The efficacy of a computer-adaptive reading program on grade 5 students' reading achievement scores

Laurie O. Campbell[1] · Cassandra Howard[2] · Glenn W. Lambie[3] · Xueying Gao[4]

Abstract

Reading in upper-elementary grades includes comprehending complex texts and learning disciplinary-specific vocabulary. This study aims to determine the effects of a computer-adaptive supplementary reading program on fifth-grade students' reading achievement. Using propensity score matching to create equivalent groups of 450 students for both control and intervention groups (N = 900), a quasi-experimental design was employed to examine changes in fifth-grade students' overall, comprehension, vocabulary, reading achievement scores who used this program as compared to those who did not. Students from eight school districts and 108 schools in a Southeastern state were matched based on demographics and initial reading ability, and the Title I status of their schools. Students in the intervention group received at least 30 min per week of supplemental computer-adaptive reading instruction between the months of September and May. Mean differences between pre- and post-test scores identified that students in the intervention group utilizing a computer adaptive reading curriculum had statistically significant higher gains than students in the control group on their overall reading, vocabulary, and reading comprehension scores with small effect sizes. The findings support the use of a supplemental computer adaptive reading program for improving overall reading and reading comprehension outcomes among these fifth-grade students.

Keywords Computer adaptive reading programs · Reading achievement · Supplemental reading program · Upper elementary

1 Introduction

The results of National Assessment of Educational Progress (NAEP, 2019) identify the continued challenges to reading achievement in the United States. Only 35%

Extended author information available on the last page of the article

of fourth graders are performing at or above proficient in reading, down from 36% to 2017. Supporting reading proficiency in upper-elementary classrooms entails responding to the differentiated needs of readers representing varying levels of proficiency (Connor et al., 2014; Connor & Morrison, 2016). National Reading Panel (2000) notes that instructional approaches for reading achievement should include a focus on (a) phonemic awareness, (b) phonics, (c) fluency, (d) comprehension, and (e) vocabulary. To support effective reading instruction, teachers require evidence of how their students are progressing in the different reading domains (Sztajn et al., 2012). Formative assessments and adaptive curriculum found in a computer adaptive reading program (CARP) may support the teaching of reading by providing teachers with data to guide instructional decisions to tailor their reading instruction to students' current abilities. Real-time knowledge of students' reading ability may be important for learners in upper-elementary grades, given research findings identify a "slump" in reading achievement for low-income students after the third grade (Campbell et al., 2019; Chall & Hirsch, 2003; Chall & Jacobs, 1983; Stockard, 2010;).

1.1 Reading instruction in the upper-elementary grades

Once students reach upper-elementary grades, a shift in classroom reading focus occurs, away from "learning to read" to "reading to learn" (Chall, 1983). Students in upper-elementary grades are tasked with comprehending increasingly complex texts which requires drawing upon a wider range of vocabulary. Shifts in reading purpose and instruction in upper-elementary grades are associated with widening reading achievement gaps – termed the "fourth-grade slump" – between low-income students and non-low-income students (Chall & Jacobs, 1983). Torgesen and colleagues (2007) identified six areas of "knowledge, skill, and aptitude," that readers in early adolescence should develop to maintain and increase reading proficiency as they advance to upper grades, including (a) fluency, (b) vocabulary knowledge, (c) content knowledge, (d) higher-level reasoning and thinking skills, (e) cognitive strategies specific to reading comprehension, and (f) motivation and engagement. A meta-analysis on students with reading difficulties and disabilities in upper-elementary grades found evidence in support of instruction in comprehension strategies, fluency interventions, vocabulary instruction, and multi-component interventions but emphasized the need for more research on interventions that address the unique needs of reading learners in upper-elementary grades (Wanzek et al., 2010). In addition, a systematic review of research studies indicated that interventions with upper-elementary grade students that change their daily teaching practices were more effective at raising reading achievement; however, the need remains for more high-quality studies on effective programs to support upper-elementary reading proficiency (Slavin et al., 2009). Therefore, research findings have identified promising practices to promote upper-elementary grade students' reading achievement. Nevertheless, additional research is warranted to distinguish effective reading programs that teachers, schools, and districts may implement to address proficiency gaps.

1.2 Computer adaptive reading and differentiated instruction

There is evidence that targeting the specific learning needs of reading students in upper-elementary classrooms improves reading achievement (Jacob et al., 2015; Kim et al., 2011; Stockard, 2010). Differentiation in general instruction is a promising method for responding to the unique needs of learners representing diverse skill sets and interests (Reis et al., 2011; Tomlinson, 2014) and may be particularly important for struggling readers. Computer-based programs have the potential to support differentiated instruction since they enable real-time data production and individualized learning profiles. Differentiation in reading instruction can be especially important when a student is missing foundational skills and needs additional instructional support. Technology-enhanced learning may increase motivation, especially when using gamification to engage students (Dicheva et al., 2015; Hong & Masood, 2014).

Research findings identify the potential of computer-assisted programs to increase students' reading achievement scores (e.g., Cheung & Slavin, 2013; Kamil & Chou, 2009). For example, an investigation into the effectiveness of a supplementary, computer-based program in Texas found statistically significant reading skill growth for fourth and fifth graders who received the technology-based intervention (SEG Measurement, 2018). In their review of 25 studies on PK-12 reading instruction and computer technology, Kamil and Chou (2009) found evidence that computer-based programs may enhance learning of vocabulary and comprehension. However, they note a need for more research on recent technological advances, such as adaptive technology. Computer adaptive technologies may be used for both reading instruction and assessment. For assessments, adaptive technology overcomes a major challenge of non-adaptive standardized tests by tailoring the assessment to examinees' individual ability, such that it is neither too challenging nor too simple to measure students' knowledge (McGlohen & Chang, 2008; Merrell & Tymms, 2007).

Adaptive technology is used in summative state assessments of students' yearly progress (Flanigan, 2014), but may be particularly useful for formative assessments of students' reading level. There are several ways formative reading assessments may be useful for elementary school teachers, including offering evidence of students' progress towards goals, providing feedback to teachers and students, identifying next steps in learning, and judging achievement (Harlen, 2012). Formative reading assessments may also help teachers predict students' performance on standardized achievement tests (Marcotte & Hintze, 2009). In reading, formative assessments such as miscue analysis inform teachers' response to both struggling and excelling students and serve as an impetus for teachers to engage students in their own learning process (Goodman & Goodman, 2014). When formative reading assessments incorporate adaptive technology, benefits include pinpointing students' specific learning needs (Bennett, 2011), which may vary in upper-elementary classrooms. Further, computer-adaptive assessments can be linked to adaptive curriculum delivered through technology or in a face-to-face environment. In some computer-adaptive programs such as the one investigated in our study; the computer delivers a supplemental computer-adaptive curriculum in addition to teacher-delivered supplemental lessons to target specific deficit skills based on learners' needs.

Theoretically, computer-adaptive technologies employed in reading support students' achievement in their potential Zone of Proximal Development (Murray & Arroyo, 2002; Navarro & Mourges-Codern, 2018; Vygotsky, 1978). As computers adapt to students' reading knowledge and skills level, new content is presented (Murray & Arroyo, 2002). In some circumstances, computers may serve as the more knowledgeable one, moving the student towards their potential (Vygotsky, 1978).

1.3 Supplemental computer-adaptive reading program

Supplementary reading curriculum increases students' reading achievement (Council et al., 2016; Gibson JR et al., 2014; Kent et al., 2017; Macaruso et al., 2006; Tyler et al., 2015; Wise et al., 2000). Supplemental reading curriculum combined with computer-adaptivity affords multiple opportunities for continuous progress monitoring to personalize instruction that enhances students' learning and supports classroom instruction (Fuchs & Fuchs, 2001).

As a supplementary computer-adaptive reading program (CARP) that provides a formative assessment tool and adaptive curriculum, the Istation Reading Program is designed to help teachers identify and respond to the diverse learning needs of their students. Within the reading program, formative assessment results are included in the interactive digital lessons scaffolded to students' individualized learning needs. Previous research investigating the reading program identified that third grade students in the most need of reading intervention have greater achievement when they practice reading using the supplemental program both in and out of class (Campbell, Sutter, & Lambie, 2019). Yet, the efficacy of the reading program and 5th grade reading achievement has not been investigated. Specifically, more research is needed to examine how CARPs may support upper-elementary school students' reading achievement (Blok et al., 2002; Jamshidifarsani et al., 2019; Wanzek et al., 2010).

1.4 The present study

To better inform educational stakeholders, the purpose of the present study was to investigate the efficacy of a CARP on fifth grade students' reading achievement. The current investigation employed a quasi-experimental design to determine the influence of the CARP (intervention) on students' reading proficiency scores. Specifically, the study addresses the following research questions:

1. What were the differences, if any, of fifth-grade students' overall reading, vocabulary, and reading comprehension scores when using a CARP as compared to students that do not use the CARP?
2. When considering all of the fifth-grade students by Academic Level, what were the differences, in the overall reading, vocabulary, and reading comprehension scores between the intervention (CARP) and control group?

2 Method

2.1 Design

To examine the efficacy of the CARP on fifth-grade students' reading scores, we used propensity score matching (PSM; Rosenbaum & Rubin, 1983) to create groups matched on key ability and background characteristics and to establish group equivalence in observational studies (Graham & Kurlaender, 2011; Rosenbaum, 2009). Specifically, groups were matched on gender, race, pre-treatment reading level, and the Title I status of their schools. The PSM approach enables researchers to address the issue of selection bias and control for confounding variables, or factors that existed before treatment that may be associated with outcomes.

PSM is recognized by the What Works Clearinghouse (U.S. Department of Education, 2017) as an acceptable approach to meet the standards for quasi-experimental designs outlined for "evidence-based practices" in Every Student Succeeds Act (2015). The Institute for Education Statistics describes specific considerations for quasi-experimental designs using PSM in its standards handbook, and the criteria for meeting those standards were consulted to support evidence of validity and reliability in the research design.

2.2 Participants

Before using PSM, the population included 14,525 fifth graders from 14 school districts and 326 different schools across in a Southeastern state. These districts and schools implemented the CARP to varying degrees of fidelity, employing distinct policies and procedures to assign, encourage, and monitor the reading program's usage. Variations in the implementation of the CARP allows us to examine the intervention as it is used under typical, instead of ideal, conditions. Common among all participants is that the reading assessment component was used to establish a pre- and posttest measure. Because the reading program recommended at least 30 min of curriculum usage per week, the intervention group consisted of students who used at least 30 min of the Istation reading curriculum per week between the months of September and May of the 2016-2017 school year. The control group consisted of fifth grade students who only used the CARP for assessment purposes. The pretest scores determined the students' academic level.

2.3 Students' academic level

Students' achievement level was measured based on their initial achievement percentiles determined on their initial formative annual assessment. Students at Level 1 include all students who initially scored at or below the 20th percentile. In other words, students at Level 1 are in the most need of reading support. Students at Level 2 initially scored between the 21st – 40th percentiles, indicating their need of reading support. Finally, students at Level 3 scored above the 40th percentile in reading, suggesting that they have sound reading proficiencies.

2.4 Control and intervention groups

A comparison of control and intervention groups before PSM identified that the groups were uneven (Table 1). Compared to the intervention group ($n = 5,548$), the control group ($n = 8,977$) had higher percentages of struggling students (Academic Level 3), male students, white, non-Hispanic students, and students from Title I schools. The results of an independent samples t-test showed that baseline scores for the overall and comprehension assessment were significantly different between groups (overall: $t(11,491)=5.7$, $p<.001$; comprehension: $t(10,869)=1.9$, $p<.001$) (Table 2).

To reduce the imbalance in the sample size of the two fifth grader student groups, nonparametric preprocessing was used to "match" students on the variables of race, gender, initial reading ability, and the Title I status of their schools using the MatchIt package in R (Ho et al., 2007). These variables were selected based on prior literature. A historical reading achievement gap has been identified in the United States related to race and socioeconomic status (SES) (Paschall et al., 2018). Moreover, other reading achievement studies have included these types of demographics (Chung et al., 2022). To prepare the data for matching, cases with missing values for assessment scores were removed. Next, a propensity score, or the probability

Table 1 Comparison of Control and Treatment Groups Before Matching

Variable	Control		Treatment	
	#	%	#	%
Tier 1	2018	22.5	1467	26.4
Tier 2	1827	20.4	1254	22.6
Tier 3	5132	57.2	2827	51
Total	8977	100.1	5548	100
Female	4194	46.7	2762	49.8
Male	4783	53.3	2786	50.2
Total	8977	100	5548	100
Black/African American	1795	20	1101	19.8
Asian	344	3.8	211	3.8
Hispanic	408	4.5	629	11.3
Other	822	9.2	562	10.1
White	5608	62.5	3045	54.9
Total	8977	100	5548	100
Title I Yes	7146	79.6	4088	73.7
Title I No	1831	20.4	1460	26.3
Total	8977	100	5548	100

Table 2 Mean Pretest Scores before Matching

Score	Control		Treatment				
	M	SD	M	SD	t	p	Cohen's d
Overall	1954.702	186.426	1976.509	190.807	5.681	<0.001	0.116
Vocabulary	1859.319	188.373	1866.501	174.355	1.906	0.057	0.039
Comprehension	2029.382	255.925	2065.139	260.629	6.434	<0.001	0.139

of receiving treatment, was estimated using logistic regression. Students from each group were then matched with the "nearest neighbor" in the corresponding group, as determined by similarities in the propensity score. Comparison of pre-matched and matched groups suggest that the matching worked very well, resulting in two similar groups of 450 students each (Table 3). On baseline pretest measures, there was no statistically significant difference between control and intervention groups on overall reading and reading comprehension scores (overall: $t(899)=0.911$, $p=.162$; comprehension: $t(899)=0.556$, $p=.056$). However, there was a statistically significant difference between groups on the vocabulary score ($t(899)=-0.421$, $p=.027$) (Table 4). To adjust for the difference in groups' vocabulary scores without upsetting the balance of the groups, a regression model was fit to remove the baseline difference.

2.5 Procedures

Students in the treatment group completed the minimum recommended reading time: 30 min per week from September to May in the 2016-2017 school year (with exceptions for winter and spring breaks, standardized assessment periods, and beginning and ending weeks). Students in the intervention group accessed the CARP from a

Table 3 Comparison of Control and Treatment Groups After Matching

Variable	Control		Treatment	
	#	%	#	%
Tier 1	160	35.6	160	35.6
Tier 2	108	24	108	24
Tier 3	182	40.4	182	40.4
Total	450	100	450	100
Female	231	51.3	231	51.3
Male	219	48.7	219	48.7
Total	450	100	450	100
Black/African American	64	14.2	64	14.2
Asian	25	5.6	12	2.7
Hispanic	4	0.9	17	3.8
Other	56	12.4	56	12.4
White	301	66.9	301	66.9
Total	450	100	450	100
Title I Yes	324	72	324	72
Title I No	126	28	126	28
Total	450	100	450	100

Table 4 Mean Pretest Scores after Matching

Score	Control		Treatment				
	M	SD	M	SD	t	p	Cohen's d
Overall	2013.264	182.67	2023.978	169.78	0.911	0.162	0.061
Vocabulary	1917.062	194.42	1911.943	169.35	-0.421	0.027	–0.028
Comprehension	2105.172	269.00	2114.821	251.34	0.556	0.056	0.037

computer or tablet on campus as a supplement to their regular language arts (inclusive of reading) instruction.

At the beginning of each month, students were prompted to take a short formative assessment inside the program that provided scores for overall reading, comprehension, and vocabulary. Teachers were afforded real-time data to evaluate students' progress and reading needs. Subsequent supplementary reading sessions included practice of specific reading skills based on students' individualized reading abilities. As students demonstrated mastery of skills, their instructional pathway would include new content and skills to learn through practice for mastery. For a typical fifth grade student with average fifth grade reading abilities, content might include informational text centered in the sciences. Regarding the teachers of the students in the study, professional development for using the CARP, its features, and the assessments were provided at the school or district level. Additionally, teachers had access to the CARP's online training and live webinars offered throughout the school year.

2.6 Measure

The CARP's assessment component, called *Istation's Indicators of Progress Advanced Reading* (ISIP-AR), is a web-delivered and computer-adaptive assessment. The ISIP-AR measures students' (grades 4-8) reading level and progress on essential elements of reading identified by Torgesen et al., 2007. The overall reading score includes four subtests: (a) word analysis, (b) word fluency, (c) vocabulary, and (d) reading comprehension. ISIP-AR provides a separate reading comprehension and vocabulary score. The reading comprehension subtest assesses known comprehension skills including: (a) main idea, (b) cause and effect, (c) inference, and (d) critical judgment. The vocabulary words chosen for the vocabulary subtest included Tier 2 words (general vocabulary) and Tier 3 words (content-specific vocabulary; Matthes, 2016).

Students' abilities are measured based on their performance on items, successfully answered questions result in increasingly more difficult items, while missed questions lead to less challenging items. To measure the CARP's influence at raising reading achievement for fifth graders, gain scores were calculated for the overall reading, reading comprehension, and vocabulary subscale scores. Simple gain scores were calculated by subtracting September scale scores from May scale scores.

2.7 Data analysis procedures

A one-way ANCOVA (analysis of covariance) was used to directly compare groups and measure the effect of treatment, with gain as a dependent variable (DV), treatment as independent variable, and initial score as a covariate. Specifically, ANCOVA is used "to adjust the means of the DV themselves to what they would be if all participants scored equally on the CVs" (covariates; Tabachnick & Fidell, 2019, p. 167). Therefore, the students' baseline scores were included as a CV to improve the precision of the estimates, and, in the case of vocabulary scores, to adjust for differences at baseline, given that there was a statistically significant difference between groups at baseline on this measure. One-way ANCOVA was also used to compare treatment and control groups for each academic level. To understand the magnitude or

the importance of the effect, *Glass's* Δ was selected as the most appropriate measure for effect size because of the differences between standard deviations for each group (Cumming, 2013; Glass, 1976; Grissom & Kim, 2012).

3 Results

The study reports findings of a quasi-experimental design using matching groups to compare reading outcomes for students who used the curricular portion of Istation, a CARP. PSM was used to match for a control and treatment group. The first research question asked: What were the differences, if any, of fifth-grade students' overall reading, vocabulary, and reading comprehension scores when using a CARP as compared to students that do not use the CARP?

Prior to data analysis, preliminary checks were conducted to test the statistical assumptions associated with ANCOVA (e.g., outliers, normality of sample distribution, linearity, homogeneity of variance, and homogeneity of regression). The results of the ANCOVA identified positive and statistically significant differences-in-differences for treatment groups for overall reading, reading comprehension, and vocabulary scores. There was a significant difference in mean gains for the overall reading score, $F(1, 897) = 11.801$, $p < .001$, $\Delta=0.21$, between treatment and control group, adjusting for pretest reading scores (Table 5). For vocabulary scores, there was a significant difference in mean gains, $F(1, 897) = 4.112$, $p = .043$, $\Delta=0.13$, between treatment and control group, adjusting for pretest scores (Table 5). For the reading comprehension scores, there was a significant difference in mean gains, $F(1, 897) = 16.348$, $p <.001$, $\Delta=0.71$, between treatment and control group, adjusting for pretest scores (Table 5). Therefore, the fifth-grade students that used the CARP had greater gains in their overall reading, reading comprehension, and vocabulary scores as compared to the fifth-grade students that did not use the CARP.

To answer research question two, when considering all of the fifth-grade students by Academic Level, what were the differences in the overall reading, vocabulary, and reading comprehension scores between the intervention (CARP) and control group? Comparable groups were determined by academic level through PSM (Table 6). For overall reading scores, results of ANCOVA identified a statistically significant dif-

Table 5 Adjusted Mean Gains by Comparison and Control Group

Measure	Control	Treatment	Difference	p	Effect Size*
Overall Gain	48.082	71.406	23.32	0.001	0.21
Vocabulary Gain	99.025	122.247	23.22	0.043	0.13
Comprehension Gain	46.568	105.182	58.61	<0.001	0.27

*Glass's Δ

Variable	Control		Treatment	
	#	%	#	%
Tier 1				
Male	80	50	80	50
Female	80	50	80	50
Black/African American	11	6.9	11	6.9
Asian	7	4.4	5	3.1
Hispanic	1	0.6	3	1.9
Other	18	11.3	18	11.3
White	123	76.9	123	76.9
Title I Yes	110	68.8	110	68.8
Title I No	50	31.3	50	31.3
Tier 2				
Male	60	55.6	60	55.6
Female	48	44.4	48	44.4
Black/African American	16	14.8	16	14.8
Asian	7	6.5	4	3.7
Hispanic	2	1.9	5	4.6
Other	13	12	13	12
White	70	64.8	70	64.8
Title I Yes	67	62	67	62
Title I No	41	38	41	38
Tier 3				
Male	91	50	91	50
Female	91	50	91	50
Black/African American	37	20.3	37	20.3
Asian	11	6	3	1.6
Hispanic	1	0.5	9	4.9
Other	25	13.7	25	13.7
White	108	59.3	108	59.3
Title I Yes	147	80.8	147	80.8
Title I No	35	19.2	35	19.2

Table 6 Participants' Demographic Data by Academic Level after Matching

Table 7 Adjusted Mean Gains by Academic Level for Comparison and Control Group

Variable	Control	Treatment	Difference	p	Effect Size*
Tier 1					
Overall Gain	46.118	66.238	20.12	0.054	0.21
Vocabulary Gain	151.188	177.11	25.922	0.231	0.13
Comprehension Gain	41.752	140.226	98.474	<0.001	0.44
Tier 2					
Overall Gain	48.946	65.818	16.872	<0.001	0.16
Vocabulary Gain	97.17	104.911	7.741	0.642	0.06
Comprehension Gain	49.481	81.889	32.408	0.254	0.15
Tier 3					
Overall Gain	49.679	78.882	29.203	0.014	0.23
Vocabulary Gain	55.166	83.405	28.239	0.105	0.16
Comprehension Gain	46.753	90.516	43.763	0.046	0.20
*Glass's Δ					

ference between control and treatment for students at all academic levels including those in Academic Level 3 (students in most need of reading support), $F (1, 361) = 6.135$, $p = .014$, $\Delta=0.23$. For reading comprehension, there was a statistically significant difference between control and treatment groups for students at Levels One $F (1, 317) = 15.501$, $p <.001$, $\Delta=0.44$ and Level Three $F (1, 361) = 4.024$, $p = .046$, $\Delta=0.20$. For vocabulary scores, differences were not statistically significant for students at any of the three Academic Levels (Table 7).

Comparing mean gains adjusted for the effect of the covariate identified that overall reading and vocabulary score gains were highest for students at Academic Level 3 and reading comprehension gains were highest for students in Academic Level 1. Therefore, when comparing the achievement as mean gains between the control and treatment matched groups (those who used the program for the recommended 30 min), the achievement was greater for students in the treatment group. Likewise, the Academic Level of the students by achievement and subtest scores provided evidence that the Academic Level of the students made a difference in students overall gains.

4 Discussion

Effective reading instruction entails a focus on fluency, comprehension and vocabulary for upper elementary education. Students in upper-elementary grades may have unique, differentiated reading needs. Previous studies indicated that computer adap-

tive reading may support personalized instruction based on students' reading needs. The current investigation was conducted to examine the efficacy of a supplementary, CARP in promoting fifth-grade students' reading achievement. The findings identified the use of the CARP for fifth-grade students as a means to promote their overall reading, reading comprehension, and vocabulary scores. Statistically significant differences were identified between fifth-grade students in the control and treatment groups for their overall reading, reading comprehension, and vocabulary score. Further analysis indicated that the fifth-grade students in Academic Level 3 (lowest quartile) using the CARP showed the greatest gains in their overall reading scores compared with students in Academic Level 1 (upper quartile) and Academic Level 2 (middle quartile). In addition, all fifth-grade students in the treatment group at all three Academic Levels demonstrated increases in their mean scores, suggesting that the Istation CARP can benefit reading achievement when used consistently (30 min per week).

Implications for schools include allocating the recommended amount of time in schools learning schedules to allow for consistent access to and use of the CARP as time can be a predictor of academic achievement (Fisher et al., 2015). With compressed instructional time due to non-instructional activities, the importance of organizing the school week to meet the recommended weekly usage may make a difference in students' reading achievement as was found in this study (Smith, 2000). Moreover, schools can support teachers by providing the necessary resources (e.g., computer access for the whole class at the same time) to optimize scheduling students' use of the CARP.

Because the program is supplemental to the primary reading curriculum, how teachers integrate the program into their reading instruction may be considered key to program effectiveness. Teachers may need time and support to make sense of how to use the data generated by the program to inform daily instructional practices. Professional development opportunities could include: (a) developing shared understanding of the purposes and guidelines of the program and how it relates to instructional goals and teaching practices and (b) supporting teachers in making sense of data generated from formative assessment to target reading skill deficits (Wayman, 2005).

Statistically significant differences in overall reading mean score gains by students' Academic Level for the treatment group supports that the CARP can promote all learners' reading scores, especially those in need fifth-grade students at Level 3. The effect size for overall gain while categorized small by Cohen's (1998) standard for 0.13–0.27, the What Works Clearinghouse (2017) considers findings as "substantively important" if the effect size is 0.25 standard deviations or greater" (p. 77). In this case, the overall effect size indicates the effect of the treatment (30 min a week for 30 weeks) would be a gain of 6-9 percentile points (Marzano, 2010) in comparison to those who did not have the treatment and have teacher instruction alone. Conversely, even though the fifth-grade students showed mean gains on their vocabulary scores, the results were not statistically significant; however, the effect size indicated 0.06–0.13, a small effect size, that translates to a 2-5 percentile increase. For the fifth-grade students at Academic Level 3, the effect sizes for their reading gains were the largest. The fifth-grade students' reading comprehension scores indicated that the

effect of the CARP was small to moderate, with percentile gains range from 6 to 17 percentiles.

The fifth-grade students in Academic Level 2 (> 20th and < 40th percentile) evidenced the least amount of growth in their overall reading, reading comprehension, and vocabulary scores. Potential reasons for students in Academic Level 2 not scoring higher on the assessment may range from instructional attention to time engaged in practicing skills. Instructional attention can be greater for students most in need of reading support (Academic Level 3) or for those who evidence sustainable growth and/or may need reading challenges (Academic Level 1). Students in the middle (Academic Level 2) may not express the same needs. More research is needed to understand the factors that may have contributed to the phenomenon observed in this study and how that may relate to the documented challenges of upper-elementary grades readers (Chall & Jacobs, 1983; Torgensen et al., 2007).

4.1 Limitations of the investigation

While PSM procedures were designed to minimize selection bias and confounding variables, it is possible that differences at the district, school, and classroom level also influenced student's reading score outcomes, suggesting an important avenue for future analysis. For example, districts may employ different policies and policy levers to promote the CARP usage, which are, in turn, implemented differently at the school-administrator level. Perhaps most importantly, how teachers "make sense" of district and school policies influences how they use CARP to drive classroom instruction (Spillane, Reiser, & Reimer, 2002). A factor unexplored in this study includes how teachers use the formative feedback generated by the CARP to inform their instruction. Previous research has found that interventions that change daily teaching practices are more effective at raising reading achievement in upper-elementary grades (Slavin et al., 2009). Investigating how CARPs alter teachers' practices may further explain findings and identify implications for practice.

5 Conclusions

The present study examined a CARP (inclusive of curriculum component and curriculum-based measure) to determine the efficacy of the use of the supplementary curriculum component for the recommended time. PSM procedures were utilized to create a comparable treatment and control group. The findings indicated that those fifth-grade students who used the Istation supplementary computer adaptive reading curriculum for a minimum of 30 min each week (30 weeks) during the school year evidenced greater gains on the ISIP-AR than those students who did not use the curriculum. While the present study provided evidence that the CARP can be used to improve achievement outcomes for fifth graders, future quantitative and qualitative studies could investigate how district, school, and teacher-level factors influence the implementation of a supplementary CARP to increase reading proficiency levels for all students.

References

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, & Practice*, 18(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, 72(1), 101–130. https://doi.org/10.3102/00346543072001101

Campbell, L. O., Sutter, C. C., & Lambie, W. G. (2019). An investigation of the summer learning effect on fourth grade students' reading scores. *Reading Psychology*, 40, 465–490. https://doi.org/10.1080/02702711.2019.1629516

Chall, J., & Jacobs, V. (1983). Writing and reading in the elementary grades: Developmental trends among low SES children. *Language Arts*, 60(5), 617–626

Cheung, A. C., & Slavin, R. E. (2013). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly*, 48(3), 277–299.

Chung, G., Kainz, K., Eisensmith, S. R., & Lanier, P. (2022). Effects of Youth Educational Aspirations on Academic Outcomes and Racial Differences: A Propensity Score Matching Approach.Journal of Child and Family Studies,1–14

Cohen, J. (1998). *Statistical power analyses for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Inc

Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute of Education Sciences Research Centers*. NCSER 2014-3000. National Center for Special Education Research. https://ies.ed.gov/ncser/pubs/20143000/pdf/20143000.pdf

Connor, C. M., & Morrison, F. J. (2016). Individualizing student instruction in reading: Implications for policy and practice. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 54–61. https://doi.org/10.1177/2372732215624931

Council, M. R., Cartledge, G., Green, D., Barber, M., & Gardner, R. (2016). Reducing risk through a supplementary reading intervention: A case study of first- and second-grade urban students. *Behavioral Disorders*, 41(4), 241–257. https://doi.org/10.17988/bedi-41-04-241-257.1

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 75–88. https://doi.org/10.2307/jeductechsoci.18.3.75. https://www.jstor.org/stable/

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (2015). Teaching behaviors, academic learning time, and student achievement: An overview. *Journal of Classroom Interaction*, 50(1), 6–24

Flanigan, R. L. (2014). Adaptive testing shaping instruction.Education Week.https://www.edweek.org/ew/articles/2014/10/22/09pl-adaptive.h34.html

Fuchs, L. S., & Fuchs, D. (2001). What is scientifically-based research on progress monitoring? *National Center on Student Progress Monitoring*. https://files.eric.ed.gov/fulltext/ED502460.pdf

Gibson Jr, L., Cartledge, G., Keyes, S. E., & Yawn, C. D. (2014). The effects of a supplementary computerized fluency intervention on the generalization of the oral reading fluency and comprehension of first-grade students. *Education and Treatment of Children*, 37(1), 25–51

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8

Goodman, K. S., & Goodman, Y. M. (2014). Helping readers make sense of print: Research that supports a whole language pedagogy. In S. E. Israel, & G. G. Duffy (Eds.), *Handbook of research on reading comprehension*. Routledge

Graham, S. E., & Kurlaender, M. (2011). Using propensity scores in educational research: General principles and practical applications. *The Journal of Educational Research*, 104(5), 340–353. https://doi.org/10.1080/00220671.2010.486082

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate approaches*. Routledge. https://doi.org/10.4324/9780203803233

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and Learning* (pp. 87–102). SAGE Publications Ltd. https://doi.org/10.4135/9781446250808.n6

Hirsch, E. D. (2003). Reading comprehension requires knowledge—of words and the world. *American Educator*, 27(1), 10–13

Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15, 199–236

Hong, G. Y., & Masood, M. (2014). Effects of gamification on lower secondary school students' motivation and engagement. *International Journal of Social, Education, Economics and Management Engineering*, 8(12), 3483–3490

Jacob, R. T., Armstrong, C., & Willard, J. A. (2015). *Mobilizing volunteer tutors to improve student literacy: Implementation, impacts, and costs of the Reading Partners program*. New York: MDRC. http://files.eric.ed.gov/fulltext/ED558508.pdf

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education*, 128, 427–451

Kent, S. C., Wanzek, J., & Al Otaiba, S. (2017). Reading instruction for fourth-grade struggling readers and the relation to student outcomes. *Reading & Writing Quarterly*, 33(5), 395–411. https://doi.org/10.1080/10573569.2016.1216342

Kamil, M. L., & Chou, H. K. (2009). Comprehension and computer technology: Past results, current knowledge, and future promises. In S. E. Israel, & G. G. Duffy (Eds.), *Handbook of Research on Reading Comprehension*. Routledge. https://doi.org/10.4324/9781315759609-25

Kim, J. S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a mixed-method literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program? Results from a randomized controlled trial of READ 180 enterprise. *Educational Evaluation and Policy Analysis*, 33(2), 183–201. https://doi.org/10.3102/0162373711399148

Marzano, R. J. (2010). *Designing & teaching learning goals & objectives*. Solution Tree Press.

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, 162–172. https://doi.org/10.1111/j.1467-9817.2006.00282.x

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*, 47(5), 315–335. https://doi.org/10.1016/j.jsp.2009.04.003

Matthes, P. (2016). *Istation's Indicators of Progress (ISIP) advanced reading technical report: Computer adaptive testing system for continuous progress monitoring of reading growth for students grade 4 through grade 8*. Istation: Dallas, Tx. https://www.istation.com/Content/downloads/studies/ar_technical_report.pdf

McGlohen, M., & Chang, H. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods*, 40(3), 808–821. https://doi.org/10.3758/BRM.40.3.808

Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. *Journal of Computer Assisted Learning*, 23(1), 27–35. https://doi.org/10.1111/j.1365-2729.2007.00196.x

Murray, T., & Arroyo, I. (2002, June). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In *International Conference on Intelligent Tutoring Systems* (pp. 749–758). Springer, Berlin, Heidelberg

National Assessment of Educational Progress. (2019). *NAEP Report Card Reading*. National Center for Education Statistics. https://www.nationsreportcard.gov/reading?grade=4

National Reading Panel (US). (2000). *National Reading Panel: Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. National Institute of Child Health and Human Development, National Institutes of Health

Navarro, J. J., & Mourgues-Codern, C. V. (2018). Dynamic assessment and computerized adaptive tests in reading processes. *Journal of Cognitive Education and Psychology*, 17(1), 70–96

Paschall, K. W., Gershoff, E. T., & Kuhfeld, M. (2018). A two decade examination of historical race/ethnicity disparities in academic achievement by poverty status. *Journal of Youth and Adolescence*, 47(6), 1164–1177

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, 48(2), 462–501

Rosenbaum, R. (2009). *Design of observational studies*. Springer Series in Statistics. Springer

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of propensity score in observation studies for causal effects. *Biometrika*, 70(1), 41–55

SEG Measurement (2018). *An evaluation of the effectiveness of Imagine Learning for improving reading skills*. Retrieved from https://cdn-websites.imaginelearning.com/corporate/sites/default/files/2019-01/Imagine%20Learning%202017-2018%20Texas%20Effectiveness%20Study%20Report%20Final.pdf

Slavin, R. E., Lake, C., Cheung, A., & Davis, S. (2009). Beyond the basics: Effective reading programs for the upper elementary grades. *Center for Data-Driven Reform in Education*

Smith, B. (2000). Quantity Matters: Annual Instructional Time in an Urban School System. *Educational Administration Quarterly*, 36(5), 652. https://doi.org.ezproxy.net.ucf.edu/10.1177/00131610021969155

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431. https://doi.org/10.3102/00346543072003387

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory-based instruction: Toward a theory of teaching. *Educational Researcher*, 41(5), 147–156. https://doi.org/10.3102/0013189X12442801

Stockard, J. (2010). Promoting reading achievement and countering the "fourth-grade slump": The impact of direct instruction on reading achievement in fifth grade. *Journal of Education for Students Placed at Risk*, 15(3), 218–240. https://doi.org/10.1080/10824669.2010.495687

Tomlinson, C. (2014). *The differentiated classroom: Responding to the unique needs of all learners* (2nd ed.). Association for Supervision and Curriculum Development

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics.* Pearson.

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., et al. (2007). Academic literacy instruction for adolescents: A guidance document from the Center on Instruction. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Tyler, E. J., Hughes, J. C., Beverley, M., & Hastings, R. P. (2015). Improving early reading skills for beginning readers using an online program as supplementary instruction. *European Journal of Psychology of Education*, 30, 281–294. https://doi.org/10.1007/s10212-014-0240-7

U.S. Department of Education (USDOE) (2017). *What works clearinghouse: Standards handbook. Version 4.0.* https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Vygotsky, L. (1978). *Interaction between learning and development. Readings on the development of children*, 23(3), 34–41

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: a synthesis of 20 years of research. *Reading and Writing*, 23(8), 889–912. https://doi.org/10.1007/s11145-009-9179-5

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk*, 10(3), 295–308. https://doi.org/10.1207/s15327671espr1003_5

Wise, B. W., Ring, J., & Olson, R. K. (2000). Individual differences in gains from computer-assisted remedial reading. *Journal of Experimental Child Psychology*, 77(3), 197–235. https://doi.org/10.1006/jecp.1999.2559

Authors and Affiliations

Laurie O. Campbell[1] · Cassandra Howard[2] · Glenn W. Lambie[3] · Xueying Gao[4]

✉ Laurie O. Campbell
  locampbell@ucf.edu

  Cassandra Howard
  cassandra.howard@ucf.edu

  Glenn W. Lambie
  glenn.lambie@ucf.edu

  Xueying Gao
  Xueying.gao@ucf.edu

[1] Learning Sciences and Educational Research, University of Central Florida, Orlando, United States

[2] Educational Leadership and Policy Studies, Florida State University, Tallahassee, United States

[3] College of Community Innovation & Education, University of Central Florida, Orlando, Fl, United States

[4] Educational Methodology, Measurement, and Analysis, University of Central Florida, Orlando, United States