

Reading Achievement and Reading Efficacy Changes for Middle School Students With Disabilities Through Blended Learning Instruction

Jesse R. Pace¹ and Daryl F. Mellard¹

Abstract

This study evaluated the effects of a blended learning instructional experience for sixth-grade students in an English/language arts (ELA) course. Students at two treatment schools participated in a blended learning instructional paradigm, and their ELA test scores were compared to one comparison school that used a face-to-face delivery. Other variables of interest were gender status, disability status, and student reading efficacy. The results of the analysis indicated that no significant changes in reading achievement were found that could be attributed solely to treatment versus comparison, to gender, or to disability status. Perhaps of greater significance to practitioners and researchers is the identification of person and programmatic-level factors that influence adoption and implementation of effective blended instruction. Implications are discussed.

Keywords

blended learning, online/web-based instruction, technology perspectives, reading, content/curriculum area, middle school, age/grade level, students with disabilities

Online education is growing rapidly: Between 2002 and 2011, the number of K-12 students enrolled in either partial or fully online schools increased from 220,000 to 1.8 million (Watson, Murin, Vashaw, Gemin, & Rapp, 2012). A review of the research literature did not identify the total number of students with disabilities enrolled in some form of online learning, but research in the state of Ohio indicates that students with disabilities may be overrepresented in online learning (Wang & Decker, 2014). Research is needed to determine the impact that online school programs have on the learning, achievement, and long-term outcomes of students with disabilities.

Online learning consists of two broad categories: blended and fully online use of computer instruction. Multiple definitions of “blended learning” exist. In the present study, the term follows Staker and Horn’s (2012) definition:

A formal education program in which a student learns at least in part through online delivery of content and instruction with some element of student control over time, place, path, and/or pace and at least in part at a supervised brick-and-mortar location away from home. (p. 3)

Staker and Horn differentiate between four models of blended learning: flex, self-blended, enriched virtual, and rotation. In a flex model, learning is customized to student needs and students move to different modalities as their individual needs

require. In a self-blended model, students take online courses which supplement their existing traditional schooling. In an enriched-virtual model, students divide their time between the brick-and-mortar school and learning remotely. In a rotation model of blended learning, students rotate learning modalities throughout the week or day. Four implementations of the rotation model are practiced: station rotation, lab rotation, flipped classroom, and individual rotation. Station rotation involves students moving from one station to the next in the same classroom to learn different subjects. Flipped classrooms involve students viewing lectures remotely, then coming to school to practice and do work. Individual rotation is similar to station rotation except that individual students are rotated to specific stations based on the student’s learning needs, not all students necessarily rotate to every station. Finally, lab rotation consists of students moving to different locations on campus to learn a subject (or subjects) predominantly online. Lab rotation

¹ The Center on Online Learning and Students with Disabilities, The University of Kansas, Lawrence, KS, USA

Corresponding Author:

Jesse R. Pace, The Center on Online Learning and Students with Disabilities, The University of Kansas, Lawrence, KS, USA.
 Email: jesse.pace@ku.edu

describes the implementation of blended learning studied in the present article.

Research Support for Blended Learning

For the general student population, blended learning may be a more effective learning environment than the traditional brick-and-mortar school. In a meta-analysis of 45 studies on blended learning, Means, Toyama, Murphy, and Baki (2013) reported that blended learning tends to be more effective than traditional face-to-face learning, and that fully online learning's effectiveness is equivalent to face-to-face instruction. Seven of the studies included focused on K-12 learners. The meta-analysis predominantly sampled students in the general population: It included only one study on students with disabilities. The one study on students with disabilities (Englert, Zhao, Dunsmore, Collings, & Wolbers, 2007) did, however, demonstrate support for the effectiveness of blended learning. Englert, Zhao, Dunsmore, Collings, and Wolbers (2007) found that a web-based instructional program produced superior improvements in writing achievement for students with disabilities compared to instruction provided using a paper-and-pencil modality.

In addition to considerations of effectiveness, the reasons that many students with disabilities enroll in online schools, including blended learning, are indicative of other potential benefits. Work by Rhim and Kowal (2008) indicates that, for students with disabilities, online instruction offers the potential for individualized instruction and appeals to parents seeking ways to optimize their child's learning (Rhim & Kowal, 2008). Burdette, Greer, and Woods (2013) interviewed state special education (SPED) directors regarding why districts are moving to more blended and fully online instruction. In discussing parent motivation, the state directors indicated that online learning holds potential for more flexibility and alternatives to traditional scheduling and instructional methods.

Disagreements and Concerns About Online Learning's Efficacy

While enrollments in online learning continue to increase, some evaluation results in national- or state-specific studies have not been positive. The Center for Research on Education Outcomes (CREDO; Woodworth et al., 2015) conducted a study in 18 states to explore the outcomes of fully online learning in charter school settings. In this study, a public school operated as a charter school (as defined by the state) that used online learning as its primary means of curriculum delivery. Woodworth et al. concluded that fully online schools overwhelmingly produced weaker achievement for students with disabilities when compared to traditional (i.e., brick-and-mortar) schools.

In Michigan, online enrollments have increased significantly among high school grade levels (Friedhoff, 2015)—the course completion rates have not. The percentage of online enrollments with a “completed/passed” outcome was 57% in 2013–2014, down 3% from the previous year. In contrast, the

same learners had completed/passed rates of 71% in their face-to-face courses. The students who did not take courses online had an 89% completed/passed rate. Thus, the opportunities afforded by online instruction have not yielded correspondingly improved outcomes.

The conclusions from a recent qualitative study conflict with the above findings in terms of online learning's efficacy for students with disabilities. Franklin, Rice, East, and Mellard (2015) interviewed five administrators of blended learning programs regarding the enrollments, persistence, progress, and achievements of students with disabilities. These program administrators indicated that, in the blended programs they oversaw, students with disabilities were outperforming their peers without disabilities in terms of growth in academic achievement. This finding is surprising given that CREDO's research on fully online charter schools showed that online programs produce weaker achievement outcomes for students with disabilities than do traditional schools and that research in traditional schools indicates that students with disabilities tend to have lower achievement levels than students without disabilities (e.g., Cortiella & Horowitz, 2014; Wagner, Cameto, & Levine, 2006). Further, research shows that the gap between students with disabilities and those students without disabilities tends to grow larger as children move into higher grades (Klein, Wiley, & Thurlow, 2006), which implies that students with disabilities' academic growth rate is slower than that of their peers without disabilities. The claim, then, that students with disabilities' achievement in blended learning is growing faster than their peers without disabilities runs contrary to what would be expected. Thus, these claims require further investigation.

The performance gap between students with and without disabilities has been most prominent in reading achievement (Wagner et al., 2006). Because reading is the area in which students with disabilities have historically demonstrated the most difficulty, this study focused on students' reading achievement growth. In order to account for important contributors to reading achievement, the design also incorporated two other variables: students' self-efficacy rating and gender status.

Variables Important to Academic Achievement

Self-efficacy, as defined by Bandura (1986, p. 391), is “people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances.” Wigfield and Guthrie (1997) concluded that reading self-efficacy was one of the strongest predictors of academic achievement. The authors also found that female students were generally more efficacious (i.e., judging themselves as more capable on reading tasks) than were male students. Because reading efficacy predicts reading achievement, a reasonable hypothesis is that males' lower reading efficacy would result in lower reading achievement. The available literature generally confirms this hypothesis. Lietz (2006) conducted a meta-analysis of 139 studies on gender differences in reading achievement at the secondary school level. Lietz concluded that female students consistently outperformed their male peers

on measures of reading achievement. This present study investigated the impact of gender on changes in reading achievement over time for both SPED and general education students in a blended learning environment. The study also examined the relationship of reading efficacy with student achievement within a blended learning curriculum.

Based on the research cited above, this study investigated the relationship among disability status, gender status, and self-efficacy in regard to reading achievement in blended learning. The research hypotheses of this study predicted finding significant group differences between categories of gender and disability status, as previous research had found these differences in traditional (e.g., nonblended) schools. Researchers also predicted the continued significance of self-efficacy as it relates to academic achievement in blended learning. Research questions included:

Research Question 1: Does the use of a supplemental blended learning curriculum lead to different student growth (reading test scores changes over time) as compared to a traditional (i.e., nonblended) classroom curriculum?

Research Question 2: Does the amount of exposure (i.e., dosage) of treatment lead to different levels of change in student reading achievement?

Research Question 3: Do students in SPED have different trends of reading growth than general education students?

Research Question 4: Are there differences in student reading growth depending on student gender?

Research Question 5: Does student reading efficacy continue to correlate with student performance in a blended learning environment?

Method

In this quasi-experimental design study, the growth of sixth-grade student English/language arts (ELA) test scores in two blended learning schools was compared to the growth of sixth-grade ELA test scores of students in one traditional school. Because these schools were not selected at random, selection effects must be considered in the analyses and findings. This study looked at growth over the school year while using baseline academic ability as a covariate rather than looking at only mean differences among schools on a single outcome time. Using the data in this design, some selection effects can be accounted for, as student growth is not confounded by previous achievement.

Selection Procedures

The school district initially identified four middle schools for this study: two comparison (i.e., face-to-face) schools and two treatment (i.e., blended) schools. Schools were identified based on many factors, including geographic proximity to each other, level of technology implementation, building-level administrator support, and demographic makeup. After implementation of

the study began, the building administrator of one comparison school chose not to participate and withdrew from the study. Due to district and project administrative concerns, including a replacement school was not feasible. Thus, comparisons were made between the three remaining schools only.

District and School Demographics

The school district from which the samples were drawn was located in a suburban/rural area. The district neighbored a large metropolitan area in the Southeastern United States. The county's population was more than 200,000 in 2013. In 2015, the school district enrolled 41,000 students at 50 attendance centers: 28 elementary schools, 11 middle schools, and 11 high schools. Fifty-three percentage of students in prekindergarten through 12th-grade enrolled in 2015 were eligible for free or reduced lunch (FRL).

Considerable differences were noted in the ethnic/racial makeup of the three schools, particularly between the two treatment schools (Blended English Language Arts [BELA1] and BELA2) and the single comparison school (teaching English language arts [TELA]). Enrollment at BELA1 had a considerably larger White population and a proportionally smaller Black population compared to BELA2 and TELA (see Table 1 for further details about the sample). Rates of FRL status varied considerably among the two treatment schools and the comparison school: BELA1's FRL rate was 48.78%, BELA2's was 49.49%, and TELA's was 78.07%.

Design of the Intervention

This intervention had two components: general education classroom instruction and online instruction. The blended learning course was designed to teach students to read critically, analyze text, and cite evidence in order to support ideas. The course also sought to improve vocabulary, listening skills, and grammar through explicit modeling and practice. Students also engaged in routine response writing activities based on the readings and more extensive essay writing.

The BELA program is a package of supplemental curricular materials for ELA in a blended classroom environment. This package of materials is commercially available and was licensed by the district from BELA, Inc. The students' specific outcomes, as stated in the BELA program overview, were being able to read complex texts at grade level; understanding and being able to analyze the structure and elements of literature from various genres; increased academic and domain-specific vocabulary; being able to use text evidence to analyze, infer, and synthesize ideas; engaging in routine writing in response to texts read and analyzed; using the writing process to complete a variety of essay writing assignments; using research skills to access, interpret, and apply information from several sources; gaining the tools for speaking and listening in discussions and presentations; and learning a variety of real-world and digital communication skills.

Table 1. Students' Demographics in Final Analysis Sample.

School	Female, %	Asian, %	Black, %	Hispanic, %	Multiethnic, %	White, %
BELA1 (<i>n</i> = 186)	56.5	1.6	21.5	7.0	4.8	65.1
BELA2 (<i>n</i> = 169)	51.5	8.3	55.6	11.2	5.9	18.9*
TELA (<i>n</i> = 140)	44.3	3.6	59.3*	12.9*	5.7	18.6*
Total (<i>n</i> = 495)	51.3	4.4	43.8	10.1	5.5	36.2

Note. Values exclude any student with less than 10% of the percentage by count variable. Values are calculated using listwise deletion. Student demographics percentages were within 2% of state Department of Education (DOE) reported demographics for the 2014–2015 year, with the exception of those marked with an asterisk. All were within 3.5%. DOE reported the White percentage at BELA2 as 22.0%. For TELA, Black: 61.6%, Hispanic: 10.13%, and White: 21.1%. BELA = blended English language arts; TELA = teaching English language arts.

The BELA program is designed to complement the physical classroom curriculum. In this study, students in the two treatment classrooms spent several 50- to 70-min periods per week throughout the school year working on BELA curricular materials in a computer lab. The BELA computer lab sessions supplemented daily (or near daily) face-to-face classroom instruction: instruction in a physical classroom with a certified ELA teacher. The study was conducted during the second year of these two schools' implementation, 2013–14 being their pilot year.

Students were informed that the blended course would require the same amount of effort as courses taught in the traditional classroom. In the blended course, students were required to participate via interactive lessons, which included direct instruction and modeling of skills in developing reading comprehension. These computer–student interactive lessons also included guided and independent reading activities. The online component incorporated a range of assignments, such as students answering comprehension questions and completing on-screen grammar exercises, short writing, and extended writing. Formative assessments included quizzes, tests, and exams, each incorporating more items and more comprehensive content reviews. The lessons were designed using best design practices in multimedia instruction to reduce cognitive load and help students learn more effectively through the use of a variety of features, such as audio narration, using two modalities for complex content, avoiding splitting attention, and breaking things into parts. The lessons also incorporated universal design for learning (UDL; Center for Applied Technology, 2011) principles. UDL principles included in the instructional design were multiple means of representation (e.g., video lectures, graphic displays, simulations, closed captioning, and text to speech), multiple means of action and expression (e.g., discussion forms, multimedia composition software, virtual manipulatives, and graphing calculators), and multiple means of engagement (e.g., self-pacing, pause and rewind, features to highlight/markup text, and tools to take notes electronically). Teachers were expected to interact with students via digital discussion, e-mail, chat, and system announcements, and students were expected to interact digitally with one another. Students were assigned 12 total units throughout the school year, grouped by quarterly themes: identity, perseverance, heroism, and community.

Students at TELA, the comparison school, did not participate in blended learning and received all of their ELA

instruction in a face-to-face classroom. TELA used the same student goals and outcomes and used the same district-adopted curriculum as the treatment schools. All teachers had flexibility in choosing supplemental materials to meet their students' needs.

Implementation of Intervention

BELA1 implemented the intervention as semester-length courses in which students spent 70 min of every school day (5 days a week) in the computer lab working on BELA curricular activities. These lab sessions were monitored by two paraprofessionals and the computer lab had no more than 70 students in it at one time. In addition to the BELA curriculum, students spent 50 min each school day in a traditional classroom receiving instruction from a certified ELA teacher.

BELA2 implemented the intervention as a series of 4- to 6-week courses in which students spent 2 days a week, for 50-min periods, in the computer lab engaged with the BELA curricular activities. The computer lab was monitored by a certified ELA teacher and the lab had no more than 35 students in it at one time. Students also spent 2 days a week, for 50-min periods, in a traditional classroom receiving instruction from a certified ELA teacher.

Students at TELA did not engage with the BELA curricular activities. Students attended 50-min ELA instruction 5 days a week, in which they were instructed face-to-face by a certified ELA teacher.

The above information and additional implementation information are presented in Table 2. Several important distinctions existed between implementation at the two BELA schools. Three key differences were: At BELA1, the BELA program was delivered as a semester-long course while at BELA2, the same content was organized in shorter topical units of 4–6 weeks in duration; the amount of time students spent at computer labs differed considerably, with BELA1's students spending 350 min a week on BELA curricular activities, compared to BELA2's total of 100 min per week; and lab sizes were much larger at BELA1 compared to BELA2—70 versus 35 students. It is worth noting, however, that BELA1's students were permitted to work on any BELA curricular material during lab time, not only ELA material, whereas BELA2's students spent the entire 100 min each week on BELA ELA material.

Table 2. ELA and BELA Instructional Opportunities.

Middle School	BELA Course Design	Computer Lab Capacity	Staff	Schedule	Professional Development Opportunities
BELA1 (treatment setting)	Semester course with lab rotation	70	Three teacher ELA team and two paraprofessionals. Three ELA teachers teach face-to-face for 50 min. Each paraprofessional monitors a computer lab for a 70-min period.	Monday–Friday students attend two 50-min ELA classroom periods. Everyday students also attend a 70-min lab period in which they can work on any of their BELA courses.	BELA training (1 day): two of three teachers attended. BELA coaching (2 days): two of three teachers attended. NWEA MAP training (2 days): two of two paraprofessionals attended.
BELA2 (treatment setting)	4–6 Week courses with lab rotation	35	Three teacher ELA team. ELA teachers teach face-to-face in a classroom for 50 min. A third certified ELA teacher monitors the BELA lab during a 50-min period.	Monday–Thursday students have two 50-min ELA classroom periods and two 50-min ELA BELA lab periods. On Friday, students attend a remedial English workshop period.	BELA training (1 day): one of three teachers attended. BELA coaching (2 days): All three teachers attended. NWEA MAP training (2 days): All three teachers attended.
TELA (comparison setting)	Semester course; face-to-face instruction	NA	Three teacher ELA team. Two teachers coteach. The third teacher teaches alone. All teachers offer face-to-face instruction in a classroom setting.	Monday–Friday students attend five 50-min ELA classroom periods.	NWEA MAP training (2 days): All three teachers attended.

Note. ELA = English language arts; BELA = blended English language arts; NWEA = Northwest Evaluation Association; MAP = measure of academic progress; TELA = teaching English language arts.

Differences between BELA1 and BELA2 also existed in professional development and instructional coaching of staff. Teachers and staff at all three schools were offered professional development on the use of the Northwest Evaluation Association (NWEA) measure of academic progress (MAP) assessments (NWEA, 2003). The professional development activities were scheduled and conducted by NWEA and by BELA for the use of their respective contributions to the treatment condition. Teacher participation in these professional development activities varied. Teachers at the two treatment schools received differential amounts of professional development in the use of BELA. At BELA2, teachers did not participate in the initial orientation sessions and implementation session. The computer lab instructor at BELA2, however, did participate in the professional development. At BELA1, two of the three teachers participated in the professional development sessions, but the computer lab supervisors did not. These differences in implementation may have led to a considerable disparity in students' exposure to and opportunities to learn the ELA curricular materials.

Student Participants

School district staff provided demographic information for 769 students across the three schools. From the total student sample, both pre- and posttest MAP scores were available for 497

students. Two of those 497 students were found to have completed less than 10% of their BELA curricular activities and were excluded from analysis. Of the remaining 495 students, 355 students were enrolled in the treatment schools and 140 were enrolled in the comparison school. See Table 1 for a breakdown of student demographics.

Ten and one half of a percentage (82 students) of the total student sample were designated as having an individual education program for SPED services. For the respective schools, BELA1's percentage of students in SPED was 6.0%; BELA2, 10.3%; and TELA, 16.2%. Forty-four SPED students were included in the analysis sample after listwise deletion. For the respective schools, the percentage of students in SPED with complete data was 2.7% at BELA1, 11.2% at BELA2, and 14.3% at TELA. The students' specific disability categories were not available to researchers.

Classroom Environment

One researcher and two BELA staff members recorded observations in the ELA classrooms among the three schools in order to describe the classrooms. Teachers at all three schools appeared to spend the majority of their time among their students (i.e., moving among the rows or work groups) or at the front of the classroom. The instructional grouping—how students were configured during instructional time—was mostly

whole-group instruction, seconded by one-on-one instruction. The students did some work in small group configurations of 2–4 students. Students worked mostly on worksheets or read from reading materials (e.g., textbooks or novels). The overall impression of the observers was that classroom management was good and that students appeared to be engaged and on task.

Teacher behavior was typically divided between three key areas: directing the students (e.g., telling students which book to use), attending to the students as they engaged in activities (e.g., monitoring students as they read silently), and communicating academic content.

Noticeable differences were observed in the computer labs between BELA1 and BELA2; lab configurations, dynamics, and number of students varied between the settings. Compared to BELA2, BELA1 had substantially higher numbers of students. BELA1 also had two different lab settings, while BELA2 had only one, and BELA1 had two staff members in the role of instructional aides. These instructional aides focused on maintaining classroom order, providing technical assistance, and addressing some content questions. At BELA2, a certified teacher, as opposed to an instructional aide, provided closer supervision and monitoring of student activities. She appeared to have fewer classroom management issues possibly due to the reduced class size.

Measures

Measures of academic progress. The efficacy of the intervention was assessed using the reading section of the NWEA MAP. The MAP reading test is a common core-aligned, computer-adaptive assessment administered to students in Grades 3–12. The MAP is adaptive in the sense that subsequent question difficulty is based on student performance on preceding items. Each MAP assessment uses the Rasch unit, an equal interval scale score, to measure student growth and determine student mastery of various defined skills within disciplines. MAP scores have no set lower or upper boundaries, although scores are typically between 150 and 300 (NWEA, 2003). Marginal reliabilities for the fall and spring MAP reading test for sixth-grade students were .94 in the validation study. Because MAP scores have norms for both fall and spring, a student may maintain the same scaled score throughout the year, yet decline in their normative percentile rank. Because of this effect, analysis in the present study was conducted using percentile ranks to better illustrate trends in normative student performance.

The MAP was administered to students at each school 3 times throughout the school year: September, January, and May. From the total student sample, as specified above, 495 students completed all three assessments.

BELA average percentage of activities completed (i.e., dosage). BELA's web administrator provided the students' task completion data. Each completed assignment on any computer-based ELA-related class activity was aggregated across the school year. This variable was used as a measure of treatment dosage,

representing how much exposure to the BELA curricular activities a given student received.

BELA average overall grade. Students were graded on several curricular activities each semester within the BELA program. The average grade of all such activities across the entire school year was used in the analysis. The average grade does not reflect student performance with their ELA classroom activities outside of the online platform. These data were not available for the comparison school as the measure is specific to online activities.

Student reading efficacy. A short survey was used to measure students' reading efficacy, and this measure was given twice, first in January and second in May. The survey used four questions from Wigfield and Guthrie's (1997) study on motivation in reading. Nine items were originally selected from the Wigfield and Guthrie Motivation for Reading Questionnaire (revised) based on researchers' appraisal of relevance to the current study. Students were given a 43-item survey which utilized these nine questions as well as 34 items from other sources which inquired about additional dimensions of students' noncognitive profiles (e.g., behavioral dissatisfaction). These other constructs were not used in the present study. A principal components analysis using a varimax rotation was conducted with the January administration sample. Individual items with poor factor loadings were successively pruned from the analysis until simple structure was obtained. The analysis resulted in a 33-item instrument with seven factors. The only factor used in the present study was reading efficacy.

Although 9 items were originally included from the Wigfield and Guthrie's (1997) reading motivation scale, only 4 items remained in the reading efficacy factor post the principal components analysis: I don't know if I will do well in reading this year, I am a good reader, I read because I have to, and I don't like reading something when the words are too difficult. Students recorded one of five fixed response choices: totally untrue, mostly untrue, somewhat true, mostly true, and totally true. On this efficacy measure, 698 students responded during the first administration and 652 responded during the second administration. Five hundred and sixty-three students completed both surveys. Cronbach's α was calculated for the 4-item scale using the January administration, as it was the larger of the two: the value was calculated at .656 with a sample of 657 students who answered all four questions.

Analysis

The analytical approach was to study changes in student MAP test scores over time, and how rates of change differed between gender, school, and SPED status. A separate correlational analysis was performed to investigate the relationship between reading efficacy and student reading test scores.

MAP scores. A repeated measures analysis of covariance (repeated ANCOVA) was performed on the results of the

January and May administrations of the MAP reading test, using the September MAP reading as a covariate; students' percentile ranks were used in the analyses. SPSS (Version 23) software was used for computing the analysis. The analysis tested for change in students' percentile rank between the January and May administrations while also examining interaction effects with school setting, SPED status, and gender. The students' mean MAP percentile rank and corresponding standard deviation for each school are included in Table 3. Table 3 also includes the aggregate of the students' MAP percentile rank scores and average grade broken out by students' SPED status.

Several indices of BELA1 and BELA2 students' work in the BELA online program are included in Table 4. These indices reflect student engagement and achievement with the online curriculum, including the average grade for completed BELA assignments, time spent completing BELA curricular activities, and percentage of BELA assignments completed. Summary statistics of these three variables are included for each school.

MAP percentile rank test scores and average grades as broken out by gender and SPED status are included in Table 5. Table 6 provides the students' MAP percentile rank scores and average grade as broken out by school and gender. The average grade for completed assignments in the BELA materials is not available for the TELA students because they did not access the online curriculum as part of the study.

Average percentage of activities completed. The impact of dosage of treatment on student achievement was tested using the percent complete variable as a covariate in the analysis. As in the previous analysis, student pretest results were also treated as a covariate and the dependent variable was the change between January and May MAP test percentile ranks. This test analyzed to what extent students' assignment or task completion data were useful for explaining variance in their achievement between the two MAP test administrations.

Reading efficacy. A separate correlational analysis was conducted to compare the results of the reading efficacy measure with student test scores. Student grades were also included in this analysis to provide a more complete picture of student achievement.

Results

Figure 1 includes the average percentile rank scores for students in the three schools for the three test administrations, September, January, and May. The scores show a general decline in achievement. Averaged across the three schools, the students' May average percentile scores were the lowest of the three administrations: 46, 42, and 35 chronologically. The evaluation of students' learning as measured by MAP reading percentile scores yielded statistically significant results for several factors in the research design. Statistical significance was determined using an α level of $\leq .05$. Two

Table 3. NWEA MAP Percentile Rank Means and BELA Average Grades by School and by Special Education (ED) Status.

School	September MAP Percentile Rank			January MAP Percentile Rank			May MAP Percentile Rank			Average Grade		
	All	Non-SPED	SPED	All	Non-SPED	SPED	All	Non-SPED	SPED	Non-SPED	SPED	SPED
BELA1	50.62 (26.61), 186	50.6 (26.8), 181	50.2 (18.9), 5	42.06 (28.16), 186	42.5 (28.1), 181	24.8 (26.1), 5	28.49 (27.21), 186	28.9 (27.4), 181	15.4 (15.2), 5	75.8 (9.4), 181	72.6 (9.9), 5	
BELA2	44.74 (28.74), 169	48.0 (27.9), 150	18.8 (22.1), 19	39.35 (27.97), 169	42.7 (27.4), 150	12.6 (16.3), 19	41.07 (27.56), 169	44.4 (26.7), 150	14.5 (18.9), 19	77.1 (8.0), 150	70.4 (12.5), 19	
TELA	42.54 (27.25), 140	46.9 (26.1), 120	16.7 (18.8), 20	44.59 (25.50), 140	48.5 (24.4), 120	21.4 (19.4), 20	38.17 (27.40), 140	42.6 (26.7), 120	11.5 (11.9), 20	NA (NA), NA	NA (NA), NA	
Total	46.33 (27.69), 495	48.8 (27.0), 451	21.4 (22.4), 44	41.85 (27.39), 495	44.2 (27.0), 451	18.0 (19.1), 44	35.52 (27.89), 495	37.7 (27.9), 451	13.2 (15.4), 44	76.36 (8.8), 331	70.8 (11.9), 24	

Note. Values are calculated using final analysis sample. Average grade is the percentage correct of students' completed assignments in BELA. BELA = blended English language arts; NWEA = Northwest Evaluation Association; MAP = measure of academic progress; TELA = teaching English language arts.

Table 4. Descriptive Statistics for Students' BELA Data.

Statistic	BELA1 (n = 186)		BELA2 (n = 169)	
	Percentage of Activities Completed	Activity Time (Minutes)	Percentage of Activities Completed	Activity Time (Minutes)
Mean	79	2,324	80	2,271
Median	97	2,265	97	1,988
Standard deviation	27	1,129	27	1,650
Skewness	-1.1	1.6	-1.4	8.3
Kurtosis	-0.3	7.8	1.7	85.8

Note. Values are calculated using final analysis sample. Mean, median, and standard deviation are rounded to the nearest integer. Skewness and kurtosis are rounded to the nearest 10th. BELA = blended English language arts.

Table 5. NWEA MAP Percentile Rank Means and Average Grades Split by Special Education Status and Gender.

Gender	September MAP Percentile Rank		January MAP Percentile Rank		May MAP Percentile Rank		Average Grade	
	Non-SPED	SPED	Non-SPED	SPED	Non-SPED	SPED	Non-SPED	SPED
	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n
Female (n = 254)	49.4 (27.0), 240	21.3 (18.9), 14	45.9 (26.8), 240	12.2 (10.7), 14	37.8 (27.2), 240	15.6 (17.7), 14	77.3 (8.1), 184	74.4 (5.3), 8
Male (n = 241)	48.0 (26.9), 211	21.4 (24.2), 30	42.2 (27.2), 211	20.7 (21.6), 30	37.6 (28.8), 211	12.1 (14.4), 30	75.2 (9.5), 147	69.0 (13.9), 6
Total (n = 495)	48.8 (27.0), 451	21.4 (22.4), 44	44.2 (27.0), 451	18.0 (19.1), 44	37.7 (27.9), 451	13.2 (15.4), 44	76.4 (8.8), 331	70.8 (11.9), 24

Note. Values are calculated using final analysis sample. NWEA = Northwest Evaluation Association; MAP = Measure of Academic Progress; SPED = special education.

Table 6. MAP Percentile Rank Means and Average Grades Split by School and Gender.

School	September MAP Percentile Rank		January MAP Percentile Rank		May MAP Percentile Rank		Average Grade	
	Female	Male	Female	Male	Female	Male	Female	Male
	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n	\bar{x} (SD), n
BELA1 (n = 186)	50.2 (27.2), 105	51.2 (26.0), 81	42.9 (28.2), 105	41.0 (28.3), 81	30.4 (27.3), 105	26.0 (27.1), 81	76.1 (8.8), 105	75.0 (10.0), 81
BELA2 (n = 169)	47.8 (28.1), 87	41.5 (29.2), 82	42.8 (28.2), 87	35.7 (27.4), 82	42.6 (26.7), 87	39.5 (28.5), 82	78.5 (6.7), 87	74.1 (10.3), 82
TELA (n = 140)	44.0 (26.8), 62	41.4 (27.8), 78	47.7 (24.1), 62	42.1 (26.4), 78	38.5 (26.0), 62	37.9 (28.7), 78	NA (NA), NA	NA (NA), NA
Total (n = 495)	47.9 (27.4), 254	44.7 (28.0), 241	44.03 (27.2), 254	39.6 (27.4), 241	36.6 (27.2), 254	34.4 (28.6), 241	77.2 (8.0), 192	74.6 (10.1), 163

Note. Values are calculated using final analysis sample. BELA = Blended English Language Arts; MAP = Measure of Academic Progress; TELA = teaching English language arts.

three-way interaction effects were statistically significant for these factors: (1) Administration Time \times Gender \times SPED status and (2) Administration Time \times Gender \times School. One statistically significant two-way interaction was also identified for the factors: Administration Time \times School. The results of this repeated ANCOVA are presented in Table 7.

A post hoc analysis of the simple effects of the three significant interaction terms was conducted to better understand these results. Specifically, the effects were analyzed for

significant change between the January and May test administrations. The Bonferroni adjustment was used to minimize Type 1 error rate (i.e., researchers conducted three post hoc tests, thus α was set to $\leq .0167$).

Administration Time \times Gender \times SPED Status

For the three-way interaction term of Administration Time \times Gender \times SPED status (Figure 2; Table 8), post hoc analysis

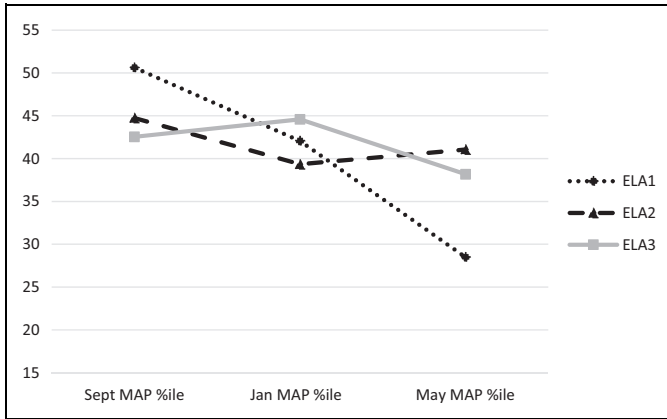


Figure 1. Average percentile ranks by school and time of test administration.

Table 7. Statistical Information From ANCOVA Analysis.

Factor	df	F	p	η_p^2
Time	1	2.920	.088	.006
Time × Pretest	1	5.80	.447	.001
Time × Gender	1	1.382	.240	.003
Time × SPED	1	0.385	.535	.001
Time × School	2	4.356	.013	.018
Time × Gender × SPED	1	6.221	.013	.013
Time × Gender × School	2	3.319	.037	.014
Time × SPED × School	2	1.261	.284	.005
Time × Gender × SPED × School	1	0.462	.497	.001

Note. SPED = special education; ANCOVA = measures analysis of covariance.

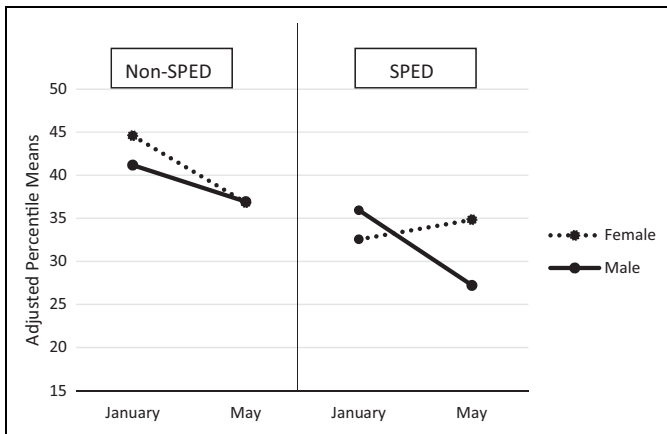


Figure 2. Interaction effect of Administration Time × Gender × SPED status.

was conducted by splitting the gender variable, then splitting the SPED status variable, and evaluating the score changes between January and May. Female and male general education students performed significantly worse on the May MAP administration compared to the January administration. Both male and female SPED students demonstrated no significant change (positive or negative) between the January and May administrations. As can be seen in Figure 2, male SPED

Table 8. Adjusted Means Used in Interaction Effect Test for Administration Time × Gender × SPED Status.

Gender	Non-SPED		SPED	
	January	May	January	May
Female (n = 254)	44.59	36.83	32.56	34.84
Male (n = 241)	41.18	36.93	35.94	27.21

Note. SPED = special education.

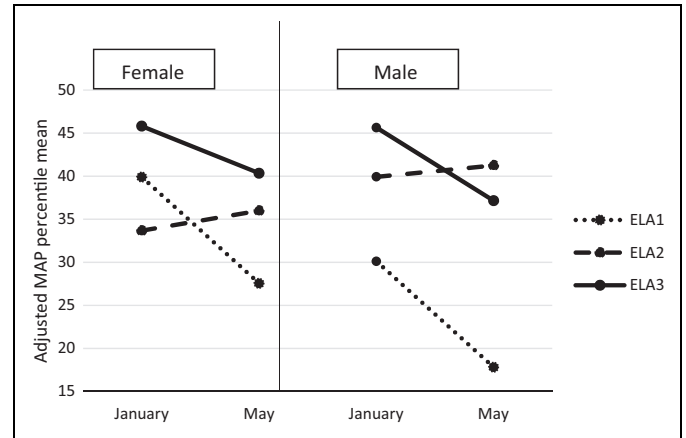


Figure 3. Interaction effect of Administration Time × Gender × School.

students' scores declined: This decline, however, was not significant at $\alpha \leq .0167$ level of significance.

Administration Time × Gender × School

For the three-way interaction term of Administration Time × Gender × School (Figure 3; Table 9), post hoc analysis was conducted by splitting the gender variable, then splitting the school variable, and examining score changes between January and May. Female students at BELA1 performed significantly worse on the May administration compared to the January administration, while female students at the other two schools demonstrated no significant change in any direction. Males at both BELA1 and TELA performed significantly worse on the May test compared to the January test, whereas male students at BELA2 demonstrated no significant change.

Administration Time × School

For the two-way interaction term of Administration Time × School (Figure 4; Table 10), post hoc analysis was conducted by splitting the school variable and looking at changes between January and May. Students at both BELA1 and TELA performed significantly worse on the May administration compared to the January administration, while students at BELA2 demonstrated no significant change in scores.

Table 9. Adjusted Means Used in Interaction Effect Test for Administration Time × Gender × School.

School	Female		Male	
	January	May	January	May
BELA1 (n = 186)	39.90	27.56	30.12	17.81
BELA2 (n = 169)	33.68	36.00	39.93	41.25
TELA (n = 140)	45.81	40.34	45.64	37.15

Note. BELA = blended English language arts; TELA = teaching English language arts.

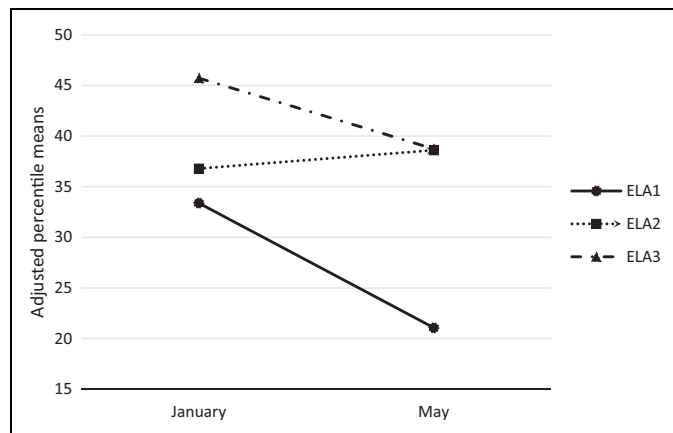


Figure 4. Interaction effect of Administration Time × School.

Table 10. Adjusted Means Used in Interaction Effect Test: Administration Time × School.

School	January	May
BELA1 (n = 186)	33.37	21.06
BELA2 (n = 169)	36.80	38.61
TELA (n = 140)	45.73	38.74

Note. BELA = blended English language arts; TELA = teaching English language arts.

The significant result of the Administration Time × School interaction indicates that, with the exception of students at BELA2, students actually did worse, normatively speaking, at the end of the year than they did in January. The meaning of this significant effect is difficult to interpret (see the Discussion section for elaboration).

Student Reading Efficacy

Correlations of students’ reading efficacy scores with MAP reading percentiles and overall BELA program grades are presented in Table 11. Correlations with test scores varied considerably depending on administration time of both the survey and the MAP test. These positive correlational values ranged from small to moderate (.191 to .417). A similar range of correlational values was found between the students reading

efficacy scores and their average grades on their BELA assignments (.167 to .324).

Discussion

The purpose of this study was to investigate the relationship among disability status, gender status, and self-efficacy in regard to reading achievement growth in blended learning. The results suggest that students experienced significant outcome effects, as measured on the MAP reading percentile ranks, depending on their school of attendance, SPED status, or gender. The conundrum is that although the results are statistically significant, the results generally reflect a significant *drop* in performance between the January and May test administrations. If the significant interactions found in this study were taken at face value, the conclusion would be that female students’ scores in the SPED programs—averaged across all schools—remained level, while both female and male students’ scores in general education declined, and male students’ scores in the SPED program declined but not significantly; that female students’ scores at BELA1 declined while female students’ scores at the other two schools stayed the same; that males’ scores at BELA2 stayed the same while males’ scores at other schools declined; and that scores at BELA1 and TELA declined while scores at BELA2 remained level. The researchers are disinclined to make these conclusions, however.

In an intervention study such as this, the expectation is that students will achieve higher levels of performance over time; yet, in this study, scores generally declined (Table 3). The results are in contrast with what is expected across treatment schools and comparison schools: Thus, the validity of these results should be questioned. The results were not an appropriate evaluation of BELA, its instructional design, or relevant human–computer interaction features. These results were likely confounded by other factors that were not assessed in the study. The declining MAP scores raise several questions about whether the students were motivated to provide an accurate indication of their skills and abilities particularly in the final test administration. In addition, several observations raised questions about the fidelity with which the BELA program was implemented.

This study’s first question concerned whether usage of the BELA ELA curriculum influenced students’ reading performance as measured on the MAP. As indicated by Table 1, the amount of instructional time for ELA and usage of the BELA curriculum varied substantially among the schools. This variation in instructional time and BELA usage was assessed in the repeated ANCOVA via the interaction term of School × Time of Administration. Because the differences were part of the overall differences between schools, the school variable incorporates the differences in instruction between the schools.

A significant effect was found for this two-way interaction, School × Time of Administration. The result, however, was not that experimental, schools performed better than the control school or vice versa. Rather, one school (BELA2) demonstrated an upward trend in student performance, while the

Table 11. Correlations of Reading Efficacy With MAP Percentile Ranks and Average Grades.

School	September MAP Percentile Rank	January MAP Percentile Rank	May MAP Percentile Rank	Average Grade
BELA1				
January reading efficacy	362	297	342	181
May reading efficacy	319	293	336	175
BELA2				
January reading efficacy	253	281	191	167
May reading efficacy	341	283	299	324
TELA				
January reading efficacy	366	301	417	NA
May reading efficacy	318	276	311	NA
Total				
January reading efficacy	319	297	300	175
May reading efficacy	319	287	320	232

Note. Decimals removed. Calculations are based on pairwise completeness. Sample sizes for the pairs ranged from $n = 145$ to $n = 266$. Test scores and grades include only those data that were at or above 10% of activities complete based on the average percentage of activities complete variable. BELA = blended English language arts; MAP = measure of academic progress; TELA = teaching English language arts.

other two schools (BELA1 and TELA, a treatment and a comparison school, respectively) demonstrated a downward trend. That only one of the treatment schools showed a positive effect indicates that the significant interaction effect cannot be attributed solely to the BELA curricular and instructional activities but instead must be attributed to other factors that were not part of the manipulation. Further, the statistically significant result was predominantly due to students' MAP score decline, despite expectations that all groups would make some detectable gains or stay level. This finding further complicates interpretation of the interaction and casts doubt on whether the effect is meaningful.

The second question, whether dosage of exposure to the BELA program related to changes, was answered in analysis using BELA's calculations of students' percent of completed assignments (i.e., of the assignments incorporated into the curriculum, what percentage a student completed). Again, no reliable effect was calculated. In this study, dosage did not contribute to a significant improvement in student test percentile ranks. This finding is particularly troubling in that one expects that the more time students spend engaged in academic learning, the more their performance should reflect improvement. Most students completed the majority of their coursework (50% of students completed more than 90% of their activities). The percent complete variable may not be a useful variable for interpreting student's dosage due to the fact that very little variability existed in the percentage values for assignment completion.

The third question, whether students in SPED showed different trends of growth than did general education students, was answered by an interaction term from the repeated ANCOVA analysis. No significant interaction of SPED Status \times Time of Administration was found. Both groups of students appeared to be progressing at a similar rate, which can be viewed as a positive outcome. Although students with disabilities performed below the level of students without disabilities, the achievement gap did not increase between the test administrations.

The fourth question, whether changes over time differ between genders, was also answered by an interaction term in the first repeated ANCOVA. Again, no significant interaction of gender by time of administration was found.

Finally, the fifth question, whether or not reading efficacy continues to correlate with reading achievement test scores, was answered by a correlation analysis. The results of this analysis demonstrated that weak to moderate correlations were found for both the blended treatment schools and the traditional comparison school. Since the MAP scores were used in this analysis also, these findings are considered very tentative.

Personal and Programmatic Influences

Regarding why most of the hypothesized effects were not found, several considerations seem plausible. Anecdotal reports from teachers, for example, indicated that students may have been fatigued around the time of the final MAP administration (in early May) due to recently having completed the state's assessment, which took place during most of April. The MAP percentile scores generally show declines in performance on the May administration. If the students were fatigued and did not fully engage in the MAP assessment, the scores may not accurately represent their learning and achievement. In addition, the classroom instruction and issues with treatment fidelity are important to consider.

Although observations of the classroom, detailed in the Method section, were intended as notes for the researchers, they revealed several dimensions that are important to consider regarding treatment fidelity. Specifically, two substantial problems with treatment fidelity (i.e., parts of the treatment that were not implemented as intended) were that students did not have access to necessary audio-playback devices (e.g., headphones) for computer instruction and that student monitoring during lab sessions was insufficient.

Dane and Schneider's (1998) research on treatment or implementation fidelity may help explain how effectiveness was compromised in the present study. Dane and Schneider

identified five dimensions of intervention fidelity: adherence, exposure, participant responsiveness, quality of delivery, and program differentiation. Adherence is defined as the extent to which specific program components were delivered as prescribed. An example of adherence is whether the correct curricular materials were used. The exposure component refers to the number, length, or frequency of instructional or practice sessions. Participant responsiveness reflects the participation and enthusiasm of participants. Quality of delivery refers to qualitative aspects of intervention and includes the interventionist's (i.e., the teacher's) preparedness.

The fifth component of treatment fidelity is program differentiation. Program differentiation safeguards against diffusion of treatments; this component ensures that students received only the planned intervention (i.e., the ELA curriculum and the BELA curriculum). One might consider this component as instructional and curricular validity. A challenge in this study is that substantial variation was noted in the ELA instruction among the three schools. As indicated in the classroom observations, substantially larger lab sessions occurred at BELA1 versus BELA2, likely leading to differences in how students experienced instruction. These differences in instruction and curricular materials created a different BELA experience for the learners depending on which school the student attended. As a consequence of this ELA course variability, the level of congruity with the BELA program and the MAP assessment may have been different between schools. Consequently, the MAP reading items may not have been equally aligned with the students' curricular and instructional activities, thus the scores may have had lower validity, that is, not accurately reflecting what the students were actually taught.

Without high levels of implementation fidelity, the evaluation is not an adequate or meaningful test of BELA's effectiveness. A common means of assessing treatment fidelity is through classroom observations. Observers' notes indicated that adherence and quality of delivery—two dimensions of fidelity—may have been unmet. Specifically, observers noted that at BELA1, blended learning activities were monitored by lab instructors rather than by certified teachers, and these lab instructors had not received the same professional development in the usage of the BELA product as had the certified teachers. As a consequence, the classroom instruction did not emphasize the students' online instruction. Researchers speculate that this disconnect between the classroom curricular emphasis and the students' blended online experience hindered their learning and achievement. The lab's physical arrangement was also challenging given the high number of students in the setting. Generally, more than 50 students were present in the lab, which made monitoring and assisting students very challenging. Further, students did not have access to earphones for listening to the computer-based teaching. Students were in an instructional setting which required them to hear the computer-based presentation, but, due to the size of the class, they needed to keep the volume on their speakers low so as to reduce the overall noise in the lab. This situation may have compromised their

ability to properly receive the intervention, thus further impacting adherence.

One of the paradoxes in the findings is that students appeared to demonstrate significant engagement with the BELA materials as indicated in the available metrics. Their usage time, percentage of task completion, and average grades were similar across the two treatment schools. In addition to potential test fatigue, the NWEA MAP is possibly not an appropriate criterion measure of ELA achievement in these schools. Students may have been engaged with the BELA supplemental materials but did not receive instruction paired directly with what was assessed on the NWEA MAP reading.

Although no significant Administration Time \times School Effect was found, the null result still provides valuable information. Despite problems with implementation, a notable finding is that the control group did not significantly outperform the experimental group. As with any comparison of a new treatment against an existing treatment (i.e., nonblended learning, in this case), a possible outcome is that the new treatment will prove less effective than the old. While one expects that BELA would elicit a marked gain in student performance, the finding that the outcome was not a significantly lower student performance provides valuable information. Also, the finding that students with disabilities' trend in academic growth was not significantly different from general education students may indicate that, at least in the environment studied in this research, students with disabilities are progressing at a rate similar to their peers in general education. Further, the results of this study shed light on some of the professional development and implementation factors that may have led to a lower quality of intervention delivery. Finally, the study found that reading efficacy continues to be an important factor in student reading achievement in this blended setting.

Limitations

This study has many limitations to consider. Despite attempts to conduct a well-designed quasi-experimental design, methods were compromised by the very small sample of students with disabilities (in the data set) who had completed all three NWEA MAP administrations. This sample contained only 5 students at BELA1 and totaled to only 44 students across the three schools. Clearly, with such a small sample, drawing conclusions is difficult and statistical power was limited for finding significant effects. Future studies would benefit from a considerably larger sample. Sampling, more broadly, was also limited in this study. Specifically, the listwise complete sample for the student NWEA MAP test scores was considerably smaller than the initial assessment. The smallest of the NWEA MAP administrations was 602 students, whereas the final sample of complete data was 495 students; more than 100 students were lost due to incomplete test administrations. As noted in the Discussion section, the study's conclusions were also limited by what appeared to be test fatigue, resulting in student scores declining unexpectedly.

Conversations with the participating schools and staff about these limitations were beneficial. The school plans to replicate the methods of the current study with more support for students in their blended work, using the MAP as a formative assessment, engaging teachers in more professional development, and changing the computer lab implementation to have fewer students per lab proctor and to ensure that the necessary technology (e.g., headphones) are provided. Along with these steps, BELA intends to perform regular fidelity checks.

Although the study results did not generally imply that the treatment was superior to the comparison condition, the findings and subsequent improvements that will be made by the participating schools in terms of implementation are likely to improve the learning opportunities of future students. Researchers interested in studying blended learning, or school officials who are interested in implementing a blended learning program, would benefit from learning from the limitations of this study so as to begin their investigations or implementations with these limitations resolved.

Authors' Note

The contents of this article were developed under a grant from the U.S. Department of Education (#H327U110011). However, the content does not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government. Project Officer is Celia Rosenquist.

Acknowledgments

We thank the school district, BELA, and the involved teachers and staff for their help with implementing the study especially with the data collection activities. We are grateful for the research partnership with the school district and with BELA, without which none of our work would have been possible. We thank BELA particularly for their help and support in data collection, professional development for staff, and for implementing the outcome measures which used for analyses. Finally, we thank BELA's staff for answering all of our many questions about their curriculum, their data sets, and for their help with making sense of the outcomes.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Burdette, P. J., Greer, D. L., & Woods, K. L. (2013). K-12 online learning and students with disabilities: Perspectives from state special education directors. *Journal of Asynchronous Learning Networks, 17*, 65–72.
- Center for Applied Special Technology. (2011). *Universal design for learning guidelines version 2.0*. Wakefield, MA. Retrieved from [http://www.udlcenter.org/sites/udlcenter.org/files/UDL_Guidelines_Version_2.0_\(Final\)_3.doc](http://www.udlcenter.org/sites/udlcenter.org/files/UDL_Guidelines_Version_2.0_(Final)_3.doc)
- Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: facts, trends and emerging issues*. New York, NY: National Center for Learning Disabilities.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45. doi:10.1016/S0272-7358(97)00043-3
- Englert, C. S., Zhao, Y., Dunsmore, K., Collings, N. Y., & Wolbers, K. (2007). Scaffolding the writing of students with disabilities through procedural facilitation: Using an Internet-based technology to improve performance. *Learning Disability Quarterly, 30*, 9–29. doi:10.2307/30035513
- Franklin, T. O., Rice, M., East, T., & Mellard, D. (2015). *Enrollment, persistence, progress, and achievement: Superintendent forum* (Report No. 1). Lawrence: Center on Online Learning and Students with Disabilities, University of Kansas. Retrieved from http://centerononlinelearning.org/wp-content/uploads/Superintendent_Topic_1_Summary_UpdatedNovember11.2015.pdf
- Freidhoff, J. R. (2015). *Michigan's K-12 virtual learning effectiveness report 2013–14*. Lansing, MI: Michigan Virtual University. Retrieved from http://media.mivu.org/institute/pdf/er_2014.pdf
- Klein, J. A., Wiley, H. I., & Thurlow, M. L. (2006). *Uneven transparency: NCLB tests take precedence in public assessment reporting for students with disabilities* (Technical Report 43). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Technical43.html>
- Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32*, 317–344. doi:10.1016/j.stueduc.2006.10.002
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*, 1–47.
- Northwest Evaluation Association. (2003). *Technical manual: For use with measures of academic progress and achievement level tests*. Portland, Oregon.
- Rhim, L., & Kowal, J. (2008). *Demystifying special education in virtual charter schools*. Alexandria, VA: TA Customizer Project, National Association of State Directors of Special Education. Retrieved from <http://www.charterschoolcenter.org/sites/default/files/Demystifying%20Sped%20in%20Virtual%20Charter%20Schools%202008.pdf>
- Staker, H., & Horn, M. (2012). *Classifying K-12 blended learning*. San Mateo, CA: Clayton Christensen Institute for Disruptive Innovation.
- Wagner, M., Newman, L., Cameto, R., & Levine, P. (2006). *The academic achievement and functional performance of youth with disabilities. A report from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2006-3000). Menlo Park, CA: SRI International. Retrieved from <http://ncser.ed.gov/pubs>
- Wang, Y., & Decker, J. R. (2014). *Examining digital inequities in Ohio's K-12 virtual schools: Implications for educational leaders and policymakers* (Paper 19). Atlanta, GA: Georgia State University, Educational Policy Studies Faculty Publications. Retrieved from http://scholarworks.gsu.edu/eps_facpub/19/

- Watson, J., Murin, A., Vashaw, L., Gemin, B., & Rapp, C. (2012). Keeping pace with K-12 online learning: An annual review of policy and practice. *Evergreen Education Group*. Retrieved from <http://www.inacol.org/wp-content/uploads/2015/03/KeepingPace2012.pdf>
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420. doi:10.1037/0022-0663.89.3.420
- Woodworth, J. L., Raymond, M. E., Chirbas, K., Gonzalez, M., Negassi, Y., Snow, W., & Van Donge, C. (2015). *Online charter school study*. Stanford, CA: Center for Research on Education Outcomes. Retrieved from <https://credo.stanford.edu/pdfs/Online%20Charter%20Study%20Final.pdf>

Author Biographies

Jesse R. Pace is a graduate research assistant with the Center on Online Learning and Students with Disabilities, as well as a doctoral student in Educational Psychology and Research at the University of Kansas. His research interests center around measurement and equity in education, including the intersection of the two.

Daryl F. Mellard is a Research Professor and director of the Center on Online Learning and Students with Disabilities at the University of Kansas. In addition to research on improving online learning outcomes for students with disabilities, his interests include adult literacy assessment and intervention and implementation of response to intervention frameworks.