Author Manuscript

The effects of teacher read-alouds and student silent reading on predominantly bilingual high school seniors' learning and retention of social studies content

Deborah K. Reed, **Elizabeth Swanson**, **Yaacov Petscher**, and **Sharon Vaughn**
Florida State University, Florida Center for Reading Research

The University of Texas at Austin, Meadows Center for Preventing Educational Risk

Abstract

Teacher read-alouds (TRA) are common in middle and high school content area classes. Because the practice of reading the textbook out loud to students is often used out of concern about students' ability to understand and learn from text when reading silently (SR), this randomized controlled trial was designed to experimentally manipulate text reading while blocking on all other instructional elements to determine the relative effects on learning content. Predominantly Spanish–English bilingual twelfth-graders (*n* = 123) were randomly assigned to either a TRA or SR condition and provided 1 week of high quality instruction in US history. Daily lessons included teaching key terms in the passage, previewing text headings, and conducting comprehension checks. Results of immediate, 1-week delayed, and 1-month delayed assessments of content learning revealed no significant differences between the two groups. Students were also asked to rate the method of reading they believed best helped them understand and remember information. Students in the SR condition more consistently agreed that reading silently was beneficial. Findings suggest low performing adolescents of different linguistic backgrounds can learn content as well when reading appropriately challenging text silently as when the teacher reads the text aloud to them.

Keywords

Read-aloud; Silent reading; Adolescents; Content learning; Bilingual

---

At all grade levels, classrooms are comprised of students representing a range of ability levels and language backgrounds. However, findings from research with middle school students reveal that, in typical practice conditions, those who have reading difficulties experience an ever widening gap in performance compared to average and above average readers (Vaughn et al., 2012). This suggests that the heterogeneity of classes increases over time and, by high school, is far greater than in the primary grades. Unfortunately, secondary teachers—particularly content area teachers—report feeling ill equipped to provide appropriate instruction for students with reading and academic difficulties (Kosanovich,

D. K. Reed (&) Y. Petscher: Florida Center for Reading Research, Florida State University, 2010 Levy Ave., Ste 100, Tallahassee, FL 32310, USA, dkreed@fcrr.org

Reed, & Miller, 2010; Platt, Harper, & Mendoza, 2003). What has been described as a low sense of self-efficacy (Tschannen-Moran & Hoy, 2001) can lead teachers to feel overwhelmed by the challenge of ensuring all students are able to meet the content standards of accountability measures (Reed, 2009).

These concerns are not unfounded. High percentages of adolescents are not demonstrating: (a) the ability to read grade-level texts proficiently (National Center for Education Statistics, 2010), or (b) preparedness for post-secondary literacy requirements (Achieve, Inc., 2005; ACT, 2009). At least a portion of the student samples analyzed for these reports were not native English speakers. Other adolescents reportedly lack the motivation or willingness to read for school purposes, despite possessing an ability to comprehend the texts (Alveramann, 2003).

In an attempt to ensure that students have access to the information in subject matter textbooks, many content area teachers in middle and high schools avoid having students read the text by spending most of the instructional time on activities or teacher read-alouds to the class (McCulley et al., 2012). In a survey of middle school teachers, 72.2 % of respondents confirmed they spent instructional time reading aloud to their students (Ariail & Albright, 2006). The authors reported those not implementing the practice were most often teachers of elective courses such as art, music, and physical education.

Theoretical basis of read-aloud

Some claim that one way to improve knowledge acquisition in the content areas is to compensate for possible deficiencies in decoding by reading aloud to students. However, this theoretical position assumes that decoding and comprehension are linear in nature and by compensating for decoding alone comprehension is improved. This ''bottom up'' model (Gough, 1972) has been proven inadequate to explain how reading comprehension develops (Rumelhart, 1977; Stanovich, 1980). Instead, according to the simple view of reading (Hoover & Gough, 1990), two components are necessary for learning new information from text: decoding and language comprehension. The two components are interrelated and develop at the same time, so compensating for one may not have the intended benefit of boosting the other. In other words, compensating for decoding difficulties alone might lead to improved comprehension, but only if language comprehension is adequate. Support for the simple view has also been found in studies with Spanish–English bilingual students (Hoover & Gough, 1990; Mancilla-Martinez, Kieffer, Biancarosa, Christodoulou, & Snow, 2011; Nakamoto, Lindsey, & Manis, 2008; Proctor et al., 2005).

Although a recent synthesis and meta-analysis found significant, positive effects of storybook read-alouds on the reading outcomes of 3- to 8-year-olds (Swanson et al., 2011), experimental research with adolescents has focused more on using read-alouds as a testing accommodation for students with identified reading disabilities (e.g., McKevitt & Elliott, 2003). Not evident in the extant literature is whether teacher read-alouds of instructional text are effective at facilitating content learning in a general education classroom.

Read-aloud in content area classes

As an alternative to having students take turns orally reading the text, content area teachers might read to the class. When used with the main textbook for the course, *teacher read-alouds* are also referred to as *listening while reading* because students are expected to follow along silently in their own books as they listen to the teacher read it (Dowhower, 1987). This format has been observed (McCulley et al., 2012) and reported (Ariail & Albright, 2006) as a common practice in middle and high school classes, particularly in English language arts and social studies classes. Those who support or criticize it offer varying rationale as described in the next sections.

Potential benefits of teacher read-alouds

Many of the supports for reading aloud to adolescents are anecdotal or descriptive in nature, often predicated on the belief that the practice enables access to content for those students with limited decoding ability. For example, surveys conducted with middle school teachers (Ariail & Albright, 2006) and middle school students (Ivey & Broaddus, 2001) indicate teacher read-alouds are believed to support students' comprehension and motivate them to complete assignments. Anecdotal reports also suggest adolescents are more engaged and can better access new vocabulary when teachers spend time reading aloud picture books related to science and social studies instruction (Albright, 2002; Braun, 2010).

Of the few intervention studies investigating the effects of read-alouds on adolescents' comprehension, four specifically sampled middle and high school students identified with special needs such as a reading disability (Schmitt, Hale, McCallum, & Mauck, 2011; Skinner, Robinson, Adamson, Atchison, & Woodward, 1998), emotional behavioral disorder (Hale et al., 2005), or moderate to severe developmental disabilities (Mims, Hudson, & Browder, 2012). Among these studies, three found positive effects favoring students following along while the teacher read the text: two quasi-experimental studies comparing the read-aloud treatment to a silent reading condition (Hale et al., 2005; Schmitt et al., 2011) and one single case design (Mims et al., 2012). However, the fourth study, a quasi-experimental design, found no significant difference between the teacher read-aloud and silent reading conditions on the overall comprehension performance of students (Skinner et al., 1998).

It should be noted that the sample sizes in all four studies described above were small, ranging from 4 to 32 students. Moreover, no intervention research has compared the reading comprehension of general education and/or bilingual adolescents assigned to a teacher read-aloud or silent reading condition. This is an interesting void in the extant literature considering the use of read-alouds has been recommended in practitioner-oriented publications (e.g., Opitz & Guccione, 2009) as a way to support the literacy development of English language learners.

Potential problems with teacher read-alouds

Similar to the support for reading aloud, much of the criticisms of the practice are conceptually rather than empirically based. For example, listening to a teacher read the text

aloud is described by Smith (2007) as disengaging and superficial because it communicates to students that the only purpose for looking at the printed page is to obtain information, rather than to recursively construct meaning by grappling with the ideas. Anecdotal support is offered for building ninth-graders' reading stamina through independent silent reading rather than whole class read-alouds (Gulla, 2012). This latter publication is representative of a line of research on sustained silent reading and self-selected reading, which are often presented as the only forms of silent reading that are more motivating and enriching alternatives to whole class read-alouds (Collins, 1980; Summers & McClelland, 1982).

Although silent reading of required materials is not consistent with allowing students to choose their own texts to read independently, it is based on a similar premise that becoming a skilled reader is associated with spending considerable time with eyes on print, practicing reading and building conceptual knowledge (Mol & Bus, 2011). Through active reading, students learn to process semantic, orthographic, and syntactic patterns (Adams, 2009). Conversely, if students can passively rely on the teacher to process the text, they might develop a sense of learned helplessness in which they maintain low expectations for their reading abilities and do not persist when encountering difficulties (Butkowsky & Willows, 1980). However, there are inconclusive results from research examining whether oral or silent reading contributes to better reading comprehension performance (Hale et al., 2011; McCallum, Sharp, Bell, & George, 2004). Moreover, the studies primarily have focused on testing rather than learning situations and have not included large numbers of students who are not native English speakers.

Purpose and research questions

Given the lack of extant literature with adolescents and those from different linguistic backgrounds, the purpose of the present research was to examine the effects that the method of text reading (teacher read alouds or student silent reading) had on the learning and retention of content among high school seniors who are predominantly Spanish–English bilingual. The research was designed to answer the questions: (a) What effect does teacher read aloud or student silent reading have on grade 12 students' social studies content learning immediately as well as at 1-week and 1-month delay? and (b) What method for reading text do students perceive to be more beneficial?

The expectation was that the findings from this study could guide content area teachers who have increasing demands for providing students with access to multiple genres of high level text (e.g., Common Core State Standards). Furthermore, if the findings from this exploratory study supported enhanced learning outcomes from teacher read-alouds, then teachers would have justification for reading text to students. If the findings were equivalent or in favor of silent reading, we would interpret these findings as supportive of students reading the text independently because active reading (Adams, 2009) is the more desirable practice as long as learning is not compromised.

METHOD

Participants

Students—The sample was drawn from six senior-level English classes, periods 1–3, taught by two different teachers at an urban high school. The campus was situated less than a mile from the US-Mexico border, and more than 90 % of the students qualified to receive free/reduced-price lunch. Of the 125[1] twelfth-graders (average age = 17.6) enrolled in the six classes, consent was received for 123 students, 77 (63 %) of whom were males and 112 (91 %) of whom were Hispanic. The remainder of the students (9 %) were identified as multi-racial.

All 123 students were considered proficient in English, 8 students (7 %) were receiving special education services, and 4 (3 %) had not yet passed the state's exit-level accountability measure that served as a gatekeeper for high school graduation. The test is only given in English. Because students routinely alternated between speaking English and Spanish, they were specifically queried as to which language they learned first. Eighty-eight students (72 %) reported that Spanish was their first language, but they considered themselves to be bilingual.[2]

Students were randomly assigned to the Teacher Read Aloud (TRA) or Silent Reading (SR) conditions. The language survey was administered after students had been assigned to treatment, so there was a difference in the language backgrounds represented in the groups. Eleven students (18 %) reported being native English speakers in the TRA condition, and 21 (34 %) reported being native English speakers in the SR condition. An examination of outcomes by first language was planned as part of the data analysis.

Random assignment at the student level necessitated moving many students to a classroom different from the one to which they usually reported for the senior English class period. The two rooms used for instruction were located in adjacent portable buildings, and a research assistant stood in between the two portables at the start of all three class periods each day to direct students to the correct room. The instructors also took roll inside the room prior to beginning instruction.

Instructors—The school requested that the regularly assigned teacher-of-record be present in the room throughout the study, but they were not actively involved in the treatment. Rather, two graduate students enrolled in a Master's level education program delivered the instruction. Both were novice teachers with <3 years of experience. One graduate student, a 44-year-old Hispanic female, was assigned to the TRA condition and the other, a 58-year-old African American male, was assigned to the SR condition. Each was trained individually by the first author on the procedures for his or her condition during a 4-h session. The graduate students (hereafter referred to as the instructors) were then given 2 days to practice

---

[1]A power analysis using Optimal Design (Raudenbush et al., 2011) indicated that a total sample size of 105 students would be needed to detect a 0.35 effect size with 80 % power using a test between means with alpha at 0.05.

[2]*Bilingual* should not be confused with *English language learner*. The students in this school ordinarily were immersed in both languages from childhood, and most of those who identified English as their first language also spoke Spanish and considered themselves to be bilingual.

on their own before meeting with the researcher again to be quizzed on the procedures. Both demonstrated 100 % accuracy in their treatment condition.

Instructional conditions and procedures

Instructional materials—Although the sample was drawn from senior English classes upon the recommendation of the school principal, the content of the lessons was US History prior to 1900. This was selected because students would not have been exposed to the content for about 4 years, based on state curriculum standards and course sequences. The research site's location on the US-Mexico border frequently resulted in a student population that had been educated at least part of the time in Mexican schools, usually up to but not exceeding middle school. Therefore, the specific events chosen for the lessons focused on the Mexican–American War of 1846–1848, which would have been content addressed in Mexican schools as well. In addition, the names and locations within the passages drew equally from US and Mexican historical figures and sites.

So that findings from this study could be interpreted based on whether differences were associated with TRA or SR, both conditions utilized the same printed materials for lessons. These consisted of a daily packet containing vocabulary previews, passage segments separated by comprehension checks, and a daily test of content knowledge. The vocabulary previews introduced one or two words per day. Each word was printed on its own page in large font with the Spanish equivalent in parentheses directly underneath. After the definition, a picture or graphic was provided to illustrate the term. For example, a picture of Stephen F. Austin appeared with the term *impresario*. Finally, there were two sentences applying the word in context and two questions for peer partners to more deeply process the vocabulary such as: ''How did Stephen F. Austin's ability so speak Spanish help him become a successful *impresario*?'' Following the vocabulary preview pages, the 400–850 word passages were printed in segments with an appropriate title and subheadings. Given the diverse language backgrounds of the students and research indicating the importance of texts being at an appropriate difficulty level to facilitate comprehension (Adams, 2009), the passages were written slightly below students' current grade-level (10–12th grade level on Flesch-Kincaid scale). At section breaks, a comprehension check appeared as a *stop and think* question with 4–6 questions included per day. A separate page was provided for recording notes and responses to each question.

After the complete passage, a set of five multiple choice and five open-response items on the content were provided. All questions were factually based and required only low level inferences. The five sessions lasted 50 min each for a total of 250 min of instructional time, or one academic week of classes.

Teacher read-aloud (TRA)—The TRA treatment was delivered by the female instructor and adhered to principles of instruction as validated in previous research (Vaughn et al. 2009). Each lesson began with the vocabulary previews during which the instructor pronounced and defined the target vocabulary word, working with one word at a time. Students were shown a picture or graphic representing the word, including a description of how the picture/graphic related to the word. The instructors then demonstrated key ideas

about the word with a few example sentences of its application. Following the instruction, students were invited to *turn and talk* to a peer partner about the questions designed for deeper processing of the vocabulary.

After pre-teaching of the vocabulary words, the instructor guided students in looking over the passage to draw attention to the subheadings. She read the subheadings and the first sentence of the paragraph under each subheading. The class then returned to the beginning of the passage, and students were directed to follow along as the instructor read the text out loud. Each section was read within a set timeframe that was designated on the teacher's version and monitored with a timer. At the end of the time/section, the instructor read the comprehension check question (e.g., How did Santa Ana rise to power?) and invited the students to talk briefly with a peer partner about the information before making notes on the provided paper. The instructor monitored student work during the 1-min time allotment and provided feedback as necessary. When all sections were completed, the instructor collected the materials before distributing the 10-item daily quizzes.

Silent reading (SR)—Instruction in the SR condition utilized the same materials and high quality lesson components as the TRA condition. The only difference was that students independently read the passage sections silently for the same amount of time established for the instructor to read aloud in the contrasting TRA condition. The instructor still directed the vocabulary pre-teaching, read the subheadings and first sentences, read the comprehension check questions for each section, provided feedback to students during their 1-min allotment for discussing the information, and monitored the daily quizzes.

Fidelity

The TRA and SR instructors wore digital recording devices during each session for the purposes of monitoring treatment integrity. A random sample of 50 % of the recordings from each condition was evaluated using a fidelity code sheet adapted from (Vaughn et al., 2011) recent work. The second author, an experienced senior researcher, and a senior research assistant listened to the audios and completed the code sheet. Scores were compared and inter-rater agreement was 100 %. Adherence to procedures in the TRA treatment was judged to be 100 %; and in the SR treatment, 90 %. Among both the TRA and SR treatment groups, fidelity ratings indicated high levels of implementation for all components of the intervention.

Measures

Pre-tests—A number of pretest measures were used to establish the baseline equivalence of the treatment groups. The measures, which are known correlates of reading comprehension, are described below.

Kaufman brief intelligence test-2 (KBIT-2; Kaufman & Kaufman, 2004): The KBIT-2 is an individually administered assessment of receptive vocabulary and general information (e.g., nature, geography). The participant is required to choose one of six illustrations that best corresponds to an examiner question. Internal consistency values for the subtests and

the composite range from 0.87 to 0.95, and test– retest reliabilities range from 0.80 to 0.95, in the age range of the students in this study.

Group reading assessment and diagnostic evaluation listening comprehension (Wilson, 2001): The GRADE is a diagnostic reading test that determines what skills students Pre-K through 12th grade have mastered. The GRADE is group-administered, norm-referenced, and based on scientific research. For the listening comprehension subtest, students listen to a sentence read orally and decide which picture best matches the sentence. Reliability coefficients for alternate form and test–retest were in the .90 range. Concurrent and predictive validity was assessed using a variety of other standardized reading assessments and was found to be adequate.

Test of silent contextual reading fluency (TOSCRF; Hammill, Wiederholt & Allen, 2006): The TOSCRF measures the speed with which students can recognize the individual words in a series of printed passages without punctuation or spaces between words. Passages become progressively longer and more difficult in context, vocabulary, and grammar. The TOSCRF reports alternate form and test–retest median reliability coefficients range from .82 to .88.

Woodcock Johnson III spelling (WJ-III; Woodcock, McGrew & Mather, 2001): The WJ-III Spelling subtest assesses students' ability to spell words of increasing complexity. Reliability coefficients exceed 0.90 for students in upper elementary.

Content/background knowledge assessment: This researcher-developed measure was designed to assess background knowledge prior to intervention and content knowledge after the 5-day treatment period. The 15 items were drawn from the content covered during the entire 5-day treatment. Ten matching items were worth one point each and five open-response items were worth two points each, totaling a minimum score of zero and a maximum possible score of 20 points. Open-response items required low level inferences, and scoring was designed to allow for partial credit (one of the two possible points) if students wrote a portion but not all of the answer. Test administration was untimed.

Testing procedures—The group-administered tests were given by the two graduate students serving as instructors. One was assigned to each classroom and tested their treatment classes of no more than 24 students per period. The instructors were trained by the first author during a 4-h training session that included practice administrations on undergraduate students. Because there was a 1-week lapse between the initial training and the test administration, the instructors had to pass a fidelity check in the morning prior to testing the high school students. Both instructors demonstrated 100 % accuracy in the procedures.

The KBIT was individually-administered by 20 college undergraduates who were pre-service teacher candidates. They were trained by the first author during a 2-h training session that included practice administration on peer partners. Because there was a 5-day delay between their training and the test administration, they were required to pass a fidelity check in the morning prior to testing the high school students. Those who did not

demonstrate accuracy were given refresher trainings and re-checks until they demonstrated 100 % accuracy in the procedures. Due to space limitations at the school, the twelfth-graders were tested in two waves each class period. Testers who were not actively administering the KBIT were either monitoring students before/after testing as they completed a demographic survey or monitoring the administrations taking place. The two graduate-level student instructors also served as monitors of the test administration. Testing occurred over 2 days. Make-up testing for students who missed one or more pre-tests was scheduled to take place in the afternoon of the original testing days or in the morning of the first day of intervention.

All test documents were double-scored by the two instructors and one undergraduate student to ensure items were marked correctly and all tabulations were accurate. A third graduate student, who served as a research assistant, converted raw scores to standard scores when appropriate and entered all scores into an electronic database used for statistical analyses.

In addition to the measures administered as described above, students' scores on the state's criterion referenced assessment of reading comprehension were obtained from the school. At the time of the study, the *exit level* test was administered under state mandated procedures in spring of the junior year as part of the high school graduation requirements. Students earning a scale score of 2,100 or greater were considered to have passed the test. Those whose scale score fell below 2,100, including four students in the study sample (3 %), were offered opportunities to retake the test at scheduled administrations until they passed. Internal consistency reliabilities were in the high .80 s to low .90 s range. Scale scores were equated using the Rash model, and the resulting classification accuracy ranged between 81.7 and 95.4 %. The archival scale scores were entered by a research assistant into an electronic database and used for equating the groups.

Post-tests—Immediately after the last treatment session and again 1-week and 1-month later, students took the group-administered test of content knowledge to evaluate their learning and retention of the lesson material. The tests were structured the same as the pre-test of content knowledge with 10 matching items worth one point each and 5 open-response items worth two points each, allowing for total scores ranging from 0 to 20. The instructors administered all post-tests to their treatment classes with no more than 24 students per period. Testing lasted 15–20 min. All documents were double-scored by the two instructors and one undergraduate student to ensure items were marked correctly and all tabulations were accurate.

Social validity—During the 1-week delayed post-test session, students were asked to complete a short survey on which they rated the method of reading (silently or teacher read-aloud) that they believed helped them understand and remember passage information best. Ratings were on a 1–6 Likert scale from *strongly disagree* to *strongly agree*. Responses were entered by the research assistant into the electronic database for analysis.

Data analytic procedures

To establish the equivalency of the TRA and SR groups prior to treatment, a series of one-way analysis of variance (ANOVA) was used to evaluate the extent to which treatment and control group participants significantly differed in their baseline performances on measures

of the KBIT, GRADE, TOSCRF, WJ-III Spelling, and the researcher developed content knowledge composite score. Next, a series of analysis of covariance (ANCOVA) were conducted to determine whether students' immediate, 1-week, and 1-month delayed post-tests of content learning differed between the TRA and SR conditions when controlling for their background knowledge of the content, and their performances on all baseline measures. Multiple AN(C)OVAs were preferred over a traditional MAN(C)OVA analysis as Huberty and Morris (1989) demonstrated that the MAN(C)OVA should only be utilized under specific, restrictive conditions. As such, with the multiple univariate AN(C)OVAs, a linear step-up procedure was used to control for the false-discovery rate (Benjamini & Hochberg, 1995). Finally, students' mean ratings of the method of reading (i.e., read-aloud or silent reading) that best helped them learn and remember content were aggregated into percentage of responses at the extremes: moderately to strongly agree/disagree.

RESULTS

Descriptive statistics and multiple imputation

Means and standard deviations for students' pre- and post-test scores for all measures prior to the missing data analysis are provided in Table 1. Although students' scale scores on the state criterion referenced test of reading comprehension suggest they are of average ability (i.e., the means are above the 2,100 cut score), the mean standard scores on KBIT, TOSCRF, and WJ-Spelling indicate this is a low performing sample. In other words, the participants exemplify many of the characteristics of students for whom teachers believe reading the text aloud is beneficial (Ariail & Albright, 2006).

A preliminary review of the data revealed that data were missing on the composite variables (the primary outcome) and the pretest covariates as follows: pretest (8 % missing), immediate post-test (20 % missing), 1-week delayed post-test (18 % missing), 1-month delayed (19 % missing), the KBIT (19 % missing), the GRADE (10 % missing), the WJ-Spelling (10 % missing), and the TOSCRF (8 % missing). Little's test on data missing completely at random (MCAR) was run using the Missing Value Analysis (MVA) package in IBM SPSS v21 (2012). The null hypothesis for the Little's MCAR test (Little, 1988) is that the data are missing completely at random, thus, a p value >0.05 indicates the data satisfy the MCAR requirement. The MVA model yielded a non-significant χ^2 value, $\chi^2(105)$ = 106.26, p = 0.447, thus verifying that the data were MCAR. As such, a multiple imputation of the data was carried out with the multiple imputation package in SAS 9.3 using 10 imputations so that cases would not be deleted for the inferential analyses. A summary of the descriptive statistics for the imputed composite score data are provided in Table 2, and demonstrated that the means and standard deviations prior to and post imputation were quite similar. Cohen's d (Cohen, 1988) statistic was used to gauge the strength of the difference between the distributions of the original and imputed data, with values ranging from d = −0.002 to 0.05 observed.

Analysis of variance

The imputed data were analyzed using a combination of general linear modeling with the MIANALYZE package in SAS 9.3. Because a multiply imputed data set requires one to

individually analyze each imputation, the MIANALYZE procedure summarizes the results across the imputations. The preliminary pre-test ANOVAs to establish baseline equivalence between the read aloud and silent reading conditions across the areas of verbal knowledge, listening comprehension, silent contextual reading fluency, spelling, and background knowledge revealed that no statistically significant differences were observed across any of the measures (Table 3). Further, the two conditions were equivalent on the immediate post-test evaluation ($p = 0.631$), the 1-week evaluation ($p = 0.881$), and the 1-month evaluation ($p = 0.958$), when controlling for the pre-test covariates (Table 4).

The lack of statistically significant differences was true for the combined score of all 10 questions on the test of content knowledge as well as the separated scores of the ten matching and five open-response items. The lack of differences on the matching items may be attributable to ceiling effects because the items were not designed to assess a range of difficulty levels. Thus, the scale of 0–10 (raw score values) was not sensitive enough to measure students' content knowledge beyond literal recognition. Despite the possible ceiling effect, variation in students' scores still existed. This indicates the matching items contributed to individual differences on the composite score; therefore, they were retained for analysis of the composite score.

There also was an interesting trend in performance at post-test. As can be seen in Fig. 1, students in both groups scored near ceiling on the matching items (mean ranged from 88 to 96 % correct) but performed less well on the open-response items (mean ranged from 42 to 58 % correct). Although the differences were not significant, students in the SR condition tended to earn slightly more points than students in the TRA condition on matching items but slightly fewer points than students in the TRA condition on open-response items. Whereas student performance on open-response items steadily decreased over time, students in the SR condition demonstrated a slight increase in performance at the 1-week delayed posttest of matching items and maintained their immediate post-test performance on the matching items 1 month later.

In addition, there were small but practically important effects favoring the TRA condition when looking at the means of the immediate post-test for the first language subgroups (Table 5). After imputing missing data, the bias-corrected Cohen's *d* effect sizes (Cohen, 1988) were small ($d = 0.18$) for the students reporting Spanish as their first language and moderate ($d = 0.46$) for the students reporting English as their first language. Effect sizes of the 1-week and 1-month delayed post-tests were less than 0.10, which were closer to the pretest effect sizes of approximate 0.10 for both languages.

More distinction between the treatment groups was found in comparing students' ratings of the method of reading they believed best helped them understand and remember information (see Table 6 for the percent moderately or strongly agreeing/ disagreeing with the social validity statements). Specifically, a higher percentage of students in the TRA condition agreed they understood and remembered information better when the teacher read the text aloud (TRA = 59.5 %; SR = 46.0 %); and a higher percentage of students in the SR condition agreed when the statement targeted reading the text silently (SR = 67.5 %; TRA = 57.5 %). Moreover, students in the TRA condition were more likely to disagree with the

statement about understanding and remembering information when reading silently (TRA = 8.2 %; SR = 4.1 %), as were students in the SR condition when the statement targeted the teacher reading aloud (SR = 10.7 %; TRA = 4.8 %). Results within the TRA group of the percent agreeing with both the teacher read-aloud and silent reading statements were roughly equivalent: 59.5 and 57.5 %, respectively. Whereas, the SR responses demonstrated more separation between the teacher read-aloud and silent reading statements: 46.0 and 67.5 %, respectively.

DISCUSSION

Because reading aloud to students is a common but understudied practice in middle and high school content area classrooms (McCulley et al., 2012), we designed this randomized controlled trial to explore the relative effects of teacher read-alouds versus students reading silently on the content learning outcomes of predominantly bilingual high school seniors who were low performing on KBIT, TOSCRF, and WJ-Spelling. In contrast to descriptive and anecdotal reports that adolescents and English language learners can better comprehend texts and access new vocabulary when teachers read aloud (Ariail & Albright, 2006; Braun, 2010; Opitz & Guccione, 2009), our results suggest students can learn and retain information and content-based vocabulary equally well when they read informational text silently as when the teacher reads the text to them. This is consistent with the findings of Skinner et al. (1998), who employed a smaller, monolingual English sample and informational narrative texts. This is also consistent with the simple view of reading that theorizes both decoding and language comprehension are necessary for comprehension. The two components develop and interact concurrently; therefore, compensating for one (i.e. decoding by reading aloud) may not necessarily improve comprehension outcomes.

The moderate effect sizes favoring the TRA condition at immediate post-test for students reporting English as their first language are interpreted with caution for a couple reasons. First, the subgroup of students demonstrating moderate effects was counterintuitive to what might be hypothesized given the common rationale for using read-alouds. As noted above, read-alouds are intended to make the text more comprehensible to students who might have difficulty with the language (Ariail & Albright, 2006; Braun, 2010; Opitz & Guccione, 2009), but it was the students who reported English as a first language that seemed to benefit more. Second, the improvements were not maintained 1 week or 1 month later and rapidly returned to a null effect.

A possible explanation for the overall comparable performance of students in the TRA and SR conditions is that both groups received high quality instruction supportive of the text (Vaughn et al., 2009). Previous studies have found students with reading difficulties can successfully comprehend when reading read silently if teachers implement practices similar to those utilized in our treatments: (a) monitor student reading, (b) monitor student understanding through questioning, (c) instruct students to provide text-based evidence for their answers, and (d) re-teach content when necessary (Applebee, Langer, Nystrand & Gamoran, 2003; Beck & McKeown, 2006; Block & Reed, 2005; McKeown, Beck & Blake, 2009; Vaughn et al. 2012). In other words, the compensatory act of reading the text aloud may not have added any benefit above what was gained by teaching students how to

comprehend and learn from that text. To the extent that actively engaging with print and practicing reading while building conceptual knowledge is important to becoming a skilled reader capable of independent learning (Mol & Bus, 2011; Smith, 2007), teacher read-alouds might be considered counterproductive rather than merely unnecessary.

However, another justification for the use of read-alouds is that students lack the motivation to read required, school-based texts silently (Albright, 2002; Ivey & Broaddus, 2001). Therefore, we also examined student perceptions of the benefits associated with the TRA and SR conditions. As opposed to being disengaged, participants in the SR group more consistently demonstrated a belief in the benefit of silent reading than the participants in the TRA group demonstrated a belief in the benefit of teacher read-aloud. In fact, the nearly equivalent percentages of agreement with teacher read-aloud and silent reading suggests the TRA students were ambivalent about which type of instruction better helped them understand and remember information. Responses were open to bias given students were not blind to the condition they experienced. However, the pattern of responses across the two conditions reflect stronger feelings of reading self-efficacy were engendered when exposed to high quality instruction accompanying silent reading. Much of the extant literature presents students' self-selected texts as the only appropriate alternative to whole class read-alouds (e.g., Collins, 1980; Gulla, 2012; Summers & McClelland, 1982). Hence, *student choice* often becomes synonymous with *student motivation*. Results presented here seem to indicate that motivation to read for school purposes is more multifaceted and, perhaps, associated with conditions in which students feel successful at understanding and learning from required texts.

A more troubling finding about student performance was evident in the differential scores by question type on the content measure. That is, students in both groups scored highly on matching items (88–96 % correct on post-tests) but exhibited difficulty with open-response items (42–58 % correct on post-tests). The open-response items included on the assessment were not overly burdensome: they required only low-level inferences and answers ranging from one to three sentences in length. For example, students read a passage about the Runaway Scrape that described the effect of the Mexican–American War on Texans. The related question on the content measure read: ''Think about the Runaway Scrape during the Texas Revolution. How were regular families in Texas affected by the war?'' To answer the question, students had to put together information provided in different parts of the text. Even with liberal scoring, allowing for partially correct responses, students only earned on average about half of the possible points. A stricter scoring of responses likely would have resulted in an average of zero points being awarded for the open-response items.

Indeed, open-ended questions place more demands on skilled reading such as the ability to synthesize information or draw conclusions based on information from the text. These *close reading* skills are expected of secondary students (see the Common Core State Standards for English Language Arts) but are rarely taught (Edmonds et al., 2009; McCulley et al., 2012; Pressley, 2000; Urquhart, 2002). Although basic inference ability may be intuited, it is evident that as children get older, drawing inferences from text becomes more difficult and necessitates formal instruction (e.g., Oakhill, Yuill, & Parkin, 1986).

To facilitate close reading, teachers can engage students in discussion that encourages students to elaborate on their inferences and then return to the text to find evidence to support their conclusions (McKeown et al., 2009). Thus, questioning becomes the key teacher skill that drives effective discussion. Applebee, Adler, and Fihan (2007) contend it is necessary to move beyond literal questions (e.g., Who was Santa Anna?) to ones that encourage making inferences, evaluating information, and articulating ideas with technical vocabulary (e.g., Think about the Runaway Scrape during the Texas Revolution. How were regular families in Texas affected by the war?). Researchers have found that more frequent practice answering deep, inferential comprehension questions coupled with a return to the text for evidence to support answers is effective at improving students' abilities (e.g., Vaughn et al., 2012). Findings from the current study indicate a need to engage students in these types of discussions and associated written responses more often.

Limitations and directions for future research

The randomized design of the study is one of its strengths, particularly as an exploratory study conducted at the high school level where true experimental research is rare. However, the sample size was not powered to detect small effects, indicating a need to replicate the study with a larger sample. In addition, the sample was drawn from a predominantly bilingual Hispanic population. Because 72 % of students reported that Spanish was their first language, we expect levels of English proficiency varied across the sample. Future studies should include English proficiency data to examine whether group differences exist among students at different proficiency levels.

We were also precluded from examining the impact of the treatments on students' general comprehension abilities because the duration of the intervention (1 week) was too short to detect a change from pre- to post-test. In order to better test the simple view of reading, an intervention of longer duration is required to better measure the broader construct of general comprehension. Given that we interpret the null findings for content learning as supportive of having students actively engaged in reading text independently, further research should be conducted to confirm the benefits of the practice for improving students' reading ability as a transferable skill irrespective of the particular content being studied. It is likely that content area teachers care first and foremost about whether students' subject matter learning is compromised by particular reading practices, but middle and high school teachers are sensitive to meeting the multiple demands of accountability assessments (Reed, 2009). Instruction that is both efficient and effective at improving various academic outcomes might be more readily embraced. To that end, teacher read-alouds need to be researched over longer durations. Although we did not find a significant difference in performance on researcher developed measures after one unit of text, it is possible that there might be a cumulative effect of implementing the practice over time or assessing outcomes with standardized instruments.

Finally, there is a need for replication of the current study to determine if findings generalize to a more natural school setting. The same slightly below grade-level reading materials were used in the TRA and SR conditions, and researchers tightly controlled implementation of the treatment. It is unknown whether results would have been different had: (a) the passages

Author Manuscript

been taken from a school-adopted textbook, or (b) the vocabulary and comprehension instruction been prepared and implemented by the teachers of record. We intentionally did not use typical practice as a comparison because we wanted to control for quality of instruction and only manipulate the method of text reading. However, teachers may not find it feasible to implement all components of our SR treatment in the natural setting (Mastropieri & Scruggs, 2001; Platt et al., 2003). In addition, future research might evaluate clustering effects due to teacher.

We conceived of this study based on observations we and others have made of the reading practices commonly used in content area classrooms so, ultimately, we are concerned about generating findings that can translate back into practice. Our initial exploration suggests educators need to challenge the assumptions underpinning the instructional status quo in their schools, particularly in the era of implementing more rigorous standards aimed at college and career readiness. We did not find in the extant literature any indication that employers or college professors read-aloud manuals or textbooks at prevalence rates near those reported for middle and high school teachers (Ariail & Albright, 2006; McCulley et al., 2012). Therefore, further research is warranted to better understand what happens when different students are exposed to teacher read-alouds or silent reading in various content area classes using unique genres of text.

Acknowledgments

References

Achieve, Inc. Rising to the challenge: Are high school graduates prepared for college and work?. Washington, DC: Achieve, Inc; 2005. Retrieved from http://www.achieve.org/files/pollreport_0.pdf

ACT. The condition of college readiness. 2009. Retrieved from www.act.org

Adams, MJ. The challenge of advanced texts: The interdependence of reading and learning. In: Hiebert, EH., editor. Reading more, reading better. New York, NY: Guilford; 2009. p. 163-189.

Albright LK. Bringing the Ice Maiden to life: Engaging adolescents in learning through picture book read-alouds in content areas. Journal of Adolescent & Adult Literacy. 2002; 45:418–428.

Alveramann, DE. Seeing themselves as capable and engaged readers. Naperville, IL: Learning Point Associates; 2003. Retrieved from http://www.craftinc.org/literacy-e-books/seeing-themselves-ascapable-and-engaged-readers.pdf.

Applebee AN, Langer JA, Nystrand M, Gamoran A. Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. American Educational Research Journal. 2003; 40:685–730.

Applebee AN, Adler M, Fihan S. Interdisciplinary curricula in middle and high school classrooms: Case studies of approaches to curriculum and instruction. American Educational Research Journal. 2007; 44:1002–1039.

Ariail M, Albright LK. A survey of teachers' read-aloud practices in middle schools. Reading Research and Instruction. 2006; 45:69–89.

Beck, IL.; McKeown, MG. Improving comprehension with questioning the author. New York, NY: Scholastic; 2006.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995; 57:289–300.

Block, CC.; Reed, K. Effects of trade book reading on comprehension, vocabulary, fluency and attitudes (Research Report No. 110224). Charlotte, NC: Institute for Literacy Enhancement; 2005.

Braun P. Taking the time to read aloud. Science Scope. 2010; 34:45–49.

Butkowsky IS, Willows DM. Cognitive-motivational characteristics of children varying in reading ability: Evidence for learned helplessness in poor readers. Journal of Educational Psychology. 1980; 72:408–422. [PubMed: 7391337]

Cohen, J. Statistical power analysis of the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.

Collins C. Sustained silent reading periods: Effect on teachers' behaviors and students' achievement. Elementary School Journal. 1980; 81:108–114.

Dowhower SL. Effects of repeated reading on second-grade transitional readers' fluency and comprehension. Reading Research Quarterly. 1987; 22:389–406.

Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK, et al. A synthesis of reading interventions and effects on reading comprehension outcomes on older struggling readers. Review of Educational Research. 2009; 79:262–287. [PubMed: 20072704]

Gough, PB. One second of reading. In: Kavanagh, JF.; Mattingly, IG., editors. Language by ear and by eye. Cambridge, MA: MIT Press; 1972. p. 331-358.

Gulla AN. Putting the ''shop'' in reading workshop: Building reading stamina in a ninth-grade literacy class in a Bronx vocational high school. English Journal. 2012; 101:57–62.

Hale AD, Hawkins R, Sheeley W, Reynolds JR, Jenkins S, Schmitt AJ, et al. An investigation of silent versus aloud reading comprehension of elementary students using maze assessment procedures. Psychology in the Schools. 2011; 48:4–13.

Hale AD, Skinner CH, Winn BD, Oliver R, Allin JD, Molloy CCM. An investigation of listening and listening-while-reading accommodations on reading comprehension levels and rates in students with emotional disorders. Psychology in the Schools. 2005; 42:39–51.

Hammill, DD.; Wiederholt, JL.; Allen, EA. TOSCRF: Test of silent contextual reading fluency: Examiner's manual. Austin, TX: PRO-ED; 2006.

Hoover WA, Gough PB. The simple view of reading. Reading and Writing: An Interdisciplinary Journal. 1990; 2:127–160.

Huberty CJ, Morris JD. Multivariate analysis versus multiple univariate analyses. Psychological Bulletin. 1989; 105:302–308.

IBM Corp. IBM SPSS statistics for windows (Version 21) [Software]. Armonk, NY: IBM Corp; 2012.

Ivey G, Broaddus K. ''Just plain reading'': A survey of what makes students want to read in middle school classrooms. Reading Research Quarterly. 2001; 36:350–377.

Kaufman, AS.; Kaufman, NL. Kaufman brief intelligence test. 2nd ed. Upper Saddle River, NJ: Pearson; 2004.

Kosanovich, ML.; Reed, DK.; Miller, DH. Bringing literacy strategies into content instruction: Professional learning for secondary-level teachers. Portsmouth, NH: RMC Research Corporation, Center on Instruction; 2010.

Little RJA. A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association. 1988; 83:1198–1202.

Mancilla-Martinez J, Kieffer MJ, Biancarosa G, Christodoulou JA, Snow CE. Investigating English reading comprehension growth in adolescent language minority learners: Some insights from the simple view. Reading and Writing: An Interdisciplinary Journal. 2011; 24:339–354.

Mastropieri MA, Scruggs TE. Promoting inclusion in secondary classrooms. Learning Disability Quarterly. 2001; 24:265–274.

McCallum RS, Sharp S, Bell SM, George T. Silent versus oral reading comprehension and efficiency. Psychology in the Schools. 2004; 41:241–246.

McCulley, L.; Swanson, E.; Wanzek, J.; Vaughn, S.; Stillman, S.; Hairrell, A., et al. Text reading in secondary English language arts and social studies classes: An observation study; Poster presented at the Pacific Coast Research Conference; San Diego, CA. 2012 Feb.

McKeown MG, Beck IL, Blake RGK. Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. Reading Research Quarterly. 2009; 44:218–253.

McKevitt BC, Elliott SN. Effects and perceived consequences of using read-aloud and teacher-recommended testing accommodations on a reading achievement test. School Psychology Review. 2003; 32:583–600.

Mims PJ, Hudson ME, Browder DM. Using read-alouds of grade-level biographies and systematic prompting to promote comprehension for students with moderate and severe developmental disabilities. Focus on Autism and Other Developmental Disabilities. 2012; 27:67–80.

Mol SE, Bus AG. To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. Psychological Bulletin. 2011; 137:267–296. [PubMed: 21219054]

Nakamoto J, Lindsey KA, Manis FR. A cross-linguistic investigation of English language learners' reading comprehension in English and Spanish. Scientific Studies of Reading. 2008; 12:351–371.

National Center for Education Statistics. The nation's report card: Grade 12 reading and mathematics 2009 national and pilot state results (NCES 2011-455). Washington, DC: Institute of Education Sciences, US Department of Education; 2010.

Oakhill J, Yuill N, Parkin A. On the nature of the difference between skilled and less-skilled comprehenders. Journal of Research in Reading. 1986; 9:80–91.

Opitz, MF.; Guccione, LM. Comprehension and English language learners: 25 oral strategies that cross proficiency levels. Portsmouth, NH: Heinemann; 2009.

Platt E, Harper C, Mendoza MB. Dueling philosophies: Inclusion or separation for Florida's English language learners. TESOL Quarterly. 2003; 37:105–133.

Pressley, M. What should comprehension instruction be the instruction of?. In: Kamil, ML.; Mosenthal, PB.; Pearson, PD.; Barr, R., editors. Handbook of reading research. Vol. 3. Mahwah, NJ: Erlbaum; 2000. p. 545-561.

Proctor CP, Carlo M, August D, Snow C. Native Spanish-speaking children reading in English: Toward a model of comprehension. Journal of Educational Psychology. 2005; 97:246–256.

Raudenbush SW, Spybrook J, Congdon R, Liu X, Martinez A, Bloom H, et al. Optimal design software for multi-level and longitudinal research (Version 3.01) [Software]. 2011 Available from www.wtgrantfoundation.org.

Reed DK. A synthesis of professional development on the implementation of literacy strategies for middle school content area teachers. Research in Middle Level Education Online. 2009; 32:1–12. Retrieved from http://www.amle.org/Publications/RMLEOnline/Articles/Vol32No10/tabid/1953/Default.aspx.

Rumelhart, D. Toward an interactive model of reading. In: Dornic, S., editor. Attention and performance. Vol. 6. Hillsdale, NJ: Erlbaum; 1977. p. 573-603.

Schmitt AJ, Hale AD, McCallum E, Mauck B. Accommodating remedial readers in the general education setting: Is listening-while-reading sufficient to improve factual and inferential comprehension? Psychology in the School. 2011; 48:37–45.

Skinner CH, Robinson DH, Adamson KL, Atchison LA, Woodward JR. Effects of different listening-while-reading rates on comprehension in secondary students with reading deficits. Special Services in the Schools. 1998; 13:115–128.

Smith, F. Reading: FAQ. New York, NY: Teachers College Press; 2007.

Stanovich K. Toward an interactive-compensatory model of individual differences in the development of reading fluency. Reading Research Quarterly. 1980; 16:32–71.

Summers EG, McClelland JV. A field-based evaluation of sustained silent reading (SSR) in intermediate grades. Alberta Journal of Educational Research. 1982; 28:100–112.

Swanson E, Vaughn S, Wanzek J, Petscher Y, Heckert J, Cavanaugh C, et al. A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. Journal of Learning Disabilities. 2011; 44:258–275. [PubMed: 21521868]

Tschannen-Moran M, Hoy AW. Teacher efficacy: Capturing an elusive construct. Teaching and Teacher Education. 2001; 17:783–805.

Urquhart I. Beyond the literal: Deferential or inferential reading? English in Education. 2002; 36(2):18–30.

Vaughn S, Martinez LR, Linan-Thompson S, Reutebuch CK, Carlson CD, Francis DJ. Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. Journal of Research on Educational Effectiveness. 2009; 2:297–324.

Vaughn S, Wexler J, Leroux A, Roberts G, Denton C, Barth A, Fletcher J. Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. Journal of Learning Disabilities. 2011; 45:515–525. [PubMed: 21512102]

Vaughn S, Swanson E, Roberts G, Wanzek J, Stillman-Spisak SJ, Solis M, et al. Improving reading comprehension and social studies knowledge in middle school. Reading Research Quarterly. 2012; 48:75–91.

Wilson, KT. Group reading assessment and diagnostic evaluation. Upper Saddle River, NJ: Pearson; 2001.

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of cognitive abilities. Itasca, IL: Riverside Publishing; 2001.
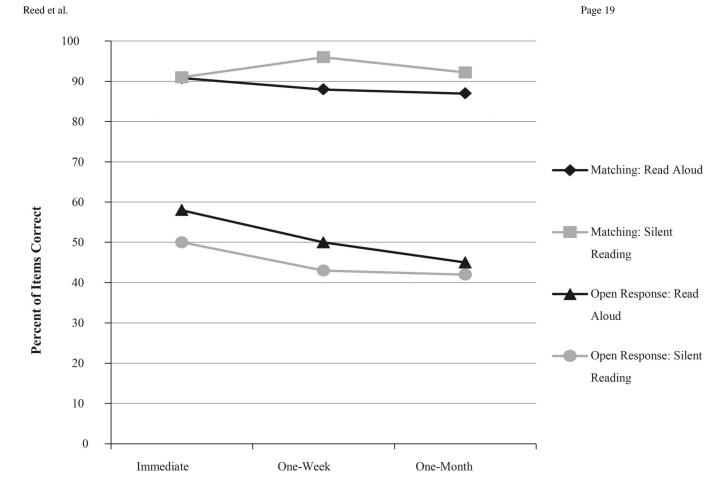
Fig. 1.
Post-test trends on matching and open-response items

Table 1

Pre- and post-test means and standard deviations by treatment group prior to imputation

	N		*M*(SD)	
	TRA	SR	TRA	SR
Pre-test only				
State criterion referenced reading test[a]	59	60	2,208.42 (90.45)	2,204.57 (105.67)
KBIT verbal knowledge[b]	50	50	60.22 (5.28)	58.98 (5.18)
GRADE listening comprehension[c]	58	52	11.52 (2.08)	10.63 (2.72)
Test of silent contextual reading[b]	59	54	85.51 (11.96)	84.35 (11.53)
Woodcock Johnson-III spelling[b]	57	53	90.46 (14.75)	90.85 (11.52)
Pre- and post-tests				
Content knowledge composite[c]				
Pre-test	59	54	8.73 (3.72)	8.70 (4.73)
Immediate post-test	50	49	14.90 (3.10)	14.06 (3.82)
One-week delayed post-test	53	48	13.87 (4.57)	13.69 (3.16)
One-month delayed post-test	52	48	13.21 (4.58)	13.25 (3.11)
Content knowledge, matching items[c]				
Pre-test	59	54	7.56 (2.79)	7.00 (3.51)
Immediate post-test	50	49	9.10 (1.88)	9.10 (2.00)
One-week delayed post-test	53	48	8.81 (2.60)	9.56 (1.17)
One-month delayed post-test	52	48	8.75 (2.47)	9.23 (1.42)
Content knowledge, open response items[c]				
Pre-test	59	54	1.17 (1.90)	1.70 (2.23)
Immediate post-test	50	49	5.80 (2.36)	4.96 (2.77)
One-week delayed post-test	53	48	5.06 (3.05)	4.33 (2.76)
One-month delayed post-test	52	48	4.46 (3.31)	4.15 (2.59)

[a]Scale score where 2,100 represents "passing"

[b]Standard scores where 100 represents the normative mean

[c]Raw score; *TRA* teacher read aloud condition; *SR* silent reading condition

Table 2

Full sample means and standard deviations before and after imputing missing data

	N	Before imputation		After imputation		d
		M	SD	M	SD	
KBIT[a]	100	59.60	5.24	59.32	5.19	0.05
GRADE[b]	110	11.10	2.43	11.07	2.42	0.01
Test of silent contextual reading (TOSCRF)[a]	113	84.96	11.72	84.99	11.59	−0.002
Woodcock Johnson-III spelling[b]	110	41.36	6.13	41.29	6.11	0.01
Pretest of content knowledge composite[b]	113	8.72	4.21	8.72	4.19	0.00
Immediate post-test of content knowledge composite[b]	99	14.48	3.48	14.36	3.27	0.04
One-week delayed post-test of content knowledge composite[b]	101	13.78	3.94	13.59	3.78	0.05
One-month delayed post-test of content knowledge composite[b]	100	13.23	3.93	13.26	3.70	0.01

[a] Standard scores where 100 represents the normative mean

[b] Raw score

Table 3

Imputation summary one-way analysis of variance: pre-tests

Outcome	Parameter	Estimate	SE	95% CI		df	t	p value
				Lower bound	Upper bound			
KBIT[a]	Intercept	58.89	0.72	57.46	60.31	96.78	82.23	<0.001
	Group	0.83	0.99	−1.15	2.82	103.47	0.83	0.497
GRADE[b]	Intercept	10.67	0.32	10.04	11.31	111.63	33.31	<0.001
	Group	0.74	0.45	−0.16	1.64	111.27	1.63	0.106
Test of silent contextual reading[a]	Intercept	84.46	1.56	81.36	87.55	109.00	54.07	<0.001
	Group	0.90	2.18	−3.43	5.22	114.35	0.41	0.682
Woodcock Johnson-III spelling[b]	Intercept	41.23	0.83	39.58	42.87	102.47	49.74	<0.001
	Group	−0.04	1.14	−2.30	2.22	114.37	−0.03	0.972
Pretest of content knowledge composite[b]	Intercept	8.71	0.57	7.59	9.84	105.78	15.39	<0.001
	Group	−0.01	0.79	−1.59	1.56	110.31	−0.02	0.985

[a]Standard scores where 100 represents the normative mean

[b]Raw score

Table 4

Imputation summary one-way analysis of covariance: post-tests of content knowledge (combined score)

Outcome	Parameter	Estimate	SE	95% CI Lower bound	95% CI Upper bound	df	t	p value
Immediate	Intercept	−4.39	4.87	−14.34	5.56	30.51	−0.90	0.375
	Group	0.31	0.64	−0.96	1.57	69.93	0.48	0.631
	KBIT[a]	0.12	0.08	−0.04	0.28	31.97	1.49	0.147
	GRADE[b]	0.28	0.13	0.02	0.54	91.33	2.12	0.037
	Test of silent contextual reading[a]	0.01	0.03	−0.05	0.08	79.82	0.43	0.666
	Woodcock Johnson-III Spelling[b]	0.14	0.07	−0.01	0.28	71.52	1.89	0.063
	Pretest of content knowledge composite[b]	0.17	0.09	−0.02	0.35	50.00	1.83	0.074
One-week	Intercept	−0.16	5.10	−10.46	10.14	41.10	−0.03	0.975
	Group	−0.10	0.67	−1.43	1.23	100.03	−0.15	0.881
	KBIT[a]	−0.04	0.08	−0.20	0.12	55.98	−0.46	0.646
	GRADE[b]	0.25	0.15	−0.04	0.55	89.28	1.72	0.090
	Test of silent contextual reading[a]	0.06	0.04	−0.02	0.13	93.65	1.51	0.136
	Woodcock Johnson-III Spelling[b]	0.16	0.08	−0.01	0.33	56.19	1.92	0.060
	Pretest of content knowledge composite[b]	0.21	0.10	0.02	0.40	76.14	2.19	0.032
One-month	Intercept	4.22	4.80	−5.35	13.78	72.17	0.88	0.382
	Group	0.04	0.69	−1.34	1.41	100.92	0.05	0.958
	KBIT[a]	−0.03	0.08	−0.20	0.13	60.15	−0.42	0.678
	GRADE[b]	0.08	0.15	−0.22	0.38	95.79	0.53	0.594
	Test of silent contextual reading[a]	0.03	0.04	−0.05	0.10	100.48	0.74	0.460

Outcome	Parameter	Estimate	SE	95% CI		df	t	p value
				Lower bound	Upper bound			
	Woodcock Johnson-III Spelling[b]	0.13	0.08	−0.03	0.30	87.68	1.63	0.106
	Pretest of content knowledge composite[b]	0.25	0.10	0.05	0.45	66.13	2.48	0.016

[a] Standard scores where 100 represents the normative mean

[b] Raw score

Table 5

Means and standard deviations by first language after imputing missing data: post-tests of content knowledge (combined raw score)

	M(SD)			M(SD)		
	TRA English (n = 12)	SR English (n = 22)	d	TRA Spanish (n = 46)	SR Spanish (n = 37)	d
Pretest	10.00 (2.89)	10.36 (3.90)	−0.10	8.37 (3.90)	7.82 (4.82)	0.13
Immediate post-test	15.46 (2.90)	14.32 (2.24)	0.46	14.52 (2.90)	13.88 (4.18)	0.18
One-week delayed post-test	13.45 (5.96)	13.30 (2.70)	0.04	13.58 (4.06)	13.71 (3.21)	−0.04
One-month delayed post-test	13.43 (6.53)	13.14 (2.92)	0.06	13.34 (3.71)	13.27 (2.99)	0.02

TRA teacher read aloud condition, *SR* silent reading condition

Table 6

Percent in each treatment group moderately of strongly agreeing/disagreeing with social validity statements

	Moderately or strongly agree		Moderately or strongly disagree	
	TRA (*n* = 53) (%)	SR (*n* = 45) (%)	TRA (*n* = 53) (%)	SR (*n* = 45) (%)
I understand and remember the information in a passage when I read it silently	57.5	67.5	8.2	4.1
I understand and remember the information in a passage when a teacher reads it aloud to me	59.5	46.0	4.8	10.7

TRA teacher read aloud condition, *SR* silent reading condition