Contemporary Educational Psychology 68 (2022) 102040

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Improving literacy instruction in co-taught middle school classrooms to support reading comprehension

Jade Wexler [a,*], Devin M. Kearns [b], Christopher J. Lemons [c], Alexandra Shelton [a,d], Marney S. Pollack [b,e], Laura M. Stapleton [a], Erin Clancy [a,1], Erin Hogan [a,f,1], Cheryl Lyon [b,1]

[a] University of Maryland, Department of Counseling, Higher Education, and Special Education, College of Education, 3942 Campus Drive, College Park, MD 20742, United States
[b] University of Connecticut, Department of Educational Psychology, Neag School of Education, 249, Glenbrook Road, Unit 3064, Storrs, CT 06269, United States
[c] Stanford University, Center for Education Research at Stanford, 520 Galvez Mall, Stanford, CA 94305, United States
[d] Johns Hopkins University, 2800 N. Charles Street, Baltimore, MD 21218, United States.
[e] Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, United States.
[f] the University of Texas at Austin, 1912 Speedway D4900, Austin, TX 78712, United States.

ARTICLE INFO

Keywords:
Adolescents
Co-teaching
Literacy
Professional development

ABSTRACT

We examined the effects of a professional development (PD) with coaching model designed to improve literacy and co-taught instruction for students with and without disabilities in co-taught content-area middle school classes. Eleven co-teaching pairs in nine schools were randomly assigned to the Content Area Literacy Instruction (CALI) condition ($n = 7$ pairs) or a business-as-usual comparison condition ($n = 4$ pairs). All 22 teachers individually completed researcher developed pre- and posttests of teacher knowledge and skills and perceived effectiveness of their personal ability and their co-teachers' ability. At pre- and posttest, students ($n = 212$) completed three measures of reading comprehension. CALI teachers demonstrated significantly higher scores than comparison teachers at posttest on a measure of knowledge and skills, perceived personal effectiveness, and perceived co-teachers' effectiveness. All CALI co-teaching pairs demonstrated high levels of fidelity. Students in the CALI classrooms demonstrated significant gains on an essential aspect of a researcher-developed measure of reading comprehension. However, the treatment effect was non-significant for the two standardized measures of reading comprehension. Results provide initial support for a model in improving teacher instructional outcomes and student academic outcomes.

1. Introduction

As students reach the secondary grades, they are expected to learn complex English language arts (ELA), social studies, and science content by reading and comprehending upper-level narrative and expository text. This text typically includes complicated concepts, text structure, and vocabulary (Gajria et al., 2007). Harmon et al. (2005) explained that reading content-area text can feel like a bombardment of unfamiliar concepts to many middle school readers. Thus, comprehending this text can be challenging for many secondary students who do not possess the basic skills necessary for reading comprehension (e.g., main idea generation; Hagaman et al., 2016). For example, in 2019, 27% of eighth-grade students performed below basic on the National Assessment of Educational Progress (USDOE, 2019).

Many secondary students, including SWD, need evidence-based vocabulary and comprehension instruction that is intensive and explicit. Sometimes, specialized teachers (e.g., reading interventionists) provide this literacy instruction through supplemental intervention in addition to core instruction in the general education classroom. However, because most students, including students with disabilities and reading difficulties, spend a majority of their day in general education classes (U.S. Department of Education [USDOE], 2018), it is especially critical at the secondary level that teachers know how to seamlessly integrate vocabulary and comprehension instruction into their content-area

* Corresponding author.
E-mail addresses: jawexler@umd.edu (J. Wexler), devin.kearns@uconn.edu (D.M. Kearns), chris.lemons@stanford.edu (C.J. Lemons), ashelt18@jhu.edu (A. Shelton), marney.s.pollack@vanderbilt.edu (M.S. Pollack), lstaplet@umd.edu (L.M. Stapleton), eclancy1@umd.edu (E. Clancy), erin.hogan@utexas.edu (E. Hogan), Cheryl.lyon@uconn.edu (C. Lyon).
[1] Please note that the 7th–9th authors contributed equally and are listed in alphabetical order.

https://doi.org/10.1016/j.cedpsych.2021.102040

Available online 24 December 2021

curriculum. In theory, by regularly providing this instruction, teachers will be able to improve students' ability to read for understanding and, thus, meet their ultimate goal of increasing student content knowledge. One service delivery model that can be used to support the literacy needs of students with and without disabilities in the general education content-area setting is co-teaching, which occurs when a content-area teacher (CAT; i.e., a general education teacher in a secondary content area) and a special education teacher (SET) work together to provide instruction for students with and without disabilities in the general education setting (Cook & Friend, 1995). In this study, we examine the benefits of professional development (PD) on co-teachers' implementation of evidence-based literacy instruction and their students' literacy outcomes. Thus, in the following sections, we first discuss co-teaching, including the difficulty that co-teachers face integrating evidence-based literacy instruction in its intended form. Next, we summarize the extant literature influencing the development of the CALI IF and PD with coaching model. We then discuss the research purpose and design as well as the method of the study. Finally, we provide the results of our study and discuss how these results fill the gap in the literature on literacy instruction in co-taught classes.

1.1. Co-teaching

In theory, a CAT and SET can combine their unique content and pedagogical expertise to provide evidence-based literacy instruction to all students, including those who struggle with reading for understanding, (Conderman & Hedin, 2014). To do this, co-teachers can use enhanced co-teaching models (see Lemons et al., 2018), which incorporate clearly established roles for both teachers and require co-teachers to co-plan instruction designed to help students access content and meet content-area literacy expectations. For example, co-teachers can capitalize on a CAT's content knowledge by ensuring that the CAT provides instruction in key concepts and terms students need to know before students read text. On the other hand, SETs can apply their specialized training in incorporating features of effective instruction (e.g., modeling; Archer & Hughes, 2011) and making data-based decisions to provide explicit literacy instruction in skills students need to perform at a basic level, thereby enhancing students' ability to comprehend the text their co-teachers assign them to read.

Despite the fact that co-teaching is used throughout the nation across grade levels (Murawski & Lochner, 2011) and provides teachers the opportunity to support students with and without disabilities in the general education content-area setting, co-teachers may face difficulty integrating evidence-based literacy instruction in its intended form. In fact, Wexler and colleagues (2018) reported a lack of text reading, limited literacy instruction, and overall poor implementation of co-teaching practices in a recently conducted observation study of 16 middle school co-teaching pairs in ELA classrooms. The authors reported that more than half of the time spent on literacy activities involved reading aloud or silently with no co-occurring literacy instruction to support reading comprehension. Furthermore, teachers primarily implemented whole-class instruction or had students working independently with little student–teacher interaction. Finally, the use of co-teaching models in which SETs spent a majority of their time supporting CATs as they led whole-class instruction was prevalent—a finding consistent with previous studies showing that SETs frequently take on a subordinate role in the co-taught classroom (e.g., Rice & Zigmond, 2000; Weiss & Lloyd, 2002). Put simply, extant data indicate that secondary co-teachers are missing opportunities to provide students with and without disabilities evidence-based literacy instruction. Therefore, it is clear that co-teachers need more opportunities for effective PD on providing evidence-based literacy instruction to all students. Thus, our research team set out to iteratively develop the CALI instructional framework (IF) and associated PD with coaching model to help middle school co-teachers seamlessly integrate evidence-based literacy instruction into their content-area classes.

1.2. Theoretical and conceptual foundations of project CALI

We considered the CALI IF and PD with coaching model to be inextricably linked. That is, we specifically designed our model to prepare co-teachers to implement the practices in the CALI IF. Conceptualizing the intervention as an inseparable package of instructional practices and PD is common in PD studies (see McMaster et al., 2018). Thus, in our Theory of Change (see Fig. 1), we hypothesized that participation in the CALI PD with coaching model, which focused on ways co-teachers could leverage the strengths of co-teaching to integrate evidence-based literacy instruction into their content areas, would have a direct effect on co-teachers' implementation of the CALI IF with high levels of fidelity. We also assumed that participation would lead to an increase in co-teachers' knowledge and skills (e.g., ability to generate main idea statements), which would also lead to high implementation fidelity. Further, teachers' high implementation fidelity would result in an improvement in their beliefs (i.e., personal perceived effectiveness and perceived effectiveness of one's co-teacher), which would encourage teachers to continue implementing the IF with high fidelity. Ultimately, co-teachers' high implementation fidelity of the CALI IF would lead to improvement of students' reading comprehension. Following is a brief summary of the theoretical background and research base informing the practices in the IF and the PD with coaching model.

1.2.1. Instructional framework

Although secondary co-teachers in the content-area setting may not be able to address all of the literacy needs of all students while also satisfying curricular goals, they can enhance comprehension of text at the micro- (i.e., units of information; Kintsch & van Dijk, 1978; McNamara & Magliano, 2009) and macrostructure levels (i.e., overall meaning of the text) in purposeful ways. Specifically, co-teachers can provide explicit instruction on a limited set of evidence-based teacher- and student-led literacy practices that target skills needed for comprehension. Thus, we designed the CALI IF to help co-teachers systematically pre-teach key concepts and vocabulary needed to understand text (Burns et al., 2011; Cromley & Azevedo, 2007) and provide instruction in evidence-based literacy strategies that students can use to help with their own cognitive processing (e.g., main idea generation; Stevens et al., 2019).

Knowing that co-teachers would be implementing the CALI IF practices, we also designed the CALI IF in such a way that would purposefully capitalize on the presumed benefits of having a CAT and a SET—each with their own unique strengths—implement it. First, the CALI IF practices lend themselves to the distinct skills that each teacher brings to the classroom. For example, the practices capitalize on the CAT's content knowledge (e.g., World and Word Knowledge; see below) and include a specialized role for the SET during a small-group station-teaching lesson (i.e., providing intensive instruction; see below). Second, features of effective instruction are intentionally integrated into the practices and highlighted through the co-taught nature of the instruction. Results from many studies suggest that students with and without disabilities benefit when teachers incorporate features of effective instruction into their evidence-based literacy instruction (Vaughn & Wanzek, 2014). These features include modeling, guided practice, and independent practice to intensify instruction (MacSuga-Gage & Simonsen, 2015; Rosenshine, 2012; Solis et al., 2014) as well as frequent opportunities for students to respond, practice, and receive immediate corrective feedback. Third, because enhanced co-teaching models include instruction that is structured to allow for differentiation (Lemons et al., 2018), we include a data-based differentiated support component for students. Therefore, informed by both theory and research, we identified five evidence-based practices to make up the CALI IF. These practices are World Knowledge, Word Knowledge, Get the Gist, Associate Gist, and Student Support.

World Knowledge and Word Knowledge. According to the Construction-Integration model of text comprehension (Kintsch, 1988),

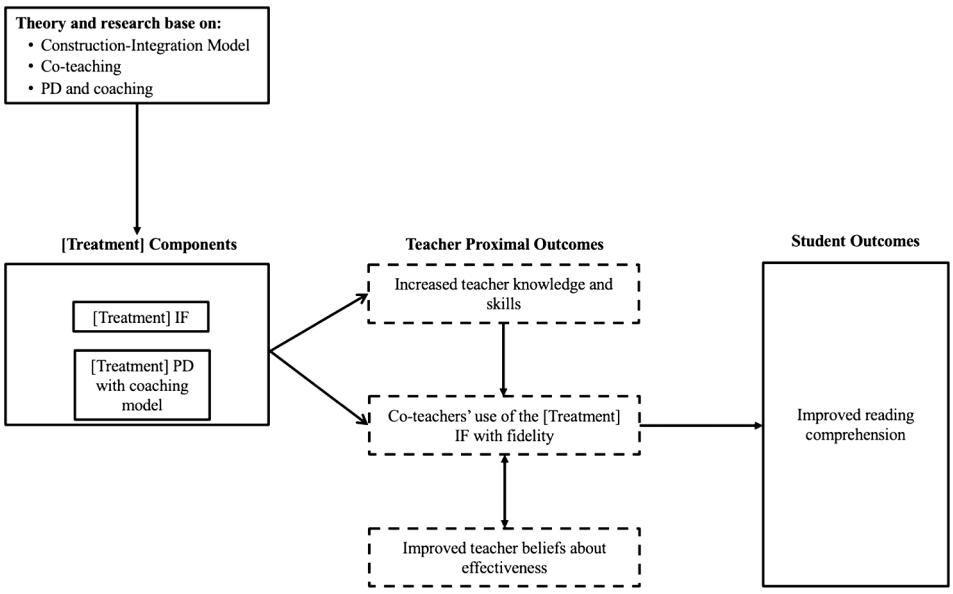

Fig. 1. CALI Theory of Change.

in order to develop a coherent understanding of a text, readers must be able to activate knowledge (including background and vocabulary knowledge), use that knowledge to create a mental representation of the text, and establish connections within their mental representation (McNamara & Magliano, 2009). Therefore, students' ability to make sense of text is enhanced when they have background knowledge about a text and when they know the meaning of essential words in a text (Elleman et al., 2009; Kearns et al., 2021). Accordingly, the first two components of the CALI IF—World and Word Knowledge—draw from the Construction-Integration model of text comprehension and compose a routine that co-teachers use to provide students with background information and vocabulary instruction before reading a text for a CALI lesson.

Get the Gist and Associate Gist. The Construction-Integration model also posits that reading comprehension at the macrostructure level is informed by comprehension of the microstructure. In a multi-paragraph text, for example, readers must understand the individual ideas the author conveys throughout the various sentences, which are organized in different paragraphs that compose the text. This process can be difficult for students who struggle with reading for understanding due to limitations with working memory and cognitive load (Cain et al., 2004). Nonetheless, extensive research reveals that explicit main idea instruction can help students comprehend individual units of information in a text in order to comprehend the overall meaning of the text (Stevens et al., 2019). For example, Get the Gist is an evidence-based strategy designed to help students generate a main idea statement (Vaughn et al., 2001). In fact, in Stevens et al. (2020), students who received explicit instruction on how to identify main ideas using Get the Gist performed better than comparison students on two measures of main idea generation. Given the strong evidence base supporting Get the Gist, we incorporated this strategy into the CALI IF as a method for supporting students' text comprehension during CALI lessons.

The CALI IF also includes Associate Gist, which builds on Get the Gist by incorporating peer-mediated practice (Wexler et al., 2018). Peer-mediated practice provides students opportunities to respond and immediate corrective feedback from peers and has evidence of effectiveness at the secondary level (Cawley et al., 2002; Wexler et al., 2015). Heterogeneous co-taught classrooms are particularly conducive to peer-mediated practice due to the ability to strategically pair a higher-level reader with a slightly lower-level reader. Additionally, peer-mediated practice is a desirable choice in a co-taught classroom because two teachers can model the peer-mediated practice routines more precisely and monitor a larger number of student pairs or groups during student practice than a single teacher.

Student Support. Some students in a heterogeneous co-taught class may continue to struggle with basic skills (e.g., main idea generation) despite the CALI IF practices described above, suggesting the need for more intensive, individualized instruction to meet their specific needs (Tomlinson & Imbeau, 2010). Meanwhile, other students who have demonstrated mastery of main idea generation may benefit from extended learning opportunities. In fact, research demonstrates that differentiated literacy instruction can lead to improved literacy outcomes (Puzio et al., 2020). Therefore, we included station-teaching Student Support (Lyon et al., 2021) as our final CALI IF practice to provide students with differentiated instruction. Co-teachers can implement Student Support to target students' needs with lessons that support the development of their foundational reading comprehension skills or extend their thinking.

1.2.2. PD with coaching model

The general purpose of PD is to improve teacher knowledge and skills so that teachers can successfully improve student performance (Yoon et al., 2007). Many effective PD opportunities also include follow-up coaching (Yoon et al., 2007). A recent *meta*-analysis of 60 studies revealed that PD in the form of group training with follow-up coaching has generally positive effects on both teachers' instruction and students' academic performance (Kraft et al., 2018). Thus, as specified in our Theory of Change, teachers' participation in the CALI PD with coaching model would both directly and indirectly (via improved knowledge and skills) lead to teachers' implementation of the CALI IF with fidelity.

The PD sessions were designed to provide teachers with structured ways to teach each practice as well as ways to plan for implementation that emphasize efficiency and delineating co-teachers' roles during implementation (Murawski & Dieker, 2004). Specifically, we incorporated features of effective PD and coaching, including a focus on content that is applicable and feasible, active learning opportunities, and a duration of at least one semester and between 20 and 100 h of contact time (Desimone, 2009; Garet et al., 2016).

The CALI PD with coaching model was also designed to include features of effective instruction to ensure that teachers were well prepared to teach students how to use the CALI IF practices. Throughout the four PD sessions that spanned the study year, we provided teachers with explicit instruction and modeling to ensure that critical concepts of the CALI IF were presented exactly as intended. This explicit guidance also

ensured that co-teachers knew how to capitalize on the benefits of co-teaching to implement the CALI IF (e.g., via models of co-planning). We also provided co-teachers with several opportunities to practice implementing the target CALI IF practices with varying levels of support. Additionally, we provided teachers feedback after each implementation from designated members of the research team (i.e., helpers; see the Method section for more information). Thus, the CALI IF was taught with clear guidance and exact models of appropriate practice so that co-teachers could follow it closely to implement the CALI IF practices and the designated co-teaching models (e.g., station teaching) effectively.

1.3. Research purpose and design

Considering that many students struggle with reading comprehension and spend most of their day in the content-area setting, it is critical that teachers are adept at providing literacy instruction. In classrooms where there are co-teachers, there is an opportunity to capitalize on this service delivery model in unique ways that allow for the integration of evidence-based literacy instruction. To do this, PD and coaching support is necessary. Thus, to achieve the goal of improving literacy achievement for secondary students in the co-taught content-area setting, we conducted a three-year project. During the first and second years of the project, we conducted a mixed-method study to investigate gaps in co-teaching and literacy PD and coaching as well as literacy instruction delivered by co-teachers in content-area classes (Sinclair et al., 2018; Wexler et al., 2018). To target needs based on our findings, we also worked closely with partner teachers to iteratively develop the CALI IF and associated PD with coaching model. During the third year of the project, we conducted the current experimental study to evaluate the feasibility and potential effectiveness of the CALI IF and PD with coaching model for improving related teacher and student outcomes. Thus, the focus of the current article is to discuss findings related to year 3 of this project.

We addressed the following research questions in the current study: (1) What is the effect of the CALI PD with coaching model compared with a comparison group on teachers' ability to generate a main idea statement from text and their perceptions of their own effectiveness and their co-teachers' effectiveness?; (2) To what extent do co-teachers who participate in the CALI PD with coaching model implement the IF with fidelity?; (3) What is the effect of the CALI IF compared with a comparison group on the reading outcomes of middle school students in co-taught classrooms?

2. Method

2.1. Setting and participants

2.1.1. Sites

Following institutional and district review board approvals, our study took place in school districts in the Mid-Atlantic and Northeastern United States. The research team worked with district personnel to identify middle schools with at least one co-teaching pair in ELA, science, or social studies. Administrators in nine middle schools in six districts, diverse in their settings and student populations, agreed to allow their teachers and students to participate in the study. See Table 1 for district demographics.

2.1.2. Teachers

Following administrator approval, we recruited eligible pairs of co-teachers. Co-teachers were eligible if (a) they co-taught at least one class period, (b) the co-taught class included at least one SWD who was proficient in English and had an Individualized Education Program (IEP) that contained a reading goal, and (c) the class also included at least one student without a disability who was proficient in English and on the prior year's state reading assessment. Eleven co-teaching pairs agreed to participate in the study. Seven schools contained one pair each, and two schools contained two pairs. Eight pairs taught ELA (5 T; 3C), and 3 pairs taught social studies (2 T; 1C). Each pair included a SET and a CAT.

The majority of teachers were female ($n = 18$). While teachers had a range of prior teaching experience, most held Master's degrees ($n = 14$) and all teachers were professionally licensed in their respective fields (e.g., ELA). During the year that the study took place, most of the pairs ($n = 6$) were co-teaching together for the first time. The remaining pairs ($n = 5$) had between two and five years of prior experience co-teaching together. See Table 2 for teacher demographics disaggregated by condition.

2.1.3. Students

All students in each co-teaching pair's participating class period were eligible for participation in the study. We obtained parental consent to conduct pretests and posttests with 212 6th-8th grade students. This represented at least 90% of students in each participating class. A majority of the students were in 7th grade ($n = 151$; 71.2%). Additionally, 47 students (22.2%) were SWD. See Table 3 for student demographics disaggregated by grade level.

2.2. Study design and overview

We used a pretest/posttest blocked random assignment control group design. Co-teaching pairs volunteered to participate in the study and were then randomly assigned to our treatment condition ($n = 7$ pairs; PD and coaching on the CALI IF) or the business-as-usual (BAU) comparison condition ($n = 4$ pairs), which is common in PD and coaching studies. For example, McMaster et al. (2020) also used random assignment of teachers (n = 19) to their treatment condition or a business-as-usual comparison condition.

For the two schools that contained two pairs, blocked random assignment was used to ensure that at least one co-teaching pair was assigned to treatment. If the pairs taught more than one eligible class period together, the pairs and research team members mutually selected one eligible class period for participation in the study, based primarily on the scheduling preference and teachers' perceived needs of the

Table 1
District Demographics 2017–2018 School Year.

				Race							
District	Region	Setting	Enrollment	White	Black	Hispanic	Other	P/A	LEP	FRL	SPED
1	Mid-Atlantic	Urban	49,117	13%	63%	20%	4%	34%	13%	78%	13%
2	Mid-Atlantic	Suburban	143,137	37%	24%	22%	16%	39%	13%	33%	12%
3	Northeast	Suburban	814	82%	2%	9%	7%	73%	1%	32%	14%
4	Northeast	Rural	375	90%	2%	2%	4%	70%	NR	14%	11%
5	Northeast	Suburban	2.128	93%	2%	2%	3%	76%	1%	8%	17%
6	Northeast	Suburban	1,224	67%	NR	13%	18%	76%	5%	25%	13%

Note: P/A = percent of students out of total district enrollment who scored proficient or advanced on the reading portion of the state test; LEP = percent of students out of total district enrollment who were considered limited in English proficiency; FRL = percent of students out of total district enrollment who qualified for free or reduced lunch; SPED = percent of students out of total district enrollment who received special education services; NR = Not reported by the district.

Table 2
Teacher Demographics by Condition.

	Comparison ($n = 8$) n (%)	Proportion of total (%)	Treatment ($n = 14$) n (%)	Proportion of total %
Licensure Status				
Professionally licensed	8 (100%)	36.4%	14 (100%)	63.6%
Not licensed for area	0	0	0	0
Highest Degree Obtained				
Master's	5 (62.5%)	22.7%	9 (64.3%)	40.9%
Bachelor's	3 (37.5%)	13.6%	5 (35.7%)	22.7%
Teaching Experience				
1–5 years	3 (37.5%)	13.6%	1 (7.1%)	4.5%
6–10 years	2 (25%)	9.1%	2 (14.3%)	9.1%
11–15 years	1 (12.5%)	4.5%	3 (21.4%)	13.6%
16–20 years	1 (12.5%)	4.5%	4 (28.6%)	18.2%
21 years or more	1 (12.5%)	4.5%	4 (28.6%)	18.2%
Co-teaching Experience for Pair[a]				
First time co-teaching together	2 (50%)	18.2%	4 (57.1%)	36.4%
2 years	1 (25%)	9.1%	1 (14.3%)	9.1%
3 years	0	0	2 (28.6%)	18.2%
5 years	1 (25%)	9.1%	0	0
Age				
20–29	1 (12.5%)	4.5%	2 (14.3%)	9.1%
30–39	3 (37.5%)	13.6%	4 (28.6%)	18.2%
40–49	1 (12.5%)	4.5%	3 (21.4%)	13.6%
50–59	3 (37.5%)	13.6%	2 (14.3%)	9.1%
60–69	0	0	3 (21.4%)	13.6%
Race/Ethnicity				
Asian Am.	0	0	1 (7.1%)	4.5%
Black	0	0	2 (14.3%)	9.1%
Latinx	0	0	0	0
Native Am.	0	0	0	0
White	8 (100%)	36.4%	11 (78.6%)	50.0%
Gender				
Female	7 (88%)	31.8%	11 (78.6%)	50.0%
Male	1 (13%)	4.5%	3 (21.4%)	13.6%

Note: [a]Co-teaching experience totals and percentages determined by number of pairs.

students in the classes. Therefore, the other classes were not included.

All treatment and comparison co-teachers participated in a study orientation meeting in mid-to-late September. The primary purpose of the meeting was to provide an overview of study procedures and expectations. CALI teachers were also provided an overview of the CALI IF and learned how a member of the research team, a 'helper' (i.e., coach), would support them. We used the term 'helper' because it suggested a collaborative coaching approach. In the two schools that included both a CALI and comparison pair, helpers explained the concept of 'intervention weep' during the orientation meeting, and CALI teachers were instructed to not share materials with their colleagues to protect against treatment contamination.

Comparison pairs continued with their typical practice content-area co-teaching instruction. They did not receive any CALI training, support, or materials. CALI pairs participated in four PD sessions, distributed across the school year, to learn all of the practices in the CALI IF. See Fig. 2 for a timeline of all key CALI activities (i.e., assessment, PD, and coaching). They also received on-going support from CALI helpers and all accompanying materials (e.g., lesson templates). We provide additional detail on the PD sessions and helper support model in the Procedures section below. At the end of the study, the research team gave comparison pairs access to the CALI materials and invited them to participate in an abbreviated training.

2.3. Procedures

2.3.1. Helpers and helper training

Three graduate students at the Mid-Atlantic site and two part-time research associates at the Northeastern site served as helpers for treatment pairs. Four of the five helpers were certified classroom teachers and collectively had over ten years of experience working as a SET or CAT in a co-taught classroom. The one helper who was not a certified classroom teacher had 3 years of experience working as a special education paraprofessional. Each co-teaching pair had a primary and a secondary helper. A primary helper led PD and coaching activities with a co-teaching pair and conducted most of the teacher fidelity checks (see Fidelity of CALI Implementation section below). A secondary helper attended PD sessions during which the secondary helper conducted PD fidelity checks and participated when modeling with two people was necessary (e.g., modeling the Associate Gist routine). The secondary helper also participated in coaching activities when schedules permitted and when modeling was required and conducted occasional teacher fidelity checks during CALI implementation (see Fidelity of CALI Implementation section below).

During a two-day training prior to the start of the school year, helpers were trained on (a) study background and logistics, (b) the CALI IF and PD with coaching model, and (c) fidelity of implementation procedures. The training included modeling and practice scenarios led by principal investigators (PIs) and the lead project coordinator from the Mid-Atlantic site. Throughout the year, helpers participated in weekly "helper calls" with PIs to problem-solve issues related to study implementation and logistics. Helpers also participated in booster training

Table 3
Student Demographics by Condition and Grade.

	Comparison ($n = 78$)				Treatment ($n = 134$)			
	Grade 6 ($n = 0$) n (%)	Grade 7 ($n = 71$) n (%)	Grade 8 ($n = 7$) n (%)	Comparison total n (%)	Grade 6 ($n = 38$) n (%)	Grade 7 ($n = 80$) n (%)	Grade 8 ($n = 16$) n (%)	Treatment total n (%)
SWD	0	16 (23%)	0	16 (20.5%)	6 (15.8%)	19 (23.8%)	6 (37.5%)	31
ELL	0	4 (6%)	0	4 (5.1%)	3 (7.9%)	4 (5.0%)	0	7
Race/ Ethnicity								
Asian Am.	0	6 (8.5%)	1 (14.3%)	7 (8.9%)	0	2 (2.5%)	0	2 (1.5%)
Black	0	11 (15.5%)	0	11 (14.1%)	10 (26.3%)	16 (20%)	0	26 (19.4%)
Latinx	0	6 (8.4%)	0	6 (7.7%)	6 (15.8%)	19 (23.8%)	0	25 (18.7%)
Native Am.	0	1 (1.4%)	0	1 (1.3%)	0	0	0	0
White	0	46 (64.8%)	6 (85.7%)	52 (66.7%)	0	43 (53.8%)	16 (100%)	59 (44.0%)
NR	0	1 (1.4%)	0	1 (1.3%)	22 (57.9%)	0	0	22 (16.4%)
Gender								
Female	0	39 (55%)	4 (57.1%)	43 (55.1%)	20 (52.6%)	43 (53.8%)	7 (43.8%)	70 (52.2%)
Male	0	32 (45%)	3 (42.9%)	35 (44.9%)	18 (47.4%)	37 (46.2%)	9 (56.2%)	64 (47.8%)

Note: SWD = Student with disability; ELL = English Language Learner; NR = Not reported.

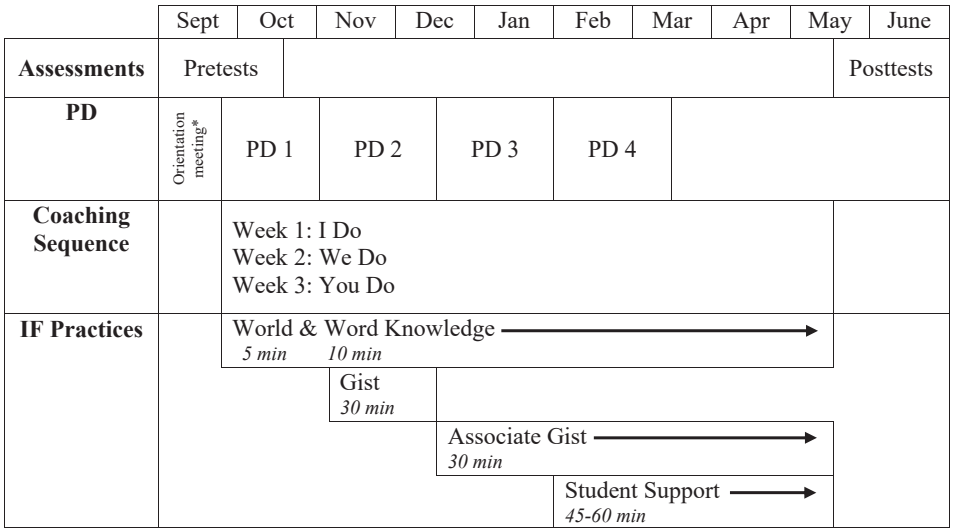

Fig. 2. CALI timeline of key activities. *For CALI and comparison teachers.

sessions with a PI prior to the delivery of each PD session to review PD content and implementation procedures.

Prior to testing in the schools, the lead project coordinator at the Mid-Atlantic site trained all helpers to administer and score pre- and post-tests. In a four-hour training, the lead project coordinator provided an overview of each assessment and reviewed scoring procedures. Next, the lead project coordinator facilitated peer administration and scoring practice with two previously developed master scripts and score sheets for each assessment. Discrepancies were resolved through discussion. Finally, each helper was required to administer and score each assessment with the lead project coordinator who used another previously developed master script. All helpers demonstrated at least 90% accuracy of administration and 90% inter-scorer agreement with the lead project coordinator for all assessments prior to testing in the schools. Agreement was calculated as the number of agreements divided by the sum of agreements and disagreements multiplied by 100.

2.3.2. *Pre- and post-testing*

Teachers completed pre- and post-tests independently (i.e., not with their co-teacher). Helpers administered pretests to teachers at the end of the study orientation meeting, prior to pairs delivering any CALI instruction. Helpers administered post-tests to teachers at the end of the school year, following completion of the study. See Fig. 2 for a timeline of assessment activities. Helpers administered paper-based pretests and posttests to all students in treatment and comparison classrooms to ensure administration accuracy with the exception of one measure that was computer-based.

2.4. *CALI IF and PD with coaching model*

We provided co-teachers PD with coaching support on the CALI IF, as described in the Introduction. Specifically, co-teaching pairs were taught at least one CALIIF practice per PD session, which was followed by coaching support from helpers. We next describe implementation procedures of the CALI IF and PD with coaching model. See www.projectcali.uconn.edu for a copy of IF and PD materials.

2.4.1. *Instruction for students: CALI instructional framework*

The practices included in the CALI IF are intentionally designed to be text and curriculum agnostic. In other words, teachers can link the CALI IF to any text, and they can implement it in any content-area classroom. Therefore, to design a CALI lesson, CALI pairs selected a text that aligned with content they were teaching using guidelines provided by helpers, per CALI text selection guidelines (see Wexler et al., 2021). Using that text, co-teachers taught a CALI lesson that included instruction in the CALI IF practices. CALI pairs were instructed to deliver one CALI lesson per week during the months of the school year when the first three PD sessions were being delivered. Following the fourth and final PD session, CALI pairs were expected to teach CALI lessons every other week due to their heightened intensity and length. Thus, CALI lessons reflected the additive nature of the PD, becoming longer as more components of the CALI IF were added. See Fig. 2, which depicts when the different IF practices were taught in each of the four PD sessions. We next describe implementation procedures for each practice in the CALI IF and the PD sequence teachers participated in to learn the different practices.

World Knowledge and Word Knowledge. The teachers introduced the text selected for the CALI lesson by providing students with essential background knowledge (i.e., World Knowledge, Kearns et al., 2021) and instruction on key vocabulary terms (i.e., Word Knowledge; Kamil et al., 2008; Kearns et al., 2021). Co-teachers used different co-teaching models (e.g., one-teach, one-assist or team teaching) to teach World and Word knowledge on a fluid basis, as determined during co-planning sessions supported by their helper (see below). CALI teachers learned to plan for and implement these practices during the first PD, *'PD1: Books and Background'* (6 h).

Get the Gist. The second step in the CALI IF is for teachers to introduce a purpose question to focus students' attention on key concepts before reading. Students then read the text and independently apply Get the Gist (Shelton et al., 2021). Specifically, students identify who or what a section of text is about and the most important information about the who or what in order to write a gist statement (i.e., main idea statement). Students then use their independently generated gist statements to write a text summary, which ultimately answers the purpose question. By engaging in this process, students hone their ability to comprehend text at the microstructure and macrostructure level. Similar to World and Word Knowledge, co-teachers used different co-teaching models on a fluid basis to implement Get the Gist, as determined during co-planning sessions (see below). During the second PD, *'PD2: Comprehension with Purpose and Gist'* (4 h), teachers learned how to (a) generate their own gist statements, (b) use their gist statements to generate a purpose question for a text, (c) teach students to use Get the Gist independently, and (d) teach students to use their generated gist statements to write a text summary.

Associate Gist. Designed to build on independent Get the Gist, Associate Gist (Shelton et al., 2021), a peer-mediated summarization activity, replaced independent Get the Gist. Co-teachers engage students in Associate Gist once most of their class becomes proficient in generating a gist statement independently. For each section of text, students

generate a gist statement independently and then work with a peer to develop an improved joint gist statement. Upon completion of Associate Gist, students independently write a text summary. Teachers learned the Associate Gist procedure, how to pair students strategically, and how to model Associate Gist by having co-teaching pairs act as student pairs—thus capitalizing on the heterogeneous nature of the co-taught classroom—during the third PD, *'PD3: Discussion of Gist with Associates'* (3 h).

Student Support. The last step in the CALI IF is for students to participate in a data-based differentiated station-teaching support lesson. Station teaching is a high-value co-teaching model that draws on the ability of co-teachers to make data-based instructional decisions about students' individual needs in order to provide instruction in rotating homogeneous small groups (Lyon et al., 2021). During one class session (day one; approx. 45 min), co-teaching pairs provide a 'full' CALI lesson (i.e., World and Word Knowledge and Associate Gist). During the next class session (day two; approx. 45 min), they provide a support lesson. Teachers diagnostically evaluated students' day-one independently generated gist statements (i.e., the gist statements each student generated prior to working with an associate) to determine which small group to place students in for the support lesson: a review group, a practice group, or an extend group. Students rotated in these small groups between CAT-led, SET-led, and independent practice stations. Students in the review group received intensive instruction focusing on generating gist statements. Specifically, students learned to use sentence-level Get the Gist (Pollack et al., 2021), a strategy that involves synthesizing sentence-level information (i.e., microstructure) in a section of text to help identify the main idea of the section (i.e., macrostructure). Students in the practice group received repeated practice on how to develop gist statements using the same procedures previously learned. Students in the extend group participated in enrichment activities related to the day one CALI lesson (e.g., designing their own World and Word Knowledge lesson for a new text related to the initial text). During this final PD session, *'PD4: Everybody Gets What They Need'* (4 h), teachers learned about the benefits of station teaching; how to implement the review group, practice group, and extend group activities using an effective station-teaching format (e.g., how co-teachers can implement group rotations); and how to design instruction for each group.

2.4.2. Instruction for teachers: CALI PD

The four PD sessions were spaced 3 weeks apart to allow teachers to learn at least one new component of the CALI IF during each session. PD sessions varied in length between 3 and 6 h, and helpers delivered all PD sessions in person to intact pairs at a time convenient for them (e.g., over one day or split between two days).

To help teachers learn how to implement the CALI IF practices during PD, video lectures, developed by the research team and video-recorded by one of the PIs, were used to ensure that critical concepts about CALI were presented exactly as intended. Second, all of the PD materials (e.g., PowerPoints) were designed using principles of effective multimedia design (Mayer, 2001). Third, helpers frequently engaged the teachers in processing activities, such as short activities to check for understanding of key ideas, activities to evaluate videotaped lessons, and activities designed to provide teachers opportunities to practice planning for and implementing the CALI IF. The goal was to reduce the cognitive load of learning a set of complex new strategies, following Martin's (2016) load-reduction instruction approach. Fourth, towards the end of each PD session, helpers provided teachers with an updated CALI lesson template (i.e., a Google Slides or PowerPoint presentation) designed to expedite co-planning and facilitate co-teaching, with customizable slides for lesson components, scripted lessons to teach the components to students, and reminders about instructional techniques taught during the PD as well as logistics. These steps maximized clarity and fidelity to help teachers successfully implement the CALI IF in their classrooms, while capitalizing on the co-taught nature of the class.

2.4.3. Fidelity of PD

Fidelity of PD was evaluated to be sure that PD sessions were implemented as intended. For each PD session, a PD implementation guide was created. Each guide includes key PD activities that must be implemented (e.g., presentation of slides, videos, and activities) as well as PD instructions and a presenter script. Because two helpers attended PD sessions, the secondary helper used the guide to record whether the primary helper implemented each activity as intended and assigned a global quality rating score on a scale of 0 (i.e., the PD session did not occur) to 3 (i.e., all PD activities were delivered as intended with high quality) at the end of each PD session.

2.4.4. Instruction for teachers: helper support

To avoid a one-time sit-and-get PD (Yoon et al., 2007), helpers provided structured coaching support after each PD session for 3 weeks. See Fig. 2 for a timeline of the coaching sequence. The weekly 'plan-implement-reflect' coaching sequence was led by a primary helper, and a secondary helper assisted when schedules permitted or modeling with two people was required, as described above. Plan sessions consisted of helpers and/or pairs working through (a) tasks, such as selecting text and filling out a lesson template, (b) instructional procedures, and (c) roles each teacher would have while implementing the subsequent lesson. An implement session was an actual CALI lesson implemented by the helpers and/or pairs. During reflect sessions, helpers and pairs used the CALI implementation checklist as a coaching tool to discuss lesson 'glows and grows'. Reflect sessions also provided an opportunity to troubleshoot or discuss new challenges (e.g., student behavior). Both plan and reflect sessions were conducted virtually (e.g., on Skype) if requested by the pairs due to scheduling conflicts. In addition to the coaching sequence sessions, primary helpers were available on a 'by request' basis and were required to 'check in' with teachers (e.g., email with follow up praise or reminders) at least once weekly.

The specific steps of the plan, implement, and reflect sessions varied based on the week of the coaching sequence. Specifically, helpers gradually released responsibility of implementing the CALI IF to each pair over the following 3 weeks using an I do, we do, you do model. During the first week of a coaching sequence (i.e., the week each PD session was provided), helpers modeled how to plan a CALI lesson using the lesson template and then modeled implementing the newly learned components of the CALI IF for each pair in their participating classroom (i.e., I do). During this modeled lesson, pairs observed the helpers and simultaneously filled out the CALI implementation checklist to use as a means of reflection during the reflect session. Helpers and pairs then participated in a helper-led reflect session after the modeled lesson. During the second week of a support sequence, helpers guided each pair by co-planning, co-implementing, and co-reflecting about the lesson (i.e., we do). Finally, during the third week of a support cycle, the pairs planned and then implemented the next lesson without assistance from a helper (i.e., you do). Helpers and pairs participated in a pair-led reflect session after the lesson. Overall, the coaching sequence allowed helpers to provide pairs with scaffolded support and active learning opportunities to ensure that their knowledge and skills adequately increased over time.

2.5. The comparison condition

Based on an end-of-year survey completed by 6 comparison teachers, comparison teachers verified that they did not participate in any PD or coaching support related to co-teaching during the school year the study took place. Additionally, only three comparison teachers reported receiving PD (but not coaching) in the area of literacy during the course of the study. In addition, the lead project coordinator at the Mid-Atlantic site and two part-time research associates at the Northeastern sites observed comparison classrooms at the beginning, middle, and end of the year, taking detailed field notes about the types of literacy instruction and co-teaching practices that occurred within these classes.

During observations, the most typical student grouping was whole-class instruction. In general, while comparison co-teachers implemented a variety of possible co-teaching models (e.g., team-teaching), one-teach, one-assist—during which CATs led instruction and SETs took on a supportive role—was most prevalent. Additionally, one pair typically facilitated project-based learning. Across all observations, co-teachers provided students with limited exposure to connected text and co-occurring evidence-based literacy instruction to support reading comprehension (e.g., explicit vocabulary instruction, main idea instruction). When students did have exposure to text, students rarely read aloud or silently. Instead, one of the co-teachers typically read to the entire class.

2.6. Measures

2.6.1. Teacher outcomes

Corresponding to our Theory of Change, we administered several measures to assess proximal teacher outcomes.

Main Idea Measure (MIM). At pre- and posttest, teachers individually completed a paper-based, timed, researcher-developed measure of teacher knowledge and skills, the MIM. The MIM tested proficiency with the familiar skill of summarizing by "getting the main idea" (Stevens et al., 2019). Teachers were provided six short passages (approximately 200 words each) at the instructional level of the students they taught (i.e., Grades 5–6; Lexile range 640–1060). Once provided the passages, teachers were given 15 min to silently read and generate a statement summarizing the main idea of each passage. The directions were to *"Read each section silently to yourself. Next, write the main idea of each section on the lines below it."* The treatment-neutral wording of the directions was designed to maximize the MIM's ecological validity.

Helpers scored each of the six main idea statements using a rubric to judge whether the main idea statements included essential information. Each main idea statement was evaluated for its identification of two items: 1) the primary subject and 2) the most important information about the primary subject. For each item, teachers received a score of 0 (incorrect) or 1 (correct). Teachers received three final MIM scores: the sum of the item 1 (primary subject) scores across the six passages, the sum of the item 2 (most important information) scores across the six passages, and the sum of the item 1 and 2 scores across the six passages (i.e., the overall main idea statement). For the overall main idea statement, which combines scores for items 1 and 2 across the six passages, teachers could receive up to 12 points (6 points per item).

Pretests and posttests were blinded and scored after their respective administrations. To score each item, helpers used anchor items established by the lead project coordinator at the Mid-Atlantic site and one of the co-PIs of the study. Each passage in the MIM had a corresponding anchor that provided acceptable answers for the two items on the rubric (i.e., primary subject and most important information). All assessments were double entered into a researcher-created database. To calculate reliability on the MIM at pre- and posttest, 30% of all responses (i.e., teacher and student responses combined; see below for description of student MIM) on the first three passages were randomly selected and scored individually by two helpers. Helper scores were compared to ensure IOA of at least 90%. Helpers then scored the remaining responses for the first three passages (i.e., the remaining 70%) independently. This procedure was repeated with the second three passages. The primary analysis was for the overall main idea statement (MIM items 1 and 2 combined). However, we also conducted analyses for the first and second items separately to understand whether the treatment effect was more strongly related to one aspect of the main idea or another.

Perceived Effectiveness Scale. Teachers' perceptions of their personal ability and their co-teachers' ability to provide evidence-based literacy instruction and meet the diverse needs of their students were measured using the Perceived Effectiveness Scale at pre- and posttest. The researcher-developed 18-item scale was adapted from Tschannen-Moran and Woolfolk Hoy's (2001) Teacher Sense of Efficacy Scale (TSES). The TSES measures teachers' perceptions of their capability on three factors: instruction (e.g., *To what extent can you craft good questions for your students?*), classroom management (e.g., *How well can you respond to defiant students?*), and student engagement (e.g., *How much can you do to get your students to think critically?*). Items for the Perceived Effectiveness Scale were adapted from the instruction items on the TSES. The Perceived Effectiveness Scale consists of nine items targeting teachers' perceptions of their personal effectiveness (e.g., *I improve vocabulary skills for all of my students*) and nine parallel items targeting their perceptions of their co-teachers' effectiveness (e.g., *My co-teacher improves vocabulary skills of all students in our co-taught class*). Teachers were provided unlimited time to individually rate each item on a scale of 0 (strongly disagree) to 100 (strongly agree), using REDCap, a secure web-based online survey system. Teachers' ratings across all personal effectiveness items were averaged to create a single personal effectiveness score, and their ratings across all co-teacher effectiveness items were averaged to create a single co-teacher effectiveness score. Internal consistency for personal perceived effectiveness was 0.75 at pretest and 0.82 at posttest. Internal consistency for co-teacher perceived effectiveness was 0.93 at pretest and 0.96 at posttest.

Fidelity of CALI Implementation. The fidelity observation tool was developed to reflect the key components of the treatment (see www.projectcali.uconn.edu for a copy of the full fidelity tool). The checklist included 15 total sections that reflected the critical content (e.g., what to teach) and pedagogy (i.e., how to teach) for each component of the CALI IF. If either or both teachers exhibited a practice on the checklist, they received credit for implementing that practice. As each pair participated in more PD sessions, the pairs were expected to implement more of the CALI IF. Thus, the number of possible observed practices increased on the fidelity sheet. Fidelity was calculated as the number of practices observed divided by the total number of possible practices. Helpers collected formal fidelity data four times each year for each CALI pair, recorded during each pair's 'you do' lesson after each PD session. The secondary helper for each pair conducted a fidelity of implementation check in one randomly chosen 'you do' lesson so that at least one fidelity check (i.e., 25% of each co-teaching pair's 'you do' lessons) was conducted by a more objective observer. Therefore, conducting fidelity checks during teachers' 'you do' lessons ensured that we were conducting checks during days the teachers were least supported by their helpers and that we continued to check on implementation fidelity of practices learned in the beginning of the year throughout the remainder of the year when they might be less focused on those practices than when originally learned.

During the two-day helper training at the beginning of the year, inter-rater reliability was established using a gold standard method (Gwet, 2001). Helpers watched videotaped lesson segments of the PIs and teachers from the previous years of the project implementing the CALI practices and independently coded the lessons on the fidelity checklist. Answers were compared with a master code sheet established by the lead project coordinator from the Mid-Atlantic site. Helpers were required to meet an inter-rater agreement of 90%. Agreement was calculated using the same method used to calculate agreement for assessment scoring as described above.

Social Validity. Teachers' perceptions of the PD were measured for all PD sessions using a PD evaluation survey. The survey consists of 8 forced-choice ratings of the value and usefulness of the PD (e.g., *The PD session was valuable*) on a scale of 1 (strongly disagree) to 5 (strongly agree) as well as two optional open-ended questions about what the teachers found helpful and what the research team could improve about each PD. Teachers' ratings across the 8 PD evaluation survey items were averaged to create a single score for each PD session.

We also administered a PD implementation survey to evaluate teachers' perceptions of the strategies covered during each PD session after 3 weeks of implementation (at the beginning of the following PD session). This survey consists of 8 forced-choice ratings related to teachers' implementation of those strategies (e.g., *The strategies I learned*

in this PD were accessible to all of my students) using the same scale of 1 to 5 and two optional open-ended questions requesting teachers' feedback on any aspects of the PD session that helped them during implementation and on what should be emphasized more or less in the PD session. Teachers' PD implementation survey ratings were also averaged to create a single score for each PD session. Internal consistency for both the PD evaluation and the PD implementation surveys ranged from 0.62 (PD 1) to 0.85 (PD 2).

2.6.2. Student outcomes

At pre- and posttest, students completed three measures of reading comprehension.

Main Idea Measure (MIM). Students were tested on the same MIM as the teachers. The test was group-administered. All scoring and reliability procedures were identical to the procedures used for teachers. Reliability for the current sample was high. Cronbach's α for student MIM scores on the 12 items was 0.78 at pretest and 0.82 at posttest. Test-retest reliability was examined using the correlation between MIM pretest and posttest scores, $r = 0.65$ ($N = 212$, $p < .001$). This indicates stability in student rank order but enough variability that a treatment effect could be observed.

We examined validity for this novel measure with correlations between MIM scores and Test of Silent Reading Efficiency and Comprehension (TOSREC) and Maze scores. The correlations were high (all $ps <$ 0.001): For pretest MIM overall, the correlations ranged from 0.61 with posttest TOSREC to 0.64 with pretest Maze. For posttest MIM overall, the correlations ranged from 0.55 with pretest TOSREC to 0.62 with posttest Maze. All correlations are shown in Table 4. With these reliability coefficients, correlations, and levels of significance, MIM was used as a reading comprehension measure with strong reliability.

In addition to examining overall MIM performance, we also examined performance of the two items separately. The first item required identifying the primary subject, a task often easily accomplished with a cursory examination of the passage. We expected less variability on this item because we expected that restricted range might reduce the possible effect. The second item required identification of the most important information in the passage, a task that was an indicator of a student's overall comprehension. We hypothesized that the intensive focus on explicit comprehension instruction in CALI would result in considerable change in students' ability to use cognitive resources when reading new passages and identifying the most important information—as Item 2 demanded.

Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al, 2010). A group-administered, norm-referenced measure of silent reading fluency and comprehension was administered to all students. Students were provided with test forms that were at their grade level (i.e., Grade 6, 7, or 8). Before the assessment began, helpers read standardized instructions to students, which directed them to read as many sentences as possible within three minutes and indicate if the statements were true or false. Students then answered two sample questions for practice. As the assessment progresses, the statements increase in difficulty and length. The measure has an alternate-form reliability of 0.79 and predictive validity ranging from 0.50 to 0.71 for the Measures of Academic Progress and from 0.45 to 0.80 for the Idaho Student Achievement Test (Wagner et al., 2010).

Acadience Reading Maze. All students completed this timed, group-administered measure of general reading comprehension for their respective grade-level. Each Maze assessment consists of a triad of passages, and in each passage, approximately every seventh word is missing. At the deletion point, students are presented a multiple-choice box with three possible word choices (with only one correct response). Students have three minutes to read each passage silently and choose the correct word in each box to complete the sentence. The Grade 6 Maze has an interrater reliability of 0.99, alternate form reliability of 0.92, and predictive validity of 0.60 (Dewey et al., 2015). Alternate-form reliability for Grades 7–8 ranges from 0.81 to 0.93, while internal consistency ranges from 0.93 to 0.98 (Abbott et al., 2020). Predictive validity ranges from 0.69 to 0.82, and concurrent validity ranges from 0.73 to 0.79 (Abbott et al., 2020).

Table 4
Bivariate Correlations Among Student Pretest and Posttest Dependent Variables.

		1	2	3	4	5	6
1	Pretest MIM (overall)	–					
2	Pretest TOSREC	**0.62**	–				
3	Pretest Maze	**0.64**	0.84	–			
4	Posttest MIM (overall)	*0.65*	0.55	0.56	–		
5	Posttest TOSREC	**0.61**	*0.86*	0.84	**0.60**	–	
6	Posttest Maze	**0.63**	0.80	*0.90*	**0.62**	0.86	–

Note. MIM = *Main Idea Measure*; TOSREC = *Test of Silent Reading Efficiency and Comprehension.* Correlations between standardized measures and researcher-designed MIM are bolded. Pretest-posttest correlations for the same measure are italicized.

2.7. Data analysis

For teacher-level data (from the MIM and Perceived Effectiveness Scale) from the 22 participating teachers, regressions of posttest scores on the pretest and intervention status were used to evaluate the effect of participation in CALI. Analyses were conducted in R (version 3.5.2; R Core Team, 2018) for teacher outcomes.

For the student-level data main-effects analysis (Research Question 3), the analysis sample included 212 students with complete pretest and posttest data nested within 11 classrooms, 7 of which were CALI and 4 comparison. For all treatment effect analyses on the outcomes (MIM, Maze, TOSREC), pretest performance for the given measure was included as a covariate. To correct standard errors for the nested data structure (i.e., students in classrooms), the PROC SURVEYREG function in SAS (version 9.4; SAS Institute Inc., 2015) provided linearized standard error estimates based on the cluster (classroom) dependency. This standard-error correction is particularly useful in small-sample studies like the current one. Given that this project included a relatively small sample of teachers and students, statistical significance was evaluated at an alpha level of 0.10.

3. Results

We next present results on PD fidelity and dosage of CALI IF implementation, teacher outcomes and social validity, and student outcomes.

3.1. PD fidelity and dosage of CALI IF implementation

Average fidelity for all PD sessions was 98.9% (range = 97.8–100%). Average fidelity scores for each PD session were 97.8% (PD1), 98.3% (PD2), 99.6% (PD3), and 100% (PD4). The average global rating score was 2.97 (range = 2.86–3.0). The average global rating score for each PD session was 3.0 except for the average score for PD3, which was 2.86. Therefore, each PD was implemented as intended and with quality across CALI pairs.

Pairs implemented between 13 and 17 CALI lessons (range = 3–4 world/word knowledge lessons; 3–4 world/word/gist lessons; 3–4 world/word/associate gist lessons; 2–6 world/word/support lessons).

3.2. Teacher outcomes and social validity

3.2.1. Main idea measure

On average, CALI teachers' posttest scores on the overall main idea statement (MIM items 1 and 2 combined) were significantly greater than the posttest scores of comparison teachers, controlling for pretest ($\widehat{\beta} =$ 1.812, $p < .002$). We also wanted to know whether the treatment effect related better to the first part of the main idea statement (primary

subject) or the second (most important information). Controlling for pretest, scores on the first item were significantly higher for CALI teachers than comparison teachers ($\widehat{\beta} = 0.378$, $p < .070$) and scores on the second item were also significantly higher for CALI teachers than comparison teachers ($\widehat{\beta} = 1.516$, $p < .002$). See Table 5 for descriptive data and Table 6 for teacher MIM estimates.

3.2.2. Perceived effectiveness scale

CALI teachers' perceived personal effectiveness at posttest was significantly higher than comparison teachers' perceived personal effectiveness at posttest, controlling for pretest ($\widehat{\beta} = 6.802$, $p = .078$). CALI teachers' perceived co-teacher effectiveness at posttest was also significantly higher than comparison teachers' perceived co-teacher effectiveness, controlling for pretest ($\widehat{\beta} = 12.774$, $p = .027$). See Table 5 for descriptive data and Table 6 for teacher Perceived Effectiveness Scale estimates.

3.2.3. Fidelity of CALI IF implementation

The average score for all CALI co-teaching pairs across all fidelity checks was 90.62%, and average fidelity ratings across pairs improved after each PD session (PD1 = 87.50%; PD2 = 90.54%; PD3 = 91.03%; PD4 = 93.3%). These average fidelity ratings are between 80% and 100%, indicating high teacher fidelity (Bryant et al., 2000).

3.2.4. Social validity

The average PD evaluation score was 4.92 (range of averages: 4.87–4.96). The average PD implementation evaluation score was 4.69 (range of averages: 4.59–4.78). These high scores suggest that both the PD with coaching model and the CALI IF have evidence of social validity among co-teachers.

3.3. Student outcomes

For Research Question 3, we examined the MIM, TOSREC, and MAZE separately. For the main effects analysis for the MIM, our analysis plan involved two ways of examining the treatment effect. First, we examined the effect of CALI on MIM performance overall (items 1 and 2 combined). There was an effect in the expected direction (Hedges' $g = 0.173$), but it was not significant ($\widehat{\beta} = 0.315$, $p = .333$). Second, we examined performance of each item separately. For item 1, CALI students' scores were not significantly better than those for the students in the comparison group, controlling for pretest, $\widehat{\beta} = -0.50$, $p = .853$. For the second item, there was a positive, significant effect of the treatment, controlling for pretest ($\widehat{\beta} = 0.304$, $p = .036$). The treatment effect was nonsignificant for both TOSREC ($p = .126$) and Maze ($p = .813$), controlling for the respective pre-test scores. See Table 7 for descriptive data for all students and SWD, in particular, and Table 8 for estimates from the treatment effect analyses for all students, in aggregate.

Table 5
Teacher Outcome Descriptive Data.

		Comparison		Treatment		
		Mean (*SD*)	*N*	Mean (*SD*)	*N*	Hedges' g^a
Pretest	MIM	9.13 (1.81)	8	9.50 (1.65)	14	
	MIM – Item 1	5.38 (0.92)	8	5.21 (0.80)	14	
	MIM – Item 2	3.75 (1.28)	8	4.29 (1.33)	14	
	Perceived Effectiveness Scale (Personal)	83.60 (8.52)	8	74.49 (9.66)	14	
	Perceived Effectiveness Scale (Co-Teacher)	82.86 (12.09)	8	68.97 (25.41)	14	
Posttest	MIM	9.38 (1.77)	8	11.43 (1.45)	14	1.072
	MIM – Item 1	5.50 (0.76)	8	5.79 (0.58)	14	0.652
	MIM – Item 2	3.88 (1.36)	8	5.64 (0.93)	14	1.061
	Perceived Effectiveness Scale (Personal)	84.28 (8.10)	8	83.71 (11.45)	14	0.768
	Perceived Effectiveness Scale (Co-Teacher)	79.84 (12.25)	8	81.28 (24.35)	14	0.245

Note. MIM = *Main Idea Measure.* [a]Each Hedges' g effect size was adjusted for small samples and calculated by subtracting the corresponding pre-post effect size of the comparison group from the corresponding pre-post effect size of the treatment group.

Table 6
Regression Results for Teacher CALI Effects.

Measure	Predictor	Estimate	*SE*	*p*
MIM	Intercept	3.494	1.430	0.025
	Pretest	0.645	0.150	< 0.001
	CALI	1.812	0.512	0.002
MIM – Item 1	Intercept	2.425	0.648	0.001
	Pretest	0.572	0.117	< 0.001
	CALI	0.378	0.196	0.070
MIM – Item 2	Intercept	2.109	0.681	0.006
	Pretest	0.471	0.159	0.008
	CALI	1.516	0.421	0.002
Perceived Effectiveness Scale (Personal)	Intercept	16.651	15.097	0.284
	Pretest	0.809	0.178	< 0.001
	CALI	6.802	3.651	0.078
Perceived Effectiveness Scale (Co-Teacher)	Intercept	12.243	10.586	0.262
	Pretest	0.816	0.118	< 0.001
	CALI	12.774	5.335	0.027

Note. MIM = *Main Idea Measure.*

4. Discussion

This article reports findings from a study examining the effectiveness of a PD with coaching model on the literacy outcomes of middle school content-area co-teachers and their students. The purpose of this study was to establish promise that the model could improve co-teachers' knowledge, skills, and beliefs; implementation of high-quality literacy instructional practices (i.e., those in the CALI IF); and students' reading achievement. We believe that the data indicate that the model accomplished this goal overall.

4.1. Effects of CALI on teacher knowledge, skills, and beliefs

Our first research question concerned teachers' ability to identify texts' main ideas and their perceptions of their own and their co-teachers' effectiveness. We observed that teachers in CALI improved in these areas relative to teachers randomly assigned not to participate. CALI teachers produced growth in the quality of their main idea statements, their perceptions of their effectiveness, and their perceptions of their co-teachers' effectiveness.

In this era of inclusive teaching, it is essential that teachers in co-taught content-area classes provide students with and without disabilities with evidence-based literacy instruction that targets their individual needs. This requires co-teachers to co-plan and co-implement enhanced co-teaching models through which they can capitalize on each other's expertise (e.g., station teaching). Thus, the improvement in teachers' perceptions of their co-teachers' effectiveness is very important. For example, when CATs improve their perceptions of their co-teachers' effectiveness (e.g., CATs believe SETs make good instructional decisions), they may be more likely to pause whole-class learning so that

Table 7
Student Outcome Descriptive Data.

		Comparison		Treatment		
		Mean (*SD*)	*N*	Mean (*SD*)	*N*	Hedges' *g*[a]
Pretest	MIM	7.67 (2.78)	78	6.54 (3.21)	134	
	MIM – Item 1	4.05 (1.59)	78	3.62 (1.70)	134	
	MIM – Item 2	3.62 (1.40)	78	2.92 (1.73)	134	
	TOSREC	28.01 (10.18)	78	25.60 (9.86)	134	
	Maze	62.74 (26.82)	78	61.09 (26.23)	134	
Posttest	MIM	8.82 (2.85)	78	8.40 (3.19)	134	0.173
	MIM – Item 1	4.66 (1.59)	78	4.38 (1.70)	134	0.056
	MIM –Item 2	4.15 (1.55)	78	4.02 (1.71)	134	0.277[b]
	TOSREC	32.67 (11.59)	78	28.51 (12.52)	134	−0.169
	Maze	77.68 (30.36)	78	74.99 (33.20)	134	−0.057
		Comparison SWD		Treatment SWD		
		Mean (SD)	*N*	Mean (SD)	*N*	Hedges' g[a]
Pretest	MIM	5.38 (3.14)	16	4.39 (2.89)	31	
	MIM – Item 1	2.94 (1.77)	16	2.61 (1.48)	31	
	MIM – Item 2	2.44 (1.55)	16	1.77 (1.67)	31	
	TOSREC	18.00 (5.24)	16	17.45 (8.52)	31	
	Maze	38.00 (15.92)	16	38.90 (17.21)	31	
Posttest	MIM	7.00 (3.12)	16	6.26 (3.80)	31	0.042
	MIM – Item 1	3.94 (1.84)	16	3.32 (1.89)	31	0.127
	MIM –Item 2	3.06 (1.57)	16	2.94 (2.10)	31	0.222
	TOSREC	20.56 (5.32)	16	17.39 (9.43)	31	−0.479
	Maze	50.50 (17.91)	16	49.03 (25.76)	31	−0.262

Note. MIM = *Main Idea Measure;* TOSREC = *Test of Silent Reading Efficiency and Comprehension* (standard score units). [a]Each Hedges' g effect size was calculated by subtracting the corresponding pre-post effect size of the comparison group from the corresponding pre-post effect size of the treatment group. [b] Indicates significance.

Table 8
Regression Results for Student CALI Effects.

Measure	Predictor	Estimate	*SE*	*p*
MIM	Intercept	3.846	0.423	<0.001
	Pretest	0.649	0.044	<0.001
	CALI	0.315	0.310	0.333
MIM – Item 1	Intercept	2.451	0.309	<0.001
	Pretest	0.547	0.047	<0.001
	CALI	−0.050	0.263	0.853
MIM – Item 2	Intercept	1.894	0.205	<0.001
	Pretest	0.625	0.061	<0.001
	CALI	0.304	0.126	0.036
TOSREC	Intercept	3.349	0.922	0.005
	Pretest	1.047	0.044	<0.001
	CALI	−1.639	0.982	0.126
Maze	Intercept	9.027	2.942	0.012
	Pretest	1.094	0.043	<0.001
	CALI	−0.885	3.635	0.813

Note. MIM = *Main Idea Measure;* TOSREC = *Test of Silent Reading Efficiency and Comprehension* (standard score units).

both co-teachers can engage in small-group instruction that targets all students' needs. This instruction includes providing students with reading difficulties and disabilities with additional explicit instruction aimed at improving their reading comprehension and providing typically achieving students with opportunities to extend their learning. We believe that when co-teachers implement the CALI IF with fidelity, their perceptions of themselves and their co-teachers improve. In turn, these perceptions of effectiveness further encourage meaningful change in the instructional environment that better aligns with expectations of evidence-based practices for teaching all students. The data support the possibility that the CALI IF and PD with coaching model can improve co-teachers' perceptions of each other's effectiveness and help teachers realize the promised value of co-teaching. Teacher self-reports bear this out. On an optional open-ended question on a social validity survey, one SET commented, "It was the first time I had done social studies as a co-teacher ever…. I really felt like [on] the CALI days, I was doing more. I was doing more with my kids and the other kids. On the other days, I was sitting back more because [the CAT] was teaching a curriculum I was not familiar with."

The fact that CALI teachers improved in their own ability to identify the main idea of a passage is interesting. For both CATs and SETs, this change is likely due to greater awareness of the importance of synthesizing text information succinctly—an essential feature of the Construction-Integration model—in order to identify important information about who or what the text is about. Additionally, this increase suggests that the teachers are better able to provide instruction related to this important foundational skill. Thus, our finding aligns with our Theory of Change. Specifically, it confirms that the CALI PD with coaching model improves teachers' knowledge and skills (e.g., main idea identification) and facilitates their implementation of the CALI IF, which includes providing students explicit instruction on main idea identification.

4.2. Teachers' content-area literacy instruction

Our second question concerned the implementation of the CALI IF, namely whether teachers could do it with fidelity. We observed that teachers could implement the CALI IF with fidelity—even when they were implementing the complete CALI IF with multiple moving parts (i.e., the CALI station-teaching support component). The finding that teachers could implement the CALI IF with high fidelity is important because it suggests that such a model—one where teachers have several practices to implement—can work in the co-taught environment.

One potential concern about implementing the CALI IF is that it requires considerable effort and time from both CATs and SETs (e.g., participating in the PD with coaching model). However, we attribute co-teachers' ability to implement the IF to the ongoing support we provided (i.e., PD with coaching model), co-teachers' increased knowledge and skills, and their improved beliefs that reinforce their fidelity of implementation. Data clearly indicate that co-teaching requires extensive support (Dieker & Murawski, 2003; Murawski & Swanson, 2001), and our model included recommended best practices in this area. We provided four sessions of PD using features of effective instruction, followed by scaffolded supports—both in planning for and implementing the CALI IF—to help teachers as they learned each new strategy. Research suggests that this kind of ongoing PD is important to change teacher behaviors, with changes associated with at least 20 contact hours in an academic year (Desimone, 2009; Garet et al., 2001). The intensity of the PD with coaching model, therefore, aligns with features of effective PD and coaching and is, therefore, no greater than what would be expected from any other effective PD structure.

Our PD implementation survey data indicate that teachers felt that the CALI PD with coaching model improved their practice: They considered it useful as part of their content-area teaching. One teacher noted on the optional open-ended question, "CALI gave us a strategy and system for co-teaching." Another teacher commented that she was

"always afraid to do station teaching…. It CALI just made me think that we can do it." These and other quotations lend additional support to the idea that the extensive ongoing PD with coaching model resulted in improved perceptions about co-teachers' effectiveness as well as instructional improvements. Thus, the results strongly align with the extant data on instructional change related to PD (e.g., Kraft et al., 2018).

4.3. Student outcomes

The CALI Theory of Change specified that teachers' improved knowledge, skills, and beliefs about effectiveness as well as their implementation of the CALI IF would ultimately lead to improved student outcomes. We investigated students' ability to identify the main ideas of passages, a summarization skill that aligns with common expectations at the middle school level. Being able to identify the main idea of a passage includes the identification of the primary subject of the passage and the most important information about that primary subject. Students in CALI classes did not significantly improve their ability to generate an overall main idea statement. However, the CALI students did significantly improve their ability to identify the most important information about the primary subject, a critical aspect of being able to identify the main idea. This finding is noteworthy for several reasons.

First, the fact that CALI students significantly improved their ability to identify the most important information about the primary subject ($g = 0.277$) is encouraging. Arguably, being able to identify the most important information in a passage (MIM item 2) is a more difficult task than identifying the primary subject (MIM item 1). The subject of a passage is often easily identified with a cursory examination of the text. Accordingly, students' pretest and posttest scores were higher for the first item than the second item. Thus, the non-significant finding from our examination of MIM item 1 separately is not surprising. On the other hand, identifying the most important information (MIM item 2) requires integration of key ideas to develop a high-quality mental representation of the text. This integration is supported by the Construction-Integration model, which posits that comprehension requires establishing connections across the text and between the text and one's knowledge. Thus, the instruction CALI students received seems to have significantly improved their ability to make these connections.

Second, in general, it is challenging to find positive effects for experimental studies of instruction for middle school students, particularly in content-area settings (Swanson et al., 2017). In addition, it is very difficult to find positive effects for experimental studies of instruction where PD is the independent variable. In the three Institute of Education Sciences (IES)-funded PD studies reviewed by the National Center for Education Statistics, the mean effects of the PD programs were effectively zero (Garet et al., 2016). This included a zero effect even for a treatment-aligned measure (albeit in a study of seventh-grade mathematics). As a result, the fact that we improved student outcomes on one aspect of a practically important skill (i.e., MIM item 2) through our PD with coaching model is especially encouraging. As noted above and aligned with our Theory of Change, teachers' improvement in their own ability to identify the main idea likely led to their ability to teach main idea identification, resulting in these gains for students.

4.4. Limitations and future directions for research

A number of important limitations are noted, which also have implications for future research. First, the PD with coaching model included three implement sessions after each PD session, necessitating almost an entire school year for teachers to learn to implement the complete CALI IF independently and with fidelity. Therefore, students did not participate in many CALI support lessons. While it is encouraging that CALI can change practice even without extensive implementation of the support lessons, additional time with CALI might have resulted in even greater growth in student outcomes for students with and without disabilities. Relatedly, while we know that intensive PD is associated with teacher change, it would be useful to know which elements are most important. For example, it is possible that a smaller number of lessons (with or without helper support) would result in similar effects. Yet, it is difficult to tease apart the active ingredients within the CALI components (i.e., the CALI IF and PD with coaching model). Still, the cost of implementing effective innovations has been a topic of increasing interest to the field (Hollands et al., 2016), so understanding the value of specific elements—especially the more costly ones (e.g., the helper support)—would be valuable. Analyses that concern the relative value of different components of providing intensive PD with coaching—specifically relative to their cost—should be a focus of future research. It would be difficult and very expensive to conduct studies of both CALI components, but it would be possible to reconfigure costly components in ways that make them less expensive and test the efficacy of the CALI components with the less expensive approach. For example, the on-site coaching might be replaced, in part, by distance-based coaching using technologies that permit coaches to observe (e.g., asynchronous videos or live web-streaming) and provide feedback via video conference (perhaps similar to the model of Amendum et al., 2011).

Second, it is important to acknowledge that we had non-significant effects on both distal measures. However, researchers have recently argued that reading comprehension studies should not be evaluated as unimportant based on the absence of generalized effects (Fuchs et al., 2018). Reading comprehension is a very complex construct that depends heavily on background knowledge, and yet, the standardized measures we used in the current study do not provide background knowledge. Furthermore, the two standardized measures we used focus on sentence-level comprehension, while the focus of CALI did not. Therefore, it is not surprising that these measures showed no effect. In a future study, researchers should use standardized measures that more closely align with CALI (i.e., ones that directly measure passage comprehension and, ideally, ones that provide background knowledge). Regardless, because the extant IES-funded PD study that included a treatment-aligned measure did not show effects on the outcome (Garet et al., 2016), we think that the effect on the MIM supports the claim that CALI appears to benefit teachers and students alike, at least as related to one high-value reading strategy.

Third, our teacher sample size was small and only included co-teachers in ELA and social studies classes. However, this limits generalizability of the findings and, thus, more research is needed with larger samples of co-teaching pairs that span additional content areas. Still, we consider it noteworthy that with only 22 teachers, our intervention resulted in statistically significant teacher treatment effects on both the MIM and Perceived Effectiveness Scales. Fourth, the size of our SWD sample was small. More research is needed with larger SWD samples so that we can examine the effects of CALI on this population, given that SWD often participate in instruction in co-taught classrooms.

Finally, we acknowledge that it would have been ideal to have the secondary helper (or another outside observer) observe each co-teaching pair for each of their fidelity of implementation checks. While we were able to accommodate this for at least 25% of each co-teaching pair's 'you do' lessons, due to limited resources, we were not able to do this for each fidelity check.

5. Conclusion

Overall, we believe the data support the idea that CALI may improve the strength of co-taught literacy instruction. Additionally, the data suggest that the use of the CALI IF and PD with coaching model can lead to increases in student achievement. The CALI components primarily draw on the PD and coaching and content-area literacy instruction literatures, and it is clear that this constellation of practices had the sought-after effect. The data are promising given the effects on both teacher and student outcomes, considering the difficulty of finding

significant effects on student achievement at the secondary level and in other PD studies.

6. Author note

This project was supported in part by Grant R324A150181 from the U.S. Department of Education, Institute of Education Sciences. Opinions expressed in the article are those of the authors and do not necessarily represent the position of the U.S. Department of Education, and no official endorsement by it should be inferred.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbott, M., Good, R. H., III, Gray, J. S., Warnock, A. N., & Powell-Smith, K. A. (2020). *Acadience® Reading 7–8 Assessment Manual*. Eugene, OR: Acadience Learning Inc.

Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. (2011). The effectiveness of a technologically facilitated classroom-based early reading intervention: The targeted reading intervention. *The Elementary School Journal, 112*, 107–131. https://doi.org/10.1086/660684

Bryant, D. P., Vaughn, S., Linan-Thompson, S., Ugel, N., Hamff, A., & Hougen, M. (2000). Reading outcomes for students with and without reading disabilities in general education middle school content area classes. *Learning Disability Quarterly, 23*, 238–252. https://doi.org/10.2307/1511347

Burns, M. K., Hodgson, J., Parker, D. C., & Fremont, K. (2011). Comparison of the effectiveness and efficiency of text previewing and preteaching keywords as small-group reading comprehension strategies with middle-school students. *Literacy Research and Instruction, 50*, 241–252. https://doi.org/10.1080/19388071.2010.519097

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 31–42. https://doi.org/10.1037/0022-0663.96.1.31

Cawley, J., Hayden, S., Cade, E., & Baker-Kroczynski, S. (2002). Including students with disabilities into the general education science classroom. *Exceptional Children, 68*, 423–435. https://doi.org/10.1177/001440290206800401

Conderman, G., & Hedin, L. R. (2014). Co-teaching with strategy instruction. *Intervention in School and Clinic, 49*, 156–163. https://doi.org/10.1177/1053451213496158

Cook, L., & Friend, M. (1995). Co-teaching: Guidelines for effective practice. *Focus on Exceptional Children, 28*, 1–12.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325. https://doi.org/10.1037/0022-0663.99.2.311

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181–199. https://doi.org/10.3102/0013189X08331140

Dewey, E. N., Powell-Smith, K. A., Good, R. H., & Kaminski, R. A. (2015). *Acadience ™ reading technical adequacy brief*. Eugene, OR: Dynamic Measurement Group Inc.

Dieker, L. A., & Murawski, W. W. (2003). Co-teaching at the secondary level: Unique issues, current trends, and suggestions for success. *The High School Journal, 86*, 1–13.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*, 1–44. https://doi.org/10.1080/19345740802539200

Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Zhang Tracy, W., … Peng, P. (2018). Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near-transfer. *Learning Disabilities Research & Practice, 33*, 11–23. https://doi.org/10.1111/ldrp.12162

Gajria, M., Jitendra, A. K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities, 40*, 210–225. https://doi.org/10.1177/00222194070400030301

Garet, M. S., Heppen, J. B., Walters, K., Smith, T. M., & Yang, R. (2016). *Does content-focused teacher professional development work? Findings from three Institute of Education Sciences Studies* [Brief]. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from the National Center for Educational Evaluation website https://ies.ed.gov/ncee/whatsnew/index.asp?Date=11%2F1%2F2016.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*, 915–945. https://doi.org/10.3102/00028312038004915

Gwet, K. (2001). *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters*. Gaithersburg, MD: STATAXIS Publishing Company.

Hagaman, J. L., Casey, K. J., & Reid, R. (2016). Paraphrasing strategy instruction for struggling readers. *Preventing School Failure: Alternative Education for Children and Youth, 60*, 43–52. https://doi.org/10.1080/1045988X.2014.966802

Harmon, J. M., Hedrick, W. B., & Wood, K. D. (2005). Research on vocabulary instruction in the content areas: Implications for struggling readers. *Reading & Writing Quarterly, 21*, 261–280. https://doi.org/10.1080/10573560590949377

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness, 9*, 30–53. https://doi.org/10.1080/19345747.2015.1055639

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices. A practice guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED502398.pdf.

Kearns, D. M., Lyon, C. P., & Pollack, M. S. (2021). Teaching world and word knowledge to access content-area texts in co-taught classrooms. *Intervention in School and Clinic, 56*(4), 208–216. https://doi.org/10.1177/1053451220944371

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163–182.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363–394.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*, 547–588. https://doi.org/10.3102/0034654318759268

Lemons, C., Vaughn, S., Wexler, J., Kearns, D., & Sinclair, A. (2018). Envisioning an improved continuum of special education services for students with learning disabilities: Considering intervention intensity. *Learning Disabilities Research & Practice, 33*(3), 131–143. https://doi.org/10.1111/ldrp.12173

Lyon, C., Hogan, E. K., & Kearns, D. M. (2021). Individualizing literacy instruction in co-taught classrooms through a station teaching model. *Intervention in School and Clinic, 56*(4), 224–232. https://doi.org/10.1177/1053451220944376

MacSuga-Gage, A. S., & Simonsen, B. (2015). Examining the effects of teacher-directed opportunities to respond on student outcomes: A systematic review of the literature. *Education & Treatment of Children, 38*, 211–239.

Martin, A. (2016). *Using Load Reduction Instruction (LRI) to boost motivation and engagement*. Leicester, UK: British Psychological Society.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

McMaster, K. L., Kunkel, A., Shin, J., Jung, P., & Lembke, L. (2018). Early writing intervention: A best-evidence synthesis. *Journal of Learning Disabilities, 51*, 1–18. https://doi.org/10.1177/0022219417708169

McMaster, K. L., Lembke, E. S., Shin, J., Poch, A. L., Smith, R. A., Jung, P. G., Allen, A. A., & Wagner, K. (2020). Supporting teachers' use of data-based instruction to improve students' early writing skills. *Journal of Educational Psychology, 112*(1), 1–21. https://doi.org/10.1037/edu0000358

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2

Murawski, W. W., & Dieker, L. A. (2004). Tips and strategies for co-teaching at the secondary level. *TEACHING Exceptional Children, 36*, 52–58. https://doi.org/10.1177/004005990403600507

Murawski, W. W., & Lochner, W. W. (2011). Observing co-teaching: What to ask for, look for, and listen for. *Intervention in School and Clinic, 46*, 174–183. https://doi.org/10.1177/1053451210378165

Murawski, W. W., & Swanson, H. L. (2001). A meta-analysis of co-teaching research: Where are the data? *Remedial and Special Education, 22*, 258–267. https://doi.org/10.1177/074193250102200501

Pollack, M. S., Shelton, A., Clancy, E., & Lemons, C. J. (2021). Sentence-level gist: Literacy instruction for students with learning disabilities in co-taught classrooms. *Intervention in School and Clinic, 56*(4), 233–240. https://doi.org/10.1177/1053451220944378

Puzio, K., Colby, G. T., & Algeo-Nichols, D. (2020). Differentiated literacy instruction: Boondoggle or best practice? *Review of Educational Research, 90*(4), 459–498. https://doi.org/10.3102/0034654320933536

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rice, D., & Zigmond, N. (2000). Co-teaching in secondary schools: Teacher reports of developments in Australian and American classrooms. *Learning Disabilities Research & Practice, 15*, 190–197. https://doi.org/10.1207/SLDRP1504_3

Rosenshine, B. (2012). Principles of education: Research-based strategies that all teachers should know. *American Educator, 36*, 12–39. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/Educational_Practices/EdPractices_21.pdf.

SAS Institute Inc. (2015). *SAS 9.4 in-database products: User's guide* (6th ed.). Cary, NC: SAS Institute Inc.

Shelton, A., Lemons, C. J., & Wexler, J. (2021). Supporting main idea identification and text summarization in middle school co-taught classes. *Intervention in School and Clinic, 56*(4), 217–233. https://doi.org/10.1177/1053451220944380

Sinclair, A. C., Bray, L. E., Wei, Y., Clancy, E. E., Wexler, J., Kearns, D. M., & Lemons, C. J. (2018). Co-teaching in content area classrooms: Lessons and guiding questions for administrators. *NASSP Bulletin, 102*(4), 303–322. https://doi.org/10.1177/0192636518812701

Solis, M., Miciak, J., & Vaughn, S. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disability Quarterly, 37*, 218–229. https://doi.org/10.1177/0731948714528806

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978–2016. *Remedial and Special Education, 40*, 131–149. https://doi.org/10.1177/0741932517749940

Stevens, E. A., Vaughn, S., House, L., & Stillman-Spisak, S. (2020). The effects of a paraphrasing and text structure intervention on the main idea generation and reading comprehension of students with reading disabilities in grades 4 and 5. *Scientific Studies of Reading, 24*(5), 365–379. https://doi.org/10.1080/10888438.2019.1684925

Swanson, E., Wanzek, J., Vaughn, S., Fall, A. M., Roberts, G., Hall, C., & Miller, V. L. (2017). Middle school reading comprehension and content learning intervention for below-average readers. *Reading & Writing Quarterly, 33*, 37–53. https://doi.org/10.1080/10573569.2015.1072068

Tomlinson, C. A., & Imbeau, M. (2010). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: ASCD.

Tschannen-Moran, M., & Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

U.S. Department of Education, Institute of Educational Sciences. (2018). *Request for applications: Special education research grants, 2018* (CFDA 84.324A). https://ies.ed.gov/funding/pdf/2019_84324A.pdf.

U.S. Department of Education, Institute of Educational Sciences, National Center for Education Statistics. (2019). *Digest of education statistics: 2019* (NCES 2018-070), Chapter 2. https://nces.ed.gov/programs/digest/d18/tables/dt18_204.60.asp.

Vaughn, S., Klingner, J. K., & Bryant, D. P. (2001). Collaborative strategic reading as a means to enhance peer-mediated instruction for reading comprehension and content-area learning. *Remedial and Special Education, 22*, 66–75. https://doi.org/10.1177/074193250102200201

Vaughn, S., & Wanzek, J. (2014). Intensive intervention in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice, 29*, 46–53. https://doi.org/10.1111/ldrp.12031

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of silent reading efficiency and comprehension*. Austin, TX: PRO-ED.

Weiss, M. P., & Lloyd, J. W. (2002). Congruence between roles and actions of secondary special educators in co-taught and special education settings. *Journal of Special Education, 36*(2), 58–68. https://doi.org/10.1177/00224669020360020101

Wexler, J., Kearns, D. M., Hogan, E., Clancy, E., & Shelton, A. (2021). Preparing to implement evidence-based literacy practices in the co-taught classroom. *Intervention in School and Clinic, 56*(4), 200–207. https://doi.org/10.1177/1053451220944369

Wexler, J., Reed, D. K., Barton, E. E., Mitchell, M., & Clancy, E. (2018). The effects of a peer-mediated reading intervention on juvenile offenders' main idea statements about informational text. *Behavioral Disorders, 42*(3), 290–301. https://doi.org/10.1177/0198742917703359

Wexler, J., Reed, D. K., Pyle, N., Mitchell, M., & Barton, E. E. (2015). A synthesis of peer-mediated academic interventions for secondary struggling learners. *Journal of Learning Disabilities, 48*(5), 451–470. https://doi.org/10.1177/0022219413504997

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.