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

Angela C. Jones 
John Carroll University 
Liane Wardlow 
Pearson and University of California, San Diego 
Steven C. Pan, Cristina Zepeda, Gail D. Heyman 
University of California, San Diego 
John Dunlosky 
Kent State University 
Timothy C. Rickard 
University of California, San Diego 

This manuscript was accepted for publication in Educational Psychology Review on July 2, 2015. This document may not exactly replicate the final version published in the Springer journal. It is not the copy of record. The final publication is available at Springer via http://dx.doi.org/10.1007/s10648-015-9330-6 

Author Note 

Angela C. Jones, Department of Psychological Science, John Carroll University. Liane Wardlow-Lane, Department of Psychology, University of California San Diego. Steven C. Pan, Department of Psychology, University of California San Diego. Gail D. Heyman, Department of Psychology, University of California San Diego. John Dunlosky, Department of Psychological Sciences, Kent State University. Timothy C. Rickard, Department of Psychology, University of California San Diego.
This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders [3RO1DC011 492]. We would like to thank Melissa Bishop, Dasom Kim, Soo Lee, and Brenda Wu for assistance with data collection.

Correspondence concerning this manuscript should be addressed to Angela C. Jones, Department of Psychological Science, John Carroll University, 1 John Carroll Blvd, University Heights, OH 44118. Telephone: 216-397-1531. Fax: 216-397-1633. Email: acjones@jcu.edu
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 two days of spelling practice, followed by spelling tests one day and five 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 ($\eta^2_p = 0.49$). In Experiments 2 (second graders, n = 16) and 3 (first graders, n = 12), students completed two days of spelling practice followed by a spelling test one day later. Results replicated Experiment 1; spelling accuracy was higher for words trained with retrieval practice compared with rainbow writing ($\eta^2_p = .42$ 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
Beyond the Rainbow: Retrieval Practice Leads to Better Spelling than Rainbow Writing

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 learning spelling (Graham, Morphy, Marris, Fink-Chorzempa, Sadler, Moran, & Mason, 2008). Spelling skill is linked to both writing and literacy outcomes (e.g., Graham & Santangelo, 2014) and may have long-term effects on skilled adult reading (e.g., Perry & 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, Harris, & Chorzempa, 2002; MacArthur, Graham, Haynes, & De La Paz, 1996; Okyere, Heron, & Goddard, 1997). Furthermore, Graham and Herbert (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 & Santangelo, 2014), the quality of writing (Graham, McKeown, Kiuhara, & Harris, 2012), and has a positive impact on phonological awareness, word reading, and reading comprehension (Graham & 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 & 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 & Bond, 2009; Veldre & Andres, 2014a; Veldre & Andrews, 2014b).

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 & 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 & Humes, 1980; McNeill & 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, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & 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 & Rawson, 2010; 2012),
and that it enhances memory for context that improves subsequent retrieval (Karpicke, Lehman,
& Aue, 2014). 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 & Verkoeijen, 2011; Fritz, Morris, Nolan, & Singleton, 2007; Gates, 1917; Lipowski, Pyc, Rawson, & Dunlosky, 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 & Walshe, 2004; Grskovic & Belfiore, 1996; McGuffin, Martz, & Heron, 1997; McNeish, Heron, & Okyere, 1992; Wirtz, Gardner, Weber, & Bullara, 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 & Thomas, 2002), and it is implemented in a popular Daily Five Word Work program (Boushey & 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 & 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 & Ryan, 1987; Patall, Cooper, & Wynn, 2010; Ryan & Deci, 2000). Third, rainbow writing modifies the commonly used copying technique (e.g., Cronnell & Humes, 1980; McNeill & Kirk, 2014) by introducing a multisensory component (Shah & Thomas, 2002). Although copying may not be as effective as retrieval practice (Grskovic & Belfiore, 1996; McGuffin, Martz, & Heron, 1997), the multisensory component of rainbow writing may make it more engaging, and thus potentially more effective in the long run, for students.

\(^1\)Our data indicate that more correct spellings of a word were produced with rainbow writing ($M = 7.4$) than with retrieval practice ($M = 3$).
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 & 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) x 2 (test: one-day retention, five-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.

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 minutes 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 minutes.

Students participated in both types of spelling practice over two consecutive days, with one 10 minute 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 one day following the practice sessions; one was given five 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 five-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 Figure 1, retrieval practice produced more learning than rainbow writing, both on the test one day following the practice sessions, $t(13) = 2.31, p = .038, d = .60$, and on the retention test five 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 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 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. 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 pretest 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) x 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 five-week retention test; the test was administered one 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

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 Figure 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. 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* = 0.88 (day 1) and *t*(13) = 2.35, *p* = .035, *d* = 0.49 (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.

---

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 participants’ data were misplaced. Thus, analyses include data from 14 students.
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 two-fold 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 (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

\(^4\)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.
**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 Figure 3). Both rainbow writing and retrieval practice produced learning, \( t(11) = 2.303, p = .042, d = .29 \) and \( t(11) = 6.50, p < .001, d = 1.04 \), 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), \( t(11) = 3.56, p = .004, d = 0.41 \) (day 1) and \( t(11) = 4.10, p = .002, d = 0.44 \) (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

---

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 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. In Experiment 1, spelling accuracy was 10% higher following retrieval practice than following rainbow writing and the benefit remained stable over a five-week delay. Experiment 2 replicated Experiment 1; spelling accuracy following 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 & 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, & Karpicke, 2014; Lipko-Speed, Dunlosky, & Rawson, in press; Metcalfe, Kornell, & Son, 2007; Roediger, Agarwal, McDaniel, & McDermott, 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 & 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 & Kirk, 2014) and many intervention studies report using multi-component spelling instruction (e.g., Berninger, Vaughan, Abbott, Begay, Byrd, Curtin, et al., 2002; Graham et al., 2002; Kirk & 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., 2014; Smith & 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 Reinhard, Haring, Levin, Patall, & Robinson, 2013).
References


Literacy Research and Instruction, 40, 143-155.


doi:10.1037/0022-0663.80.4.437


Shah, T., & Thomas, A. (2002, May 1). Improving the Spelling of High Frequency Words in Daily Writing through the Use of Multiple Intelligence Centers.


Acknowledgements

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders [3RO1DC011 492]. We thank Melissa Bishop, Dasom Kim, Soo Lee and Brenda Wu for assistance with data collection.
Appendix A. Materials included in Experiments 1, 2, and 3.

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>List 1</td>
<td>List 2</td>
<td>List 1</td>
</tr>
<tr>
<td>Fact</td>
<td>Began</td>
<td>Watch</td>
</tr>
<tr>
<td>Since</td>
<td>Title</td>
<td>Rough</td>
</tr>
<tr>
<td>Able</td>
<td>Easy</td>
<td>Young</td>
</tr>
<tr>
<td>Stranger</td>
<td>Display</td>
<td>Pattern</td>
</tr>
<tr>
<td>Sketch</td>
<td>Empty</td>
<td>Flight</td>
</tr>
<tr>
<td>Eager</td>
<td>People</td>
<td>Carry</td>
</tr>
<tr>
<td>Hoping</td>
<td>Explain</td>
<td>Together</td>
</tr>
<tr>
<td>Approach</td>
<td>Believe</td>
<td>Current</td>
</tr>
<tr>
<td>Tongue</td>
<td>Rougher</td>
<td>Raise</td>
</tr>
<tr>
<td>Happier</td>
<td>Laziest</td>
<td>Leave</td>
</tr>
</tbody>
</table>
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.
Figure 1. Spelling accuracy for students in Experiment 1. Error bars are standard error.
Figure 2. Spelling accuracy for students in Experiment 2. Error bars are standard error.
Figure 3. Spelling accuracy for students in Experiment 3. Error bars are standard error.
### Table 1. Outcomes of Omnibus Analyses of Variance for Experiments 1 – 3.

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F</th>
<th>MSE</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1 – 2 (test: post-test vs. retention) x 2 (practice: quizzing vs. rainbow writing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect of test</td>
<td>1,13</td>
<td>16.64</td>
<td>73</td>
<td>.001</td>
<td>.56</td>
</tr>
<tr>
<td>Main effect of practice</td>
<td>1,13</td>
<td>12.07</td>
<td>96</td>
<td>.004</td>
<td>.49</td>
</tr>
<tr>
<td>Interaction</td>
<td>1,13</td>
<td>0.07</td>
<td>103</td>
<td>.797</td>
<td>.01</td>
</tr>
<tr>
<td>Experiment 2 – 2 (test: pre- vs. post-) x 2 (practice: quizzing vs. rainbow writing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect of test</td>
<td>1,15</td>
<td>32.38</td>
<td>243</td>
<td>&lt; .001</td>
<td>.68</td>
</tr>
<tr>
<td>Main effect of practice</td>
<td>1,15</td>
<td>.47</td>
<td>163</td>
<td>.504</td>
<td>.03</td>
</tr>
<tr>
<td>Interaction</td>
<td>1,15</td>
<td>11.04</td>
<td>62</td>
<td>.005</td>
<td>.42</td>
</tr>
<tr>
<td>Experiment 3 – 2 (test: pre- vs. post-) x 2 (practice: quizzing vs. rainbow writing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect of test</td>
<td>1,11</td>
<td>33.00</td>
<td>164</td>
<td>&lt; .001</td>
<td>.75</td>
</tr>
<tr>
<td>Main effect of practice</td>
<td>1,11</td>
<td>5.00</td>
<td>220</td>
<td>.047</td>
<td>.31</td>
</tr>
<tr>
<td>Interaction</td>
<td>1,11</td>
<td>19.81</td>
<td>88</td>
<td>.001</td>
<td>.64</td>
</tr>
<tr>
<td><strong>Retrieval Practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 2 – 2 (day: first vs. second) x 2 (retrieval attempt: first vs. second)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect of day</td>
<td>1,13</td>
<td>12.63</td>
<td>96</td>
<td>.004</td>
<td>.49</td>
</tr>
<tr>
<td>Main effect of retrieval attempt</td>
<td>1,13</td>
<td>12.55</td>
<td>205</td>
<td>.004</td>
<td>.49</td>
</tr>
<tr>
<td>Interaction</td>
<td>1,13</td>
<td>5.44</td>
<td>84</td>
<td>.036</td>
<td>.30</td>
</tr>
<tr>
<td>Experiment 3 – 2 (day: first vs. second) x 2 (retrieval attempt: first vs. second)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect of day</td>
<td>1,11</td>
<td>20.73</td>
<td>79</td>
<td>.001</td>
<td>.65</td>
</tr>
<tr>
<td>Main effect of retrieval attempt</td>
<td>1,11</td>
<td>22.30</td>
<td>84</td>
<td>.001</td>
<td>.67</td>
</tr>
<tr>
<td>Interaction</td>
<td>1,11</td>
<td>0.00</td>
<td>45</td>
<td>1.000</td>
<td>.00</td>
</tr>
</tbody>
</table>
Table 2. Descriptive Statistics for the Self-Report Questionnaires for Experiments 1 – 3.

<table>
<thead>
<tr>
<th></th>
<th>Future Practice</th>
<th>Liking</th>
<th>Learning</th>
<th>Liking</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainbow</td>
<td>28.60%</td>
<td>7.10%</td>
<td>28.60%</td>
<td>2.71</td>
<td>1.29</td>
</tr>
<tr>
<td>RP</td>
<td>71.40%</td>
<td>92.90%</td>
<td>71.40%</td>
<td>3.79</td>
<td>1.86</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>31.30%</td>
<td>25%</td>
<td>37.50%</td>
<td>1.81</td>
<td>1.87</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>68.80%</td>
<td>75%</td>
<td>62.50%</td>
<td>2.87</td>
<td>2.87</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>50%</td>
<td>33.30%</td>
<td>33.30%</td>
<td>2.25</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>66.70%</td>
<td>66.70%</td>
<td>3.25</td>
<td>3.58</td>
</tr>
</tbody>
</table>

Note. “RP” stands for retrieval practice.
Table 3. Spelling accuracy during retrieval practice.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Day 1</th>
<th>Day 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RP1</td>
<td>RP2</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>71%</td>
<td>91%</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>52%</td>
<td>64%</td>
</tr>
</tbody>
</table>

*Note.* “RP” stands for retrieval practice.