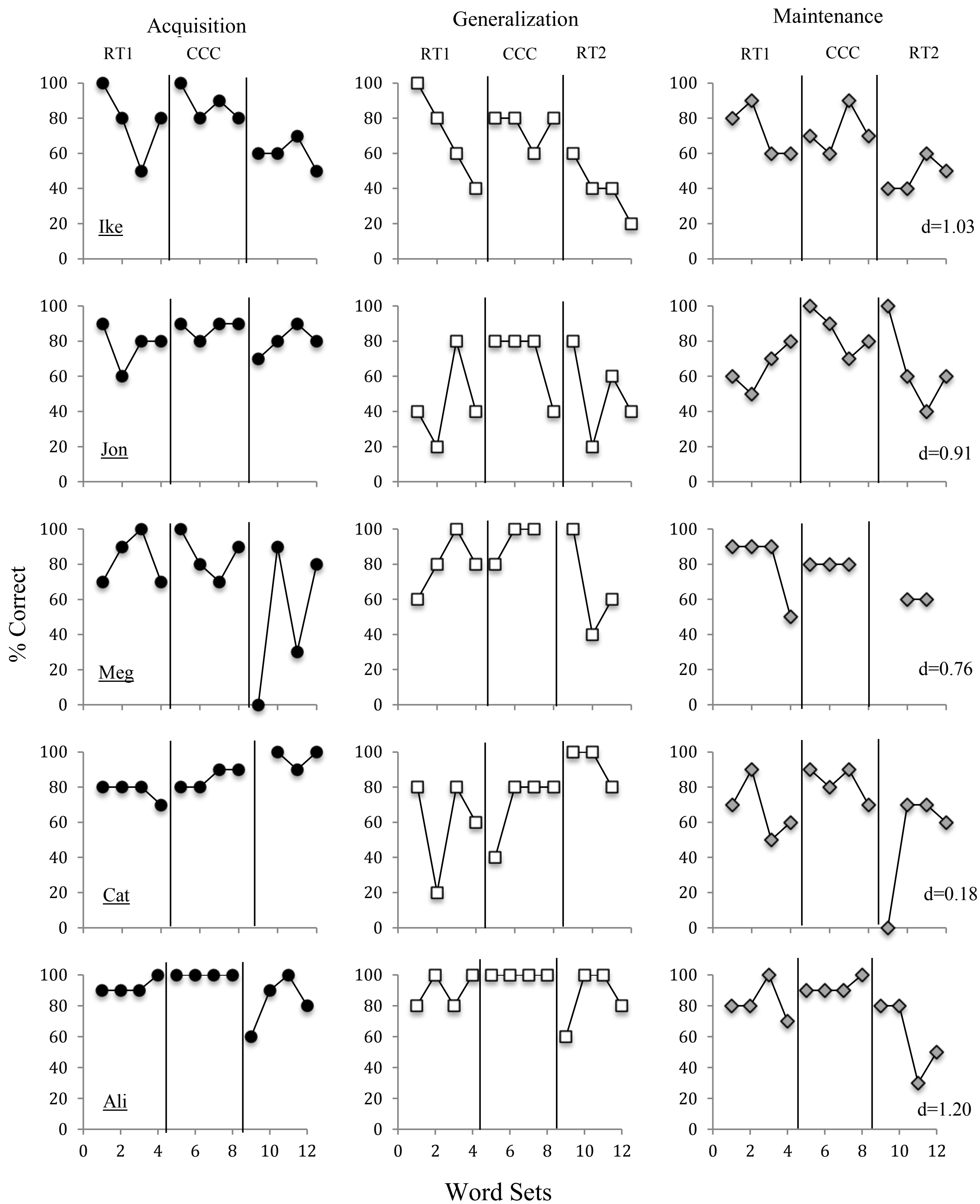

Figure 3. Percent correct on acquisition tests (left column), generalization tests (middle column), and maintenance tests (right column) for five students. Effect sizes for each participant (Cohen's *d*) are shown in the far right column.

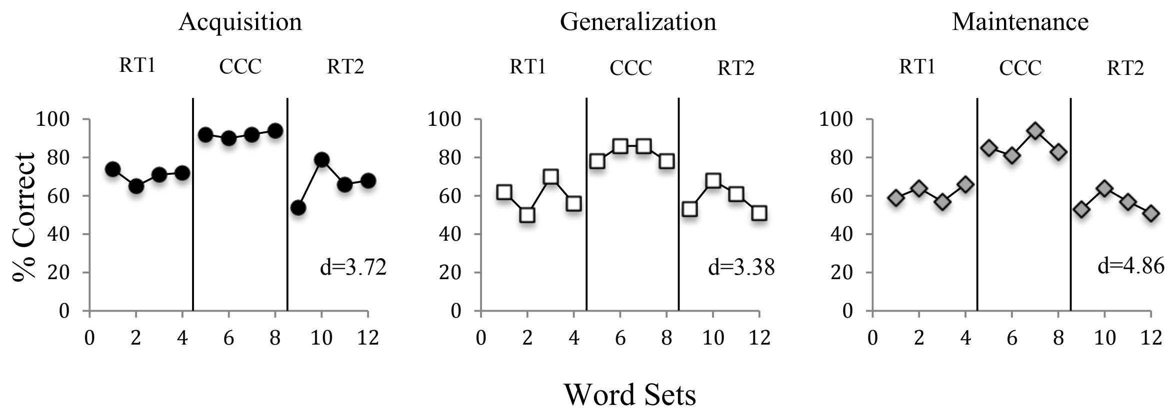

Figure 4. Class-wide mean percent correct for all ten participants on acquisition (left panel), generalization (middle panel), and maintenance (right panel) tests.

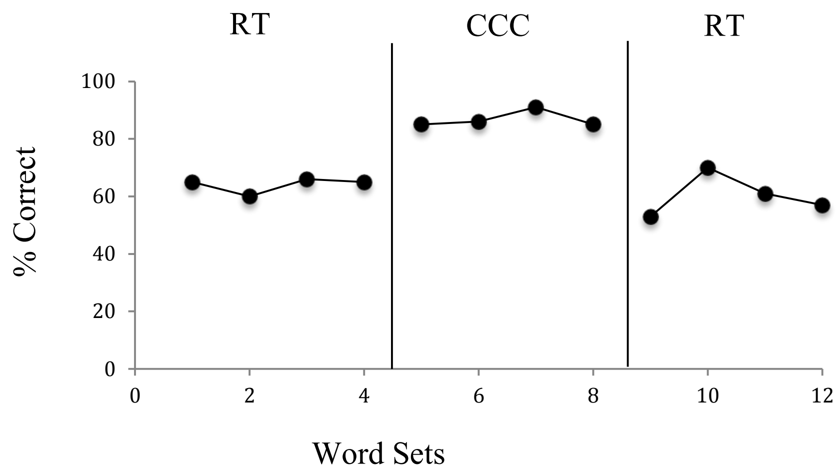

Figure 5. Classwide mean composite score (average of the percent correct on acquisition, generalization, and maintenance) for each word set.

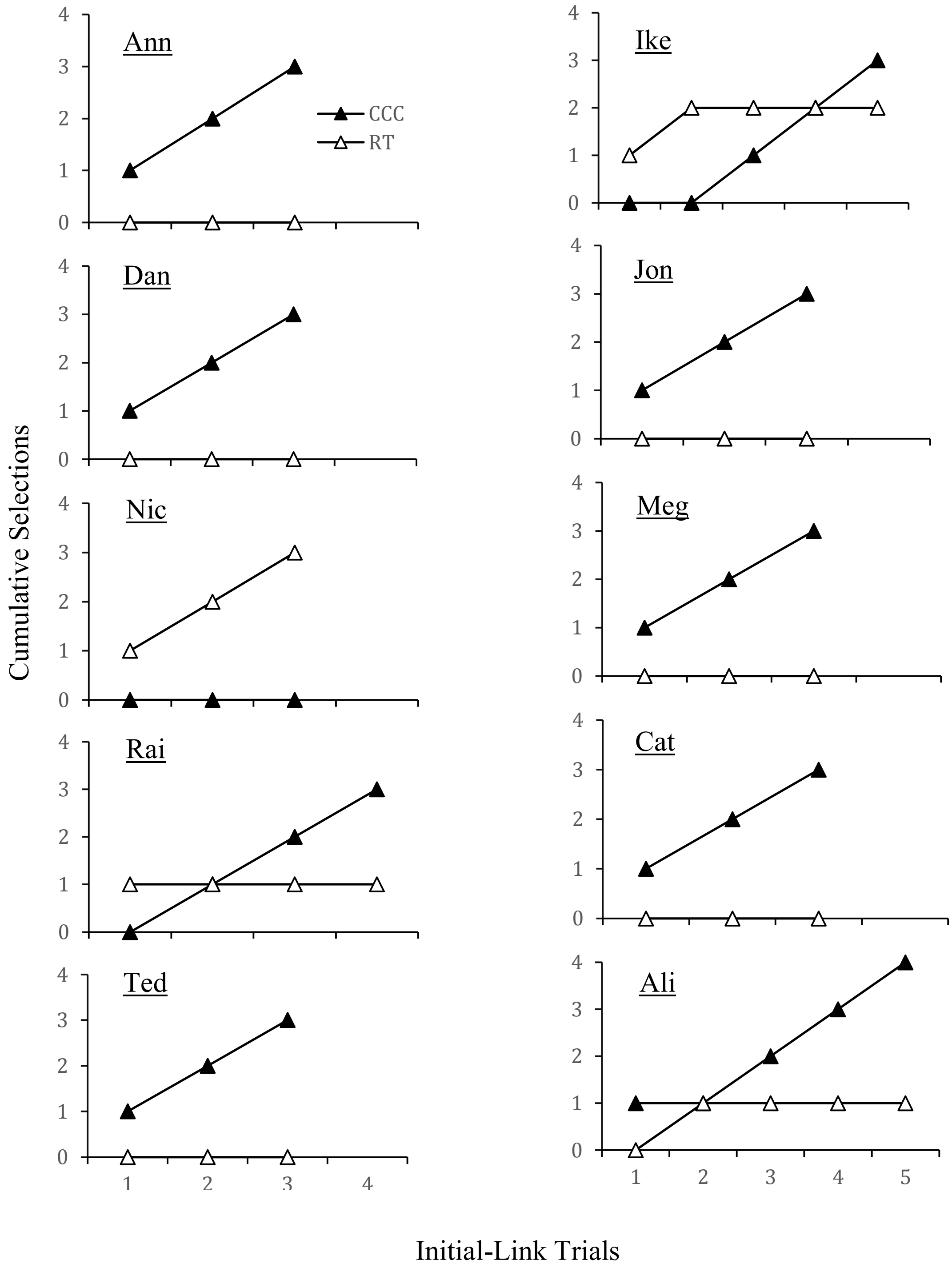

Figure 6. Figure 6 shows the cumulative number of selections in the initial-link trials of the preference assessment.

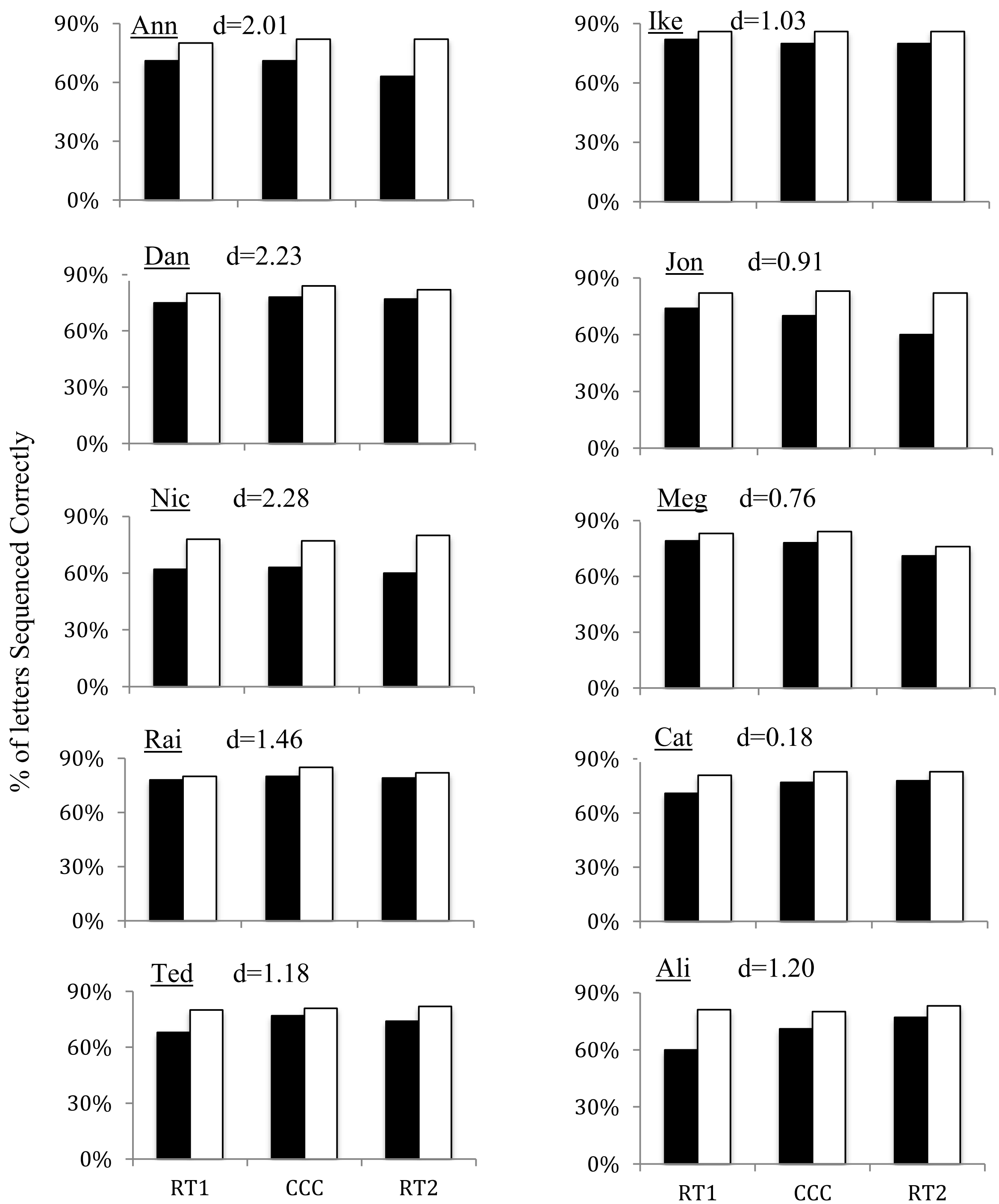

Figure 7. Figure 7 displays the average percentage of letters sequenced correctly in misspelled words on each student's pretest across condition (black bars). The white bars show the percentage of letters they would have spelled correctly if they had made only 1 letter error per word. The difference in height between the bars suggests the degree of spelling error.

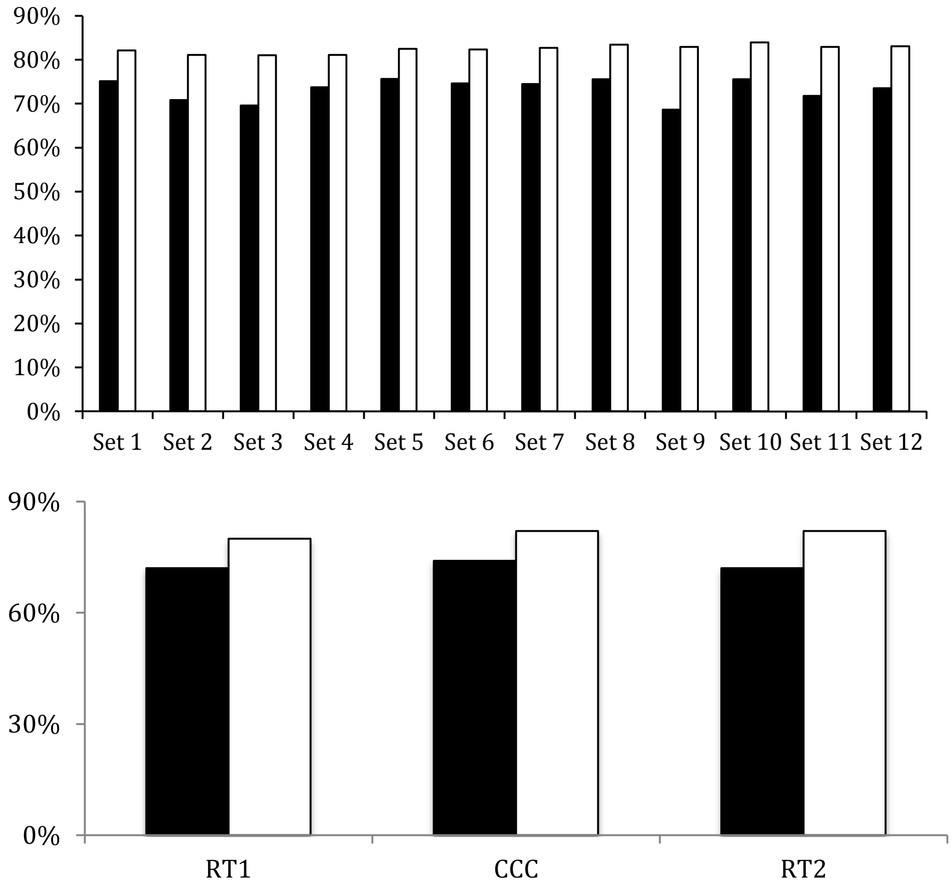

Figure 8. The top panel shows the mean percent of letters sequenced correctly for all ten students across word sets (black bars) and the mean percent of letters they could have sequenced correctly by making only 1 letter sequence error per word (white bars). The difference between the black and white bars suggests the degree of spelling error. The bottom panel shows mean scores aggregated by condition.

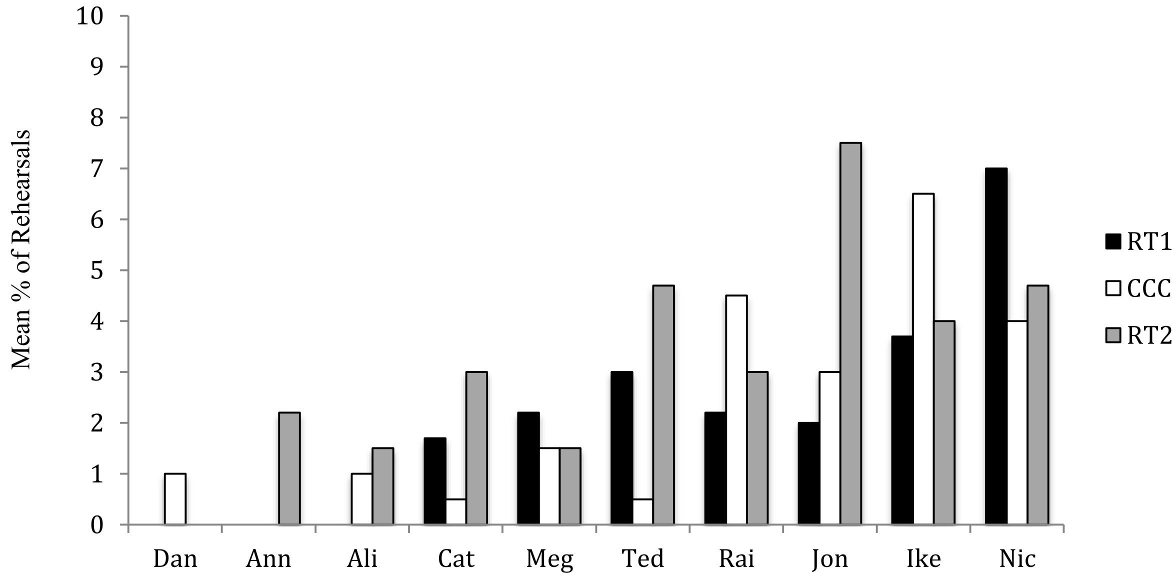

Figure 9. Figure 9 displays the mean percent of rehearsals with an error that occurred during practice sessions for each student for each condition: RT1 (black bars), CCC (white bar), and RT2 (gray bar).

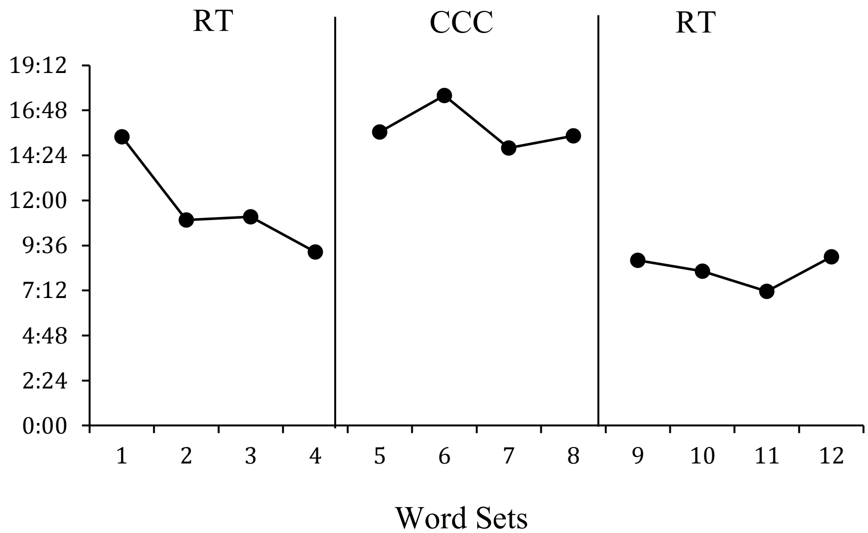

Figure 10. Figure 10 depicts classwide data for mean amount of time for practice sessions across each condition.