Hammill Institute on Disabilities

Remedial and Special Education
2018, Vol. 39(6) 341–352

DOI: 10.1177/0741932517744662
rase.sagepub.com
SAGE

The Application of Empirically Supported Practices in Middle School Social Studies Classrooms

Shawn C. Kent, PhD[1], Jeanne Wanzek, PhD[2], and Leticia Martinez, PhD[3]

Abstract

Material encountered in social studies has been shown to be challenging for students with reading difficulties, including those with disabilities. This study first examined the amount of time eighth-grade U.S. History teachers implemented empirically supported instructional practices. Second, this research investigated the association between instruction and student growth in content knowledge. The sample included 188 students identified with low reading ability, of which 20% were students with disabilities. Instruction in 30 classrooms was coded for the presence of learning strategy instruction, study aids, graphic/spatial organizers, peer-mediated instruction, computer-aided instruction, and mnemonics. Utilization of these practices was limited (~20% of total instruction), with the use of spatial/graphic organizers and peer-mediated activities most prevalent. Time allocated to these practices was not significantly associated with student gains in content knowledge. Additional research to determine how to best impact content learning for students at-risk for poor learning outcomes is warranted.

Keywords

evidence-based practice, content area, instruction, academic achievement

Yecke (2005) proclaimed that "middle schools are where student academic achievement goes to die" (p. 1), citing national and international assessments of achievement that demonstrated a precipitous decline in academic performance in the middle school years for students in the United States. Recent assessments have continued to reveal the generally middling performance of students in the United States in comparison with other nations (Kelly et al., 2013; Provasnik et al., 2012). One area where U.S. students have particularly struggled is in the social studies. Specifically, less than one third of eighth graders achieved at the proficient level on the most recent National Assessment of Education Progress report (National Center for Education Statistics [NCES], 2014) in the areas of U.S. History, Civics, and Geography. Social studies are particularly important for the promotion of students' civic competence and their ability to be active and engaged participants (National Council for the Social Studies [NCSS], 2010). Furthermore, this content area also provides an avenue for students to increase their ability to answer meaningful questions and solve problems (National Center for History in the Schools, 1996).

Social studies courses at the secondary level frequently involve conceptually rich content that requires critical thinking for understanding. Social studies also present other potential challenges to the many adolescent students with poor literacy skills, including students with disabilities (NCES, 2013). As highlighted in a report from the Carnegie Corporation (Lee & Spratley, 2010), textbooks are the most frequent source for reading about historical topics covered in these classrooms. This is particularly troublesome given that such textbooks are frequently written at advanced levels and contain ambiguous or indirect references, information irrelevant to the main idea, and an overall density of ideas that often make comprehension of material quite difficult (Lee & Spratley, 2010). In addition, adolescents with disabilities have been noted to have trouble with answering content-area questions that go beyond simple definitions for key vocabulary and/or basic facts and details, with aspects of study skills and self-regulation, and with long-term and working memory (e.g., Boyle, 2010; Newman et al., 2011; Scruggs, Mastropieri, Berkeley, & Graetz, 2010; Vaughn,

[1]University of Houston, TX, USA
[2]Vanderbilt University, Nashville, TN, USA
[3]The University of Texas at Austin, USA

Corresponding Author:
Shawn C. Kent, Department of Educational Leadership and Policy Studies, University of Houston, 318D Farish Hall, Houston, TX 77204-5023, USA.
Email: sckent@uh.edu

Gersten, & Chard, 2000). Furthermore, the fast pace of instruction and expectations for independent learning in the secondary grades pose a challenge for knowledge acquisition for low-performing students (Mastropieri & Scruggs, 2001). In sum, the often heavy reliance on the reading of textbooks and the above-mentioned characteristics of students with reading/literacy difficulties, including students with learning disabilities, serves as a potential barrier for students to gain the requisite information to both access and succeed in the general education content-area classroom.

In recent years, there has been a concerted effort toward increased standards and expectations across the curriculum, including in the social studies, with the goal of college and career readiness (e.g., Common Core State Standards Initiative, 2010; NCSS, 2013). This heightens the need for effective instruction for all students to meet these rising expectations. Fortunately, the extant literature contains direct evidence for the potential efficacy of a variety of instructional strategies in promoting positive literacy and learning outcomes for adolescents with reading and other related learning difficulties within the content areas. Scruggs et al. (2010) completed a meta-analysis of instructional studies in the areas of science and social studies for students identified as having learning disabilities. Among the instructional strategies identified as effective were peer-mediated learning, computer-assisted instruction, mnemonic instruction, study aids, spatial or graphic organization, learning strategy instruction, explicit instruction, and hands-on activities. The overall effect size (ES) for these instructional methods when utilized with social studies content was 1.00. The ES for the 15 studies conducted within the general education setting was 0.81. When instruction was implemented by general education teachers, the ES remained in the moderate range at 0.62.

Within the meta-analysis by Scruggs and colleagues (2010), studies addressing the teaching of learning strategies (ES = 1.11) involved specific strategies that students can generalize and apply within and across content such as self-questioning techniques, comprehension monitoring, summarization, analyzing text structure, and note-taking strategies. Research on providing students study aids (ES = 0.94) included studies that addressed supports such as study guides, advance organizers, and text outlines (e.g., guided notes). Meanwhile, instruction that included the use of spatial and graphic organizers (ES = 0.93) involved studies that provided visual organization and structure to promote learning of content (e.g., semantic maps, Venn diagrams). Research on instruction utilizing mnemonic strategies (ES = 1.47) focused on specifically targeting verbal associative learning most often via keyword or pegword methods. Peer-mediated instruction research (ES = 0.86) included instruction that involved classmates, either partners or small groups, engaged in peer teaching, reciprocal teaching, cooperative learning, and/or discourse-based activities. Studies that focused on computer-assisted strategies (ES = 0.63) included activities involving the presentation of content via technological applications or multimedia formats (e.g., online tutorials). Explicit instruction (ES = 1.68) involved segmenting complex skills/content into smaller instructional units and following a sequence of teacher modeling followed by guided and independent practice opportunities. Finally, instruction utilizing hands-on activities (ES = 0.63) was characterized by students working directly with relevant material rather than learning content through the reading of text and/or teacher presentation (Scruggs et al., 2010). It should be noted that one or none of the studies targeted explicit instructional methods or hands-on activities in the area of social studies.

A more recent meta-analysis of instructional research for supporting literacy using social studies content for students with learning disabilities across Grades K–12 found that studies involving graphic organizers and mnemonic instruction were most prevalent, again with moderate to large ESs (E. Swanson et al., 2014). Despite the available evidence demonstrating the potential promise of these instructional practices, to date, we are aware of no studies that have specifically investigated whether such practices have translated into typical educational practice in the classroom. This is particularly relevant given previous research has demonstrated that typical instruction in social studies classes was marked by passive reading of informational texts via course textbooks in addition to teacher lecture, with the primary goal of rote learning of vocabulary and key factual information (Bulgren, Deshler, & Lenz, 2007; Okolo, Ferretti, & MacArthur, 2007; Paxton, 1999; Scruggs & Mastropieri, 2003). In fact, teachers generally have identified lecture-based activities as their most effective teaching method (Bolinger & Warren, 2007). As noted by Scruggs et al. (2010), however, this type of instruction has been routinely outperformed when compared with other instructional practices included in their meta-analysis. Furthermore, despite a heavy reliance on lecture and note-taking during social studies instruction, students at the secondary level have indicated that such methods are not preferred for helping them learn content (Chiodo & Byford, 2004). Similarly, though the ability to effectively and efficiently process content-area informational text is certainly crucial to students making academic progress (Cervetti, Bravo, Hiebert, Pearson, & Jaynes, 2009), instruction limited to the passive or independent reading of text is not likely to benefit the many students exhibiting reading deficits. Wanzek et al. (2014) specifically noted the need for further examination of secondary content-area instruction within general education classrooms to determine the instructional supports students most at-risk require for meaningful learning gains.

Numerous researchers have addressed the perceived gap between practices supported empirically through educational research and actual educational practices (e.g., Biesta, 2007; Broekkamp & van Hout-Wolters, 2007; Cook, Smith, & Tankersley, 2012; Riley-Tillman, Chafouleas, Eckert, &

Kelleher, 2005). This issue came to the forefront with specific language in both the No Child Left Behind Act (NCLB; 2002) and the Individuals With Disabilities Education Act (2004) that stressed the role of scientific research in determining effective instructional practices. Such emphasis on practices supported by empirical research has continued with the recently enacted Every Student Succeeds Act (ESSA; 2015). An argument has been made that broad implementation of instructional methods and practices with demonstrated efficacy for improving student outcomes should result in increased student learning (Cook et al., 2012; Slavin, 2008). Given that a majority of students with poor literacy skills, including those with learning disabilities, are likely to be enrolled in general education content-area courses (H. L. Swanson & Deshler, 2003; U.S. Department of Education [USDOE], 2014), there is a growing need for application of instructional methods that meet the needs of students of diverse ability levels.

There were two primary aims of the current study. First, we sought to provide observational data on the extent to which instructional practices with at least moderate empirical support for improving the learning of students within the content areas were evident in general education, content-area classrooms at the secondary level. Second, we aimed to determine the relationship between the amount of time teachers spent in these specific instructional practices and students' acquisition of content in these classrooms.

Method

Participants

The student and teacher samples in the present study were drawn retrospectively from participants in a series of randomized control trials (RCT), conducted over the 2012–2013 and 2013–2014 school years, examining the efficacy of an instructional framework for improving content-knowledge acquisition in the middle school U.S. History classrooms (Vaughn et al., 2017; Vaughn et al., 2015). As part of the larger RCT studies, all eighth-grade U.S. History teachers ($n = 37$) in participating schools were selected to participate. U.S. History class sections were randomly assigned, blocking on teacher, to the treatment or comparison (i.e., typical instruction) conditions. Each teacher had at least one treatment and one comparison class. The current study utilized a sample of audio recordings of teacher instruction from comparison condition classes. Given the goal of examining typical practice in these U.S. History classrooms, only students from these audio-recorded comparison classes were included in the sample of this present study.

Students. The sample was selected to include those students with reading difficulties, which was defined as performance at or below the 30th percentile in comparison with same-age peers on the reading comprehension subtest of the *Gates–MacGinitie Reading Test* (GMRT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006). These students were specifically targeted to capture students with learning disabilities and struggling readers, both of whom may be at-risk for poor outcomes given the potential incompatibility of their learning characteristics and the content-area learning environment (Scruggs et al., 2010). These criteria resulted in a total sample of 188 students. Of this sample of students with demonstrated reading difficulty, 35 (19%) had been identified as a student with a disability. The specific disability category was not provided by schools/districts, although it is important to note that all were enrolled in a general education U.S. History course. Anecdotal reports from teachers suggest that most students were identified with high-incidence disabilities such as a Learning Disability or Speech/Language Impairment. The majority of students were female (55.3%), and 60.6% identified as Hispanic. Sample race was approximately half White (49.5%), with the majority of Hispanics identifying as White race. Sample race also included 18.6% African American, 19.1% American Indian, 3.7% mixed race, and 2% or less each of Asian and Pacific Islander. Data on race were missing for approximately 5% of the sample. Nearly one half of the sample was enrolled in free or reduced price lunch programs and just over one quarter (28%) were designated Limited English Proficient or had been within the last 2 years.

Students demonstrated a mean standard score of 81.09 ($SD = 7.72$) on the GMRT comprehension assessment, indicative of below average reading ability. The bottom quartile of this sample included students with a standard score of 75 and below and the upper quartile included students with standard scores of 87 and above. On the 44-item pretest measure of content knowledge (assessment of social studies knowledge [ASK]; Vaughn et al., 2013; see "Measures" section), students averaged nearly 13 correctly answered questions ($M = 12.71$, $SD = 4.30$). On average, students had 16 correct responses ($M = 15.91$, $SD = 5.85$) at posttest.

Teachers. The sample of students represented 30 different general education U.S. History teachers. These teachers were in 14 different schools in districts across two southern states. There were 16 female teachers. The sample of teachers was racially homogeneous with 87% being White. Four teachers (13.3%) identified as Hispanic. All teachers were certified to teach U.S. History and held bachelor's degrees, and 10 teachers had a master's degree. The mean teaching experience for these teachers was 12.3 years ($SD = 12.7$). One teacher was in her first year teaching whereas the most experienced teacher had taught for 38 years.

Procedures

As noted, all participating teachers taught both treatment and comparison classes. In comparison classes, teachers were instructed to utilize their typical instructional practices to teach the content. Teachers taught three U.S. History

units for 3 weeks (i.e., approximately 15 daily lessons) each. All daily classroom instruction in one comparison class for each teacher was recorded by providing the teacher with a recorder that they kept on their person during the entire class period. Teachers were specifically instructed to record any and all activities occurring during the scheduled class period including delivery of content instruction, testing/assessment, videos, group work, and project-related tasks. Although all teachers were trained in the experimental instructional methods, fidelity observations conducted as part of the larger study demonstrated that teachers did not utilize those specific methods in their comparison classrooms (Vaughn et al., 2017; Vaughn et al., 2015).

To specifically examine the presence of empirically supported instructional practices, a random selection of 20% of the recordings for each unit of instruction was sampled for each teacher. This resulted in a total of 243 recordings of typical classroom instruction. Although we sampled 20% of recordings per teacher, the number of recordings eligible for sampling was quite variable across the 30 teachers. This was primarily due to teacher differences in the actual number of days of instruction across the three units. More specifically, four teachers' classes utilized block scheduling with generally 90-min periods that met 2 to 3 times per week, whereas the remaining classes employed more typical daily class periods of 45 to 50 min. Furthermore, although instruction for the three units was generally completed over the course of 9 weeks (i.e., 45 instructional sessions), one teacher taught the units over the course of 80 days. In addition, differences in the number of available recordings reflected the variability in fidelity of recording (i.e., teacher forgot to record class period). Finally, some recordings of instruction were either inaudible or less than 10 min in length (due to teacher and/or machine error) and were excluded ($n = 20$). The final number of recordings coded across teachers was between three and 16 instructional periods ($M = 7.26$). However, five to 10 instructional sessions were coded for the majority of teachers (21/30; 70%). Four teachers, all of whom taught in a block schedule, had fewer than five instructional sessions coded whereas 11 or more instructional sessions were coded for five teachers.

All recordings of instruction were coded by trained research staff utilizing a codesheet specifically designed for this study (see "Measures" section). Prior to formal coding, three research staff were trained to use the observation tool, including an overview of the codesheet and the accompanying rubric with descriptors for the various instructional categories. Furthermore, staff conducted several practice sessions that involved individual coding of recordings of instruction, a discussion of coding decisions among staff, and the rationale for such decisions. To establish reliability, recordings were coded by each rater and compared with a standard of reliability established by the project coordinator. The recordings used for training, practice, and reliability purposes were all randomly selected recordings of U.S. History instruction that were not included in the current study.

Reliability was calculated as the percentage agreement for each minute of instruction coded (e.g., if the recording was 50 min in length and coders agreed on instructional activities for 47 min, agreement is 47/50 or 94%). Each rater was required to obtain a minimum reliability (i.e., agreement) of 90% prior to beginning formal coding; initial agreement was 100%. Furthermore, reliability checks were conducted twice during the coding process (after approximately one third and two third of audios coded). Overall reliability was 90.0%. In addition, 20% of the sample of audios were randomly selected and double-coded to compute interrater agreement. Reliability was generally good with 92.1% agreement among coders across all minutes of instructional activities coded.

All students were assessed on a measure of content knowledge (ASK; Vaughn et al., 2013) covering topics from the three units of instruction prior to and at the completion of the study. Reading comprehension was also assessed at the beginning of the study. All pretesting and posttesting took place within 2 weeks of the commencement and the conclusion of the three units of instruction in each respective classroom. Assessments were group administered by trained research assistants over the course of 2 days. Student assessments were double-scored after administration and double-entered during data entry to ensure the accuracy of the data. All discrepancies in scoring were discussed until consensus was reached.

Measures

Classroom instruction. An observational tool was developed by the first author to allow for coding of instructional activities occurring during the recorded lessons. Based on the previous syntheses and meta-analyses by Scruggs et al. (2010) and E. Swanson et al. (2014), the decision was made to focus on instructional practices with at least moderate empirical evidence in support. Although we did not conduct a rigorous review of study methodology and quality, we selected those practices that were supported as effective by at least four studies with positive effects across content areas and two or more studies with positive effects when implemented in social studies (using the Scruggs et al. and E. Swanson et al. meta-analyses for reference; see Gersten et al., 2005; What Works Clearinghouse, 2014). Thus, we coded six distinct instructional practices: (a) learning strategy instruction, (b) study aids, (c) spatial/graphic organizers, (d) mnemonic instruction, (e) peer-mediated learning, and (f) computer-assisted instruction. Again, given the scarcity of research on explicit instruction and hands-on activities in this content area (see Scruggs et al., 2010), we did not include these instructional methods in the current study.

Table 1. Sample Descriptors for Coding of Instructional Practices.

Instructional practice	Descriptors
Learning strategy instruction	• Teaching of text structures found in History (e.g., compare/contrast, sequencing, problem-solution) • Strategies for note-taking • Cognitive strategies for before reading (e.g., previewing) • Strategies for during text reading (e.g., self-questioning, self-monitoring) • Strategies for after reading (e.g., summarization)
Study aids	• Advance organizers (e.g., outline/overview of important concepts from lesson and/or text) • Study guides • Guided notes (e.g., Teacher prepared notes with blank spaces for students to enter key concepts/ vocabulary)
Spatial/graphic organizer	• Graphic organizers/concept map for documenting and organizing concepts and content • Semantic maps for learning new vocabulary and depicting relationships among concepts/vocabulary • Map(s) depicting relevant content to be taught
Mnemonic instruction	• Acronyms or acrostics to aid in recall of lists of information • Providing "keyword" or acoustically similar, familiar word to link to new unfamiliar word via picture (e.g., picture of a tiger in battle to remember Fort Ticonderoga)
Peer-mediated learning	• Peer tutoring/reciprocal teaching of content • Pairs of students engage in think-pair-share to discuss content • Partner reading of text (e.g., students read text in pairs and ask each other specific questions) • Two more students engage in focused discussion/debate of content • Multiple students work jointly on assigned task
Computer-assisted instruction	• Students use software that involves graphic/cognitive organizer of content • Multimedia presentation of content (NOT PowerPoint to present lecture) • Tutorials presented on computer • Computerized maps/study guides

Learning strategy instruction was coded when the teacher engaged in any explicit instruction in strategies for processing and studying content-area information and text. The study aids category was coded when teachers provided students with support(s), in text form, that organized the content in some way for the students before learning the content (e.g., study guides, outlines). Instruction was coded as spatial/graphic organizer when the presentation and/or learning of content included the provision of some type of visual organization to facilitate learning (e.g., Venn diagram, map). Any instruction that involved the use of specific strategies for verbal recall combined with the modified presentation of content was coded as mnemonic instruction. For example, providing an acoustically similar, familiar word to link to new unfamiliar word via a picture (e.g., a picture of a tiger in battle to remember Fort Ticonderoga) was coded as mnemonic instruction. The peer-mediated learning code was utilized when teachers employed classmates in cooperative learning groups (e.g., peer tutoring, discussion, reciprocal teaching). Finally, instruction was coded as computer-assisted in cases where instruction involved students' use of technological application to aid learning (e.g., specialized software involving presentation/organization of content, computerized tutorial programs). A null code was also used for any instruction that did not fall into one of these six categories. More specific descriptors used for coding each of these instructional activities can be found in Table 1. These descriptors are fairly general due to the fact that we were more broadly interested in whether these instructional practices were used in typical content-area instruction rather than examining specific, discrete components of a given practice or the quality of implementation.

The codesheet required coders to listen to a recording of teacher instruction and utilize continuous duration recording methods to document the beginning and end time for any of the targeted empirically supported practices evident during the instructional period. More specifically, coders documented the exact time duration (and total number of minutes) for all empirically supported instructional activities occurring during the entire recorded class session along with a brief summary of the activity. Once an instructional activity(ies) was evident during instruction (see Table 1 for descriptors utilized during coding), coding for that activity was initiated and continued until such a point that it was clearly no longer being utilized. For example, at the point a teacher referenced a study aid (e.g., study guide) or a spatial/graphic organizer that students were requested to access and/or complete, coding of this practice commenced and was coded continuously even if the teacher did not continue to reference the specific tool. Once the teacher made a clear transition to another instructional activity, coding for the previous activity was discontinued. The codesheet further allowed, when applicable, for coding of multiple instructional activities for each minute. For example, instruction

that involved the teacher asking students to complete a graphic organizer through discussion with a partner or small group of peers would be an example of multiple activities, peer-mediated instruction, and spatial/graphic organizer. For instances when an individual coder was uncertain of the presence of a particular instructional activity, another coder listened to the recording in question and discussed until a consensus was reached. The resulting data included the total duration the teacher utilized each instructional activity during the class period.

ASK. The ASK (Vaughn et al., 2013) assessment is a researcher-developed measure of vocabulary and content knowledge in the three units that comprised the intervention (Colonial America, Road to Revolution, Revolutionary War). The measure is untimed and consists of 44 multiple-choice items. Items on the ASK assess either knowledge of key vocabulary/concepts or required content knowledge of important historical events, facts, and issues. Items generally consisted of one sentence statements with four possible responses—for example, one reason the Mayflower Compact was written was to establish (a) principles of government for the new colony, (b) a charter for a joint-stock company, (c) strict military control over the colonists, (d) a system of trade with other colonies. Students were presented with the test the same way they are presented with tests in their class and/or state tests. If accommodations were required due to a student's Individualized Education Program (IEP), those were implemented accordingly (i.e., test read aloud to student). The questions were specifically aligned with identified content to be taught during the units but many required the application of the material. The test was not designed for mastery (similar to *Advanced Placement* [AP] tests) to avoid ceiling effects.

To develop this assessment, an item bank was assembled that consisted of a large pool of content-relevant items with known difficulty parameters from released state and AP tests in U.S. History. Each sponsoring agency (e.g., College Board; State Department of Education) granted permission to use selected items and relevant psychometric information. A series of pilot tests were run to validate the provided difficulty parameters, refine instructions, and estimate the time needed for administration. A final sample of items that performed comparably well across the pilot studies and the item developer's large-scale administrations, was feasible for administration, and was maximally sensitive to change in the topics of interest was selected.

To refine and finalize the group of selected items, a series of item-level confirmatory factor analysis were conducted to evaluate model fit and to estimate item parameters. Observed data (n = 715) from a single-factor, 44-item assessment fit the model very well, χ^2 = 934.492, df = 902, p = .22, confirmatory fit index (CFI) = .97, root mean square error of approximation (RMSEA) = .009, suggesting that the selected items measured the targeted constructs with the greatest information available in the average ability distribution (θ). Results indicated that the peak of the test information function (information = .91) occurred at −0.25 on the ability distribution, about one quarter of a standard deviation below average θ, fixed at 0 by default. The information function was above 0.80 for ability levels ranging from −2.0 to 2.0, thus indicating reasonable coverage across the range of θ. Test information indices indicate the precision of measurement for persons at different levels of the underlying latent construct, with higher information denoting more precision (Guion, 2011). Test information indices represent an item response theory (IRT)–based analog for estimates of reliability used in classical test models. Across all sample members, Cronbach's alpha was .95.

GMRTs. The Gates–MacGinitie (MacGinitie et al., 2006) is a group-administered, norm-referenced reading test with various levels from kindergarten through adults. We administered the *Reading Comprehension* subtest. Students are provided with expository and narrative reading passages, ranging from single to multiple paragraphs in length, followed by multiple-choice questions related to each passage. Questions addressed facts, inferencing, and drawing conclusions. Students had 35 min to complete this assessment. Internal consistency reliability coefficients are between .91 and .93 and alternate form reliability is reported as between .80 and .87. We utilized student data from the pretest assessment to identify students considered as low-performing readers for inclusion in the study sample.

Data Analysis

For the first research question (To what extent were instructional practices with at least moderate empirical support for improving the learning of students within the content areas evident in general education, content-area classrooms at the secondary level?), we computed descriptive statistics (means, standard deviations, ranges) for the instructional variables observed and coded during classroom instruction in these eighth-grade U.S. History classrooms. To obtain teacher-level means, we calculated a simple average across all instructional periods coded for each teacher. These data are reported separately for each of the six categories of instruction. We obtained overall means for each category by taking the mean of each of the 30 teacher-level means. For interpretation purposes, we quantified utilization of these instructional practices in terms of minutes of instruction and percentage of class time. Furthermore, we also report the range of time (in minutes) each practice was observed across all observations.

To answer the second research question (What is the relationship between the amount of time teachers spent in these specific instructional practices and students' acquisition of content in these classrooms?), we specified a two-level model (students nested within teacher) to predict

Table 2. Teacher Implementation of Instructional Practices.

Instructional practice	Total minutes			Percentage of class period		
	M[a]	*SD*	Range[b]	*M*[a]	*SD*	Range[b]
Learning strategy instruction	0.19	0.67	0–3.13	0.32	1.23	0–6.46
Study aids	1.24	2.74	0–11.33	2.74	6.22	0–27.42
Spatial/graphic organizer	4.04	4.13	0–3.40	8.82	9.11	0–28.23
Mnemonic instruction	0	0	NA	0	0	NA
Computer-aided instruction	0.28	1.07	0–4.60	0.58	2.21	0–9.44
Peer-mediated instruction	4.37	3.98	0–12.44	10.11	9.49	0–31.46

Note. Mean length of instructional period was 45.49 min (*SD* = 10.81).
[a]Mean of teacher-level means. [b]Range of teacher-level means.

students' learning outcomes (ASK) from the amount of time teachers, on average, utilized the targeted instructional practices. In addition to examining the relation between the individual instructional practices and student content knowledge, we also ran a model to investigate how teachers' total time spent across instructional practices (average per class period) was associated with outcomes to examine the cumulative effect of teachers employing these practices. Multilevel modeling was appropriate given the nested nature of the data (Hox, 2002). To control for student's prior knowledge, students' pretest ASK scores were grand mean centered at Level 1 in each multilevel model (Enders & Tofighi, 2007). We utilized the Hierarchical Linear Model statistical package (HLM7; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) for analyses using all cases and full information maximum likelihood to estimate model parameters.

Results

Instructional Practices

Descriptive data on instructional practices coded in these classrooms are provided in Table 2. The total length of instruction in these U.S. History classrooms averaged just over 45 min (*M* = 45.49, *SD* = 10.81) with a range of 30 to just over 91 min. Again, this range in class session length was primarily due to the fact that some teachers utilized block scheduling. To aid interpretation, the findings are reported for both the mean number of minutes and mean percentage of class time that each instructional practice was observed and coded across teachers. For additional context on the typical implementation of a respective practice within a class period, we report the range of times each practice was observed across all 223 coded observations. The most frequently coded instructional practice across teachers was the use of peer-mediated learning (*M* = 4.37 min; *SD* = 3.98; range across all observations = 0–46 min). On average, this represented just over 10% of teachers' typical instructional practice during class sessions. A total of 23 teachers utilized this practice at least once. The use of spatial and/or graphic organizers to support instruction was also implemented by 23 of the teachers and was evident for an average of 4 min (*SD* = 4.13; 8.81% of instructional time; range = 0–39). The U.S. History teachers in this study provided their students with study aids during classroom instruction for just over 1 min per class period (*M* = 1.24; *SD* = 2.74; range = 0–47). A total of nine teachers utilized this practice at least once and, on average, these study aids accounted for just under 3% of classroom instructional time.

The remaining instructional practices were coded as occurring less than 1 min per class period. Specifically, two teachers provided students the opportunity for computer-aided instruction of content for an average of 0.28 min (*SD* = 1.07; range = 0–46) per class period. The implementation of instruction in learning strategies across these teachers was also minimal, with a mean of less than 1 min per period (*M* = 0.19; *SD* = 0.67; range = 0–25). Four teachers were observed implementing learning strategy instruction during at least one class period. Both of these practices were utilized for less than 1% of typical instruction in these 30 classrooms observed. Finally, there was no evidence of any teacher employing the use of mnemonic strategies to aid in content learning in their U.S. History classroom. In sum, when examining teacher utilization of any of the empirically supported practices, the teachers in this study averaged just over 10 min (*SD* = 7.19) of implementation per class period, or just over 22% of total instructional time.

Association Between Instructional Practices and Content Knowledge

To examine how teachers' utilization of these practices related to student acquisition of content knowledge from the three units of instruction, we first ran a baseline (null) model to determine the amount of variance in the outcome that could be partitioned at the student and teacher level. The intraclass correlation (ICC) from the null model was .21, indicating that 21% of the variance in the ASK posttest was at the teacher level. A nested model test comparing this

Table 3. Fixed and Random Effects for Models Predicting ASK Posttest Outcome.

Parameter	Model 1 (null)	Model 2 (covariate)	Model 3 (full model)
Fixed effects			
Intercept	15.99 (.70)	15.98 (.58)	16.15 (.053)
Level 1 (student)			
ASK pretest		0.40* (.10)	0.42* (.10)
Level 2 (teacher)			
Learning strategies			0.98 (.89)
Study aids			0.20 (.20)
Spatial/graphic organizer			−0.21 (.13)
Peer-mediated			0.11 (.13)
Computer-aided			0.21 (.52)
Random parameters			
Level 2	7.35* (2.71)	3.56* (1.89)	3.03 (1.74)
Level 1	27.23 (5.22)	26.83 (5.18)	27.07 (5.20)
Model fit			
–2 log likelihood	990.09	949.71	949.22

Note. Standard errors for fixed effects and standard deviations for random parameters are in parentheses. ASK = assessment of social studies knowledge.
*$p < .05$.

original null model with a model with a nonrandom intercept suggested that the Level 2 (teacher) random intercept should be retained, $\chi^2(df = 1) = 15.21$, $p < .001$. We also examined a second model that accounted for student's initial level of performance by including the ASK pretest as a covariate (Level 1). Students pretest performance on the ASK was significantly related to their posttest ASK score ($p < .001$). Results from this model demonstrated that 12% (ICC = .117) of the variance in the outcome could be attributed to teachers after taking initial content knowledge into account. Next, all of the teacher-level (Level 2) predictors—that is, the average amount of instructional time devoted to each of the instructional practices (e.g., peer-mediated instruction)—were added to the model. The mnemonic instruction predictor was not included given that no teachers in this study implemented the practice. None of the instructional predictors were significant in this model. However, the ASK pretest variable was significant ($p < .001$), suggesting that although the amount of time teachers engaged in any of the targeted practices was not related to posttest, initial content knowledge prior to instruction did predict the outcome. The fixed and random effects from all of these models are provided in Table 3.

Finally, we examined the association between the combined amounts of time teachers utilized these instructional practices with posttest content knowledge. No significant effect was found ($p = .92$) though pretest content knowledge on the ASK remained significant.

Discussion

General education content-area courses at the secondary level are likely to include a sizeable percentage of students with poor reading/literacy skills and students identified as having a disability (e.g., NCES, 2013; USDOE, 2014). The fact remains, however, that material encountered in these courses, such as in the social studies, can be particularly challenging for these students (e.g., Boyle, 2010; Lee & Spratley, 2010; Newman et al., 2011; Scruggs et al., 2010; Vaughn et al., 2000). Initial findings from nearly 225 observations across 30 teachers revealed generally limited utilization of empirically supported practices in classrooms where students with reading difficulties, including students with disabilities, were enrolled. When considering a typical 45-min class period, results suggest that teachers spent just over 1 out of every 5 min of available classroom instruction implementing these practices. The average percentage of class time devoted to empirically supported practices ranged from 1% to 58% across teachers. To our knowledge, this was the first study to explicitly examine how instructional practices supported by experimental research have been manifested in middle school social studies classrooms, so the present results cannot be situated within prior findings. The present findings are in line with a recent synthesis examining observation studies of classroom reading and mathematics instruction provided to students with learning disabilities, which demonstrated a relative absence of empirically validated practices (McKenna, Shin, & Ciullo, 2015). Furthermore, our results do provide continued support for the notion of a gap between educational research and current educational practice (e.g., Cook et al., 2012; Riley-Tillman et al., 2005).

Of the practices examined, teachers providing students with spatial/graphic organizers for organizing content being learned and the use of peer-mediated, cooperative learning activities were most prevalent. To put this in perspective, in

a typical class period of 45 min, this equates to just over 4 min of instructional time for each of these activities. These practices were observed at least once in more than two thirds of classrooms. By comparison, the present findings suggest that the provision of study aids (e.g., advance organizers, outlines, guided notes), computer-aided instruction, and learning strategy instruction (e.g., explicit instruction in note-taking, processing of text) would generally be evident for 1 min or less during a typical length class period; each of these practices was evident in less than one third of all classrooms in this study. Mnemonic instructional methods were not evident in any observation of classroom instruction in these U.S. History classes.

One possible explanation for the higher prevalence of spatial/graphic organizers and having students work with peers in class is that these practices may require less teacher planning and preparation than the other practices. For example, the use of graphic organizers typically involved providing students with a blank organizer and directing students to complete it during lecture or text reading. Likewise, when peer-mediated learning was evident, it generally involved teachers asking students to read a passage of text or complete a worksheet and to discuss with a partner or small group of classmates. In contrast, when teachers utilize study aids, they typically must spend time planning and developing the advance organizers, guided notes, or outlines based on course content. Similarly, engaging in learning strategy instruction requires teacher time for determining the most appropriate strategy and developing a plan for implementation, and can require considerable amounts of class time when first introducing the skill or strategy.

A second potential explanation for the limited utilization of these practices may simply be that general education teachers lack knowledge of and/or have received minimal professional development in the specific practices that have been demonstrated as effective with the most at-risk learners. In their article addressing the process of scaling up and sustaining effective practices, Klingner, Boardman, and McMaster (2013) discussed the need for effective—sustained and ongoing supported—professional development of staff, as well as school- and district-level leadership that helps facilitate the scaling of such practices. In essence, it is critical for teachers to learn and be supported in how to implement specific empirically supported practices and see how such practices can help facilitate their instruction. Future research that surveys teachers regarding their awareness and training in these specific practices may help extend the current study and possibly identify a critical need in scaling these methods from research to practice.

Finally, other barriers to implementing empirically supported instructional practices in education settings such as educators finding research findings untrustworthy or unusable; teachers preferring their own, practical evidence; and systemic issues resulting in a lack of leadership, training, and support for implementing new practices (see Cook et al., 2012) may underlie our findings. If an instructional practice such as the use of mnemonic strategies, with a solid research base for nearly 30 years, is to ultimately be integrated into teacher practice, researchers should consider their role in overcoming these potential barriers. Smith, Richards-Tutor, and Cook (2010) highlighted the importance of including not only theory and data when reporting research findings but also including narrative accounts from educators who have effectively utilized such practices (i.e., incorporating local data). Although commonplace in practitioner-oriented journals, this type of information may be helpful in other outlets. Furthermore, the influence of the school and school leadership on teacher practices must also be considered if researchers are to assist in reducing the research-to-practice gap (e.g., Klingner et al., 2013).

Our second primary aim was to examine the relation between time spent utilizing these specific instructional practices and students' acquisition of content knowledge. It is important to first highlight the relatively low improvement (i.e., 3 points) in performance on the content-knowledge assessment following three units of instruction. When reflecting on this minimal gain for this sample of struggling readers, we must consider the appropriateness of the ASK measure. Considering that the assessment consisting of individual questions generally one sentence in length was untimed, and students were allowed any accommodations per their IEP, we felt that this measure represented a challenging, yet fair assessment of content. This is further supported by the fact that the ASK was constructed specifically to align with the intended curriculum in these classes and that test information indices for the ASK suggested reasonable coverage across ability levels. In sum, although a specific reason(s) for this minimal improvement may not be known, these descriptive results do serve to emphasize the need to consider current instructional practices in social studies for students with reading difficulties, including those with disabilities.

Although Cook et al. (2012) theorized that increased utilization of empirically supported instruction results in increased student achievement, analyses showed that the duration of implementation of these instructional strategies were not associated with increased student outcomes in the present study. One possible explanation for these null effects is that the amount of total time devoted to these practices was not sufficient to produce desired learning gains. In the Scruggs et al. (2010) meta-analysis, the average length of sessions in the studies reviewed was approximately 42 min with all sessions at least 10 min in length. In the present study, no teacher averaged more than 14 min for any of these instructional practices. Furthermore, even though in some cases a practice may have been observed for up to 46 min during an individual class session, the frequency of use was low. For example, only five and nine teachers, respectively, utilized peer-mediated instruction and spatial/graphic organizers during at least three of their instructional

sessions across these three units. Thus, it may be that to have the desired impact, increased intensity of instruction is required. Given it is not plausible to expect teachers to spend 42 min on a particular strategy—as reported in Scruggs et al.—research examining the impact of different dosages of these instructional strategies within the general education setting is warranted.

Second, teachers may not have sufficient training in how to implement these practices effectively (Klingner et al., 2013). For example, elements of instructional quality include how well the implementation of the specific method supports critical thinking and concept development, teacher allocation of available instructional time, the use of corrective feedback to support learning, and overall management and organization of students during instruction (e.g., Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). Effective implementation involves not only quality but also adherence, or fidelity, to the specific components of an instruction practice. Fidelity has been shown to have a strong effect on the relative effectiveness of an intervention (Hulleman & Cordray, 2009). In the present study, we did not explicitly code for the fidelity of implementation of specific components or quality of instruction during teacher implementation of these specific practices. Thus, it is possible that the low quality and/or fidelity of implementation underlie the lack of association between time utilizing the empirically supported practices and student performance.

Finally, it is important to note that the empirical support for the practices in this study was demonstrated via studies with students with mild disabilities (Scruggs et al., 2010). The sample of students in this study was not limited to students with disabilities but specifically focused on those with reading difficulties. This may be another potential explanation for the obtained results and highlights the need for caution when making interpretations about the lack of association between these instructional approaches and content learning in this study.

Limitations

Several limitations should be considered in the present study. First, as we focused explicitly on coding implementation of these specific practices, we did not collect specific data on instruction that occurred when these practices were not utilized. Second, our use of audio recordings of instruction presents another limitation. Although these recordings allowed for efficient data collection, they rely on verbalizations from the teacher or students to identify the specific instructional method or activity. For example, it is possible that students were presented with study aids and graphic/spatial organizers during instruction without the coders explicitly knowing and, thus, no coding would have occurred. Furthermore, the lack of visual input may have decreased the reliability of coding teachers' use of empirically supported strategies. However, we allowed coders to confer with one another when they were unclear, and interrater reliability across coders was generally high (above 90%) in this study. Third, although the content assessed on the ASK measure reflected the content intended to be taught, often per district curriculum standards, we did not collect data that examine the actual alignment of instruction and assessment items. Finally, due to the nature of our study utilizing existing recordings of classroom instruction, the scope of this research was limited to eighth-grade U.S. History classrooms and should not be generalized to other content areas or grade levels. Additional observational research at this level would help validate the current findings, and extending the research to other grade levels and content areas is warranted.

Conclusion

The present study provides preliminary data on the extent to which instructional practices that have promising empirical evidence for improving content knowledge for students with learning difficulties are used in the typical practice of secondary content-area teachers. In the case of the general education U.S. History classes observed in this study, such practices do not appear to be routinely implemented. Furthermore, the amount of time teachers spent in such instruction was not significantly associated with student content acquisition. In general, students' increase in knowledge over three units of instruction was minimal. The findings suggest that more research is needed to determine how to best impact content learning through supporting effective, quality implementation and maintenance of these and other practices with the potential to support those students most at-risk for poor learning outcomes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by Grant R305F100013 from the Institute of Education Sciences, U.S. Department of Education. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Biesta, G. (2007). Bridging the gap between educational research and educational practice: The need for critical distance. *Educational Research and Evaluation*, *13*, 295–301. doi:10.1080/13803610701640227

Bolinger, K., & Warren, W. J. (2007). Methods practiced in social studies instruction: A review of public school teachers' strategies. *International Journal of Social Education*, *22*, 68–84.

Boyle, J. R. (2010). Note-taking skills of middle school students with and without learning disabilities. *Journal of Learning Disabilities*, *43*, 530–540. doi:10.1177/0022219410371679

Broekkamp, H., & van Hout-Wolters, B. H. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, *13*, 203–220. doi:10.1080/13803610701626127

Bulgren, J. A., Deshler, D. D., & Lenz, B. K. (2007). Engaging adolescents with learning disabilities in higher-order thinking about history concepts using integrated content enhancement routines. *Journal of Learning Disabilities*, *40*, 121–133. doi:10.1177/00222194070400020301

Cervetti, G. N., Bravo, M. A., Hiebert, E. H., Pearson, P. D., & Jaynes, C. A. (2009). Text genre and science content: Ease of reading, comprehension, and reader performance. *Reading Psychology*, *30*, 487–511. doi:10.1080/02702710902733550

Chiodo, J. J., & Byford, J. (2004). Do they really dislike social studies? A study of middle school and high school students. *Journal of Social Studies Research*, *28*, 16–26.

Common Core State Standards Initiative. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Cook, B. G., Smith, G. J., & Tankersley, M. (2012). Evidence-based practices in education. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues* (pp. 495–527). Washington, DC: American Psychological Association. doi:10.1037/13273-017

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. doi:10.1037/1082-989x.12.2.121

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149–164. doi:10.1177/001440290507100202

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York, NY: Routledge. doi:10.4324/9780203836767

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88–110. doi:10.1080/19345740802539325

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. 1400 et seq. (2004)

Kelly, D., Xie, H., Nord, C. W., Jenkins, F., Chan, J. Y., & Kastberg, D. (2013). *Performance of U.S. 15-year-old students in mathematics, science, and reading literacy in an international context: First look at PISA 2012* (NCES 2014-024). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children*, *79*, 195–211. doi:10.1177/001440291307900205

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York: Carnegie Corporation of New York.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2006). *Gates–MacGinitie Reading Tests* (4th ed.). Rolling Meadows, IL: Riverside Publishing.

Mastropieri, M. A., & Scruggs, T. E. (2001). Promoting inclusion in secondary classrooms. *Learning Disability Quarterly*, *24*, 265–274. doi:10.2307/1511115

National Center for Education Statistics. (2013). *The nation's report card: A first look: 2013 mathematics and reading* (NCES 2014-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2014). *The nation's report card: U.S. History*. Retrieved from http://www.nationsreportcard.gov/hgc_2014/#history

National Center for History in the Schools. (1996). *History standards*. Retrieved from http://www.nchs.ucla.edu/history-standards/historical-thinking-standards

National Council for the Social Studies. (2008). *Curriculum guidelines for social studies teaching and learning: A position statement of the National Council for the Social Studies*. Retrieved from http://www.socialstudies.org/standards

National Council for the Social Studies. (2013). *The college, career, and civic life (C3) framework for social studies state standards: Guidance for enhancing the Rigor of K-12 civics, economics, geography, and history*. Retrieved from www.socialstudies.org/C3

Newman, L., Wagner, M., Huang, T., Shaver, D., Knokey, A.-M., Yu, J., . . . Cameto, R. (2011). *Secondary school programs and performance of students with disabilities. A special topic report of findings from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2012-3000). Washington, DC: National Center for Special Education Research, U.S. Department of Education.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Okolo, C. M., Ferretti, R. P., & MacArthur, C. A. (2007). Talking about history: Discussions in a middle school inclusive classroom. *Journal of Learning Disabilities*, *40*, 154–165. doi:10.1177/00222194070400020601

Paxton, R. J. (1999). A deafening silence: History textbooks and the students who read them. *Review of Educational Research*, *69*, 315–339. doi:10.3102/00346543069003315

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*, 365–397. doi:10.3102%2F0002831207308230

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2013-009 Revised). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Riley-Tillman, T. C., Chafouleas, S. M., Eckert, T. L., & Kelleher, C. (2005). Bridging the gap between research and practice: A framework for building research agendas in school psychology. *Psychology in the Schools*, *42*, 459–473. doi:10.1002/pits.20085

Scruggs, T. E., & Mastropieri, M. A. (2003). Science and social studies. In H. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 364–382). New York, NY: Guilford Press.

Scruggs, T. E., Mastropieri, M. A., Berkeley, S., & Graetz, J. E. (2010). Do special education interventions improve learning of secondary content? A meta-analysis. *Remedial and Special Education*, *31*, 437–449. doi:10.1177/0741932508327465

Slavin, R. E. (2008). Evidence-based reform in education: What will it take? *European Educational Research Journal*, *7*, 124–128. doi:10.2304/eerj.2008.7.1.124

Smith, G. J., Richards-Tutor, C., & Cook, B. G. (2010). Using teacher narratives in the dissemination of research-based practices. *Intervention in School and Clinic*, *46*, 67–70. doi:10.1177/1053451210375301

Swanson, E., Hairrell, A., Kent, S., Ciullo, S., Wanzek, J. A., & Vaughn, S. (2014). A synthesis and meta-analysis of reading interventions using social studies content for students with learning disabilities. *Journal of Learning Disabilities*, *47*, 178–195. doi:10.1177/0022219412451131

Swanson, H. L., & Deshler, D. (2003). Instructing adolescents with learning disabilities: Converting meta-analysis to practice. *Journal of Learning Disabilities*, *36*, 124–135. doi:10.1177/002221940303600205

U.S. Department of Education. (2014). *35th annual report to Congress on the Implementation of the Individuals With Disabilities Education Act, 2013*. Washington, DC: Office of Special Education and Rehabilitative Services, Office of Special Education Programs.

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research. *Exceptional Children*, *67*, 99–114. doi:10.1177/001440290006700107

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology*, *109*, 22–34.

Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A., & Stillman-spisak, S. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review*, *27*, 31–50.

Vaughn, S., Swanson, E., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, *48*, 77–93. doi:10.1002/rrq.039

Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., . . . Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness*, *7*, 183–204. doi:10.1080/19345747.2013.836765

What Works Clearinghouse. (2014). *Procedures and standards handbook: Version 3.0*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Yecke, C. P. (2005). *Mayhem in the middle: How middle schools have failed America—And how to make them work*. Washington, DC: Thomas B. Fordham Institute.