The Effects of Cooperative, Collaborative, and Peer-Tutoring Strategies on English Learners’ Reading and Speaking Proficiencies in an English-Medium Context: A Research Synthesis

Shifang Tang	extsuperscript{1}, Beverly J. Irby	extsuperscript{1,2}, Fuhui Tong	extsuperscript{1,3}, and Rafael Lara-Alecio	extsuperscript{1}

Abstract
We conducted a research synthesis to examine the impact of cooperative, collaborative, and peer-tutoring strategies on elementary English learners’ (EL) reading and oral proficiencies in studies from the United States. Seven studies were included in the analysis and fully examined regarding the characteristics of sample, intervention, design, and outcome. Effect sizes were reported by the outcome. We found that cooperative/collaborative/peer-tutoring (CCP) strategies boost elementary ELs’ reading comprehension, reading fluency, and phonemic awareness. Ongoing professional development (PD) and coaching help teachers to improve the quality of strategy implementation. The findings indicated that the quality of implementation, the actual amount of time of these strategies are used in the classroom, and the instructional impact of CCP strategies utilization to improve ELs’ English reading and speaking should be the focus of future research.

Keywords
English learners, cooperative/collaborative/peer-tutoring strategies, reading, speaking, elementary schools

Introduction
The term “English learner” (EL) refers to an individual whose first language is not English and who is in the process of actively learning English (Bardack, 2010). According to the Condition of Education (McFarland et al., 2019), the total public school enrollment of ELs in the United States increased from 8.1% (or 3.8 million) in 2000 to 9.6% (or 4.9 million) in 2016, with a greater percentage of ELs in lower-grade levels. However, EL students greatly lagged behind their non-EL peers in reading and subject content, which has been consistently observed in the past decade (McFarland et al., 2019). The National Research Council (1998) stated that “reading is essential to success in our society” (p. 17). Learning to read is a challenging process, especially for children in early-grade levels (Rubin, 2016). This challenge is even more daunting for ELs whose first language is not English (Ross & Begeny, 2011). Due to the difficulty in acquiring reading skills, ELs have been placed at a disadvantage in learning content subjects (August & Shanahan, 2006; Tong et al., 2017), and they have a higher grade-retention rate (U.S. Department of Education, 2016).

On the other hand, oral language proficiency is a strong predictor of students’ reading proficiency (Miller et al., 2006; Zhang et al., 2013). According to Peregoy and Boyle (2005), comprehensible input and social interaction are the two most essential elements of oral language development. All students, including ELs and non-ELs, need to be surrounded by spoken English to build their English oral language skills (Echevarria et al., 2000). The increasing percentage of ELs, existing achievement gaps between ELs and non-ELs, and ELs’ specific learning

\textsuperscript{1}Center for Research and Development in Dual Language and Literacy Acquisition, Department of Educational Psychology, College of Education and Human Development, Texas A&M University, College Station, TX, USA

\textsuperscript{2}Education Leadership Research Center, Department of Educational Administration and Human Research Development, College of Education and Human Development, Texas A&M University, College Station, TX, USA

Corresponding Author:
Shifang Tang, Center for Research and Development in Dual Language and Literacy Acquisition, Department of Educational Psychology, College of Education and Human Development, Texas A&M University, College Station, TX, USA
Email: shifangtang03@exchange.tamu.edu
needs make finding effective instructional strategies more imperative (Bowman-Perrott et al., 2016). However, literacy instruction for ELs is too often teacher-led, whole-class instruction, which limits ELs’ engagement in English oral language practice and academic activities (Zhang et al., 2013). Compared with one-on-one instruction, the use of cooperative, collaborative, and peer-tutoring strategies (CCP strategies) has been recommended as a cost-effective practice to support interaction among students with various needs (August et al., 2014; Fuchs et al., 1997; Greenwood et al., 1989). Teacher use of CCP strategies allows individualization of instruction at different levels of content among pairs or small groups of students and provides students with opportunities to engage in classroom activities simultaneously (Bowman-Perrott et al., 2016; Short et al., 2012).

When CCP strategies are applied in EL classrooms, students are more likely to receive one-on-one feedback, teacher correction, as well as gain social support from their peers (Fuchs et al., 1997; Greenwood et al., 1989). Therefore, students have more opportunities to engage in academic activities without adding more instructional time (Mathes et al., 2003). When carefully utilized, these strategies have repeatedly lead to improvement in both ELs’ and non-ELs’ academic achievement (Bowman-Perrott et al., 2016; Short et al., 2012; Zhang et al., 2013).

Although there has been an increasing number of empirical studies on the effectiveness of instructional practice over the last decade, few studies have been published in peer-reviewed journals in which researchers have investigated the impact of teachers’ application of CCP strategies in EL classrooms in the United States, especially at the elementary school level (Cole, 2014). Therefore, the purpose of this research synthesis (Cooper, 2016; Parson et al., 2018) was to extend the CCP strategies literature by providing the first synthesis of the impact of CCP strategies on elementary ELs’ English reading and oral language proficiency in the United States.

**Theoretical Framework**

CCP strategies, as research-based English as a second language (ESL) strategies, have been used to support student learning of academic content (King et al., 1998). These are strategies for quality instruction (Topping et al., 2017), which creates optimal environments in which students’ differences in knowledge are not seen as a problem but as an opportunity for them to learn (Stainback & Stainback, 1992). The classroom application of CCP strategies to improve ELs’ English language proficiency and academic knowledge is supported by a number of theories, including socio-cognitive theory (Piaget, 1932), sociocultural theory (Vygotsky, 1978), the four-dimensional transitional bilingual pedagogical theory (Lara-Alecio & Parker, 1994), and second language acquisition theory (Cummins, 1980; Krashen, 1985).

Sociocultural Theories and Socio-cognitive Theory of Language Development, CCP Strategies, and ELs

Vygotsky defined the zone of unassisted performance, which is the stage of content mastered in past development, and the zone of assisted performance, which refers to the process of mastering content in current development (Cole, 2014). When applying Vygotsky’s sociocultural theories to second language learning, researchers specified that language scaffolding could also be provided to EL peers when they work together in a cooperative/collaborative manner, which might exceed what is possible when students work individually (Lantolf, 2000). Piaget (1932) emphasized the importance of cooperation among peers. He pointed out that the cognitive conflict between what the child has in his/her mind, and the new information that he/she comes across with his/her peers may prompt him/her to remove his/her misconceptions and switch to a more accurate conception, which is the process of cognitive development.

In teacher-dominated classrooms, ELs were not provided with enough opportunities to practice their language (Ramirez et al., 1991). Based on the works of Piaget (1932) and Vygotsky, (1978), Klingner and Vaughn (2004) summarized that for ELs, peer discourse could stimulate cognitive growth by creating conflicts during peer interaction in mutual problem-solving. CCP strategies provide student-centered instruction or dialogical instruction to ELs to improve their literacy outcomes (Barton, 2009; Cole, 2014; Lee & Smagorinsky, 2000; Topping et al., 2017). Furthermore, compared with assistance from adults, students prefer to receive support from their peers (Vaughn et al., 1995).

Second Language Learning Theories, CCP Strategies, and ELs

The CCP strategies are also linked to Cummins (1980) theory of language proficiency, in which he proposed that language proficiency encompasses two levels: Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP). The CCP strategies were recommended to provide cognitively demanding instruction to ELs for both the BICS (Baca & de Valenzuela, 1998) and CALP level (Sáenz et al., 2005). Lara-Alecio and Parker (1994) proposed that BICS and CALP can be further broken down. CALP can be defined as two layers, the dense cognitive content and the light cognitive content, and BICS as social exchanges and academic routines, which further clarifies the levels of language content used in EL classrooms. Lara-Alecio and Parker (1994) argued, “the dichotomy between CALP and BICS has obscured the large amount of classroom communication which exists on a continuum between BICS and CALP” (p. 122). How teachers allocate instructional time (e.g., spending the majority of instructional time focused on the light and dense cognitive content that
develops CALP) in the EL classroom is an essential indicator of the quality of instruction students receive (Irby et al., 2007; Lara-Alecio et al., 2009). Therefore, it is meaningful to examine how teachers apply instructional practices/strategies in which language content (e.g., using CCP strategies in CALP [dense or light cognitive content] vs. in BICS [social exchange or academic routine]).

According to Krashen’s (1985) second language acquisition theory, teachers can make language and content-area input more comprehensible for ELs via instructional activities and strategies (Krashen, 1985). As compared with whole-class undifferentiated instruction, one-on-one instruction or small group instruction has been recommended as one of the most effective instructional practices for students with a diverse linguistic and cultural backgrounds (Elbaum et al., 2000). Mathes et al. (2003) also argue that applying CCP strategies is better than one-on-one or small-group instruction because it can accommodate every student in the classroom instead of leaving some students without support during one-on-one or small group instruction.

Taken together, researchers from sociocultural, sociocognitive, applied linguistics, and psycholinguistic fields suggested that CCP strategies can scaffold and improve ELs’ language learning and cognitive development more than traditional teacher-centered or individualized instruction. Therefore, ELs who have been struggling with English language learning and academic achievement could use the support of these strategies to help them meet the demands of school.

Description of CCP Strategies

The CCP strategies are broadly defined as “the instructional use of small groups in which students work together to maximize their own and each other’s learning” (Johnson & Johnson, 1999, p. 73). Cole (2014) summarized the ESL cooperative learning strategy as three varieties: (a) cooperative strategy, (b) collaborative strategy, and (c) peer-tutoring strategy. Three related varieties are noted in this study as the cooperative/collaborative/peer-tutoring (CCP) strategies in an effort to encompass the nature of collaborative learning strategies for ELs. The cooperative strategy was defined as a strategy that emphasizes students’ role in carefully structured group instruction, which was referred to by Slavin (1996) as “one of the greatest success stories in the history of education research” (p. 43). The collaborative strategy is very similar to the cooperative strategy (Cohen, 1994) with a lighter emphasis on students’ role in instruction (Cole, 2014; Topping et al., 2017). The peer-tutoring strategy varies widely, but in general, it is highly academically structured and involves pairing an older or more capable student with a younger or less academically successful student (Cole, 2014; Topping et al., 2017). Although each of the three strategy varieties is distinct in the supporting literature, most researchers treat them as similar terms (Bowman-Perrott et al., 2016; Cohen, 1994; Slavin, 1996; Swain et al., 2002).

Previous WWC Reports, Systematic Reviews, and Meta-Analyses on CCP Strategies With ELs

We included three What Works Clearinghouse (WWC, 2007a, 2007b, 2010) reports, two systematic reviews (i.e., Bowman-Perrott et al., 2016; Pyle et al., 2017), and two meta-analyses (Cole, 2013, 2014). Three WWC reports evaluated eight intervention studies, from which the evaluators concluded that CCP strategies had potentially positive effects on ELs’ reading achievement.

Cole (2013, 2014) examined ELs’ achievement outcomes at all school levels. In Cole’s (2013) meta-analysis, he examined the effect of peer-assisted strategies, including peer-tutoring, collaborative, and cooperative strategies, on ELs’ oral and writing proficiency. He included 32 studies from 12 different countries at the elementary and secondary school levels and found that overall peer-assisted learning positively mediates ELs’ oral and writing proficiencies. However, only six were conducted at the elementary level in the United States and published in peer-reviewed journals. Moreover, it is unclear which studies were included in the meta-analysis since this information was not specified in the study. Furthermore, in the included studies, the peer-assisted strategy was not always the intervention-focused strategy in the classroom, which indicated that the effect of the peer-assisted strategy might be confounded with other instructional strategies applied in the same studies.

Cole (2014) examined 28 studies to test the effect of peer-mediated strategies, including cooperative, collaborative, and peer-tutoring strategies on ELs’ literacy outcomes. The overall finding was that the peer-mediated strategy was more effective for ELs than individualized or teacher-centered instruction, particularly at the elementary level. Moreover, compared with collaborative and peer-tutoring strategies, the cooperative strategy had the highest effect size. He suggested future studies investigate the intervention process and pedagogical factors that support teachers’ application of these strategies. It is worth noting that among nine included studies that were conducted at the elementary level in the United States, only two were published in peer-reviewed journals; those two studies were coded as low quality in this meta-analysis.

Bowman-Perrott et al. (2016) examined the impact of peer-tutoring on ELs’ academic, social, and linguistic outcomes. They examined 17 studies from pre-K to grade 12 published in the last 40 years, with 12 being conducted at the elementary level. It was found that cross-age and same-age peer-tutoring had a positive impact on ELs’ English language proficiency. The authors suggested school level should be examined in future systematic reviews as ELs have a different level of need regarding the content vocabulary and concepts, especially between elementary and secondary levels. Further, the duration was also recommended to be investigated in future studies since the relation between duration and impact of peer-mediated strategies on ELs’ outcomes remains unclear. Bowman-Perrott et al. (2016) focused more on the impact of CCP strategies on ELs’ academic achievement than
language proficiency. The authors did not report country and grade level of the included studies.

Pyle et al. (2017) examined the impact of peer-mediated interventions (PMIs) on ELs from Kindergarten to grade 12; their review included 14 peer-reviewed journal publications from 1983 to 2013. The authors analyzed selected studies regarding intervention characteristics, the effect of PMIs on ELs’ academic outcomes, methodological quality, and the effectiveness of pairing and cooperative/collaborative strategies under the umbrella of PMIs. Pyle et al. (2017) used fidelity of implementation (FOI) as an important quality indicator and concluded only eight studies had high methodological quality. They found that PMIs could support ELs’ development in phonemic awareness, vocabulary, and comprehension. The authors called for future studies to investigate the impact of PMIs on improving ELs’ language proficiency, including speaking, listening, reading, and writing. A limitation of the Pyle et al. (2017) study was that their research synthesis did not follow any standard protocol or procedure, creating concerning gaps.

To summarize, previous reviews did not address whether and how professional development was provided in the intervention studies for participating teachers to improve the quality of implementation (i.e., Cole, 2013, 2014; Pyle et al., 2017). FOI was also neglected in the meta-analyses (i.e., Cole, 2013, 2014). In all the previous reviews, the researchers focused more on ELs’ achievement instead of language skills (i.e., speaking, reading, listening, and writing).

**Purpose of the Research Synthesis**

The purpose of this research synthesis (Cooper, 2016; Parson et al., 2018) was to extend the CCP strategies literature by providing the first synthesis of U.S. studies on the impact of CCP strategies on elementary ELs’ reading and oral proficiencies. This work addresses a gap in the CCP strategies literature by providing needed information about the effectiveness of these strategies for elementary-age ELs’ English reading and oral proficiencies.

**Method**

We followed the research synthesis procedure proposed by Cooper (2016) and also referred to Parson et al.’s (2018) application of Cooper’s procedure. The stages included: (a) formulating the problem, (b) searching the literature, (c) gathering information from studies, (d) evaluating the quality of studies, (e) analyzing and integrating the outcomes of studies, and (f) interpreting the evidence. In the following sections, we describe how we engaged in each step of the procedure.

**Stage 1: Formulating the Problem**

In the problem formulation stage, we reviewed and synthesized past systematic reviews, WWC reports, and meta-analyses on CCP strategies with ELs, as demonstrated in our literature review. We proposed the following two research questions:

1. How did empirical CCP studies on developing English reading and oral proficiencies with ELs in kindergarten through sixth-grade settings vary in their study characteristics (i.e., sample characteristics, intervention features, design characteristics, and outcome characteristics)?
2. What was the impact of CCP strategies on ELs’ reading and oral proficiencies?

**Stage 2: Searching the Literature**

In this stage, four databases, including Educational Resources Information Center (ERIC), PsycINFO, Education Full Text, and Academic Search Ultimate, were employed. The timeframe for the inclusion of literature ranged from January 1, 2000 to January 31, 2019; this timeframe was selected due to wanting to include all literature after the 21st century began until the time when the current study was conducted. Search terms and variable combinations included *cooperative strategy, collaborative strategy, peer-tutoring, elementary school, English language learners, ESL, bilingual, reading, and oral proficiency*. Furthermore, the Scopus database was used for an additional subject search (Sampson et al., 2006). With this step, we were able to include the relevant studies that cited the studies identified in the previous four databases. The guidelines laid out in Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were adopted for reporting the search protocol (See Figure 1).

**Stage 3: Gathering Information From the Studies**

To get the information from the included studies, we indexed 71 articles and collected their logistic information (e.g., year of publication, title), abstract, research questions, grade level, and outcomes. All studies were coded by two of the authors. We excluded 59 studies after the title/abstract screening and another five studies after the full-text screening by applying the inclusion/exclusion criteria (see Figure 2 for criteria; see Table 1 for the screening results). The literature reviews and meta-analyses related to the topic were summarized in the previous synthesis section, but were not included in the coding procedure.

**Results**

In this section, we included stage 4: evaluating the quality of studies and stage 5: analyzing and integrating the outcomes of studies. The coding and analyzing results are presented in the Appendix.

**Stage 4: Evaluating the Quality of Studies**

The studies in this review were published from 2000 to 2019 (Mean = 2007.7, SD = 4.86). As stated in Research Question 1, we analyzed the included studies based on the following study features: sample characteristics, intervention features, design
Sample characteristics. A total of 961 ELs participated across seven studies, with a range of 75 to 351 participants per study ($\text{Mean}=137.3$, $\text{SD}=96.4$). The terms used to describe the participants across the studies were English language learners and English learners. These referred to students whose first language was not English and were identified as ELs based on a home language survey (McMaster et al., 2008; Zhang et al., 2013), district level information (Liu & Wang, 2015; Sáenz et al., 2005), and/or if they did or did not meet a certain level of English language proficiency (Almaguer, 2005; Calhoon et al., 2007; Sáenz et al., 2005). In one study, the researchers (e.g., Greenwood et al., 2001) did not disclose how ELs were defined. The participants in the included studies received instruction in a variety of program models, including ESL (Greenwood et al., 2001), transitional bilingual (Almaguer, 2005; Sáenz et al., 2005), two-way bilingual (Calhoon et al., 2007; McMaster et al., 2008), both bilingual and mainstream programs (Zhang et al., 2013), and in two studies, the researchers did not specify the program model (Liu & Wang, 2015; McMaster et al., 2008).

Grade levels. The distribution of age groups for the included studies was prescreened and limited to the elementary level. In six studies, the instruction was delivered in a single grade level: Kindergarten (McMaster et al., 2008),
Table 1. Screening Results After Inclusion/Exclusion Criteria Applied.

| Reason for inclusion/exclusion                      | Number |
|-----------------------------------------------------|--------|
| Included in the final analysis                      | 7      |
| Review or introduction                              | 6      |
| Not elementary level                                | 2      |
| Not a journal paper                                 | 15     |
| Not ELs                                             | 6      |
| Not CCP strategies                                   | 27     |
| Not about reading or speaking                       | 2      |
| Not in the U.S.                                     | 5      |
| Not in the year range (2000–2019)                   | 1      |

In six studies (e.g., Calhoon et al., 2007; Almaguer, 2005; Säenz et al., 2005; Zhang et al., 2013) conducted in higher grades (grades 3–5) focused on higher-level English proficiency skills, like reading comprehension. Oral reading fluency was measured through lower and upper grades (e.g., Almaguer, 2005; Greenwood et al., 2001; McMaster et al., 2008; Säenz et al., 2005; Zhang et al., 2013).

**CCP instructional time (duration, frequency, and intensity).** The amount of instructional time devoted to CCP strategies varied in each study based on its duration, frequency, and intensity. The duration referred to the number of weeks in which CCP strategies-embedded instruction was delivered. The researchers (e.g., Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Säenz et al., 2005) reported the duration of the implementation ranged from 4 weeks to an academic year. The intervention in four studies (Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Säenz et al., 2005) was implemented from 15 to 20 weeks. The intervention in one study (Zhang et al., 2013) was implemented for only 4 weeks and in one study (e.g., Almaguer, 2005) for 9 weeks. Within each week, the instruction was delivered between 2 and 5 days a week (Mean = 3.5, SD = 1.05). The intervention was implemented two times per week in one study (n = 1; Zhang et al., 2013), three times per week in two studies (e.g., Calhoon et al., 2007; Säenz et al., 2005), four times per week in two studies (e.g., Greenwood et al., 2001; McMaster et al., 2008), and five times per week in one study (e.g., Almaguer, 2005). The intensity of instruction was defined as minutes of CCP strategies-embedded instruction per day. The researchers of six included studies reported the intensity of CCP strategies-embedded instruction. The intensity ranged from 20 to 35 minutes (Mean = 27, SD = 6.41).

**Group size.** In six studies (e.g., Almaguer, 2005; Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Säenz et al., 2005; Zhang et al., 2013), the instruction was organized in a whole group or an entire class. In Liu and Wang’s (2015) study, the authors reported the frequency of three types of grouping: small group, pair work, and independent reading in large-scale data sets.

**Measures of the level of exposure/FOI.** Besides CCP strategies duration, frequency, and intensity, researchers also...
used multiple methods, such as observation protocols and checklists, to document whether teachers were implementing CCP strategies as intended (fidelity measures). In five studies (e.g., Almaguer, 2005; Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005), the researchers checked FOI to document how much CCP strategies were implemented with ELs during the intervention. In one study, the researcher (e.g., Almaguer, 2005) used student attendance records of intervention as FOI, which was above 91% during the period of intervention. In other four studies (e.g., Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005), classroom observation checklists were adopted to record fidelity of implementation, with FOI ranging from 80% to 96%. McMaster et al. (2008) only included the teachers whose FOI was 0.9 and above, and reported the results of the sub-sample of a large-scale randomized control trial.

Professional development. Traditional/face-to-face professional development (PD) components were provided in six intervention studies (e.g., Almaguer, 2005; Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005), the teachers were offered PD training and coaching before and throughout the intervention. To be more specific, in Greenwood et al.’s (2001) study, participating teachers received an individualized one-hour PD session five to seven times, then biweekly PD consultation. In Calhoon et al.’s (2007) study, ongoing feedback and support were provided three times a week via classroom observation.

Design characteristics. To avoid potential bias, we evaluated all the included studies’ design characteristics based on the following: random assignment, attrition rate, baseline equivalence, comparison group, and data analysis.

Random assignment. Five studies involved comparison groups, with three studies (i.e., Almaguer, 2005; McMaster et al., 2008; Zhang et al., 2013) using a quasi-experimental design, in which the researchers administered pre- and post-test, but did not randomly assign multiple groups to different instructional approaches. Two studies (e.g., Calhoon et al., 2007; Sáenz et al., 2005) were identified as cluster randomized control studies, in which teachers were the randomized unit and students were tested before and after the intervention.

Attrition rate. In two studies, the researchers (e.g., Calhoon et al., 2007; Sáenz et al., 2005) provided attrition information. Based on the formula from the WWC’s (2015) Standard Brief for Attrition, the attrition rate was 23.7% in Calhoon et al. (2007) and 10.9% in Sáenz et al. (2005). Since neither study reported attrition rate by condition, we could not examine the differential attrition rate.

Baseline equivalence. Due to the potential lack of baseline equivalence among conditions, researchers conducting quasi-experimental studies should assess whether the equivalent baseline was achieved between treatment and control conditions (WWC, 2015). In two randomized experimental studies (Calhoon et al., 2007; Sáenz et al., 2005), the researchers examined the baseline equivalence and reported there was no statistically significant difference between treatment and control conditions. Zhang et al. (2013) is the only quasi-experiment study that reported and established baseline equivalence. In the other two quasi-experimental studies (Almaguer, 2005; McMaster et al., 2008), they used a statistical technique, Analysis of Covariance (ANCOVA), to

---

Table 2. Categories and Subcategories for Analyzing the Quality of Included Studies.

| Study feature          | Category     | Subcategory                  |
|------------------------|--------------|------------------------------|
| Sample characteristics | Grade level  |                              |
|                        | Size         |                              |
|                        | Language groups |                            |
|                        | SES background |                            |
| Study/intervention features | Focus of study/intervention |                      |
|                        | Instructional time |                      |
|                        | Group size     |                              |
|                        | Fidelity of implementation |                |
|                        | Professional development |               |
| Design characteristics | Random assignment |                       |
|                        | Pre- and post-testing |                     |
|                        | Attrition rate                  | Rate across condition |
|                        | Baseline equivalence             |                            |
|                        | Comparison group type            | Demographic type         |
| Outcome characteristics | Assessment                | Ability                     |

---
include pretest as a covariate, but did not report the baseline equivalence.

**Comparison group type.** In four studies (e.g., Almaguer, 2005; Calhoon et al., 2007; McMaster et al., 2008; Sáenz et al., 2005), the researchers reported and compared demographic data and instructional practice in both treatment and control conditions. Zhang et al. (2013) reported and compared the demographic data of treatment and control conditions, but did not report the instructional practices in the control condition. The comparison group received the same instructional time as the treatment group in all studies except for Almaguer’s (2005) study, in which an extra 30 minutes of instruction was offered to treatment students.

**Data analysis.** To investigate the impact of the intervention or CCP strategies on ELs’ reading and oral proficiencies, seven studies included a range of statistical analyses, including Analysis of Variance (ANOVA; e.g., Calhoon et al., 2007; Sáenz et al., 2005), ANCOVA (e.g., Almaguer, 2005; McMaster et al., 2008; Zhang et al., 2013), correlation analysis (e.g., Liu & Wang, 2015), and chi-square analysis (e.g., Greenwood et al., 2001). Single-level analyses at the student level were adopted in five studies (e.g., Almaguer, 2005; Calhoon et al., 2007; Liu & Wang, 2015; McMaster et al., 2008; Zhang et al., 2013). An aggregated class mean was used in two studies (e.g., Greenwood et al., 2001; Sáenz et al., 2005).

**Outcome characteristics.** In this review, outcome measures that were used to assess ELs’ English reading and oral language proficiencies were examined by assessment type and target language proficiency. Description of the outcome measures included assessment types used (i.e., standardized, high-stakes, curriculum-based, or researcher-created) and language proficiency measured (i.e., phonological awareness, reading comprehension, oral expression, or reading fluency). Reliability and validity of outcome measures were also examined.

**Assessment types and abilities measured.** The researchers adopted curriculum-based measures to assess ELs’ phonological awareness (e.g., Greenwood et al., 2001) and oral reading fluency (e.g., McMaster et al., 2008). Standardized measures were used to evaluate ELs’ letter-word identification and word fluency (e.g., Calhoon et al., 2007; McMaster et al., 2008), reading comprehension (e.g., Liu & Wang, 2015; Zhang et al., 2013), and reading fluency (Calhoon et al., 2007). In four studies, the researchers used a researcher-created assessment to measure students’ reading comprehension (e.g., Almaguer, 2005; Sáenz et al., 2005; Zhang et al., 2013), phonemic awareness (e.g., McMaster et al., 2008), oral expression (Zhang et al., 2013), oral reading fluency (e.g., Almaguer, 2005; McMaster et al., 2008), and reading comprehension (e.g., Almaguer, 2005).

The same measures were adopted for measuring both pretest and posttest scores (e.g., Almaguer, 2005; Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005). The researchers used different measures for pretest and posttest (e.g., McMaster et al., 2008; Zhang et al., 2013). In one study, there was one-time testing administered (Liu & Wang, 2015). Moreover, in three studies (e.g., Calhoon et al., 2007; McMaster et al., 2008; Zhang et al., 2013), the researchers used multiple measures assessing ELs’ different abilities, including reading comprehension, phonemic awareness, oral reading fluency, and oral expression. There were four studies involving only a single measure to assess one aspect of ELs’ language proficiency, such as reading comprehension (e.g., Almaguer, 2005; Liu & Wang, 2015; Sáenz et al., 2005) or reading vocabulary (e.g., Greenwood et al., 2001).

**Reliability and validity of outcome measures.** In two studies (e.g., Calhoon et al., 2007; McMaster et al., 2008), the researchers reported the reliability of their standardized measures, ranging from 0.9 to 0.96. Among these two studies, McMaster et al. (2008) reported concurrent validity as 0.9, and Calhoon et al. (2007) reported a reliability of 0.95. It was reported in two studies (e.g., Almaguer, 2005; Sáenz et al., 2005) that the test-retest reliability of researcher-created measures was 0.93 to 0.95, and concurrent validity was 0.91. Zhang et al. (2013) reported their researcher-created measures with test-retest reliability between 0.47 and 0.87, without validity reported. The researchers of three studies (e.g., Greenwood et al., 2001; Liu & Wang, 2015; McMaster et al., 2008) did not report the reliability and validity of their curriculum-based or researcher-created measures.

**Stage 5: Analyzing and Interpreting the Outcomes of Studies**

In this stage, we focused on analyzing and interpreting the outcomes of the included studies to address research question 2 (See Table 3 for reporting categories). Cohen’s $d$ was reported as the indicator of effect size. The effect sizes were not calculated if: (a) the authors did not provide sufficient or accurate data or (b) there was no significant difference between the treatment and control groups. The effect sizes were calculated according to formulas published by Lipsey and Wilson (2001) with an online calculator developed by Wilson (n.d.). The interpretation of Cohen’s $d$ was based on the recommendations of Cohen (1988): small, $d=.2$; medium, $d=.5$; and large, $d=.8$.

**Oral language.** In one study (Zhang et al., 2013), the researchers investigated the impact of collaborative reasoning (CR) on ELs’ English oral language development. They found that CR significantly improved participants’ (both ELs and
native-English speakers) coherent narratives in the storytelling tasks (Cohen’s \(d=0.37, 95\% \text{ CI } [-0.09, 0.83]\)). The authors reported that ELs produced significantly longer stories than native-English speakers did, but spoke more slowly and with more pauses. There were no available data for calculating the effect sizes for these two subcategories (i.e., length of story and storytelling fluency). There was no interaction effect between the CR intervention and the program type (bilingual vs. mainstream). The authors did not report the comparison between the CR ELs and control ELs regarding their English oral language development. Therefore, it was unclear how much CR learning impacted ELs’ English oral language development.

**Oral reading fluency.** The researchers of three studies (e.g., Almaguer, 2005; Calhoon et al., 2007; McMaster et al., 2008) examined the impact of CCP strategies on improving ELs’ reading fluency. There were mixed results in the treatment-control comparison regarding students’ achievement in oral reading fluency. Almaguer (2005) found that CCP reading activities significantly improved grade 3 ELs’ reading fluency with a large effect (Cohen’s \(d=2.53, 95\% \text{ CI } [1.70, 3.37]\)) based on Cohen’s (1988) criterion. However, Calhoon et al. (2007) and McMaster et al. (2008), working with early grades, found that the application of CCP strategies did not significantly improve ELs’ oral reading fluency as compared with ELs in the control condition.

**Reading comprehension.** In four studies (e.g., Almaguer, 2005; Liu & Wang, 2015; Sáenz et al., 2005; Zhang et al., 2013), the researchers measured the impact of CCP strategies on ELs’ reading comprehension. The authors (e.g., Almaguer, 2005; Sáenz et al., 2005) of two studies adopted reading multiple-choice comprehension questions to measure the impact of CCP strategies on ELs’ reading comprehension outcome. Only in two studies (e.g., Almaguer, 2005; Sáenz et al., 2005) did the researchers provide accurate and accurate data for calculating effect sizes. Based on Cohen’s (1988) criterion, we found CCP strategies had a positive impact on ELs’ reading comprehension (measured by multiple-choice questions) with effect sizes; Cohen’s \(d=1.03, 95\% \text{ CI } [0.37, 1.69]\) in Almaguer (2005) and Cohen’s \(d=2.67, 95\% \text{ CI } [1.11, 4.22]\) in Sáenz et al. (2005).

Furthermore, in two studies (e.g., Almaguer, 2005; Zhang et al., 2013), the researchers used a cloze reading assessment to measure ELs’ reading comprehension. There were mixed results found in the impact of CCP strategies on ELs’ reading comprehension as measured by the cloze test. Almaguer (2005) found that CCP activities improved ELs’ reading comprehension with a small to medium effect size (Cohen’s \(d=0.33, 95\% \text{ CI } [-0.29, 0.96]\)). Zhang et al. (2013) found there was no statistical difference between the treatment ELs who conducted collaboration reasoning activities and the ELs in the control condition regarding their improvement in reading comprehension as measured by cloze assessment.

In Liu and Wang (2015), the negative impact of small-group and pair-work reading instruction on ELs’ reading comprehension was identified. They found that for the fourth-grade ELs, the more small-group (Cohen’s \(d=0.92, 95\% \text{ CI } [0.29, 0.72]\)) and pair-work reading instruction (Cohen’s \(d=0.34, 95\% \text{ CI } [0.13, 0.55]\)) used in the classroom, the lower ELs’ reading achievement was. They concluded that independent learning worked best for grade 4 ELs’ improvement in reading.

**Phonemic awareness, letter-word identification, and word fluency.** Researchers of two studies (e.g., Calhoon et al., 2007; McMaster et al., 2008) examined ELs’ phonemic awareness, letter-word identification, and word fluency. Mix results were identified. Calhoon et al. (2007) found that the application of CCP strategies significantly impacted grade 1 ELs’ nonsense word fluency (Cohen’s \(d=0.75, 95\% \text{ CI } [0.28, 1.22]\)) and letter naming fluency (Cohen’s \(d=0.67, 95\% \text{ CI } [0.20, 1.14]\)) with large effect sizes, while no statistical significance was found in the improvement in ELs’ phoneme segmentation as compared with ELs in the control condition. McMaster et al. (2008) found for Kindergarten ELs, CCP strategies better supported the development of rapid letter naming (Cohen’s \(d=1.08, 95\% \text{ CI } [0.17, 2.00]\), phonemic awareness, including segmentation (Cohen’s \(d=1.00, 95\% \text{ CI } [0.10, 1.92]\)) and blending (Cohen’s \(d=1, 95\% \text{ CI } [0.12, 1.93]\)).

**Professional development and ELs’ reading vocabulary.** In Greenwood et al. (2001), the researchers investigated the effects of consultation on improving ELs’ reading vocabulary in class-wide peer-tutoring classrooms from grades 1 to 5. They found that when teachers were provided with ongoing individualized consultation, more students in their classes achieved significantly better in vocabulary learning (Cohen’s \(d=6.32, 95\% \text{ CI } [0.21, 1.34]\)).
Synthesis and Discussion of the Findings

In this section, following Cooper’s step 6, we critique and interpret the evidence and discuss the findings. Exposure to CCP activities develops ELs’ reading comprehension and oral reading fluency in varied ways. To understand the impact of CCP strategies on ELs’ English reading and speaking proficiencies, researchers need to examine the research design, treatment characteristics, and outcome variables of previous empirical studies.

To answer the research questions, we conducted a multi-layered analysis to identify which relevant studies met the inclusive criterion. Further, an investigation was conducted to examine the included studies’ characteristics of sample, design, outcome, and the level of effectiveness. In this research synthesis, we placed a particular emphasis on elementary school ELs’ English reading and oral development. The findings across randomized control trial, quasi-experimental design, and single-subject design research studies provided the first summary of English reading and oral proficiency improvement for elementary ELs engaged in CCP learning in the past 20 years.

Intervention Features and Design Characteristics of Included Studies

Intervention design, sample, and duration. Five included studies (e.g., Almaguer, 2005; Calhoon et al., 2007; McMaster et al., 2008; Sáenz et al., 2005; Zhang et al., 2013) with a RCT or quasi-experimental design were considered to have high-methodological quality. For the studies using a quasi-experimental design (e.g., Almaguer, 2005; McMaster et al., 2008; Zhang et al., 2013), the investigators either examined the baseline equivalence or included students’ pretest achievement as covariance or both. The researchers of five intervention studies (e.g., Almaguer, 2005; Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005) reported the total intervention time exceeded 20 hours over a year, which was more than the recommended dosage of CCP strategies for EL classrooms as suggested by Rohrbeck et al.’s (2003) meta-analysis. Further, ELs with different language backgrounds benefited from the application of CCP strategies on one or more aspects of English reading and oral language proficiency.

FOI. Measuring FOI was an essential design feature. In the previous reviews, only Pyle et al. (2017) and Bowman-Perrott et al. (2016) examined the fidelity of implementation in their syntheses. The findings in this synthesis were consistent with theirs in that for most studies, FOI was above 90%, except for Greenwood et al. (2001) with FOI at or above 80%. It was also observed in the included intervention studies, that an observation checklist was the most common instrument for examining FOI (e.g., Calhoon et al., 2007; Greenwood et al., 2001; McMaster et al., 2008; Sáenz et al., 2005). Researchers employed a FOI observation checklist to examine both teachers’ and students’ implementation of this strategy. Student attendance records used as a fidelity measure was observed in only one study (Almaguer, 2005).

Measurements. In previous CCP syntheses on EL achievement (e.g., Bowman-Perrott et al., 2016; Cole, 2013, 2014; Pyle et al., 2017), only Pyle et al. (2017) and Cole (2014) examined the types of measures in each study. In this review, it was noted that curriculum-based measures, standardized measures, and published researcher-created measures were adopted to test ELs’ reading comprehension, oral reading fluency, oral language/expression, and letter-word identification. We found that standardized assessments were the most frequently used measure, which accounted for half of the included studies. This finding was consistent with the synthesis conducted by Pyle et al. (2017). Cole (2013) found in his meta-analysis review that the standardized measures were less likely to be adopted in high-quality studies, but the findings in this review did not support his argument. Moreover, large effect sizes were generated from both researcher-created and standardized measures. Therefore, which types of measures are more sensitive to examine the impact of CCP strategies on ELs’ English language proficiency still needs to be investigated.

Professional development. Professional development was provided via face-to-face delivery in all included the intervention studies. No technology-integrated components or online sessions were involved in the PD of included studies. No researchers examined and analyzed PD component in their syntheses (e.g., Bowman-Perrott et al., 2016; Cole, 2013, 2014; Pyle et al., 2017). Out of six included intervention studies, researchers provided PD to the treatment teachers in five studies. However, only in two studies (e.g., Calhoon et al., 2007; Greenwood et al., 2001) were treatment teachers supported with ongoing PD. As pointed out by Lonigan et al. (2011), well-designed and intensive PD programs did not show a significant impact on improving students’ achievement. It has been emphasized that more PD sessions should be required and offered in conjunction with coaching (Hamre et al., 2017; Piasta et al., 2017). The findings in this synthesis confirmed that there was a positive impact of ongoing, individualized coaching on ELs’ reading vocabulary learning (e.g., Greenwood et al., 2001).

Study Outcomes and Effect Sizes

Reading comprehension. We found a consistently significant positive impact of CCP strategies on ELs’ English reading comprehension as measured by multiple-choice reading comprehension questions with large effect sizes (Cohen’s ds > 0.9). This finding was consistent with Pyle et al.’s (2017) conclusion that CCP strategies support ELs’
development in reading comprehension. However, the mixed results were found in the included studies where cloze reading assessment was used to measure ELs' reading comprehension. For example, Calhoon et al. (2007) found that there was a positive impact of CCP strategies with a small effect size and Zhang et al. (2013) found there was no significant impact of CCP strategies on ELs' reading comprehension. This discrepancy may be because more ask/answer activities involved in CCP learning promoted ELs' reading proficiency of understanding and answering questions. Reading comprehension questions, therefore, seem to be a more sensitive or stable measure to examine the impact of CCP strategies on improving ELs' English reading comprehension. We encourage more researchers to explore the effect of CCP strategies with ELs on their reading comprehension proficiency, as measured by reading multiple-choice comprehension questions and cloze reading assessment.

It is worth noting that in studies (e.g., Liu & Wang, 2015; Zhang et al., 2013) where there was no intervention involved or no PD provided to ensure the quality of the application of CCP strategies, no impact or even negative impact was found by merely pairing or grouping ELs to work together (e.g., Liu & Wang, 2015; Zhang et al., 2013). These findings suggested that the intervention duration may be a critical factor for the effectiveness of CCP strategies on ELs' reading comprehension. According to Rohrbeck et al. (2003), students who received less than 19 hours of CCP intervention were expected to demonstrate no improvement compared with the control group. The findings in this review were in line with other studies indicating no impact from collaborative reasoning, as suggested in a study by Zhang et al. (2013) that only had 2.67 hours' intervention. Furthermore, we found that the grade level was an important factor. The CCP strategies worked significantly better with ELs in grades K-3. However, we found no positive or negative impact of CCP strategies on ELs' reading proficiency in grade 4. Analysis based on large-scale public datasets also indicated that grade 4 ELs might benefit more from independent learning (Liu & Wang, 2015). This finding was consistent with previous studies that CCP strategies may work more effectively with younger students (Cole, 2014; Rohrbeck et al., 2003).

Phonemic awareness and letter-word identification. Among the seven included studies, we found CCP strategies significantly and positively impacted EL development in phonemic awareness and letter-word identification across grade levels with medium to large effect sizes. Since we identified only two studies in which researchers examined the impact of CCP strategies on ELs’ phonemic awareness and letter-word identification, it would be beneficial if scholars further explored this area in the future.

Oral reading fluency. The seven included studies contained mixed results concerning CCP and oral reading fluency. Almaguer (2005) found use of CCP strategies significantly improved ELs' oral reading fluency with a large effect size, while Calhoon et al. (2007) and McMaster et al. (2008) determined there was no significant positive impact on ELs' oral reading fluency. There are two possible explanations for this discrepancy. First, in the studies conducted by McMaster et al. (2008) and Calhoon et al. (2007), both treatment and control students received the same amount of instructional time, while in the study conducted by Almaguer (2005), students received extra intervention time in English literacy. Second, both researcher-created and standardized measures were adopted in the included studies. It seems like researcher-created measures are more sensitive to the effects of CCP strategies; possible reason might be researcher-created measures are designed to be more aligned with intervention. Although Cole (2014) claimed that researcher-created measures were more likely to be used in high-quality studies measuring the intervention’s effect, we suggest that future researchers investigate which type of measures can best examine the effects of CCP strategies on ELs' oral reading fluency.

Oral language and oral expression. We identified only one study (e.g., Zhang et al., 2013) with evidence of CCP strategies positively supporting ELs' development of English oral language/expression. In this study, CCP strategies positively influenced students' development of coherent English narratives for both ELs and mainstream students. No difference was detected between the ELs and mainstream students who received the same CCP intervention, which indicated that CCP strategies supported both language groups. Since there was only this one study on the effects of CCP strategies on ELs’ English oral language development, we recommend that more researchers should explore these strategies with ELs.

Professional development. Among the seven studies, we found that teacher PD was delivered face-to-face prior to the interventions. The lack of information reported on ongoing PD requires additional attention. In one study (Greenwood et al., 2001), we found that individualized, ongoing coaching positively impacted teachers’ ability to increase ELs’ reading vocabulary learning. Furthermore, on the basis of the evidence of this research synthesis, it seems fair to suggest that if not intervention of PD was provided to ensure the quality of teachers’ application of CCP strategies, no impact or even negative impact might be found by merely pairing or putting ELs work together.

Implications and future research

In this research synthesis, we provided a base for future researchers who are interested in the use of CCP strategies with ELs. However, there are several limitations. First, our results were drawn from the available information, as reported by the authors of the included studies. Therefore,
the details provided in each study varied, which may have led to some bias in the studies with less information included regarding their intervention design or study characteristics. Secondly, the sample sizes in the included intervention studies were relatively small. A finding of no difference or mixed results did not necessarily mean that no difference existed. In the future, researchers should consider involving a large sample in well-designed experimental or quasi-experimental studies to continue to evaluate the effectiveness of CCP strategies in improving ELs’ English language proficiency.

There are several implications regarding the curriculum and intervention design in future studies. First, CCP strategies can be embedded into the existing curriculum to provide more opportunities for ELs to engage in academic-content learning via reading, speaking, listening, and writing (Bowman-Perrott et al., 2016). Second, applying CCP strategies in the classroom should take place for at least 20 hours over a year to be effective, a recommendation that is also supported by the previous review (Rohrbeck et al., 2003). Third, future researchers can examine the most effective duration of CCP usage. It would be even more beneficial to the field if scholars report the intensity, frequency, and duration of intervention, which was a suggestion from Pyle et al. (2017). Fourth, we agree with the previous review authors (i.e., Bowman-Perrott et al., 2016; Pyle et al., 2017) that future researchers should focus on the quality of the application of CCP strategies, which can be a more refined index than simply the description of intensity, frequency, and duration. It should be the exact minutes/frequency that teachers and students participate in CALP activities when CCP strategies are applied. Finally, given the benefits of involving ongoing teacher PD/coaching and an FOI monitoring process during the intervention to ensure the quality of implementation (Tang et al., 2020), future studies might focus the effect of how ongoing PD/coaching and FOI can impact the effectiveness of teachers’ application of CCP strategies in instruction for elementary ELs.

To conclude, to our best knowledge, this research synthesis is the first to systematically synthesize research studies on the usage of CCP strategies for ELs and examine the quality and effectiveness of the CCP intervention. This research synthesis provides evidence of teacher implementation of CCP strategies in the United States, which has 4.9 million ELs. Findings from this synthesis can shed light on the best practices for EL instruction. In addition, by scrutinizing sample characteristics, study/intervention features, design characteristics, and outcomes, this study provides research and instructional implications and suggestions on how to design a CCP intervention with consideration of dosage of CCP strategies and teacher professional development on using such strategies. Policymakers, practitioners, and researchers should encourage teachers to use CCP strategies to enhance ELs’ learning experience and outcomes.

**Appendix. Research Synthesis Coding Summary.**

| Studies             | Grade level | Size | Language groups                      | SES background          |
|---------------------|-------------|------|--------------------------------------|-------------------------|
| Almaguer (2005)     | 3           | 80   | Homogenous Spanish-speaking ELLs      | High-poverty            |
| Calhoon et al. (2007)| 1           | 94   | Heterogeneous language samples       | High-poverty            |
| Greenwood et al. (2001) | 1–5         | 117  | Heterogeneous language samples       | NA                      |
| Liu and Wang (2015) | 4           | 351  | Heterogeneous language samples       | NA                      |
| McMaster et al. (2008)| k          | 112  | Heterogeneous language samples       | NA                      |
| Sáenz et al. (2005) | 4           | 132  | Homogenous Spanish-speaking ELLs      | NA                      |
| Zhang et al. (2013) | 5           | 75   | Homogenous Spanish-speaking ELLs      | Low to middle SES       |
| Studies                        | Focus of study/intervention                                      | Instructional time | Fidelity of implementation | Professional development |
|-------------------------------|------------------------------------------------------------------|--------------------|----------------------------|--------------------------|
| Almaguer (2005)               | Oral reading fluency, reading comprehension                      | 9 weeks            | 5 times/week               | 22.5 hours               |
|                               |                                                                  | Whole group        | Student attendance records | 91%                      |
| Calhoon et al. (2007)         | Phoneme awareness, letter-word identification, and word fluency  | 20 weeks           | 3 times/week               | 30–33.3 hours            |
|                               |                                                                  | Whole group        | Classroom observation checklist | 94.9% to 98.7%, on average 96.4% |
| Greenwood et al. (2001)       | Professional development supporting ELLs' reading vocabulary     | 15–21 weeks        | 4 times/week               | 20–28 hours              |
|                               |                                                                  | Whole group        | Classroom observation checklist | 80% or above |
| Liu and Wang (2015)           | Reading comprehension                                            | NA                 | NA                         | NA                       |
| McMaster et al. (2008)        | Oral reading fluency, Phoneme awareness, letter-word identification, and word fluency | 18 weeks           | 4 times/week               | 24–36 hours              |
|                               |                                                                  | Whole group        | Classroom observation checklist | 90% or above |
| Sáenz et al. (2005)           | Oral reading fluency, reading comprehension                      | 15 weeks           | 3 times/week               | 26.25 hours              |
|                               |                                                                  | Whole group        | Classroom observation checklist | 90% or above |
| Zhang et al. (2013)           | Reading comprehension; oral expression                           | 4 weeks            | 2 times/week               | 2.67 hours               |
|                               |                                                                  | Whole group        | Not report                 | Not report               |
### Design characteristics

| Studies            | Random assignment | Pre- and post-testing | Attrition rate | Baseline equivalence | Comparison group type | Data analysis     |
|--------------------|-------------------|-----------------------|----------------|----------------------|-----------------------|-------------------|
| Almaguer (2005)    | Quasi-experimental| Pre- and post-test    | Not reported   | Not reported         | Not examined, use ANCOVA to adjust the baseline equivalence | Reported | Reported | ANCOVA |
| Calhoon et al. (2007) | Randomized control | Pre- and post-test | 23.70%         | Not reported         | Examined and achieved | Reported | Reported | ANOVA  |
| Greenwood et al. (2001) | Single-subject design | Pre- and post-test | Not reported | Not reported | NA | NA | Chi-square analysis |
| Liu and Wang (2015) | NA                | Pre- and post-test    | Not reported   | NA                   | Not examined, use ANCOVA | Reported | Reported | ANCOVA |
| McMaster et al. (2008) | Quasi-experimental | Pre- and post-test | Not reported   | Not reported         | Not examined, use ANCOVA | Reported | Reported | ANCOVA |
| Sáenz et al. (2005) | Randomized control | Pre- and post-test    | 10.90%         | Not reported         | Examined and achieved | Reported | Reported | ANOVA  |
| Zhang et al. (2013) | Quasi-experimental | Pre- and post-test    | Not reported   | Not reported         | Examined and achieved | Reported | Not reported | ANCOVA, MANCOVA |
### Outcome characteristics

| Studies           | Type                                      | Ability                                      | Reliability and validity of outcome measures                      |
|-------------------|-------------------------------------------|----------------------------------------------|-------------------------------------------------------------------|
| Almaguer (2005)   | Curriculum-based measures; published measures | Phonological awareness; reading comprehension | Reported for published measures; not reported for curriculum-based measure |
| Calhoon et al. (2007) | Standardized measures                     | Letter-word identification, and word fluency | Reported for standardized test                                     |
| Greenwood et al. (2001) | Curriculum-based measures                | Reading vocabulary                            | Not reported for curriculum-based measure                          |
| Liu and Wang (2015) | Standardized measures                     | Reading comprehension                          | Not reported for standardized test                                |
| McMaster et al. (2008) | Curriculum-based measures; standardized measures | Oral reading fluency; letter-word identification, and word fluency | Reported for standardized test; not reported for curriculum-based measures and published measures |
| Sáenz et al. (2005) | Published measures                        | Reading comprehension                          | Not reported for published measures                                |
| Zhang et al. (2013) | Standardized measures; published measures | Reading comprehension; oral expression        | Not reported for standardized test; reported for published measures; |

### Impact the CCP strategies on ELs' reading and speaking proficiency

| Studies           | Oral language (Cohen's d) | Oral reading fluency (Cohen's d) | Reading comprehension (Cohen's d) | Phoneme awareness, letter-word identification, and word fluency (Cohen's d) | Professional development supporting ELs' reading vocabulary (Cohen's d) |
|-------------------|---------------------------|----------------------------------|----------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------|
| Almaguer (2005)   | Positive (2.53)           | Positive (reading comprehension question, 1.03); Positive (reading comprehension cloze, 0.33) | Positive (word fluency, 0.75; letter-naming fluency, 0.67); No difference (Phoneme segmentation) | Individualized consultation positively impact ELs' reading vocabulary (0.63) |
| Calhoon et al. (2007) | No difference            |                                    | Positive (reading comprehension question, 1.03) | | |
| Greenwood et al. (2001) | Negative for small group instruction (0.5) and pair work instruction (0.34) | | | | |
| Liu and Wang (2015) | No difference            |                                    | Positive (letter naming, 1.08; phonemic awareness (segmentation, 1.00; blending, 1.02) | | |
| McMaster et al. (2008) | No difference            |                                    | Positive (reading comprehension question, 2.67) | | |
| Sáenz et al. (2005) | Positively impact students’ coherent narratives (0.37) | No difference (reading comprehension cloze) | | | |
| Zhang et al. (2013) | Positively impact students’ coherent narratives (0.37) | No difference (reading comprehension cloze) | | | |

### Declaration of Conflicting Interests

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

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The open access publishing fees for this article have been covered by the
The input hypothesis: Issues and implications. Longman.

Lantolf, J. P. (Ed.) (2000). Sociocultural theory and second language learning. Oxford University Press.

Lara-Alecio, R., & Parker, R. I. (1994). A pedagogical model for transitional English bilingual classrooms. Bilingual Research Journal, 18(3–4), 119–133.

Lee, C. D., & Smagorinsky, P. (Eds.) (2000). Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry. Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. SAGE.

Liu, S., & Wang, J. (2015). Reading cooperatively or independently? Study on ELL student reading development. Reading Matrix: An International Online Journal, 15(1), 102–120.

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children’s emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. Reading and Writing, 24(3), 305–337.

Mathes, P. G., Fuchs, D., Fuchs, L. S., & Henley, A. M. (1994). Increasing strategic reading practice with Peabody classwide
peer-tutoring. *Learning Disabilities Research and Practice*, 9(1), 44–48.

Mathes, P. G., Torgesen, J. K., Clancy-Menchetti, J., Santi, K., Nicholas, K., Robinson, C., & Grek, M. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *The Elementary School Journal*, 103(5), 459.

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., & Barmer, A. (2019). *The condition of education 2019 (NCES 2019–144)*, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

McMaster, K. L., Kung, S. H., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A “tier 1” approach to promoting English learners’ response to intervention. *Exceptional Children*, 74(2), 194–214.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice*, 21(1), 30–43.

National Research Council. (1998). *Preventing reading difficulties in young children*. National Academy Press.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Parsons, A. W., Pierzynski, M., & Allen, M. (2018). Teachers’ instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242.

Perego, S., & Boyle, O. (2005). *Reading, writing, and learning in ESL: A resource book for k-12 teachers* (4th ed.). Pearson/Allyn and Bacon.

Piaget, J. (1932). *The moral judgment of children*. Routledge & Kegan-Paul.

Piasta, S. B., Justice, L. M., O’Connell, A. A., Mauck, S. A., Weber-Mayer, M., Schachter, R. E., Farley, K. S., & Spear, C. F. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Educational Effectiveness*, 10(2), 354–378.

Pyle, D., Pyle, N., Lignugaris/Kraft, B., Duran, L., & Akers, J. (2017). Academic effects of peer-mediated interventions with English language learners: A research synthesis. *Review of Educational Research*, 87(1), 103–133.

Ramirez, J. D., Yuen, S. D., Ramey, D. R., Pasta, D. J., & Billings, D. K. (1991). Final report: Longitudinal study of structured immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children. Aguirre International.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Education & Psychology*, 95, 240–257.

Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners’ reading fluency: The effects of small-group and one-on-one intervention. *Psychology in the Schools*, 48(6), 604–618.

Rubin, D. I. (2016). Growth in oral reading fluency of Spanish ELL students with learning disabilities. *Intervention in School and Clinic*, 52(1), 34–38.

Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children*, 71(3), 231–247.

Sampson, M., McGowan, J., Cogo, E., & Horsley, T. (2006). Managing database overlap in systematic reviews using Batch Citation Matcher: Case studies using Scopus. *Journal of the Medical Library Association JMLA*, 94(4), 461–463.

Short, D. J., Fidelman, C. G., & Louguit, M. (2012). Developing academic language in English language learners through sheltered instruction. *TESOL Quarterly*, 46(2), 334–361.

Slavin, R. E. (1996). Cooperative learning in middle and secondary schools. *Clearing House*, 69(4), 200–204.

Stainback, S., & Stainback, W. (1992). *Curriculum considerations in inclusive classrooms*. Paul H. Brookes.

Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 22, 171–185.

Tang, S., Tong, F., Irby, B. J., Lara-Alecio, R., & Guerrero, C. (2020). Fidelity of implementation in a randomized controlled trial study: The effect of virtual professional development on bilingual teachers. *Bilingual Research Journal*, 43(1), 111–124.

Tong, F., Luo, W., Irby, B. J., Lara-Alecio, R., & Rivera, H. (2017). Investigating the impact of professional development on teachers’ instructional time and English learners’ language development: a multilevel cross-classified approach. *International Journal of Bilingual Education and Bilingualism*, 20(3), 292–313.

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). Effective peer learning: From principles to practical implementation. Routledge.

U.S. Department of Education. (2016). *Educational experiences of English learners: Grade retention, high school graduation, and GED attainment*, 2011–12. https://www2.ed.gov/rschstat/eval/title-iii/crdc-retention-completion.pdf

Vaughn, S., Schum, J. S., Klingner, J., & Saumell, L. (1995). Students’ views of instructional practices: Implications for inclusion. *Learning Disability Quarterly*, 18(3), 236–248.

Vygotsky, L. (1978). Interaction between learning and development. *Annual Review of Educational Research*, 23(1), 34–41.

What Works Clearinghouse. (2007a). *Bilingual cooperative integrated reading and composition*. http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED4995727&site=ehost-live

What Works Clearinghouse. (2007b). *Peer-tutoring and response groups*. http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED4992966&site=ehost-live

What Works Clearinghouse. (2010). *Peer-assisted learning strategies*. http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED512003&site=ehost-live

What Works Clearinghouse. (2015a). *WWC standard brief for attrition*. What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Docs/referenceResources/wwc_brief_attrition_080715.pdf

Wilson, D. B. (n.d.). *Practical meta-analysis effect size calculator [Online calculator]*. https://www.campbellcollaboration.org/this-is-a-web-based-effect-size-calculator/explore/this-is-a-web-based-effect-size-calculator

Zhang, J., Anderson, R. C., & Nguyen-Jahiel, K. (2013). Language-rich discussions for English language learners. *International Journal of Educational Research*, 58, 44–60.