Research Paper

ROE

ECNU Review of Education
2021, Vol. 4(1) 84–107

DOI: 10.1177/2096531120972709
journals.sagepub.com/home/roe

SAGE

ECNUP

A Quasi-Experimental Study of a Web-Based English Literacy Tool for Grade 3 Students in China

Hui Gu（顾慧） **and Jijun Yao**（姚继军）
Nanjing Normal University

Longjun Zhou（周龙军）
Jiangsu Second Normal University

Alan C. K. Cheung
The Chinese University of Hong Kong

Philip C. Abrami
Concordia University

Abstract

Purpose: This study explores the effectiveness of a A Balanced Reading Approach for Children Always Designed to Achieve Best Results for All (ABRACADABRA; hereinafter, ABRA)—a web-based literacy program designed by Concordia University in Canada—on third-grade students in Nanjing, China.

Design/Approach/Methods: Participants comprised 999 students from three treatment schools (N = 711) and three control schools (N = 288). Three different approaches were used in the treatment schools: namely, a computer laboratory once a week, noontime study after lunch, and single-game instruction (SG) during every English lesson. Interviews were also conducted with teachers, producing qualitative data.

Corresponding author:
Jijun Yao, Nanjing Normal University, Nanjing 210024, China.
Email: yaojijun_njnu@163.com

Findings: Following 20 weeks of intervention, the overall effect size was +0.05. The SG group reflected the smallest effect size (d = -0.52). The noontime study group produced an effect size of 0.39, and the laboratory group an effect size of 0.55. This study conducted interviews with teachers to gain a qualitative understanding of the differential impacts. In doing so, this study found that teachers in the SG group were poorly motivated due to a lack of school support and heavy workload, resulting in passive roles and low ABRA program intensity.

Originality/Value: The results of this study indicate that ABRA is an effective means of improving Chinese students' English literacy skills. Results also underscore the need for critical measures to encourage teachers to actively participate in the program.

Keywords

A web-based literacy tool, ABRACADABRA, China, foreign language learning, teaching strategies

Date received: 18 February 2020; accepted: 20 October 2020

Introduction

In foreign language learning, the learning medium is essential for reducing the negative transfer effect from the mother language and bridging the two language systems (Su, 2000; Yang & Hu, 2004). Chinese textbooks employ the International Phonetic Alphabet (IPA) as the medium of foreign language learning at the middle school level. However, this medium is unsuitable for primary school students—that is, beginner level English language learners—because the different pronunciations and forms of writing may confuse students, negatively impacting their ability to memorize and retain language learning. Moreover, the IPA is constantly updated, requiring the frequent modification of textbooks and teaching strategies to adapt to these changes. As such, there is no appropriate and stable medium of language learning for primary students in China (Zhu, 2018).

Phonics is a popular learning medium in English-speaking countries. Blevins (2017) encouraged the use of phonics in English language learning as it reveals the corresponding relationship between phonetics and forms and makes it easier for students to spell words. Existing studies show that phonics instruction benefits students in vocabulary learning (Aro & Wimmer, 2003; Avalos et al., 2007; Ehri, 1995; M. C. Wood, 1996). For instance, demonstrating the stability of the medium for the learning of vocabulary, particularly in primary school-level reading materials, Adams (1994) showed that children can spell around 500 words using the 37 most common sounds. At the beginning of the 21st century, Taiwan region and Hong Kong Special Administrative Region (SAR) introduced phonics in English language learning (Zhu, 2018), revealing that phonics had a positive transfer effect on spelling, writing, and reading comprehension because it facilitated students' ability to decode words, allowing them to memorize specific word forms (H. Cheung, 1999; Chu & Chen, 2014).

Phonics instruction has also become popular in Chinese mainland in recent years. However, the majority of studies have focused on brief instruction in phonics as well as some theoretical strategies for phonics-based teaching (Y. Guo, 2014; X. Wu, 2016). Only a few empirical studies have found that the use of phonics improves students' spelling, phonological awareness, and interest in English (Hou, 2013; Liu, 2019). For instance, comparing the use of phonics and IPA as a learning medium, Zhou (2017) showed that phonics had a greater impact on improving students' spelling, reading, and interest in learning. However, with no more than 200 samples, extant studies on phonics in Chinese mainland are relatively limited and may have a positive bias. Chinese teachers use traditional teaching methods to teach students how to distinguish between letter sounds, as well as other phonics-related skills. Studies indicate that computer-based supplementary phonics programs also improve students' reading skills, especially low-performing students (Macaruso et al., 2006). Different teaching methods have produced similarly positive results, raising the question of whether computer-based phonetics-based teaching enhances student learning. Information communication technology (ICT) has played a key role in transforming teaching strategies, particularly insofar as technology's integration of visual and audio stimuli serve as learning aids in teaching children to read (Mak et al., 2017). Studies suggest that game-based mobile learning can enhance students' attention and aid vocabulary retention in English classes (T. Wu & Huang, 2017). In view of this, this study examines the effectiveness of phonics-based teaching supported by ICT for English language learning in China. In addition to complementing our understanding of the value of ICT in language learning, the results of this study may help in identifying the best use of phonics for language learning.

A Balanced Reading Approach for Children Always Designed to Achieve Best Results for All (ABRACADABRA, hereinafter ABRA) is an excellent and comprehensive tool for English learning as it focuses on developing students' understanding of phonics while enhancing their phonological awareness, fluency, comprehension, and writing abilities. The tool, according to a report by the National Reading Panel, promotes the importance of alphabetic fluency, comprehension, and writing and argues that effective reading interventions must be comprehensive and well-balanced (Mak et al., 2017). ABRA is a web-based tool providing comprehensive resources to enhance students' concentration and motivation in an enriched learning environment. Research evidencing the positive effects of computer-aided learning tools on students' achievements indicates that ABRA is the most effective learning software available (E. Wood et al., 2014). Accordingly, this study examines the impact of ABRA, a web-based literacy skills teaching aid, on Chinese students new to both phonics and an ICT teaching environment.

Regardless of the effectiveness of ABRA and phonics-based teaching, it is important to recognize the role of teachers' pedagogical strategies and attitudes on student learning. For instance, a meta-analysis examining 75 samples found that the relationships between second language learning achievement and motivation are uniformly higher than those between achievement and other

factors, including integrativeness and attitudes toward the learning situation (Masgoret & Gardner, 2003). These findings underscore the need to improve learner motivation to achieve effective second or foreign language teaching. Bernaus and Gardner (2008) showed that there is a significant relationship among student motivation, language achievement, and teachers' didactic strategies and that certain strategies—such as applying new technology, group work, and satisfying student needs—are more effective than others in enhancing motivation (Madrid, 2002). However, research indicates that teachers lose their motivation the longer that they teach, influencing their teaching strategies and student achievements (X. Wu et al., 2003). Therefore, this study considers the implementation strategies employed by different teachers in its examination of the impact of ABRA on phonics-based teaching among Chinese students.

Literature review

ABRA is a freely available, evidence-based literacy program developed by the Center for the Study of Learning and Performance (CSLP) at Concordia University. Activity content is derived directly from systematic reviews of what works in reading and spelling (Abrami et al., 2016). The content is suitable for English beginners, with half of the activities dedicated to teaching the alphabet and how to apply this knowledge to read words.

Pedagogically, the software is based on the comprehensive method, which seeks to improve reading through the development of decoding skills, vocabulary, and comprehension in a literature-rich context. More specifically, the tool comprises three modules: a student module comprising 17 games and other activities related to fluency, comprehension, and writing; a teacher module consisting of teaching materials and instructions for student assessment; and a parent module providing support, advice, and at-home reading resources. The activities in the student module are integral to improving students' basic literacy skills. For instance, among the activities offered, 21 stories in a gaming context promote fluency and comprehension skills by having students practice replicated approaches while maintaining their interest.

Any teacher, student, and parent can participate in games and organize different learning experiences. Significantly, ABRA supports both teachers and students in learning and professional development. In this respect, teachers are provided with teaching materials, instruction on how to apply ICT in the classroom, and tools for student assessment. Students receive frequent feedback and apply their skills under the guidance of games, thereby improving learning independence.

The efficacy of ABRA

Several studies have examined the effectiveness of ABRA in improving the literacy skills of students from different backgrounds, confirming that the program significantly improved student literacy. For instance, Savage et al. (2008) conducted a classroom-level randomized control trial in

three Canadian provinces, demonstrating that the tool significantly impacted children's sight-reading ability and phonological awareness. In 2013, the researchers conducted a scaled-up randomized control trial intervention involving 1,069 K-2 children in Canada, finding that a 20-hr ABRA intervention resulted in significant improvement in students' phonological blending, letter-sound knowledge, and phoneme segmentation fluency compared to those taught under regular instruction. Moreover, these effects were still evident during posttests (Savage et al., 2013).

Similar beneficial effects were observed among students with poor attention spans and lower socioeconomic status. In another Canadian study, Deault et al. (2009) compared the ABRA approach with regular teaching using a sample of 144 participants. Results show that inattention accounted for the greater variance in control groups but cannot explain students' achievements after ABRA support (Deault et al., 2009). The researchers thus concluded that ABRA appears to influence the association between literacy and attention and may support students with lower reading ability and attention issues. Examining the efficacy of ABRA instruction among disadvantaged urban kindergarten students, Comaskey et al. (2009) found that the program improved students' word blending and articulation of final consonants.

Positive results have been similarly observed in other countries, including English as a second or foreign language learning in Australia, China, and Kenya (Abrami et al., 2016; A. Cheung et al., 2016; Mak et al., 2017; Wolgemuth et al., 2011, 2013). For instance, Wolgemuth et al. (2013) compared the program's impact on indigenous and nonindigenous students, comparing indigenous students and nonindigenous students in control group and treatment group separately. Results show that students in the treatment group performed significantly better than their peers in the control group in terms of phonological awareness ($d = 0.37$) and phoneme-grapheme knowledge ($d = 0.37$). Additionally, indigenous students gained significantly more per hour of instruction than their nonindigenous peers in terms of phonological awareness and early literacy skills, although this may be a result of the ceiling effect (Wolgemuth et al., 2013). These findings indicate that students lacking basic English skills may benefit significantly from the intervention.

In the Kenyan context, Abrami et al. (2016) demonstrated that a 13-week intervention significantly improved the reading comprehension of students in the treatment group; students also achieved higher results in their final subject exams for English, Mathematics, Science, and Social Studies. Additionally, the effect size of phoneme-grapheme correspondence (0.22), phoneme segmentation (0.46), and nonsense word fluency (0.22) indicates that the program is beneficial for students in non-English language environments (Abrami et al., 2016).

Abrami et al. (2015) conducted a meta-analysis of the extant literature, summarizing the overall effect of ABRA. In doing so, the scholars classified the outcomes in accordance with the definitions provided by the National Reading Panel. According to the results, the overall average independent effect size was $g = .17$ ($p < .01$), and distributions were homogeneous in most of the categories,

Table 1. Effect size and I^2 of reading abilities according to Abrami et al.'s (2015) meta-analysis of extant studies.

Reading skill	Number of outcomes	Average effect size	I^2
Phonemic awareness	20	.32	0.00
Phonics	19	.19	0.01
Fluency	6	.08	73.21***
Vocabulary knowledge	15	.14	50.38**
Reading comprehension	6	.07	0.00
Listening comprehension	6	.38	23.17
Overall (random-effects model)	73	.17	50.38

***$p < .01$; **$p < .05$; *$p < .1$.

indicating the consistency of the application and the universal applicability of the ABRA software (Abrami et al., 2015). Table 1 presents the average effect size and I^2 of reading skills observed in the extant literature. I^2 is an important index for judging heterogeneity. When $I^2 \geq 50\%$ or $p < .1$ and $I^2 \geq 25\%$, it can be inferred that there is statistical heterogeneity (Higgins & Green, 2009).

The role of teachers in ABRA's efficacy

Teachers play an integral role in the efficacy of ABRA as their teaching methods and strategies, as well as their personal characteristics, influence student outcomes (Marzano et al., 2001). Comparing the effects of an ABRA intervention between the two groups categorized by a phoneme-based synthetic or rime-based analytic phonics approach, Deault et al. (2009) found that the program impacts the association between attention and reading outcomes. Comaskey et al. (2009) observed qualitatively different effects on students' phonological development, with ABRA students under synthetic instruction developed their word blending and articulation of final consonants abilities, while ABRA students in the analytic group developed their ability to articulate shared rimes.

Conducting teacher interviews after an ABRA experiment, Mak et al. (2017) found that teachers were consistently positive about ABRA, praising the program as easy, convenient, and helpful. In this respect, Mak et al. emphasized the importance of teachers' attitudes, with the positive results of phoneme-grapheme correspondence, phoneme segmentation, and nonsense word fluency reflecting the impact of teachers' positive attitudes. Meanwhile, Savage et al. (2010, 2013) ranked teachers' ABRA teaching level into three levels: entry, adoption, and adaption. Teachers at the entry level make time-consuming mistakes and seldom alter instruction. They cannot execute a lesson plan or lead a session effectively. At the adoption level, teachers can blend technology into teaching practices, but they cannot link various teaching forms like collaborative learning with technology. At the adaption level, teachers develop ABRA

lesson plans and strategies to ensure that students are engaged in ABRA activities. The studies concluded that adaption-level teachers were most beneficial to students in terms of improving their literacy skills (Savage et al., 2010, 2013).

Study rationale and research questions

As such, extant research indicates that ABRA has a positive impact on the learning outcomes of both native and nonnative English speakers. In the Chinese context, two studies in Hong Kong SAR (Abrami et al., 2016; Mak et al., 2017) and one study in Hunan (X. Guo, 2018) evidenced the effectiveness of the program for Chinese students. However, how ABRA functions in Chinese schools requires further investigation for two reasons. First, differences in class size and teaching styles result in different learning environments, and it is unclear how this impacts the efficacy of ABRA as a teaching tool. For example, unlike the participants in studies conducted in Canada and Hong Kong SAR, teachers and students in Chinese mainland are more accustomed to teacher-led teaching (Mak et al., 2017; Savage et al., 2013). Moreover, class sizes in China typically exceed 30 or even 40 students, resulting in less attention and one-on-one instruction by teachers. Accordingly, teachers' attitudes and behavior play a significant role in the teaching process.

Second, there is little data on phonics teaching using ICT in Chinese mainland. Indeed, Guo's unpublished experiment in Hunan comprises 217 treatment students and 122 control students and uses hierarchical linear modeling to identify the effect of the program on students' reading abilities. Further investigation is required in more economically advanced provinces, like Jiangsu, to prove the effectiveness of this medium for language learning. Extant research has also focused on students' reading improvements, overlooking the impact of teachers' teaching methods and attitudes on student outcomes. Given the diversity of conditions in Chinese mainland, more research is required to identify the effect of such interventions on students in China.

Accordingly, this study examines the effect of ABRA on the learning outcomes of primary school students in Nanjing, China, and examines teachers' attitudes and behavior in using this new tool. In doing so, this study seeks to answer the following research questions:

(1) Do ABRA students outperform control students in terms of English literacy skills?
(2) Do teachers' attitudes and teaching behavior influence the results of the ABRA intervention?

Methodology

Research design and participants

This study employs a pre-post control group design to examine the effect of ABRA on the learning outcomes of Chinese students at the primary school level. Conducted in six Nanjing primary

Table 2. English starting grade of participating schools.

	Treatment schools	Start year of English lesson	Number of students in the pretest	Control schools	Start year of English lesson	Number of students in the pretest
SG	School A	Grade 1	339	School D	Grade 3	79
NS	School B	Grade 1	213	School E	Grade 3	88
CL	School C	Grade 3	162	School F	Grade 3	64

Note. SG = single-game instruction; NS = noontime study after lunch; CL = computer laboratory.

schools from March 2018 to November 2018, this constituted the first quasi-experimental study of the large-scale use of ABRA in Chinese mainland.

Three pairs of schools located in the capital of Jiangsu province, Nanjing, were selected for the experiment: three for the treatment group, namely, School A, School B, and School C, and three for the control group, namely, School D, School E, and School F. Researchers first contacted the school leaders of all participating schools, and at least two teachers from each school were involved in the experiment. As most Chinese primary schools follow the advice of China's Ministry of Education (2011) and provide English language classes from the third grade, this study selected participant teachers and students from the third grade. Neither the teachers nor the students had received comprehensive phonics training prior to the experiment.

In regard to the treatment group, this study selected 11 Grade 3 English classes from three schools. In total, 12 experimental teachers and 945 students were selected to participate in this study's experiment, with 9 teachers (4 in School A, 3 in School B, and 2 in School C) and 711 students in the treatment group and 3 teachers and 231 students in the control group. More specifically, 339 students from School A, 213 students from School B, and 162 students from School C participated in this study. Control groups were paired according to the participating students, school size, teacher–student ratio, and school area. Table 2 presents the details of the participant schools, including the grade in which students started receiving English language training.

ABRA intervention was provided under daily school conditions and routines. Treatment teachers added ABRA intervention to their English language classes, while control teachers provided traditional instruction as usual. All participating schools continued their English language instruction according to the national curriculum. Each school provided a mode of ABRA intervention according to their facilities and the amount of time available in the daily teaching schedule: School A provided single-game instruction (SG), School B provided noontime study after lunch (NS), and School C provided computer laboratory (CL). An online chat room was available to all participants in the treatment group. English teachers of SG groups and participating teachers were responsible for exchanging ideas.

In terms of socioeconomic environment, the majority of public schools in Nanjing are attended by students living in proximity to the school. Accordingly, better schools tend to be situated in areas with higher property values. As one of the best primary schools in Nanjing, School A enrolls students from a better socioeconomic status, the area's housing price averaging at about CNY 50,000 (approximately US$7,258) per square meter. The housing prices around the other two schools average at approximately CNY 40,000 (US$5,806) per square meter.

ABRA training

ABRA training was provided to teachers in the treatment group. In August 2017, nine teachers with the potential to participate in the experiment attended a 2-hr training workshop, where they were introduced to ABRA and made aware of its function and value. The workshop aimed to identify participant schools interested in the program, and three pairs of schools were selected based on an on-the-spot investigation. In January 2018, teachers and administrative personnel from the six schools attended a workshop detailing the general curriculum, as well as the phonics and ICT used in ABRA classes. Teachers in control groups were not allowed to use ABRA in their classes.

Following a pretest in March 2018, treatment teachers were considered ready to conduct ABRA lessons. Training was conducted in three ways: online training, face-to-face workshops, and in-school training. More specifically, between August 2017 and November 2018, teachers in the treatment group participated in both online and off-line training. They received at least four intensive off-line training sessions and one or two in-school sessions. Online training primarily involved teachers sharing experiences and discussing problems and did not have a fixed time. Teachers were also encouraged to share their ideas using an online chat room. In training, teachers tried to simulate possible classroom scenarios to get hand-on experience of game activities. Participants also raised questions, including concerns regarding the integration of ABRA and existing curriculum, how to better organize teaching sections in single-game or laboratory settings, and how to monitor student achievement. Each teacher received at least 15 hr of training.

A workshop comprising members of this study's research team, members of CSLP, the teachers who had participated in previous ABRA experiments conducted in Hong Kong SAR and Hunan, and the Nanjing teachers participating in this study was held at the end of October 2018. CSLP members provided detailed presentations on the use of and future updates to ABRA programs. Teachers who had participated in previous ABRA experiments shared their experiences, lesson plans, and game activities. The workshop thus provided valuable insights for the Nanjing teachers, who were encouraged to make final improvements before the end of the experiment.

Intervention

The treatment schools started using ABRA in March 2018 and were required to provide 20 hr of ABRA intervention. As noted, each school implemented ABRA according to their facilities and amount of time available for the ABRA intervention in the daily teaching schedule.

More specifically, School A implemented SG in English class, introducing one or two games to students at a time. Teachers usually focused on the national curriculum system and tried to link the words students were required to master to the words or phonics that appeared in ABRA games. Teachers spent 5–10 min on ABRA games in a class. Not every student was able to practice ABRA games. Rather, students watched the teacher or some of their classmates play the game and followed along by reading aloud, memorizing, or playing the games in their minds.

School B implemented noontime study in the classroom (NS), with ABRA conducted in the classroom for about 30 min after lunch and before formal courses in the afternoon. Students were shown the games and stories on a screen, after which English teachers demonstrated the gameplay for about 5–10 min. Some students then played the game, while the rest of the students followed the "leader," memorized the phonics or words, and experienced the procedure. Teachers monitored students' gameplay to correct any mistakes and strengthen their use or understanding of important words.

School C implemented ABRA through a CL, with the phonics curriculum forming part of the information and technology coursework. In the first class, English teachers showed students how to log in/out and return to the main page of ABRA and provided an overview of the available games and how to use them. Students were required to familiarize themselves with the software at home. In each class, two students each introduced a game and then played it with their classmates. Teachers monitored gameplay, stepping in to fix mistakes and explain any unfamiliar words. The overall process for game instruction usually lasted about 10 min, while the remaining 30 min was dedicated to playing the two games. Games were not introduced more than once, but students could select their favorite games for any after-school play.

Over the course of the experiment period (excluding summer vacation), NS and CL schools guaranteed 30–40 min of intervention per week. SG schools did not give sufficient attention to the intervention in the early stages of the experiment, providing less time for ABRA use than the other groups. Following the research team's communication with School A's leaders and teachers toward the end of May 2018, the amount of time allocated to the intervention was increased slightly to 20 min per week—still significantly less than the other groups. Arguably, the difference in intervention time reflects the difference in the attention and interest of the schools and teachers; this is discussed in greater detail later in this article.

In regard to the control schools (Schools D–F), English language classes proceeded as before. Teachers focused on students' vocabulary knowledge through repeated reading, picture-vocabulary

Table 3. School conditions.

	Group SG		Group NS		Group CL	
	Treatment School A	Control School D	Treatment School B	Control School E	Treatment School C	Control School F
School region in Nanjing	Gulou	Gulou	Jianye	Qinhua	Xuanwu	Xuanwu
Total number of classes in the school	52	18	20	15	30	24
School average scores[a]	1,129.62	1,008.09	1,100.54	1,044.65	1,169.05	1,189.03
Student/teacher ratio	0.053	0.077	0.062	0.069	0.059	0.062
Student/school area ratio (per meter square)	4.34	13.77	9.43	5.47	5.35	11.31

Note. SG = single-game instruction; NS = noontime study after lunch; CL = computer laboratory.

[a]Scores are based on the standardized test results for Jiangsu province for 2016; the standardized test has been held every 2 years since 2006.

correspondence, dialogue, and silent writing. There was little reference to or information about phonics or phonological awareness during the experiment. Students were not introduced to the use of ICT for English language learning. Table 3 summarizes the conditions of the participating schools.

Measures

The *Group Reading Assessment and Diagnostic Evaluation* (GRADE; Williams, 2001) is a diagnostic reading test determining a student's mastered ability and potential for improvement in the future. The test does not follow the criteria of the Common European Framework of Reference for Languages but determines ability according to 11 reading levels ranging from prekindergarten to adult. This study selected the GRADE kindergarten level for Grade 3 students for the following two reasons. First, as the Ministry of Education (2015) advocated the initiation of English language lessons from Grade 3, third-grade students in Nanjing were at the beginner level in terms of their English learning. Second, as presented in Table 2, students in Schools A and B started English learning in Grade 1 and other students started in Grade 3. The experiment began in March 2018 and it is the second semester for students in Grade 3. So, all students had received English language learning for at least half a year. Accordingly, the kindergarten level was appropriate for participating students.

This study adopts GRADE as this reading readiness test evaluates phonological awareness, early literacy skills, and phoneme-grapheme awareness. More specifically, the phonological awareness items measure sound matching and rhyming, while early literacy skills measure students' abilities to recognize letters and find the same or different words. Meanwhile, phoneme-grapheme

Table 4. MDES with eight outcomes.

Outcome	Same words	Different words	Rhyming	Sound matching begins with	Sound matching ends with	Letter recognition	Phoneme-grapheme initial	Phoneme-grapheme final
R^2	.898	.473	.868	.681	.866	.739	.742	.637
MDES	.063	.142	.071	.111	.072	.1	.099	.118

Note. MDES = minimum detectable effect size.

assessment asks students to match the right grapheme of words to their initial or final phoneme. Accordingly, this study tests eight basic scales or outcomes: same words, different words, rhyming, sound matching begins with, sound matching ends with, letter recognition, phoneme-grapheme initial, and phoneme-grapheme final. GRADE has been shown to have strong internal consistency (.95–.99), high alternate form reliability (.81–.94), and high test–retest reliability (.80; Mak et al., 2017). All subtests were used in the pre- and posttests.

This study also conducted interviews with the teachers using open-ended questions to explain the difference in the effect size. Questions were designed to obtain information regarding a teacher's teaching behavior in using ABRA, as well as their perceptions regarding their own professional development and student improvement. The questions also probed teachers' resistance to the experiment and recommendations for ABRA implementation (Appendix A). Of the nine participating teachers, three agreed to a formal interview, while the remaining six agreed to informal interviews. Of the former, two teachers were interviewed via WeChat, an online instant messaging tool, on March 24, 2019, and March 25, 2019, respectively; the other teachers were interviewed via phone on March 26, 2019. In regard to informal interviews, six teachers were informally interviewed during and after the intervention. All interview content was recorded with the teachers' consent.

Data coding using a back-to-back method was conducted by three members of the research team. Interview records were individually coded to categorize data by subject, which were then connected and aggregated according to the correlation and meaning of the codes.

Power estimates

The minimum detectable effect size (MDES) was calculated for the study using PowerUp! (Dong & Maynard, 2013). This study comprises a sample of 945 students (711 treatment, 231 control) with two-tail test producing an α of .05, power equal to .8, and a pretest–posttest correlation calculation ranging from .47 to .89 for eight outcomes. The MDES was 0.06–0.14. Table 4 presents the correlation of each outcome and corresponding effect size.

Results

The analytical sample for the control group was 288 in the pretest and 271 in the posttest; 10 samples were deleted after the pretest and 27 after the posttest due to absence or a score of 0 points. Regarding the analytical sample of the treatment group, 106 samples in the pretest and 142 in the posttest were deleted for the same reasons. However, students' scores were not deleted in matching pretest–posttest scores as their scores could present an important part of the overall student reading level. Therefore, effect size data are calculated using Cohen's d based on the formula recommended by Lipsey and Wilson (2001), where $\overline{M_{GT}}$ and $\overline{M_{CT}}$ represent the mean difference for the treatment group and control group, respectively, and S_{pooled} reflects the pooled standard deviation (Hedges & Olkin, 1985):

$$\overline{ES} = \frac{\overline{M_{GT}} - \overline{M_{GC}}}{S_{pooled}}. \quad (1)$$

To compare the outset of the treatment and control students, independent two-sample tests of the pretest were completed for the overall pretest scores. Analysis of all the treatment and control students revealed no significant difference in four basic scales, namely, same words, different words, letter recognition, and phoneme-grapheme final ($p > .05$). However, student scores for rhyming, $t(937) = 7.08$, $p < .00$; sound matching begins with, $t(937) = 4.05$, $p < .00$; sound matching ends with, $t(937) = 3.78$, $p < .00$; and phoneme-grapheme initial, $t(937) = 2.16$, $p < .05$, differed between the two groups, with student scores in the treatment group significantly higher than those of the control students. Accordingly, this study conducted detailed two-sample tests of a pretest for the three groups of treatment students (i.e., SG, NS, and CL) separately to explore the critical difference.

The SG group showed no significant difference in three basic scales, including same words, different words, and phoneme-grapheme initial ($p > .05$). Treatment students performed better than control students in rhyming, $t(414) = 11.82$, $p < .00$; sound matching begins with, $t(102) = 6.07$, $p < .00$; sound matching ends with, $t(96) = 2.17$, $p < .05$; and phoneme-grapheme final, $t(414) = 4.47$, $p < .00$. However, control students performed significantly better in terms of letter recognition, $t(218) = -2.18$, $p < .05$.

The NS group showed no significant differences in six basic scales, including same words, different words, rhyming, sound matching begins with, letter recognition, and phoneme-grapheme initial ($p > .05$). Treatment students scored higher in sound matching begins with, $t(118) = 3{,}46$, $p < .05$, while control students scored higher in phoneme-grapheme final, $t(297) = -2.87$, $p < .05$.

The CL group showed no significant differences in seven basic scales, namely, same words, different words, rhyming, sound matching begins with, sound matching ends with, letter recognition, and phoneme-grapheme initial ($p > .05$). Control students performed better than treatment students in phoneme-grapheme final, $t(222) = 2.35$, $p < .05$.

Table 5. Results of different items.[a]

		Control		Treatment	
Outcome	ES	Pretest ($N = 231$)	Posttest ($N = 271$)	Pretest ($N = 714$)	Posttest ($N = 675$)
Same words	−0.13	4.96 (0.27)	4.98 (0.15)	4.98 (0.34)	4.97 (0.34)
Different words	0.00	3.98 (0.19)	3.97 (0.21)	3.99 (0.18)	3.98 (0.18)
Rhyming	−0.14	7.11 (2.56)	9.82 (2.12)	8.48 (2.04)	10.78 (2.04)
Sound matching begins with	1.64	4.70 (1.34)	4.87 (1.18)	5.07 (0.91)	5.38 (0.91)
Sound matching ends with	0.27	4.90 (1.47)	4.44 (1.26)	5.25 (0.99)	5.18 (0.99)
Letter recognition	0.13	10.53 (1.51)	10.35 (1.35)	10.44 (1.28)	10.46 (1.28)
Phoneme-grapheme initial	−0.27	7.05 (1.42)	7.27 (1.10)	7.28 (1.01)	7.16 (1.01)
Phoneme-grapheme final	0.38	5.32 (1.50)	6.68 (1.36)	5.36 (1.18)	7.12 (1.18)
Random	0.05				

[a]Pretest–posttest scores are provided with mean scores and standard errors in parentheses.
Note. ES = Effect Size.

As such, treatment and control schools in the CL group showed the greatest equivalence, while those in the SG group showed the greatest difference, indicating that the overall difference between treatment and control students was primarily due to students in the SG group. Scales relating to phonics and phonological awareness showed greater variation between treatment and control schools.

Overall student achievement

To identify whether ABRA impacted students' literacy skills, this study compared the test scores of the treatment and control students. Table 5 presents a summary of the students' scores for all items. Some students in the pretest–posttest provided multiple or missing choices in most of the items in the paper; this is not reflective of their ability but that they did not understand the test's instructions. Moreover, some students do not appear to have taken the test seriously, some doodling on the test paper. This was more common toward the end of the experiment, when participant school appeared to be treating the intervention more casually. The results of these pre- or posttest papers are excluded, but paired scores are retained.

Results indicate that the effects favor ARBA (treatment) students in some items, while the overall effect size of $d = 0.05$ in the randomized model ($Q = 75.90$, $p < .01$) indicates little effect according to Cohen's criteria (Cohen, 1988). Analysis revealed little difference between the two groups in terms of the change in their scores in subscales related to early literacy skills, including same words, different words, letter recognition, and rhyming, for which the effect size ranged from

−0.15 to 0.15. However, results show that treatment students improved their scores in phonics-related or phonological awareness at a higher rate than control students. More specifically, treatment students outperformed control students in sound matching begins with ($d = 1.64$), sound matching ends with ($d = 0.27$), and phoneme-grapheme final ($d = 0.38$) but made less improvement than control students in phoneme-grapheme initial ($d = -0.27$).

Differences between groups

This study also examined whether the effects of ABRA differed between groups. Table 6 summarizes the differences between treatment and control groups in terms of student gain scores. Results show that treatment students in NS and CL groups improved more than control students with a medium overall effect sizes of 0.39 and 0.55, respectively. However, the scores of treatment students in the SG group exceeded the overall effect size of −0.52. This means the overall effect for three groups was significantly influenced by the effect size of group SG because approximately half of the treatment students were in the SG group.

In regard to the SG group, the ABRA tool had little effect on treatment students. Treatment and control students performed similarly in phonics-related or phonemic-awareness-related subscales, including rhyming, sound matching begins/ends with, and phoneme-grapheme initial/final. Students also scored similarly in all early literacy skills scales except different words ($d = 0.31$). Meanwhile, in terms of the NS group, treatment students showed progress in all dimensions of early literacy skills, phonics, and phonemic awareness except in same/different words and rhyming. The effect size was particularly large in sound matching begins with (0.90), sound matching ends with (0.70), and phoneme-grapheme final (1.59). Finally, while the CL group showed a similar pattern of progress to the NS group, both treatment and control students in the CL group evidenced higher gains in all three dimensions. Effect size was highest in rhyming (1.15), sound matching begins with (1.18), sound matching ends with (0.71), and phoneme-grapheme final (1.29). As such, students in the CL group showed the greatest gains in all of the tested abilities except different words, letter recognition, and phoneme-grapheme initial. While students in the NS and CL groups showed similar outcomes, they also evidenced a significant difference in scores in same words, rhyming, and letter recognition.

Teacher interviews

As such, results indicate that ABRA has a positive influence on students' achievements, although the relatively small effect size in the SG group and the medium effect size on the NS and CL groups are somewhat surprising. This study controlled the treatment and control groups in terms of the number of participating students, school size, teacher–student ratio, and location. As such, the difference between the SG, NS, and CL groups may have resulted from differences within the

Table 6. Results of different groups.

			Control		Treatment	
	Outcome	ES	Pretest ($N = 79$)	Posttest ($N = 105$)	Pretest ($N = 339$)	Posttest ($N = 314$)
SG	Same words	0.05	5.00 (0.00)	4.99 (0.10)	4.99 (0.18)	4.96 (0.41)
	Different words	0.31	4.00 (0.00)	3.95 (0.25)	3.99 (0.18)	3.99 (0.06)
	Rhyming	−0.92	6.35 (2.26)	9.90 (1.93)	9.66 (2.22)	11.32 (1.99)
	Sound matching begins with	−1.11	4.32 (1.29)	5.43 (1.00)	5.28 (1.05)	5.14 (1.03)
	Sound matching ends with	−0.33	4.95 (1.50)	4.93 (0.97)	5.34 (1.07)	4.97 (1.07)
	Letter recognition	−0.11	10.53 (1.32)	10.68 (0.68)	10.09 (2.47)	10.16 (1.76)
	Phoneme-grapheme initial	−1.04	6.78 (1.35)	7.54 (0.66)	7.17 (1.69)	6.75 (1.16)
	Phoneme-grapheme final	−0.97	4.69 (1.40)	6.90 (1.13)	5.56 (1.58)	6.78 (1.46)
	Random	−0.52				
	Outcome	ES	Pretest ($N = 88$)	Posttest ($N = 99$)	Pretest ($N = 213$)	Posttest ($N = 212$)
NS	Same words	−0.18	4.92 (0.41)	4.97 (0.17)	4.96 (0.24)	4.95 (0.37)
	Different words	−0.24	3.95 (0.26)	3.97 (0.22)	3.99 (0.07)	3.96 (0.30)
	Rhyming	−0.11	7.38 (2.38)	10.16 (2.29)	7.80 (2.11)	9.97 (1.88)
	Sound matching begins with	0.90	5.01 (1.24)	4.53 (1.20)	5.13 (0.99)	5.57 (0.76)
	Sound matching ends with	0.70	4.82 (1.44)	3.77 (1.24)	4.40 (0.95)	5.33 (0.91)
	Letter recognition	0.29	10.62 (1.22)	9.96 (2.00)	10.88 (0.54)	10.83 (0.46)
	Phoneme-grapheme initial	0.17	7.45 (1.21)	6.92 (1.44)	7.68 (0.69)	7.52 (0.77)
	Phoneme-grapheme final	1.59	5.84 (1.25)	6.36 (1.62)	5.40 (1.15)	7.39 (0.86)
	Random	0.39				
	Outcome	ES	*M* (pre-T) ($N = 64$)	*M* (post-T) ($N = 67$)	*M* (pre-C) ($N = 162$)	*M* (post-C) ($N = 149$)
CL	Same words	0.10	4.97 (0.18)	4.97 (0.17)	4.99 (0.08)	5.00 (0.00)
	Different words	−0.12	4.00 (0.00)	4.00 (0.00)	4.00 (0.00)	3.99 (0.08)
	Rhyming	1.15	7.70 (2.90)	9.16 (2.01)	6.89 (2.91)	10.80 (2.08)
	Sound matching begins with	1.18	4.75 (1.44)	4.52 (1.06)	4.55 (1.44)	5.59 (0.69)
	Sound matching ends with	0.71	4.956 (1.45)	4.67 (1.24)	4.86 (1.45)	5.42 (0.83)
	Letter recognition	−0.07	10.40 (1.99)	10.42 (0.58)	10.63 (1.99)	10.58 (0.84)
	Phoneme-grapheme initial	0.18	6.84 (1.63)	7.34 (0.94)	6.97 (1.63)	7.48 (0.65)
	Phoneme-grapheme final	1.29	5.38 (1.64)	6.81 (1.14)	4.86 (1.64)	7.46 (0.71)
	Random	0.55				

Note. ES = Effect Size; SG = single-game instruction; NS = noontime study after lunch; CL = computer laboratory.

groups. Therefore, this study uses participant teacher interviews to explore the different effect sizes between groups.

Interview design and interviewees

The three treatment schools were treated equally and received the same amount of training, indicating that the differences in results may be due to differences in the attitudes or methods of the subject teachers or leaders. This study thus conducted interviews to investigate these differences; all interviews were completed via an online messaging application or telephonically. Interviewees were selected from teachers who participated in the treatment group. Treatment teachers in the same group shared a similar teaching and administration context and had similar outcomes. The leader of the English subject for the third grade was responsible for guiding the teachers in the application of the ABRA tool; these teachers also have a firm grasp of the teaching methods used by the teachers in their school. Therefore, this study randomly selected a teacher or an English subject leader from every treatment school to explain the management process. The characteristics of three randomly selected interviewees are similar in gender, marital status, and teaching years. They are all married females and have been a teacher for more than 5 years and less than 10 years. The teacher in treatment School A in the Group SG graduated from a National 211 Project University, which means she graduated from one of the highest-ranked universities in the country.

Qualitative analysis

Interviews revealed that teachers held ambiguous attitudes toward ABRA and had a passive role in the experiment. Interview data indicate that experiment outcomes were influenced by three issues related to teachers' attitudes or teaching method: namely, teachers' ambiguous attitudes and complaints about heavier workloads, the shortage of teaching hours, and teachers' passive roles in the experiment.

First, teachers' attitudes toward ABRA were found to be ambiguous, particularly in the light of heavy workloads. Before the experiment, teachers' attitudes toward ABRA were positive and teachers were optimistic regarding the beneficial impact of the program on students' English language progress. However, heavy daily workloads were overly taxing, making it impossible for them to prepare external courses or activities and expand their learning of phonics. For example, a teacher from School C, who had approximately 5 years of experience working as a teacher, stated that the curriculum plan was too narrow and did not allow for the preparation of extra activities. In addition to their unfamiliarity with phonics-based teaching and the ABRA website, teachers found it difficult to teach systematic phonics knowledge in games. One of the teachers explained that

some students were accustomed to following the radio or teachers rather than speaking out loud after decoding a word because certain phonics skills were too difficult for them.

Significantly, teachers' attitudes in this regard improved after further training, which provided a better understanding of phonics and phonics teaching and helped teachers teach phonics more systematically. Teachers in the NS and CL groups posed several questions and sought to reform some of their teaching strategies to alleviate students' confusion regarding the IPA, Hanyu Pinyin, and phonics. A teacher in School C stated that she tried to make some changes in her daily English class. In contrast, the teachers in School A appeared less interested in participating in the experiment and did not attempt to revise their strategies. Rather, School A teachers maintained the view that they were unfamiliar with phonics and unable to match ABRA games with textbook teaching because they were already too busy trying to understand the new edition of the textbook.

Second, interviews raised the shortage of teaching hours as an issue, noting that it was difficult to integrate phonics knowledge into daily lessons because the Chinese English language curriculum focuses on text comprehension and vocabulary. Consequently, teachers found little time in the daily class schedule to set aside for ABRA games. In regard to club classes, schools could arrange between two and four class hours a week; one class hour is typically 40–45 min in length. Some schools focused on establishing a featured curriculum. For example, School A encourages "a happy childhood" and provides courses in calligraphy, swimming, roller skating, and Chinese traditional musical instruments, while School C includes opera in their curriculum. As such, schools found it difficult to balance the curriculum. While this responsibility lies with school leaders, the cooperation of teachers and students is also necessary. In this respect, teachers grappling with already tight teaching schedules found it impossible to set aside enough time for the ABRA program. For instance, a teacher from School B claimed that she was already worried about teaching the required knowledge from the textbook in the allotted time and she thought such a large workload for such young students may have led to a less satisfying result. It should also be noted that the treatment schools chose different times for ABRA. The school in the SG group chose to use some of the time available in the English lesson, while the school in the NS group chose a period between lunch and formal afternoon classes and the school in the CL group chose club classes. Consequently, treatment students in the CL group used ABRA for the longest time, their results indicating that program intensity is a critical factor in its effectiveness (A. C. Cheung & Slavin, 2012).

Third, interviewees reflected that teachers played a passive role in the experiment. In terms of daily teaching, the experiment was executed by teachers rather than a member of the research team. However, as a result of heavy workload and lack of phonics knowledge, participating teachers revealed different degrees of passivity with most regarding the ABRA program as a burden. In Schools B and C, this passivity was rooted in a sense of powerlessness; teachers were

willing to implement the program and improve their ability to do so but faced challenges in terms of their heavy workload and inability to solve problems in the classes themselves. For instance, a teacher from School C recalled experiencing difficulty encouraging students to read aloud in high-frequency words, with many preferring the previous game. She also noted her desire to try new games or higher-level ones, but these ways seemed too difficult for students. She was further discouraged by her inability to identify useful strategies to enhance students' interest in the learning process. While teachers from the NS and CL groups experienced problems in teaching itself, the treatment teachers in the SG group were concerned about their lack of phonics knowledge, different teaching habits, and less systematic teaching resources. In this respect, a teacher from School C advised inviting institutions well-versed in phonics to teach phonics to students in club classes rather than relying on the teachers themselves. Certainly, cooperation between colleges and schools could improve teachers' professional abilities. Moreover, as resources are developed in the teaching process rather than researchers' theories, the passive attitude of the teachers in the SG group hinders their development.

Discussion

This study sought to examine the effectiveness of the ABRA program on the literacy skills of third-grade students in Nanjing. Results show that after about 20 weeks of ABRA intervention, treatment students performed better than control students, with effect sizes of 0.39 and 0.55 in the NS and CL groups, respectively. These positive results are consistent with prior ABRA experiments conducted in Canada (e.g., Comaskey et al., 2009; Di Stasio et al., 2012), Australia (Wolgemuth et al., 2011, 2013), and Kenya (Abrami et al., 2015) and indicate that the findings of such studies are applicable to Chinese students. However, there are some stumbling blocks hindering the effective application of ABRA in the Chinese context.

Teachers' workloads are too heavy to make any changes without the full support of the schools. In this study, the overall effect size of 0.05 was influenced by the effect size of the SG group, particularly insofar as interviewees revealed the passive attitude of the teachers in the SG group as well as their lack of knowledge regarding the ABRA program. Although SG teachers approved the positive effects of ABRA, they struggled to embrace the new teaching method and make revisions to their strategies because they had no time for ABRA plans or instruction. Teachers fundamentally influence students' behavior, providing both instruction and feedback (Becker & Ravitz, 2001). As such, based on interviews, this study argues that teachers' attitudes may account for the difference in results between the three groups.

Accordingly, the school's support plays an important role in the effective application of ABRA. However, most of China's public schools are conservative and seek to maintain their existing achievements. Public schools, especially top-ranked ones like the treatment school in the SG

group, face greater pressure to cultivate well-rounded students with high scores and have thus already established a curriculum for physical and mental development that does not leave room for a foreign, new, and time-consuming teaching program. Therefore, it is likely that the teachers in the SG group were influenced by the attitudes of the school leaders.

To resolve these issues, ABRA intervention could be implemented as informal learning during students' after-school activities. Universities or colleges can provide school–parent workshops, instructing parents on the use of ABRA, how to answer questions from children, and how to avoid internet addiction. Accordingly, the responsibility of ABRA implementation would rest with the parents rather than the primary school, thereby reducing the burden on teachers to learn and implement ABRA in their daily English lessons.

This study found that the control groups in the NS and CL groups performed significantly better in phoneme-grapheme scale in the posttest than the treatment groups, with an effect size of 1.59 and 1.29, respectively. In contrast, treatment students performed significantly higher in rhyming, sound matching begins with, sound matching ends with, and phoneme-grapheme final, although there was no significant difference between the treatment and control students in the SG group. The treatment students in the SG group began learning English in the first grade, which means that they had spent more time learning English than the control students. Studies comparing the second language learning period of students between the ages of 6 and 7 and 8 and 9 show that students between the ages of 6 and 9 develop similarly (Becker & Ravitz, 2001; Lenneberg, 1967). This may explain why the treatment students in the SG group did not retain an advantage in phonics after a semester; in other words, despite learning English for a longer time, the treatment students were in the same period of second language acquisition as their peers in the control group.

It is also worth noting that students in the NS and CL groups received enough ABRA instruction to make improvements in their phonics and phonological skills. Overall, the amount of per student ABRA exposure in this study was below that of the interventions in other experiments. According to Savage et al. (2010), students should have an hour of ABRA exposure per week, although some interventions ensured that the software was used 2 hr a week. However, none of the treatment groups in this study met these criteria, with weekly ABRA application ranging between the SG group's 20 min and CL group's 40 min. As a result of the limited equipment at both the schools and students' homes, not every student could use the software by themselves. Consequently, the students' time using ABRA was allotted by the teacher.

Nonetheless, results indicate the significant value of ABRA for phonics instruction, with the effect size of nearly 1.00 for sound matching begins with, sound matching ends with, and phoneme-grapheme final. Although ABRA is a comprehensive early literacy program, approximately half of its activities are related to alphabetic practice. While Chinese teachers are familiar with reading, understanding, and writing skills, they are new to phonics as an efficient word-

learning method. Teachers tend to spend more time preparing and introducing phonics to students than changing their teaching methods on reading and writing. Moreover, the teaching objectives for the first to third grades are focused on words or letter recognition. There is a consensus among teachers that reading, understanding, and writing skills are more suitable for older students and too difficult for children in this period.

Conclusion

This study demonstrates that ABRA had a positive effect on students' English language literacy in Nanjing, China. These results strengthen the value of the widespread use of this program in English learning countries, regardless of whether it is applied in an English or Chinese language context. However, teachers are critical to the success of this initiative, particularly in terms of their teaching behaviors. This study found the majority of teachers were too busy with their daily workload to prepare ABRA lessons. Accordingly, ABRA can only be effective if school leaders provide adequate support for teachers and if teachers are willing to dedicate the necessary time to ABRA and innovate their teaching strategies. This can be facilitated by greater communication between primary schools and universities. In this respect, this study suggests that greater support be given to the less-advantaged schools rather than schools that are already successful. Additionally, school leaders need to recognize the importance of making a difference. Therefore, researchers should provide training resources to and discuss potential measures with teachers, including those that may help them reduce their workload, schedule enough time for ABRA, and make ABRA more suitable for the Chinese context.

This study has several limitations. Firstly, there was not enough supervision of and communication with teachers due to the limited number of researchers and funds. Although there were several training sessions and visits to schools, most of the communication was conducted online. Some teachers, especially those in the SG group, prioritized their daily teaching content over ABRA instruction, leaving less time for the latter. Second, teachers could not provide equal attention to all students, with classes typically over 30 students in size, with some over 40. Teachers thus spend more time maintaining order, often overlooking disadvantaged students. There are likely to be difference in student development according to gender, cognitive level, and socioeconomic status. Third, the testing instrument was not designed for second language learners. To control the difficulty of the subject matter, third-grade students used a test paper for kindergarten students who learn English as the first language, resulting in some bias of students' abilities. Future studies could provide labs for students to receive similar exposure from games or teachers could use more cooperative teaching methods with small learning groups to provide more communication opportunities. It is also advisable to design standardized questionnaire for primary school English beginners in China.

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: This project was funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Contributorship

Hui Gu, Jijun Yao, Longjun Zhou, Alan C. K. Cheung and Philip C. Abrami concieived the presented idea. Cheung, Yao and Abrami designed the experiment. Zhou, Gu, Yao, and Cheung implemented the experiment and contributed the data. Gu and Yao wrote the initial draft of the manuscript. Gu, Yao, Zhou, Cheung, and Abrami revised the manuscript.

ORCID iD

Hui Gu https://orcid.org/0000-0001-8018-3430

Supplemental Material

Supplemental material for this article is available online (https://journals.sagepub.com/doi/suppl/10.1177/2096531120972709).

References

Abrami, P.C., Borohkovski, E., & Lysenko, L. (2015). The effects of ABRACADABRA on reading outcomes: A meta-analysis of applied field research. *Journal of Interactive Learning Research*, *26*(4), 337–367.

Abrami, P. C., Wade, C. A., Lysenko, L., Marsh, J., & Gioko, A. (2016). Using educational technology to develop early literacy skills in sub-Saharan Africa. *Education and Information Technologies*, *21*(4), 945–964.

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print*. MIT Press.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635.

Avalos, M. A., Plasencia, A., Chavez, C., & Rascón, J. (2007). Modified guided reading: Gateway to English as a second language and literacy learning. *The Reading Teacher*, *61*(4), 318–329.

Becker, H. J., & Ravitz, J. L. (2001, March). Computer use by teachers: Are Cuban's predictions correct (pp. 10–14). *Annual Meeting of the American Educational Research Association*, Seattle, WA.

Bernaus, M., & Gardner, R. C. (2008). Teacher motivation strategies, student perceptions, student motivation, and English achievement. *The Modern Language Journal*, *92*(3), 387–401.

Blevins, W. (2017). *Teaching phonics & word study in the intermediate grades*. Scholastic Professional Books.

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*(3), 198–215.

Cheung, A., Mak, B., Abrami, P., Wade, A., & Lysenko, L. (2016). The effectiveness of the ABRACADABRA (ABRA) web-based literacy program on primary school students in Hong Kong. *Journal of Interactive Learning Research*, *27*(3), 219–245.

Cheung, H. (1999). Improving phonological awareness and word reading in a later learned alphabetic script. *Cognition*, *70*(1), 1–26.

Chu, M. C., & Chen, S. H. (2014). Comparison of the effects of two phonics training programs on L2 word reading. *Psychological Reports*, *114*(1), 272–291.

Cohen, J. (1988). *Statistical power analysis for the behaviors science* (2nd ed.). Laurence Erlbaum.

Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomised efficacy study of web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading*, *32*(1), 92–108.

Deault, L., Savage, R., & Abrami, P. (2009). Inattention and response to the ABRACADABRA web-based literacy intervention. *Journal of Research on Educational Effectiveness*, *2*(3), 250–286.

Di Stasio, M. R., Savage, R., & Abrami, P. C. (2012). A follow-up study of the ABRACADABRA web-based literacy intervention in Grade 1. *Journal of Research in Reading*, *35*(1), 69–86.

Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2), 116–125.

Guo, X. (2018). Examining the effects of ABRACADABRA, a web-based literacy program on primary school students in China: A mixed method study [Unpublished doctoral dissertation]. The Chinese University of Hong Kong.

Guo, Y. (2014). The application strategy of natural spelling in primary school English teaching [in Chinese]. *Modern Primary and Secondary Education*, *8*(8), 72–76.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Higgins, J. P., & Green, J. (2009). *Cochrane handbook for systematic reviews of interventions*. Elsevier's Scopus.

Hou, J. (2013). *An empirical study on using natural spelling to promote primary school students' introduction to English reading* [Doctoral dissertation, in Chinese]. Suzhou University.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67.

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

Liu, C. (2019). *Applied research of natural spelling in junior English vocabulary teaching* [Doctoral dissertation, in Chinese]. Yili Normal University.

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading*, *29* (2), 162–172.

Madrid, D. (2002). The power of the FL teacher's motivational strategies. *Cauce*, *25*, 369–422.

Mak, B. S., Cheung, A. C., Guo, X., Abrami, P. C., & Wade, A. (2017). Examining the impact of the ABRACADABRA (ABRA) web-based literacy program on primary school students in Hong Kong. *Education and Information Technologies*, *22*(6), 2671–2691.

Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. ASCD.

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*(S1), 167–210.

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, *105*(2), 310.

Savage, R., Abrami, P., Piquette-Tomei, N., Wood, E., & Deleveaux, G. (2008). *ABRACADABRA: A study in the development, implementation and effectiveness of a web-based literacy resource*. Executive Report for the Canadian Council of Learning. https://www.researchgate.net/publication/242170831_ABRACADABRA_A_study_in_the_development_implementation_and_effectiveness

Savage, R. S., Erten, O., Abrami, P., Hipps, G., Comaskey, E., & van Lierop, D. (2010). ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in Grade 1 language arts programs. *Computers & Education*, *55*(2), 911–922.

Su, L. (2000). The impact of mother tongue transfer on second language learning [in Chinese]. *Journal of Beijing International Studies University*, *42*(4), 44–52.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. American Guidance Service.

Wolgemuth, J. R., Savage, R., Helmer, J., Harper, H., Lea, T., Abrami, P. C., Kirby, A., Chalkiti, K., Morris, P., Carapetis, J., & Louden, W. (2013). ABRACADABRA aids indigenous and non-indigenous early literacy in Australia: Evidence from a multisite randomized controlled trial. *Computers & Education*, *67*, 250–264.

Wolgemuth, J. R., Savage, R., Helmer, J., Lea, T., Harper, H., Chalkit, K., Bottrell, C., & Abrami, P. (2011). Using computer-based instruction to improve indigenous early literacy in Northern Australia: A quasi-experimental study. *Australasian Journal of Educational Technology*, *27*(4), 727.

Wood, E., Grant, A., Gottardo, A., & Savage, R. (2014). Evaluating online and offline reading software for young learners. *Global Journal on Technology*, *5*. http://www.world-education-center.org/index.php/PITCS/article/viewArticle/3054.

Wood, M. C. (1996). (*U.S. Patent No. 5,511,980)*. Talking phonics interactive learning device. U.S. Patent and Trademark Office.

Wu, T., & Huang, Y. (2017). A mobile game-based English vocabulary practice system based on portfolio analysis [in Chinese]. *Journal of Educational Technology & Society*, *20*(2), 265–277.

Wu, X., Zeng, L., Qin, X., & Zheng, Q. (2003). Research on the status quo and related factors of job burnout of primary and secondary school teachers [in Chinese]. *Psychological and Behavioral Studies*, *18*(4), 262–267.

Wu, X. (2016). On the application of natural spelling in primary school English teaching [in Chinese]. *China's Extracurricular Education*, *14*(2), 96–97.

Yang, W., & Hu, L. (2004). Interlanguage theory and error analysis in second language teaching [in Chinese]. *Journal of Chongqing Three Gorges University*, *20*(3), 45–48.

Zhou, J. (2017). *The application of natural spelling in junior English vocabulary teaching* [Doctoral dissertation, in Chinese]. Shaanxi Normal University.

Zhu, J. (2018). The development of international phonetic alphabet system and its embodiment in middle school English textbooks [in Chinese]. *Overseas English*, *380*(16), 249–250.