CrossMark

INTERVENTION STUDY

Beyond the Rainbow: Retrieval Practice Leads to Better Spelling than does Rainbow Writing

Angela C. Jones[1] · Liane Wardlow[2] · Steven C. Pan[3] ·
Cristina Zepeda[3] · Gail D. Heyman[3] · John Dunlosky[4] ·
Timothy C. Rickard[3]

Abstract In three experiments, we compared the effectiveness of rainbow writing and retrieval practice, two common methods of spelling instruction. In experiment 1 ($n=14$), second graders completed 2 days of spelling practice, followed by spelling tests 1 day and 5 weeks later. A repeated measures analysis of variance demonstrated that spelling accuracy for words trained with retrieval practice was higher than for words trained with rainbow writing on both tests ($\eta_p^2=.49$). In experiments 2 (second graders, $n=16$) and 3 (first graders, $n=12$), students completed 2 days of spelling practice followed by a spelling test 1 day later. Results replicated experiment 1; spelling accuracy was higher for words trained with retrieval practice compared with rainbow writing ($\eta_p^2=.42$ and $.64$, respectively). Furthermore, students endorsed both liking and learning from retrieval practice at least as much as (and sometimes more than) rainbow writing. Results demonstrate that retrieval practice is a more useful (and as engaging) training method than is rainbow writing and extend the well-established testing effect to beginning spellers.

Keywords Spelling · Instruction · Retrieval practice · Testing effect · Emergent literacy

Spelling is a crucial skill that students learn during the first several years of formal schooling. Despite teachers endorsing the importance of spelling instruction and reporting an average of 90 classroom minutes per week on it, they report that more than 25 % of students struggle with

✉ Angela C. Jones
  acjones@jcu.edu

[1] Department of Psychological Science, John Carroll University, 1 John Carroll Blvd, University Heights, OH 44118, USA

[2] Pearson and Department of Psychology, University of California San Diego, La Jolla, CA, USA

[3] Department of Psychology, University of California San Diego, La Jolla, CA, USA

[4] Department of Psychological Sciences, Kent State University, Kent, OH, USA

🖄 Springer

learning spelling (Graham et al. 2008). Spelling skill is linked to both writing and literacy outcomes (e.g., Graham and Santangelo 2014) and may have long-term effects on skilled adult reading (e.g., Perry and Ziegler 2000). In particular, students who are poor spellers are poor readers (e.g., Ehri 1987) and struggle with writing (e.g., Juel 1988). Poor spellers use simpler terms in their writing, forget ideas they want to express, and write less than students who are strong spellers (e.g., Graham et al. 2002; MacArthur et al. 1996; Okyere et al. 1997). Furthermore, Graham and Hebert (2011) found that teachers judge the quality of ideas in papers containing spelling errors more harshly than the same papers with no spelling errors. Importantly, spelling instruction improves both writing and reading skills; recent meta-analyses demonstrate that spelling instruction improves spelling accuracy during writing (Graham and Santangelo 2014), the quality of writing (Graham et al. 2012), and has a positive impact on phonological awareness, word reading, and reading comprehension (Graham and Santangelo 2014).

The findings that improvements in spelling skill transfer to both writing and reading skills are consistent with the lexical quality hypothesis (see Perfetti 2007, for a review). According to this hypothesis, readers who have well-specified word representations are able to devote cognitive resources to higher level text comprehension tasks as opposed to word decoding; one way of indexing lexical quality (or well-specified word representations) is through assessing spelling skill (e.g., Andrews and Bond 2009). Similarly, others suggest that strong spelling skills allow writers to focus more on writing processes that include, for example, planning and revising (Berninger 1999; Graham 1999).

Further demonstrating the importance of good spelling skills, the impact of spelling difficulty during childhood may persist. For example, Perry and Ziegler (2000) found that skilled adult readers were slower to identify words that are difficult for first graders to learn how to spell (even when controlling for other factors commonly known to influence word identification abilities in adult readers). Furthermore, skilled adult spellers outperform those who are less skilled in different reading measures, even when reading comprehension skills are equivalent (Andrews and Bond 2009; Veldre and Andrews 2014a, b).

For all these reasons, investigating which instructional methods lead to the strongest spelling skills is important. In the current study, we examined both the efficacy of and student engagement in two common methods of direct spelling instruction. Direct instruction involves explicit training in word spelling; students engage in activities (either teacher led or independently) designed to improve spelling for assigned lists of words.

The utility of direct instruction has been downplayed by some educational researchers (e.g., Bean and Bouffler 1987; Brown 1990; Krashen 1989, Wilde 1990) who endorsed a spelling-is-caught approach. For this approach, students learn to spell in an incidental fashion during reading and writing activities. Indeed, students do learn new word spellings following reading and writing, without direct spelling instruction (for reviews, see Graham 2000; Krashen 1989). Even so, other educational researchers endorse the use of direct spelling instruction and the majority of teachers report dedicating classroom time to direct spelling instruction activities (e.g., Graham et al. 2008). Perhaps most important, although spelling-is-caught approaches do improve spelling, Graham and Santangelo (2014) discovered in a meta-analysis of 23 studies that direct spelling instruction leads to more learning than do spelling-is-caught approaches.

Because direct spelling instruction outperforms spelling-is-caught approaches, we chose to investigate two common methods of direct instruction used in the classroom—rainbow writing and retrieval practice. Rainbow writing is a relatively new method that involves repeatedly copying spelling words in different colors, creating a rainbow effect. It is related to another

commonly used and older technique, copying, in which spelling words are written without changing colors (Cronnell and Humes 1980; McNeill and Kirk 2014). Retrieval practice involves taking practice quizzes and then checking produced spellings against correct spellings.

The selection of these two methods was not arbitrary. We selected retrieval practice because experimental studies conducted in the laboratory, involving undergraduate participants for the most part, provide strong evidence that it is an effective way to promote learning in other domains (for reviews of the retrieval practice literature, also known as the testing effect, see Dunlosky et al. 2013; Roediger and Butler 2011). Several hypotheses have been proposed to explain the benefits of retrieval practice, including that retrieval enhances semantic elaboration (Carpenter 2011), that it increases the likelihood learners will use better strategies to encode the correct responses (Pyc and Rawson 2010; 2012), and that it enhances memory for context that improves subsequent retrieval (Karpicke et al. 2014b). The mechanisms described in these hypotheses are not mutually exclusive and the positive evidence for each hypothesis suggests that all proposed mechanisms may contribute to retrieval practice benefits in some circumstances. Despite evidence in support of these hypotheses and numerous demonstrations of the robust benefits of retrieval practice, only four studies have examined the influence of retrieval practice on memory in younger elementary school children (Bouwmeester and Verkoeijen 2011; Fritz et al. 2007; Gates 1917; Lipowski et al. 2014) and results have been mixed. These studies have investigated name learning, list learning, and nonsense syllable learning; none have investigated the influence of retrieval practice in an authentic classroom setting with spelling materials.

In the spelling instruction literature, several studies report a benefit following a variant of the traditional retrieval practice paradigm (Alber and Walshe 2004; Grskovic and Belfiore 1996; McGuffin et al. 1997; McNeish et al. 1992; Wirtz et al. 1996). Whereas the traditional retrieval practice paradigm involves retrieval practice followed by restudy, these studies used retrieval practice plus self-correction. That is, students engaged in retrieval practice were shown correct spellings and then rewrote any misspelled words. Because these studies always used self-correction, it is unclear whether retrieval practice or rewriting the misspelled words correctly improved spelling. Furthermore, all of these studies used very small samples ($n=5$ to 6) of students with learning disabilities or identified by teachers as at-risk spellers in grades three or higher—and only descriptive statistics were reported. As such, it is unclear whether retrieval practice is an effective instructional method for typically developing, younger elementary school children when formal spelling instruction begins.

In contrast to the literature indicating that retrieval practice may be beneficial in teaching children to spell, we are unaware of any empirical investigations evaluating the effectiveness of rainbow writing. Literature searches on ERIC, PsycINFO, and Web of Science produced no results for empirical investigations of the efficacy of rainbow writing. Even without an evidence base, however, rainbow writing appears to be a popular instructional method. It is recommended for use by the Scholastic Corporation (e.g., Taylor 2011; Wagstaff 2009) and as part of a spelling instruction curriculum aimed at appealing to multiple intelligences (Shah and Thomas 2002), and it is implemented in a popular Daily Five Word Work program (Boushey and Moser 2006, 2014). Indeed, of the three authors of this paper who have children in elementary school, all have seen their children complete rainbow writing exercises in school. Remarkably, then, it appears that rainbow writing, and likely other instructional tasks, are routinely used in schools with no evidence of efficacy.

Importantly, rainbow writing may be as (or even more) effective than retrieval practice. First, exposure to correctly spelled words increases the likelihood of producing a correct spelling, whereas exposure to an incorrectly spelled word increases the likelihood of producing a misspelling (e.g., Jacoby and Hollingshead 1990). Because rainbow writing virtually precludes spelling errors, it may be more effective than retrieval practice, which is likely to include spelling errors.[1] Second, rainbow writing allows students a level of freedom and choice (i.e., color selection, self-paced) not provided with retrieval practice; student choice is associated with improved motivation and learning outcomes (Grolnik and Ryan 1987; Patall et al. 2010; Ryan and Deci 2000). Third, rainbow writing modifies the commonly used copying technique (e.g., Cronnell and Humes 1980; McNeill and Kirk 2014) by introducing a multisensory component (Shah and Thomas 2002). Although copying may not be as effective as retrieval practice (Grskovic and Belfiore 1996; McGuffin et al. 1997), the multisensory component of rainbow writing may make it more engaging, and thus potentially more effective in the long run, for students.

Because empirical investigations of the efficacy of retrieval practice and rainbow writing for learning spelling are absent in the literature, we conducted three experiments, with typically developing first- and second-grade children, that directly compared the two instructional methods. In addition to assessing their effectiveness as instructional techniques, we assessed the extent to which they engaged children during practice. Based on the rationale above, one prediction is that both techniques will be equally effective but that rainbow writing will be more interesting to students and hence, they will find it more engaging and enjoyable to do. Of course, given the power of retrieval practice over repetition in the larger literature (Dunlosky et al. 2013; Roediger and Butler 2011), another plausible outcome is that retrieval practice will be more effective. Because competing predictions can be made, we evaluated these predictions by exploring the relative efficacy of these two techniques across multiple experiments.

Experiment 1

Method

Participants and Design Fourteen second-grade students (six girls, eight boys) enrolled at an elementary school in Southern California participated in experiment 1. The sample size used in this experiment (and in the following experiments) was based upon the number of available students in participating classrooms; all available students were included in the sample. The design was a 2 (spelling practice: rainbow writing, retrieval practice)×2 (test: 1-day retention, 5-week retention) within-participant design.

Materials and Procedure Materials included 20 words which were selected from a list provided by the teacher (see Appendix A). These words came from materials that students were to be taught and tested on after the experiment was complete, later in the academic year. The 20 words were split into two 10-word lists; one list was practiced with rainbow writing and one with retrieval practice. Students were randomly split into two groups so that list assignment to type of spelling practice was counterbalanced.

---

[1] Our data indicate that more correct spellings of a word were produced with rainbow writing ($M=7.4$) than with retrieval practice ($M=3$).

For rainbow writing, students were provided with worksheets with the practice words printed and several crayons of varying colors. They were instructed to complete the worksheets by writing each word several times, in the colors of their choosing. The research assistant demonstrated the procedure on a white board, writing a single word several times in various colors before students began. They were given 10 min to complete the rainbow writing activity and they worked continuously at their own pace throughout each trial.

For retrieval practice, the research assistant explained the procedure and then read a list of 10 words, one at a time (each repeated once). Students wrote each word on worksheets provided by the research assistant. After all 10 words were presented, the research assistant wrote the correct spellings of the words on a white board for students to view. Students marked any incorrectly spelled words with an X and marked correctly spelled words with a check mark. Students then turned their worksheets over, words were erased from the white board, and the process began again, repeating until time elapsed. Retrieval practice, like rainbow writing, lasted for 10 min.

Students participated in both types of spelling practice over two consecutive days, with one 10-min block for each practice activity on each day. In addition to counterbalancing word list across practice type, the order of practice was counterbalanced across practice days and each student group.

Students completed two retention tests. One was given 1 day following the practice sessions; one was given 5 weeks later. For the tests, students had blank paper and pencils. All 20 practiced words were presented in random order. Similar to the retrieval practice phase, each word was read twice for students to spell.

In addition to the retention tests, students also completed a questionnaire regarding each of the practice methods immediately following the first retention test (see Appendix B). The questionnaire contained three questions in which students chose between rainbow writing and retrieval practice; they concerned which task the students would choose to do in the future, which task was more fun, and which task helped them learn more. The questionnaire also contained four questions in which students used a 5-point Likert scale for responding; two questions were about rainbow writing and two were about retrieval practice. Students rated how much they liked and learned from each method. In all experiments, if students had a question about the questionnaire, a research assistant answered them to ensure they understood the task.

Results

In the interest of focusing on statistical analyses necessary to answer our specific research questions, below, we report only the planned comparisons. However, outcomes of the repeated measures analyses of variance (ANOVAs) for all experiments are reported in Table 1.

Spelling Accuracy As illustrated by Fig. 1, retrieval practice produced more learning than rainbow writing, both on the test 1 day following the practice sessions, $t(13)=2.31$, $p=.038$, $d=.60$, and on the retention test 5 weeks later, $t(13)=2.75$, $p=.017$, $d=.48$.

Questionnaires Descriptive statistics for the self-report questionnaires are listed in Table 2. When asked to choose between retrieval practice and rainbow writing, a larger percentage of students chose retrieval practice over rainbow writing; they preferred it as a future instructional

Table 1 Outcomes of omnibus analyses of variance for experiments 1–3

	df	F	MSE	p value	ηp2
Test performance					
Experiment 1: 2 (test delay: 1-day vs. 5-week)×2 (practice: retrieval practice vs. rainbow writing)					
Main effect of test delay	1,13	16.64	73	.001	.56
Main effect of practice	1,13	12.07	96	.004	.49
Interaction	1,13	.07	103	.797	.01
Experiment 2: 2 (test: pre- vs. post-)×2 (practice: retrieval practice vs. rainbow writing)					
Main effect of test	1,15	32.38	243	<.001	.68
Main effect of practice	1,15	0.47	163	.504	.03
Interaction	1,15	11.04	62	.005	.42
Experiment 3: 2 (test: pre- vs. post-)×2 (practice: retrieval practice vs. rainbow writing)					
Main effect of test	1,11	33	164	<.001	.75
Main effect of practice	1,11	5	220	.047	.31
Interaction	1,11	19.81	88	.001	.64
Retrieval practice					
Experiment 2: 2 (day: first vs. second)×2 (retrieval attempt: first vs. second)					
Main effect of day	1,13	12.63	96	.004	.49
Main effect of retrieval attempt	1,13	12.55	205	.004	.49
Interaction	1,13	5.44	84	.036	.3
Experiment 3: 2 (day: first vs. second)×2 (retrieval attempt: first vs. second)					
Main effect of day	1,11	20.73	79	.001	.65
Main effect of retrieval attempt	1,11	22.3	84	.001	.67
Interaction	1,11	0	45	1	0

method, endorsed learning more from it, and liked it more. However, note that the sample sizes in this experiment as well as in experiments 2 and 3 were relatively small, so none alone
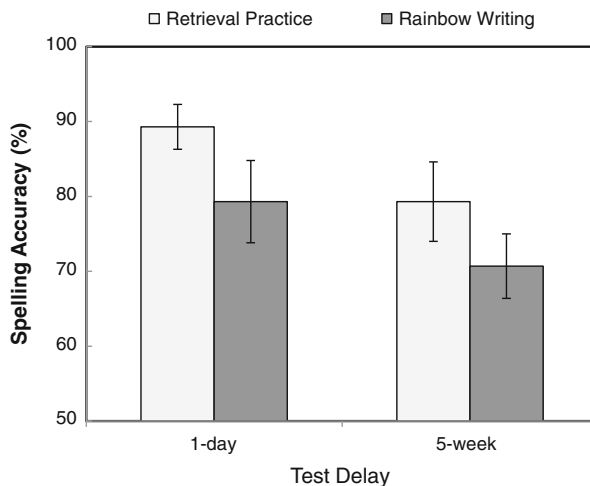

Fig. 1 Spelling accuracy for students in experiment 1. *Error bars* are standard error

Table 2 Descriptive statistics for the self-report questionnaires for experiments 1–3

	Choose one method						Likert scale questions			
	Future practice		Liking		Learning		Liking		Learning	
	Rainbow (%)	RP (%)	Rainbow (%)	RP (%)	Rainbow (%)	RP (%)	Rainbow	RP	Rainbow	RP
Experiment 1	28.60	71.40	7.10	92.90	28.60	71.40	2.71	3.79	1.29	1.86
Experiment 2	31.30	68.80	25.00	75.00	37.50	62.50	1.81	2.87	1.87	2.87
Experiment 3	50.00	50.00	33.30	66.70	33.30	66.70	2.25	3.25	1.92	3.58

RP for retrieval practice

provided enough power to reveal a significant effect using this relatively insensitive measure based on binary (yes/no) responses; thus, we combined data for the first three questionnaire questions across experiments and report the inferential statistics in the "General Discussion" section. When asked to report the degree to which they liked and learned from each instructional method on its own (i.e., not as a comparison between methods), retrieval practice was rated higher than rainbow writing for both liking, $t(13)=2.11$, $p=.055$, $d=.98$, and learning, $t(13)=3.31$, $p=.006$, $d=1.14$.

Experiment 2

The purpose of experiment 2 was to replicate and extend results from experiment 1 with a new sample of second graders from a different geographic region. Again, we were interested in comparing the efficacy of retrieval practice and rainbow writing. In this experiment, we included a pre-test to assess the amount of learning that occurred during training.

Method

Participants and Design Sixteen second graders (eight girls, eight boys) enrolled in an elementary school in Northeast Ohio participated in experiment 2. The design was a 2 (test: pre-training, post-training)×2 (spelling practice: rainbow writing, retrieval practice) within-participant design.

Materials and Procedure As in experiment 1, spelling materials were selected from a list provided by the teacher (see Appendix A); items were selected from a list of words used by the school district that were expected to be learned during the next academic year. The questionnaire was identical to that used in experiment 1. The procedure was identical to that of experiment 1, with the following three exceptions. First, to assess baseline performance before spelling practice began, students were administered a pre-test of all 20 spelling words in randomized order. Second, there was no 5-week retention test; the test was administered 1 day following the practice sessions. Third, the order in which items were presented during each day's practice phase was randomized to reduce the possibility that students had more practice with some words than others. Because rainbow writing was self-paced, it is possible that the difference in accuracy observed between rainbow writing and retrieval practice in experiment

1 was due to students only focusing on the first several words of the list and not practicing the end items.[2]

Results

Spelling Accuracy Comparisons of pre-training and post-training test performance indicate that both rainbow writing and retrieval practice produced learning, $t(15)=3.74$, $p=.002$, $d=.60$ and $t(15)=6.32$, $p<.001$, $d=1.13$, respectively. Furthermore, as illustrated in Fig. 2, the benefits of retrieval practice were larger than the benefits of rainbow writing, $t(15)=2.78$, $p=.014$, $d=.47$.

Retrieval Practice We also retained students' retrieval practice data to examine spelling accuracy during training.[3] On each training day, students completed two retrieval practice attempts. On both days, accuracy improved from the first to second retrieval attempt (see Table 3), $t(13)=3.51$, $p=.004$, $d=.88$ (day 1) and $t(13)=2.35$, $p=.035$, $d=.49$ (day 2).

Self-Report Questionnaires Similar to experiment 1, students rated retrieval practice as high as or higher than rainbow writing (see Table 2). A larger percentage of students again preferred it as a future instructional method, endorsed learning more from it, and liked it more. When asked to report how much they liked and learned from each instructional method on its own, the differences between retrieval practice and rainbow writing were nonsignificant, although in the same direction as in experiment 1, $t(15)=1.38$, $p=.189$, $d=.59$, $t(15)=1.41$, $p=.178$, $d=.58$, respectively.

Experiment 3

In experiment 3, we sought to replicate and extend our findings with a younger sample of students for two reasons. First, baseline performance on the pre-training test was relatively high in experiment 2 ($M=62.8$ %, SD=31).[4] As such, the 28.8 % improvement in performance following retrieval practice (compared with 15.6 % following rainbow writing) may reflect an underestimate of the advantage for retrieval practice given possible ceiling effects. In fact, not only was overall performance at 90 % after retrieval practice, but 56.3 % of the students in experiment 2 had perfect accuracy on items following retrieval practice. Thus, the nearly twofold benefit of retrieval practice over rainbow writing we observed may not accurately reflect the advantage of retrieval practice over rainbow writing. Because first grade is when formal spelling instruction typically begins, we expected baseline performance to be lower. With lower baseline performance, we may observe an even greater benefit of retrieval practice over rainbow writing. Second, only two studies have examined the effects of retrieval practice

---

[2] Only one student failed to practice all items with rainbow writing; two items were unpracticed. On average, each item was practiced nine times. Thus, observed differences in accuracy for retrieval practice and rainbow writing on the final test are unlikely due to insufficient practice with all items in the rainbow writing condition.

[3] Two students' retrieval practice data were misplaced. Thus, analyses include data from 14 students.

[4] Using the first round of retrieval practice in experiment 1 as a baseline measure, performance was relatively high ($M=66.9$%, $SD=22$), similar to experiment 2.
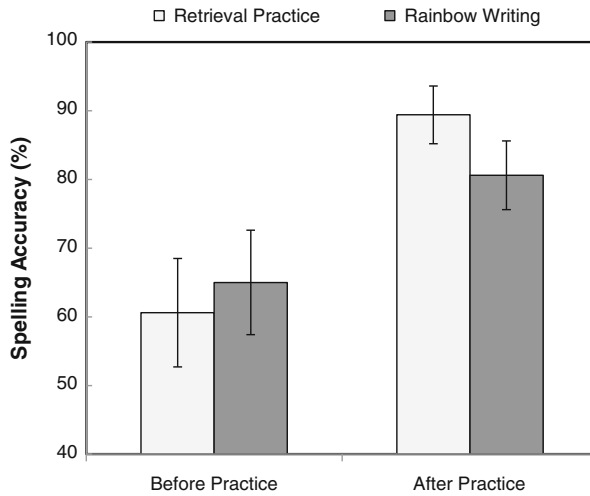
Fig. 2 Spelling accuracy for students in experiment 2. *Error bars* are standard error

(both in non-spelling task domains) with first graders (Gates 1917; Lipowski et al. 2014) and results were mixed. Thus, it is unclear whether retrieval practice would yield the same benefits for first graders as was established for second graders in the first two experiments.

Method

Participants and Design Twelve first graders (nine girls, three boys) from the same school in Northeast Ohio (as reported in experiment 2) participated in experiment 3. The design was identical to that of experiment 2.

Materials and Procedure Again, spelling materials were selected from a list provided by the teacher (see Appendix A) and the questionnaire was identical to those used in experiments 1 and 2. The procedure was identical to that of experiment 2.

Results

Spelling Accuracy Results replicated experiment 2 (see Fig. 3). Both rainbow writing and retrieval practice produced learning, $t(11)=2.303$, $p=.042$, $d=.29$ and $t(11)=6.50$, $p<.001$, $d=$

Table 3 Spelling accuracy during retrieval practice

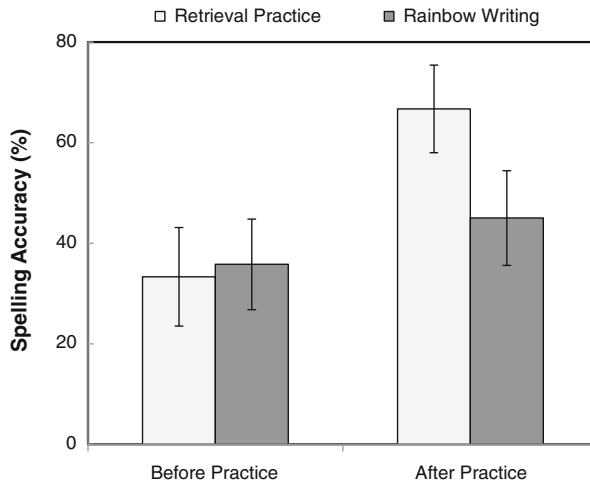|	Day 1		Day 2	
	RP1 (%)	RP2 (%)	RP3 (%)	RP4 (%)
Experiment 2	71	91	86	94
Experiment 3	52	64	63	76

RP stands for retrieval practice

Fig. 3 Spelling accuracy for students in experiment 3. *Error bars* are standard error

1.04, respectively. Furthermore, the benefits of retrieval practice (34 % gain) were again larger than the benefits of rainbow writing (9 % gain), $t(11)=5.61$, $p<.001$, $d=.69$.[5]

Retrieval Practice As in experiment 2, spelling accuracy increased from the first to second retrieval practice attempts on both days (see Table 3), $t(11)=3.56$, $p=.004$, $d=.41$ (day 1) and $t(11)=4.10$, $p=.002$, $d=.44$ (day 2).

Self-Report Questionnaires Although an equal proportion of students chose each method when asked to choose between rainbow writing and retrieval practice as a future instructional method, a larger percentage of students endorsed both liking and learning more from retrieval practice (see Table 2). When not choosing between methods, but rating how much they liked and learned from each method, students endorsed more learning from retrieval practice than rainbow writing, $t(11)=2.87$, $p=.015$, $d=1.23$. As in experiment 2, although retrieval practice was rated numerically higher than rainbow writing, this difference in student ratings of liking was not significant, $t(11)=1.30$, $p=.220$, $d=.63$.

General Discussion

In three experiments, we consistently found that retrieval practice promotes student learning more than rainbow writing does. In experiment 1, spelling accuracy was 10 % higher following retrieval practice than following rainbow writing and the benefit remained stable over a 5-week delay. Experiment 2 replicated experiment 1; spelling accuracy following

---

[5] Similar to experiment 2, only one student failed to practice all items with rainbow writing; two items were unpracticed. On average, items were practiced with rainbow writing 6.5 times. Thus, the nearly threefold gain produced by retrieval practice over rainbow writing is unlikely due to a failure in practicing items during rainbow writing.

retrieval practice was 9 % higher than following rainbow writing. Finally, in experiment 3, with reduced baseline performance, spelling accuracy following retrieval practice was 22 % higher than following rainbow writing. These results constitute the first empirical investigation regarding the relative efficacy of rainbow writing and retrieval practice as instructional methods for spelling. Furthermore, the benefit in learning afforded by retrieval practice did not come at the expense of children's enjoyment; students rated retrieval practice to be as or more preferable than rainbow writing.

Our results also uniquely contribute to the testing effect literature. Publications about the testing effect with young elementary school children represent a small, but growing, literature (Bouwmeester and Verkoeijen 2011; Fritz et al. 2007; Gates 1917; Lipowski et al. 2014). Adding to the several demonstrations of the testing effect with older elementary and middle school children using applied, course-relevant concepts (e.g., Karpicke, Blunt, Smith and Karpicke 2014; Lipko-Speed et al. 2015; Metcalfe et al. 2007; Roediger et al. 2011), our results constitute the first demonstration of the benefits of retrieval practice for young elementary students in an authentic classroom setting. These data also provide a somewhat unusual and surprising *concordance* between performance and metacognitive awareness, which is not common in the adult testing effect literature (e.g., Roediger and Karpicke 2006). In experiments 1 and 3, students rated learning as higher from retrieval practice than from rainbow writing; the trend, although nonsignificant, was in the same direction in experiment 2. Furthermore, when combining questionnaire data from all three experiments, when students had to choose between retrieval practice and rainbow writing, retrieval practice was chosen as the superior learning method, $\chi^2=4.67$, $p=.031$. In the only other study with young elementary students that assessed beliefs about learning (Lipowski et al. 2014), only third graders believed retrieval practice was superior to restudying; first graders endorsed learning more from restudying than retrieval practice, similar to findings in the adult testing effect literature. Here, though, both the first and second graders endorsed retrieval practice as the superior learning method. (Across the three experiments, students also chose retrieval practice over rainbow writing as a preferred future practice method, $\chi^2=3.43$, $p=.064$, and endorsed liking it more, $\chi^2=13.71$, $p<.001$.)

Our findings have straightforward implications for improving spelling instruction; namely, retrieval practice promotes better learning than rainbow writing. However, teachers report using a variety of spelling activities (Graham et al. 2008; McNeill and Kirk 2014) and many intervention studies report using multi-component spelling instruction (e.g., Berninger et al. 2002; Graham et al. 2002; Kirk and Gillon 2009). Aside from the instructional methods we investigated, other common methods include alphabetizing, writing-saying, finding the missing letter, unscrambling the letters, word searches, and dictionary work. As such, programmatic and parametric research is needed to examine which instructional activities among the variety used promote the best learning as well as which combination(s) of activities promote the best learning. For example, retrieval practice may produce more learning if preceded by another activity like rainbow writing or unscrambling the letters, so as to ensure some successful retrieval of word spellings during practice. That is, retrieval practice followed by feedback may not be as effective when retrieval performance during practice is low (Karpicke et al. 2014a, b; Smith and Karpicke 2014). In the present case (Table 3), the students did show retrieval success during practice trials, but an

activity which promotes initial learning gains prior to engaging in retrieval practice may provide the best outcomes, particularly for students struggling to learn spelling.

More generally, in terms of educational practice, these results demonstrate the importance of empirical evidence for instructional techniques. When teachers seek to maximize the efficiency and effectiveness of classroom practices, they are often faced with a wide array of choices, sometimes with little more than their intuitions to guide them. Presumably, the use of rainbow writing has become popular precisely because it is believed to be more enjoyable for students. Our results indicate that is not the case. These findings serve as an important reminder that even when teaching methods have been developed to be fun and innovative, claims that they are educationally beneficial and that children find them appealing require (causal) empirical support (see Reinhart et al. 2013).

Appendix A

Materials included in experiments 1, 2, and 3

Experiment 1		Experiment 2		Experiment 3	
List 1	List 2	List 1	List 2	List 1	List 2
Fact	Began	Watch	Solve	Means	Clown
Since	Title	Rough	Energy	Because	Over
Able	Easy	Young	Every	Turn	Around
Stranger	Display	Pattern	Group	Years	Globe
Sketch	Empty	Flight	Reach	Place	History
Eager	People	Carry	White	Broom	Think
Hoping	Explain	Together	Family	Follow	Little
Approach	Believe	Current	Simple	Elect	Learn
Tongue	Rougher	Raise	Quickly	Letter	Ocean
Happier	Laziest	Leave	Meaning	Paint	Letter

Appendix B

Appendix B. Self-report questionnaires used in Experiments 1, 2, and 3.

Students' questionnaires contained the following text:

1. (a) Rainbow Writing   (b) Quizzing

2. (a) Rainbow Writing   (b) Quizzing

3. (a) Rainbow Writing   (b) Quizzing

The following three questions were read aloud to the class:

1. If I came back next week to help you learn some more spelling words, which activity would you like to do – Rainbow Writing or Quizzing? If you would rather do rainbow writing, circle the first choice. If you would rather do quizzing, circle the second choice.

2. Do you think you learned to spell more words when you did rainbow writing or quizzing? If you think you learned more during the rainbow writing activity, circle the first choice. If you think you learned more during quizzing, circle the second choice

3. Which did you like doing more – Rainbow Writing or Quizzing? If you liked rainbow writing more, circle the first choice. If you liked quizzing more, circle the second choice.

The students' questionnaires contained the following graphics:

The following four questions were read aloud to the class:

1. How much did you like the Rainbow Writing activity? If you didn't like rainbow writing at all, circle the very sad face. If you liked it a lot, circle the big smiley face. Or choose one of the faces in between if you felt in between about how much you liked rainbow writing.

2. How much did you feel like you learned from the Rainbow Writing activity? If you feel like you didn't learn anything at all from it, circle the very sad face. If you feel like you learned a lot from rainbow writing, circle the big smiley face. Or choose something in between if you felt in between about how much you learned from rainbow writing.

3. How much did you like the Quizzing activity? If you didn't like quizzing at all, circle the very sad face. If you liked it a lot, circle the big smiley face. Or choose one of the faces in between if you felt in between about how much you liked quizzing.

4. How much did you feel like you learned from the Quizzing activity? If you feel like you didn't learn anything at all from it, circle the very sad face. If you feel like you learned a lot from quizzing, circle the big smiley face. Or choose something in between if you felt in between about how much you learned from quizzing.

References

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

Andrews, S., & Bond, R. (2009). Lexical expertise and reading skill: bottom-up and top-down processing of lexical ambiguity. *Reading and Writing, 22*, 687–711.

Bean, W., & Bouffler, C. (1987). *Spell by writing*. Portsmouth: Heinemann.

Berninger, V. (1999). Coordinating transcription and text generation in working memory during composing: automatic and constructive processes. *Learning Disability Quarterly, 22*, 99–112.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Byrd, K., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition along and together: implications for the simple view of writing. *Journal of Educational Psychology, 94*, 291–304. doi:10.1037/0022-0663.94.2.291.

Boushey, & Moser. (2006). *The daily 5: fostering literacy independence in the elementary grades*. Portland: Stenhouse Publishers.

Boushey, & Moser. (2014). *The daily 5: fostering literacy independence in the elementary grades* (2nd ed.). Portland: Stenhouse Publishers.

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language, 65*, 32–41.

Brown, A. (1990). A review of recent research on spelling. *Educational Psychology Review, 2*, 365–397.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1547–1552.

Cronnell, B., & Humes, A. (1980). Elementary spelling: what's really taught. *The Elementary School Journal, 81*, 59–64.

Dunlosky, J., Rawson, K. A., Marsh, E. G., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4–58.

Ehri, L. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*, 5–31.

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: an effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology, 60*, 991–1004.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6* (Serial No. 40).

Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: a review. *Learning Disability Quarterly, 22*, 78–98.

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

Graham, S., & Hebert, M. (2011). Writing to read: a meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*, 710–744.

Graham, S., & Santangelo, T. (2014). Does spelling instruction makes students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing, 27*, 1703–1743. doi:10.1007/s11145-01409517-0.

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing and reading of poor spellers. *Journal of Educational Psychology, 94*, 669–686.

Graham, S., Morphy, P., Harris, K., Fink-Chorzempa, B., Saddler, B., Moran, S., et al. (2008). Teaching spelling in the primary grades: a national survey of instructional practices and adaptations. *American Educational Research Journal, 45*, 796–825.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879–896. doi:10.1037/a0029185.

Grolnik, W. S., & Ryan, R. N. (1987). Autonomy in children's learning: an experimental and individual difference investigation. *Journal of Personality and Social Psychology, 52*, 890–898.

Grskovic, J. A., & Belfiore, P. J. (1996). Improving the spelling performance of students with disabilities. *Journal of Behavioral Education, 6*, 343–354.

Jacoby, L. L., & Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology, 44*, 345–358. doi:10.1037/h0084259.

Juel, C. (1988). Learning to read and write: a longitudinal study of 54 children from first through fourth grade. *Journal of Educational Psychology, 80*, 437–447. doi:10.1037/0022-0663.80.4.437.

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014a). Retrieval-based learning: the need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition, 3*, 198–206. doi:10.1016/j.jarmac.2014.07.008.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014b). Retrieval-based learning: an episodic context account. *Psychology of Learning and Motivation, 61*, 237–284.

Kirk, C., & Gillon, G. T. (2009). Integrated morphological awareness intervention as a tool for improving literacy. *Language, Speech, and Hearing Services in Schools, 40*, 341–351.

Krashen, S. (1989). We acquire vocabulary and spelling by reading. Additional evidence for the input hypothesis. *The Modern Language Journal, 73*, 440–464.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2015). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition.*

Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50*, 994–1000.

MacArthur, C. A., Graham, S., Haynes, J. B., & De La Paz, S. (1996). Spelling checkers and students with learning disabilities: performance comparisons and impact on spelling. *Journal of Special Education, 30*, 35–57.

McGuffin, M. E., Martz, S. A., & Heron, T. E. (1997). The effects of self-correction versus traditional spelling on the spelling performance and maintenance of third grade students. *Journal of Behavioral Education, 7*, 463–476.

McNeill, B., & Kirk, C. (2014). Theoretical beliefs and instructional practices used for teaching spelling in elementary classrooms. *Reading and Writing, 27*, 535–554. doi:10.1007/s11145-013-9457-0.

McNeish, J., Heron, T. E., & Okyere, B. (1992). Effects of self-correction on the spelling performance of junior high students with learning disabilities. *Journal of Behavioral Education, 2*, 17–27.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology, 19*, 743–768.

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.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102*, 896–915.

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383. doi:10.1080/10888430701530730.

Perry, C., & Ziegler, J. C. (2000). Linguistic difficulties in language and reading development constrain skilled adult reading. *Memory & Cognition, 28*, 739–745.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science, 330*, 335.

Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 737–746.

Reinhart, A. L., Haring, S. H., Levin, J. R., Patall, E. A., & Robinson, D. H. (2013). Models of not-so-good behavior: yet another way to squeeze causality and recommendations for practice out of correlational data. *Journal of Educational Psychology, 105*, 241–447.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science, 15*, 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests to improve long-term retention. *Psychological Science, 17*, 249–255.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382–395.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Shah, T., & Thomas, A. (2002). Improving the spelling of high frequency words in daily writing through the use of multiple intelligence centers.

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory, 22*, 784–802. doi:10.1080/09658211.2013.831454.

Taylor, S. (2011). A kindergarten writing center in action. Retrieved from http://www.scholastic.com/teachers/2011/12/kindergarten-writing-center-action.

Veldre, A., & Andrews, S. (2014a). Lexical quality and eye movements: individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology, 67*, 703–727.

Veldre, A., & Andrews, S. (2014b). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000039. Advance online publication.

Wagstaff, J. (2009) *Using name walls to teach reading and writing (Scholastic Teaching Resources)*. New York, NY: Scholastic Inc. Available from: http://www.scholastic.com/teachers/article/collateral_resources/pdf/83/9780545269483.pdf.

Wilde, S. (1990). Spelling textbooks: a critical review. *Linguistics and Education, 2*, 259–280.

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