RESIDENTIAL TREATMENT FOR CHILDREN & YOUTH
2018, VOL. 35, NO. 1, 60–91
https://doi.org/10.1080/0886571X.2018.1448739

Routledge
Taylor & Francis Group

Check for updates

The efficacy of a literacy intervention for incarcerated adolescents

David E. Houchins[a], Joseph C. Gagnon[b], Holly B. Lane[b], Richard G. Lambert[c], and Erica D. McCray[b]

[a]Department of Educational Psychology, Special Education, and Communication Disorder, Georgia State University, Atlanta, GA, USA; [b]Department of Special Education, University of Florida, Gainesville, FL, USA; [c]Educational Leadership, University of North Carolina-Charlotte, Charlotte, NC, USA

ABSTRACT

This randomized trial with repeated measures investigated the efficacy of a literacy intervention for low-performing readers in a juvenile corrections setting over a 31-month period. Students were randomly assigned to treatment (i.e., Read 180®) or treatment-as-usual (i.e., typical literacy program with computer instruction) conditions. Students were scheduled to receive 110-min of daily literacy instruction. The average total amount of instruction across both groups was 137 h. This article examines treatment effects using curriculum-based measures of comprehension, oral reading fluency, and spelling and standardized diagnostic reading and language assessments. Significant differences in favor of the treatment group were found with reading comprehension and language. Findings are discussed in the context of literacy instruction for male students in small long-term juvenile correctional facilities.

KEYWORDS

literacy; juvenile justice; evidence-based practices; randomized control trial; reading instruction

A substantial portion of students in juvenile corrections (JC) function below their peers in reading (Davis et al., 2014). On average, these students read two to four grade levels below their peers (Krezmien, Mulcahy, & Leone, 2008). Common deficits are in phonological processing, fluency, and reading comprehension. Similarly, incarcerated students possess language disorders (Snow, Sanger, Caire, Eadie, & Dinslage, 2015). Such deficits may be related to the fact that the prevalence of students with disabilities is four to five times greater in JC as compared to typical schools (U.S. Department of Education, 2014a) with the majority having emotional disturbance (ED; 47.7%) or learning disabilities (LD; 38.6%).

Closely related to the overrepresentation of students of disabilities, particularly those with ED, is the disproportionate number students in JC with mental health issues. Sixty percent of students in JC have three or more comorbid mental disorders (Wasserman, McReynolds, Ko, Katz, & Carpenter, 2005). Mental health issues include anxiety disorders (Wasserman et al., 2005), depression, post-traumatic disorders (Abrams, 2013), and substance abuse

CONTACT David E. Houchins dhouchins@gsu.edu Georgia State University, Atlanta, GA, 30302-3965 USA
© 2018 Taylor & Francis

(McClelland, Elkington, Teplin, & Abrams, 2004). This may account for why a disproportionate number of these students receive psychopharmacologic treatment (Soller, Karnik, & Steiner, 2006).

The combination of low academic abilities and an overrepresentation of students with disabilities and mental health issues contributes to the challenge of providing incarcerated students with literacy instruction (Gagnon, Houchins, & Murphy, 2012). The provision of literacy instruction is further complicated by a JC system that is structured around security rather than educational need (Houchins, Jolivette, Shippen, & Lambert, 2010). Additionally, there has been a national trend toward smaller facilities housing fewer than 200 students (Abrams, 2013), which has resulted in a reduction in the number of available teachers within each facility. A reduced teaching staff decreases a facility's ability to schedule tiered reading instruction. Thus, there is a need for teachers to have even greater access to literacy programs and interventions that can simultaneously accommodate the diverse educational needs of the students. Yet, despite a national need for evidence-based JC literacy instruction research, scant methodologically sound reading research is available.

Reading research in juvenile corrections

Wexler, Pyle, Flower, Williams, and Cole (2014) synthesized 16 peer-reviewed JC literacy intervention studies published between 1970 and 2012. The authors evaluated the studies using group experimental and quasi-experimental quality design indicators (U.S. Department of Education, 2009) and single-case design research indicators (Horner et al., 2005) to determine the rigor of the studies. Fifteen of those studies included reading as part of the intervention and most focused on remedial or functional reading. Of the 15 studies, just two (Calderone, Bennett, Homan, Dedrick, & Chatfield, 2009; Houchins, Jolivette, Krezmien, & Baltodano, 2008) were experimental, while three were quasi-experimental (Mayer & Hoffman, 1982; Scarlato & Asahara, 2004; Shippen, Morton, Flynt, Houchins, & Smitherman, 2012). Studies included a limited number of participants (~65), a small number of instructional sessions (5–76), and restricted durations (~13 weeks). Three studies used random assignment (Calderone et al., 2009; Houchins, et al., 2008; Shippen, et al., 2012), two accounted for implementation fidelity (Houchins et al., 2008; Shippen, et al., 2012), and three used multi-component outcome measures (Houchins et al., 2008; Scarlato & Asahara, 2004; Shippen et al., 2012). Only Houchins et al. included instrumentation reliability data and findings across multiple states. Furthermore, just two studies (Calderone et al., 2009; Shippen et al., 2012) examined computer instruction while none used more advanced methodologies (e.g., growth modeling).

Of the remaining studies reviewed by Wexler et al., four studies used a single-case design (Allen-DeBoer, Malmgren, & Glass, 2006; Drakeford, 2002; Heward, McCormick, & Joynes, 1980; Murph & McCormick, 1985) and five used a single-group design (Coulter, 2004; Hill, Minifie, & Minifie, 1984; Malmgren & Leone, 2000; Platt & Beech, 1994; Sinatra, 1984). All of these studies examined reading and only one studied spelling and writing (Hill et al., 1984). None of the single-case studies meet quality indicator criteria (Horner et al., 2005). Wexler et al. (2014) concluded that it was difficult to determine what literacy interventions work in JC due to collectively poor design and measurement issues.

In addition to Wexler et al., one additional but non-peer reviewed study (Loadman et al., 2011) exits. The Loadman et al. study is relevant since their intervention (e.g., Read 180®; Scholastic Inc., 2007) is used in the current study with incarcerated students. Loadman et al. examined the effect of Read 180® across 56 months of implementation in seven Ohio JC facilities with 1,982 students. A series of cross-sectional, longitudinal Intent-To-Treat, and growth modeling analyses were conducted using Scholastic Reading Inventory (SRI; Scholastic, 2006) scores as the dependent measure. Unfortunately, the study included numerous methodological problems, such as: (a) not being able to report the actual age of students; (b) having unequal amounts of instruction in the experimental and control groups; (c) having high rates of student attrition; and (d) assigning students to groups without accounting for previous control group instruction. Such extensive design issues make it difficult to draw study conclusions. Yet, the Loadman et al. study does provide a context for understanding the complexities of conducting JC research and indicate that Read 180® has been used in JC.

Overall, the existing literature base includes a limited number of rigorously conducted JC literacy studies. Previous studies were mostly of short duration, included small sample sizes, and had methodological issues that limited the interpretation of results. Additionally, the majority of studies investigated only the basic reading needs of students. While large-scale high-quality literacy studies are challenging due to numerous methodological and practical contextual issues, there is a critical need for rigorous experimental studies in JC.

Study purpose

The purpose of this study was to determine the effectiveness of Read 180® (Scholastic Research, 2010) for incarcerated students. Thus, the primary research question was "For incarcerated adolescents, is there any difference in the growth rate between the treatment group and the control condition, when controlling for demographic covariates, across a range of literacy

outcomes?" Literacy outcomes included reading skills (i.e., decoding, fluency, passage comprehension) and language abilities.

Method

Setting

This study was conducted over a 31-month period in one rural private medium security, long-term JC residential facility in a southeastern state where supervision was provided by trained staff 24-hours per day, 7-days per week. The private company provided educational and security services. A maximum of 150 male students ages 12–18 years old at any given time participated in the study. The typical stay was 6–9 months. Educational services were provided five days per week on a schedule similar to a typical school. Students were scheduled to receive 110-min of daily literacy instruction across two sequential class periods during the fall, spring, and summer semesters with a two-week break between semesters. Students were placed in literacy classes based on SRI scores (Scholastic, 2006) as prescribed by Read 180®.

Participants

Sixteen teachers participated, eight in the treatment condition and eight in the control condition. High teacher turnover was a characteristic of this school. One teacher and one substitute teacher who started in the control condition moved to treatment classrooms. The move did not affect instructional dosage. No teachers moved from treatment to control. The teachers had a range of teaching experiences, but few had prior formal preparation for teaching reading. Teacher demographic characteristics are summarized in Table 1.

Four hundred sixty-four (n = 464) male students between the ages of 12 and 18 years were included in the study. Exclusion criteria for participation in the study included passing the state reading assessment or having received a high school diploma or passed the General Educational Development (GED) test. The treatment condition was comprised of 225 students, while the control condition had 239 students. Student characteristics are summarized in Table 2.

Interventions

Treatment condition

Read 180®, a blended literacy intervention program (Houghton Mifflin Harcourt, 2017), was selected for several reasons. First, Read 180® is a

Table 1 Teacher Demographic Characteristics

Teacher	Sex	Race	Highest Degree	Area(s) of Certification	Highly Qualified	Reading Endorse	ESOL	Years Teaching	Years at Facility
T1	M	W	Bachelor's	Professional, English 6-12, ESE, Middle Grade Intermediate (5-9)	Y	Y	Y	2	2
T2	F	W	Bachelor's	Professional, Elementary K-6, English 6-12, ESE	Y	Y	Y	7	4
T3	F	W	Bachelor's	Professional, English 6-12, ESOL	Y	N	Y	5	4
T4	F	A	Master's	Professional, English 6-12, ESE	Y	N	N	21	.2
T5 (sub)	F	W	Bachelor's	Temporary, Business Ed.	N	N	N	1	1
T6	M	W	Bachelor's	Professional, English 6-12	Y	Y	Y	1.5	1.5
T7	F	W	Bachelor's	Temporary, Business Ed.	N	N	N	1	1
T8	F	W	Bachelor's	Temporary	N	N	N	0	0
C1	M	W	Bachelor's	Professional, English 6-12	Y	Y	Y	1.5	1.5
C2	F	W	Bachelor's	Professional, English 6-12, ESE, Middle Grade Intermediate (5-9), Biology	Y	Y	Y	7	7
C3 (sub)	F		Bachelor's	N/A	N	N	N	N/A	6
C4	M	W	Master's	Professional, English 6-12	Y	N	N	10	.25
C5	M	W	Doctorate	Professional, English 6-12, ESE, Middle Grade Intermediate (5-9)	N	Y	Y	22	4
C6	M		Bachelor's	Math (5-9)	N	N	Y	5	5
C7	F		Master's	Professional, Middle Grade Intermediate (5-9)	Y	Y	Y	4	2
C8 (sub)	M	B	Bachelor's	N/A	N	N	N	N/A	7

Note: T = Treatment, C = Control; ESE = Exceptional Student Education; ESOL = English for Speakers of Other Languages

comprehensive literacy program comprised of aged-appropriate materials that targets struggling readers in 4th through 12th grade with evidence of its effectiveness in improving comprehension and general literacy achievement for non-incarcerated students in grades 4–9 (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2009). The second reason Read 180® was selected was because of its technology components. Placement techology components allow the teacher to group students of similar abilities together and monitor student daily progress while computerized instructional materials may allow JJ teachers to address the instructional diversity of students. Third, Read 180® is programmed instruction, reducing the need for teachers to select or develop literacy materials.

Specifically, Read 180® included textbooks, trade books, and computer software, along with supplemental worksheets. Treatment teachers used an online data management system that allowed them to monitor student progress data and plan for differentiated instruction. The initial 20-min of the class were dedicated to whole-group instruction where the teacher provided direct instruction to the entire class on academic vocabulary, reading strategies, grammar, and writing skills intended to prepare students for differentiated small-group instruction later in the period. An anchoring video was used for motivational background information.

After whole group instruction, the class was divided into three ability-based groups, and the groups completed three 20-min rotations: (a) small-group instruction, (b) computer-based instruction, and (c) independent reading. During small group instruction, the teacher provided direct instruction based on student needs. All students used the Read 180® computer-based instructional program to develop individual literacy skills. A small percentage (5–10%) initially used System 44 computer materials (Scholastic, 2007), in accordance with program requirements. System 44 provided phonics instruction for students whose skills were insufficient to access Read 180® computer materials. Then, during independent reading time, all students read books with or without audiobook support. Finally, the last 10-min involved whole-group wrap-up where the teacher summarized the content from the day's lesson.

Control condition

The control condition was the treatment-as-usual facility literacy program. Teacher-led instruction included grade-level Language Arts materials provided by other schools in the county in which the study took place. Common activities included journal writing, review of the previous day's instruction, silent reading, round-robin oral reading, teacher-generated comprehension questions, worksheet activities, and writing assignments. Additionally, students were provided with 30-min of daily reading instruction using New

Century Learning System (New Century Education Foundation, 2010) software. The diagnostic-prescriptive program assessed each student's skills in decoding, vocabulary, and comprehension, provided customized self-paced instruction that targeted the student's weaknesses, assessed progress, and adjusted instruction. The teacher monitored student progress and provided one-on-one assistance while students worked independently. A management system tracked students' progress and provided performance summaries.

Outcome measures

The SRI (Scholastic Inc, 2006) is a computer-adaptive placement assessment used as part of the Read 180® program. The SRI yields a Lexile score determined by passages that are read and understood by the student with 75% comprehension. The SRI test–retest reliability is .89 (Morsy, Kieffer, & Snow, 2010).

The *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) Vocabulary and Matrix Reasoning subtests scores were used. These two subtests provided an estimate of general cognitive functioning through the Full-Scale IQ-2 (FSIQ-2). The FSIQ-2 has an internal consistency of .96, a test–retest reliability of .88, and inter-rater reliability of .98 (Wechsler, 1999).

The *Woodcock-Johnson-III* (WJ-III; Woodcock, McGrew, & Mather, 2001) is a norm-referenced academic achievement test. The following WJ-III subtests were used: Letter-Word Identification, Word Attack, Reading Fluency, Passage Comprehension, Oral Comprehension, and Spelling. The Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests make up the WJ-III's Broad Reading cluster. All subtests have acceptable test–retest scores ranging from .83 to .95. The Broad Reading cluster is a general measure of reading ability comprised of decoding, fluency, and comprehension subtests with a median reliability of .93.

The *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner, & Rashotte, 1999) is a norm-referenced test with two subtests: (a) Sight Word Efficiency, which measures automatic word recognition; and (b) Phonemic Decoding Efficiency, which measures decoding automaticity. Reliability coefficients for both subtests exceed .90.

The *Peabody Picture Vocabulary Test* (PPVT-4; Dunn & Dunn, 2007) is a norm- referenced receptive vocabulary assessment. The PPVT-4 has shown internal consistency of .94–.95 and a test–retest reliability of .93.

A battery of AIMSweb curriculum-based measures (CBMs; www.aimsweb.com; Shinn & Good, 1992) were administered. The maze comprehension passage, Oral Reading Fluency (ORF) passages, and spelling measures were used. AIMSWeb CBMs have acceptable validity and reliability scores for research purposes.

Fidelity of implementation

Twenty-percent of all instructional periods, evenly distributed across conditions, teachers, and classrooms, were randomly selected, videotaped, and then evaluated by trained graduate students (Durlak & DuPre, 2008; O'Donnell, 2008). Students were trained to mastery (≥90% inter-rater agreement; IOR) on a sample videotape. Similar mean implementation occurred across conditions (treatment group = 2.5; control condition = 2.18). No contamination was reported. Further, 20% of observations were checked for IOA. The IOA absolute agreement within one score using inter-class correlations ranged from 96% to 99%. A 4-point Likert-type scale was used to document implementation levels (4 = highest rate of target behaviors; 1 = lowest rate of target behaviors).

Procedures

University Institutional Review Board approval, facility approval, teacher consent, guardian consent, and student assent were obtained prior to conducting research. Treatment condition teachers attended a one-day Read 180® training session prior to implementation. Reading coaches observed each teacher at least weekly using an instructional quality instrument and provided follow-up coaching involving modeling, collaborative planning, instructional suggestions, reflective conversations, professional readings, and supplementary material guidance. Reading coaches were graduate students with experience as reading teachers and were supervised by a special education reading professor.

Upon entry into the facility, students were randomly assigned to the treatment or control condition using a computer-generated list of numbers, unless they were ineligible for the study. Thirty-nine students were ineligible and excluded from the study and randomization procedures because they passed the GED test or the State-required reading test. Each student was assessed within three school days of arrival using the SRI (Scholastic Inc, 2006) and assigned a low, medium, or high rating based on his Lexile score (Low = 0 to 715; Medium = 716–960; and High = 961 and above) based on program guidelines. Each student was placed into a reading class based on their Lexile scores (Smith, 2000).

The SRI (Scholastic Inc, 2006) and WASI (Wechsler, 1999) were individually administered at intake. Several other standardized measures were administered multiple times, including at intake, after 140 instructional hours, and either within two weeks immediately preceding the student's facility exit or when the student reached 280-h of instruction, whichever came first. Measures administered multiple times included: (a) the WJ-III (Woodcock et al., 2001); (b) the TOWRE (Torgesen et al., 1999); and (c) PPVT-4 (Dunn & Dunn, 2007).

A battery of AIMSweb CBMs (www.aimsweb.com; Shinn & Good, 1992) were administered at intake and exit, and on a monthly basis. The battery included: (a) two group administered maze passages, each taking three minutes; (b) three individually administered ORF passages where the student orally read each passage for one minute as the assessor recorded errors and correct words read; and (c) one individually administered spelling probe taking about two minutes. The highest maze score, the median ORF passage score, and the correct number of words on the spelling probe were used for analyses purposes. All students were initially administered both third-grade and sixth-grade CBM probes. To mitigate ceiling effects, students who read a complete ORF passage with 100% accuracy on the initial sixth grade probes were subsequently given eighth-grade passages instead. All students were administered either third- and sixth-grade probes or sixth- and eighth-grade probes.

Research design and analyses

The research design for this study was a completely randomized trial with repeated measures. The study had no attrition. To ensure that the treatment and control group were statistically similar, the groups were compared relative to their race, disability status, whether or not they were on psychotropic medication(s), their psychological functioning, and instructional minutes. Groups were equivalent with respect to ethnicity ($\chi^2_{(3)} = 1.682$, $p = .794$), whether the student was taking psychotropic medication(s) ($\chi^2_{(1)} = .384$, $p = .536$), and whether the student had an Individualized Education Plan (IEP) ($\chi^2_{(1)} = .802$, $p = .370$) (see Table 2). Both groups were comprised of approximately 48% African-American, 40% White, and 10% Hispanic students. The majority of each group took psychotropic medication(s) (treatment group = 59.4%; control group = 56.3%). Over 40% of each group had an IEP (treatment group = 40.6%, control group = 45.2%). The mean ages for the two groups were nearly the same ($t_{(370)} = 1.230$, $p = .220$) as were the group mean Global Assessment of Functioning (GAF; American Psychiatric Association Task Force on DSM-IV, 2000) scores ($t_{(370)} = .674$, $p = .501$) with mean scores within the "Serious Symptoms range (treatment group = 45.42; control group = 45.96).

The two groups received statistically similar amounts of instructional minutes (see Table 3; $t_{(334)} = 1.042$, $p = .298$) with a small mean difference between the groups (~8 h). Similarly, the small mean difference between the groups as measured in minutes for the CBM (256.4 min) amounted to less than three class periods and was not statistically significant ($t_{(370)} = .594$, $p = .553$). The full distribution of scores was similar. Both groups received approximately one-half of an academic year of instruction in their literacy class, had some students that received the equivalent of a full year or more of

Table 2 Demographic characteristics by treatment group

		Treatment	Control	Total
		n=175	*n*=197	*n*=372
Ethnicity	Caucasian	41.0%	37.8%	39.3%
	African American	47.4%	49.7%	48.6%
	Hispanic	9.8%	11.4%	10.7%
	Other	1.7%	1.0%	1.4%
On medication	Yes	59.4%	56.3%	57.0%
	No	40.6%	43.7%	43.0%
IEP Status	Yes	40.6%	45.2%	43.0%
	No	59.4%	54.8%	57.0%
Age	Mean	16.12	15.82	15.96
	SD	2.73	2.00	2.37
Global functioning	Mean	45.42	45.96	45.71
	SD	8.21	7.28	7.73
Grade level	5th	0.0%	0.6%	0.3%
	6th	5.8%	8.5%	7.3%
	7th	11.0%	18.1%	14.8%
	8th	27.3%	31.6%	29.6%
	9th	27.3%	23.7%	25.4%
	10th	19.5%	9.6%	14.2%
	11th	8.4%	7.9%	8.2%
	12th	0.6%	0.0%	0.3%

instruction, and were allocated 110 min of daily literacy instruction for five days each week. This equated to approximately 2,200 possible instructional minutes per month.

To test for treatment effects, growth curve modeling was used. Multiple assessments over time were nested within student. The level one model included instructional exposure (measured in hours or minutes) as the predictor variable and the score for each assessment occasion was the dependent variable. The intercept for this model represents initial status, or score at intake, and the slope associated with instructional minutes represents the growth rate.

The level one model estimated the initial status and growth for each student by nesting measurement occasions within student. The level two model contained treatment group (0 = control group, 1 = treatment group), age, whether the student was taking psychotropic medication(s),

Table 3 Instructional exposure by treatment condition

	Treatment Group	Control Group	All Participants
For measures on hourly schedule	*n*=181	*n*=155	*n*=336
Mean	141.198	133.373	136.982
SD	70.848	66.706	68.656
Minimum	11.000	0.000	0.000
25th	77.767	87.333	80.583
Median	149.167	134.750	141.125
75th	197.083	180.708	188.833
Maximum	312.670	301.530	312.670
For the CBM analyses, in minutes	*n*=175	*n*=197	*n*=372
Mean	8024.54	7768.14	7888.76
SD	4315.91	4006.20	4151.10
Minimum	285.00	0.00	0.00
25th	4074.00	4682.50	4483.75
50th	8150.00	7620.00	7817.50
75th	11340.00	10670.00	10908.75
Maximum	18760.00	18092.00	18760.00

special education classification, African American or Latino ethnicity. The level two models treated both initial status (intercept) and growth rate (slope) from the level one models as outcomes across students.

The next in running the statistical analyses was to form unconditional models for each of the outcomes. Given the use of growth curve modeling, these models were not completed unconditional at level one as the timing of the measurement occasion was entered as a predictor. The level two models were completely unconditional and were tested as random intercepts and slopes models. The results of the unconditional models indicated that there was sufficient between student variance in the intercepts (initial status) to model those as random for all outcomes measures. Therefore, as shown in Table 4, error terms (r_{0i}) were added for each outcome to form randomly varying intercept models at level two. There was not sufficient between student variance in the slopes to model those as randomly varying, and therefore errors terms were not added to the slopes as outcomes models. There was sufficient variability

Table 4 Implementation fidelity by treatment condition

		Treatment Group	Control Group	*t*
Quality	mean	3.33	2.83	7.27***
	SD	0.62	0.67	
Differentiation	mean	2.72	1.79	17.42***
	SD	0.54	0.44	
Contamination	mean	1.10	1.00	2.81**
	SD	0.41	0.00	
Exposure	mean	3.05	2.78	3.34***
	SD	0.63	0.92	
Adherence	mean	2.50	2.18	4.99***
	SD	0.65	0.53	
Responsiveness	mean	2.71	2.40	5.17***
	SD	0.50	0.62	

Note. * - $p < .05$, ** - $p < .01$, *** - $p < .001$.

in slopes across groups to model, however given the insufficient between student variance in slopes, non-randomly varying slopes models were chosen. Furthermore, the non-randomly varying slopes models were chosen so that the treatment effects on growth rates could be tested.

From the WJ-III (Woodcock et al., 2001), W-Ability scores were used from the following: Basic Reading, Brief Reading, Broad Reading, Letter-Word Identification, Oral Comprehension, Passage Comprehension, Reading Fluency, Spelling, and Word Attack. Both age and grade levels at the time of entry to the facility were used as covariates. Many of the students did not enter the facility at their age-appropriate grade and therefore age and grade level were not highly correlated ($r = .321$). Additional covariates included whether the student was taking psychotropic medication(s), special education classification, and African-American or Hispanic ethnicity. White and mixed-race was the baseline condition for ethnicity.

The level one model for the WJ-III outcomes included an intercept as an estimate of initial status and a slope as an estimate of hourly growth rate. The level two models were used to examine the associations between treatment group membership and both initial status and growth rate after controlling for the covariates. Age and grade level were entered into the models as group mean centered and all remaining variables were entered as uncentered

indicator (0/1) variables. For the age standardized measures, the PPVT standard score and the TOWRE Phonemic Decoding, Sight Word Efficiency, and Reading Efficiency measures, a similar two-level hierarchical linear growth model was used. The same covariates were used, with the exception of age and grade. The statistical models for the CBMs were similar to the models used for the WJ-III,with the exception of the removal of grade level as it is embedded within the measures. See Table 4 for WJ-III, PPVT, TOWRE, and CBM models.

Results

Woodcock-Johnson measures

With respect to initial status for the WJ-III measures, grade level was statistically significantly associated with all of the outcomes except Oral Comprehension (see Tables 5 and 6). Once grade level at intake was included in the model, age was not associated with any of the outcomes. Being on psychotropic medication(s) was associated with lower scores on the Reading Fluency and Word Attack scores. Special education classification was associated with lower scores at initial status across all the measures. African American students scored lower, on average, on all of the measures at intake and these differences were statistically significant on all measures except Spelling. Latino students also scored lower, on average, at initial status, yet only the differences on Oral Comprehension and Passage Comprehension were statistically significant. There were no statistically significant differences between the groups at initial status.

Both treatment and control groups made statistically significant growth on the following measures: Brief Reading, Broad Reading, Letter-Word Identification, Oral Comprehension, Passage Comprehension, and Reading Fluency (see Table 5 and 6). There were few statistically significant differences between the treatment groups in growth rate. Special education classification was associated with slightly faster hourly growth rates for Basic Reading, Letter-Word Identification, and Word Attack. African American students grew, on average, slightly slower on the Reading Fluency measure. The only statistically significant group difference in growth rate was found for the Passage Comprehension measure where the treatment group grew faster. This effect, when projected across a full academic year (270 h of instruction) and converted to a standardized mean difference effect size (ES) measure by dividing by the pooled standard deviation, was .283. This method was used for all of the growth rate ES, follows the strategy outlined by Raudenbush and Lui (2001) and Feingold (2009), and involves multiplying the growth rate advantage for the treatment group by the amount of time that would simulate an entire treatment delivery cycle (academic year).

Table 5. Model results for the Woodcock Johnson W Ability Reading scores

	Basic Reading		Brief Reading		Broad Reading		Reading Fluency	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model								
Intercept	527.081***	2.850	523.720***	2.361	520.325***	2.001	513.662***	1.697
Grade	3.020**	1.047	2.756***	0.866	2.756***	0.734	2.312***	0.623
Age	-1.345	1.155	-0.809	0.956	-0.866	0.811	-0.853	0.688
On Medication	-4.097	2.386	-2.614	1.978	-2.806	1.677	-3.29*	1.421
Special Education Placement	-13.305***	2.291	-11.14***	1.899	-9.992***	1.61	-7.801***	1.364
African American	-6.769**	2.476	-6.968***	2.054	-5.998***	1.741	-4.038**	1.475
Latino	-0.662	4.000	-4.623	3.311	-4.132	2.806	-3.260	2.381
Treatment Group	0.869	2.263	1.101	1.874	1.030	1.589	0.612	1.347
Effect Size	0.045		0.068		0.075		0.058	
Growth Rate Intercept								
Intercept	0.012	0.009	0.020***	0.006	0.029***	0.005	0.044***	0.009
Grade	-0.004	0.003	-0.003	0.002	-0.002	0.002	0.002	0.003
Age	0.002	0.003	0.000	0.002	0.000	0.002	0.001	0.003
On Medication	0.004	0.007	0.000	0.005	-0.001	0.004	-0.001	0.007
Special Education Placement	0.022***	0.007	0.007	0.005	0.005	0.004	0.001	0.007
African American	0.003	0.008	-0.002	0.005	-0.007	0.004	-0.018*	0.007
Latino	0.005	0.012	0.011	0.008	0.003	0.007	-0.011	0.011
Treatment Group	-0.001	0.007	0.005	0.005	0.003	0.004	0.001	0.007
Effect Size	-0.010		0.089		0.067		0.032	

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 6. Model results for the Woodcock Johnson W Ability scores

	Letter Word Recognition		Oral Comprehension		Passage Comprehension		Spelling		Word Attack	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model										
Intercept	534.453***	3.279	517.751***	1.905	512.737***	1.820	518.682***	2.810	519.667***	2.676
Grade	3.782**	1.047	1.258	0.701	1.981**	0.668	3.945***	1.033	2.18**	0.983
Age	-1.353	1.329	-0.253	0.773	-0.274	0.737	-1.449	1.139	-1.264	1.085
On Medication	-3.185	2.745	0.211	1.597	-1.827	1.525	-1.867	2.355	-4.999*	2.241
Special Education Placement	-16.306***	2.635	-5.529***	1.534	-5.674***	1.465	-11.429***	2.261	-10.417***	2.152
African American	-7.172*	2.849	-10.377***	1.657	-6.421***	1.584	-1.423	2.443	-6.338**	2.326
Latino	-0.455	4.601	-14.007***	2.672	-8.449***	2.551	-1.213	3.943	-0.820	3.755
Treatment Group	1.140	2.603	0.654	1.514	0.862	1.445	0.738	2.232	0.606	2.125
Effect Size	0.051		0.059		0.073		0.041		0.035	
Growth Rate Intercept										
Intercept	0.024*	0.009	0.021*	0.011	0.016*	0.008	0.018	0.011	-0.003	0.012
Grade	-0.006	0.003	-0.006	0.004	-0.001	0.003	-0.002	0.004	-0.002	0.004
Age	0.000	0.004	0.005	0.004	0.000	0.003	0.001	0.004	0.003	0.005
On Medication	-0.008	0.008	0.001	0.009	0.007	0.007	0.006	0.009	0.016	0.010
Special Education Placement	0.021**	0.008	0.015	0.009	-0.006	0.006	-0.001	0.009	0.025**	0.009
African American	0.004	0.008	0.000	0.009	-0.008	0.007	-0.003	0.010	0.003	0.010
Latino	0.000	0.013	0.021	0.015	0.015	0.011	0.022	0.015	0.010	0.016
Treatment Group	0.002	0.007	-0.002	0.009	0.012*	0.006	0.002	0.009	-0.002	0.009
Effect Size	0.021		-0.055		0.283		0.032		-0.025	

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

This method provides an ES equivalent to the standardized model-estimated mean differences between the treatment and control groups at the end of the study, adjusted for baseline differences between the groups.

Age standardized measures

Students with special education classification scored lower at initial status on all of the age- standardized measures (Table 7). African American and Latino students scored lower at initial status on the PPVT than did the reference group (White and mixed race). All students, across groups, showed average statistically significant growth on the TOWRE Phonemic Decoding measure. Students with special education placements grew faster on the PPVT and slower on the TOWRE Phonemic Decoding measure. African American students grew faster on the PPVT measure than the reference group (White and mixed race). The only statistically significant group difference in growth rate was found for the PPVT standard score where the students in the treatment group grew faster than those in the control condition. This effect, when projected across a full academic year (270 h of instruction) and converted to a standardized mean difference ES measure by dividing by the pooled standard deviation, was .276.

Curriculum-based measures

The coefficients in Table 8–11 are displayed to six decimal points because the time metric for the growth model was scaled in minutes. Thus, the growth rates for each additional minute of instruction are very small and would round to zero at two or three decimal points. Table 8. displays the results of the model for the AIMSweb maze comprehension measure for third, sixth, and eighth grades. The treatment groups were approximately equal at initial status across all three measures. Age was not associated with initial status. Students taking psychotropic medication(s) scored approximately three points lower at initial status on all three measures and these differences were statistically significant. Students with a disability classification scored statistically significantly lower across all three measures than did typically developing students. African American students scored statistically significantly lower than White students on all three measures and Latino students scored lower than White students on the third and sixth-grade measures. The growth rate models indicated that the treatment group grew at a statistically significantly faster rate than the control group on both the third and sixth-grade measures, but not on the eighth-grade measure.

In order to enhance interpretability, the statistical models for third- and sixth-grades AIMSweb maze comprehension assessment, the group differences in growth rate were converted to a standardized mean difference ES

Table 7. Model results for the age standardized measures

	PPVT Standard Score		TOWRE Phonemic Decoding		TOWRE Sight Word Efficiency		TOWRE Reading Efficiency	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model								
Intercept	95.335***	1.560	90.584***	2.033	88.872***	1.612	87.947***	2.109
On Medication	2.505	1.288	-1.907	1.681	-1.553	1.331	-1.872	1.744
Special Education Placement	-6.013***	1.261	-6.574***	1.646	-4.741***	1.302	-7.125***	1.708
African American	-13.501***	1.349	-2.874	1.760	-0.541	1.395	-2.555	1.827
Latino	-10.398***	2.146	0.878	2.876	-1.030	2.281	-0.344	2.984
Treatment Group	-0.762	1.230	1.796	1.609	0.071	1.274	0.758	1.669
Effect Size	-0.073		0.131		0.007		0.055	
Growth Rate Intercept								
Intercept	-0.008	0.007	0.019*	0.008	0.007	0.006	0.014	0.009
On Medication	0.003	0.005	0.008	0.007	0.004	0.005	0.007	0.007
Special Education Placement	0.017**	0.005	-0.016*	0.007	-0.007	0.005	-0.013	0.007
African American	0.013*	0.006	-0.013	0.007	-0.004	0.005	-0.009	0.007

(*Continued*)

Table 7. (Continued).

	PPVT Standard Score	TOWRE Phonemic Decoding		TOWRE Sight Word Efficiency		TOWRE Reading Efficiency		
Latino	0.003	0.009	-0.019	0.011	-0.010	0.009	-0.014	0.012
Treatment Group	0.011*	0.005	-0.002	0.006	0.004	0.005	0.001	0.007
Effect Size	0.276		-0.040		0.102		0.020	

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 8 Treatment effects for the AIMS Web Maze Comprehension measure by grade level

	Grade 3		Grade 6		Grade 8	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model						
Intercept	32.200793***	1.528905	33.965463***	1.617683	33.162116	1.965665 ***
Treatment group	0.225735	1.244280	0.257567	1.317470	1.136477	1.6515
Effect size	0.021		0.022		0.129	
Age	0.438291	0.264474	0.378823	0.275530	0.212731	0.280025
On medication	-3.371871**	1.295713	-3.213052*	1.372027	-5.953219	1.753222***
Special education status	-5.372398***	1.275339	-5.868416***	1.350248	-2.994785	1.709328
African American	-5.801721***	1.356729	-6.108663***	1.432043	-7.264168	1.815095***
Latino	-4.914990*	2.149850	-5.123314*	2.272283	0.433628	2.607802
Growth Rate Model						
Intercept	0.000116	0.000116	0.000021	0.000112	-0.000224	0.000196
Treatment group	0.000187*	0.000094	0.000290**	0.000093	0.000087	0.000157
Effect size	0.284		0.408		0.161	
Age	-0.000021	0.000020	-0.000020	0.000021	0.000002	0.000037
On medication	0.000003	0.000102	0.000248*	0.000098	0.000713	0.000171***
Special education status	0.000105	0.000096	0.000098	0.000095	-0.000019	0.000164
African American	0.000122	0.000103	0.000061	0.000100	0.000187	0.000168
Latino	-0.000164	0.000165	-0.000027	0.000165	-0.000690	0.000264**

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

evaluated at 16,200 min, which represents an academic year of treatment. For the third-grade level maze measure, this value was .284. A similar pattern was found for the sixth-grade comprehension maze where the ES was .408.

For the ORF measures in Table 9, there were no statistically significant differences between the treatment conditions in either initial status or growth rate. Students with a disability classification scored statistically significantly lower at intake than typically developing students on all three grade level measures and African American students started lower than White students on the sixth- and eighth-grade measures. There were only a few statistically significant differences in growth rate. Older students grew slower than younger students on the sixth and eighth-grade measures and African American students grew slower on third and sixth-grade measures.

For the AIMSweb spelling measure, there were no statistically significant differences between the treatment conditions in either initial status or growth rate (see Table 10). Older students started statistically significantly higher than younger students on the sixth- and eighth-grade measures and students with a disability classification scored lower at intake on all three grade level measures. There were no statistically significant differences in growth rate across any of the comparisons or grade levels. The intercepts for the growth rate models were not statistically significantly different from zero for the sixth and eighth grade measures, indicating that students on average in neither the treatment group nor control group showed growth.

The Spelling Correct Letter Sequence measures represented in Table 11 showed a very similar pattern. There were no statistically significant differences between the treatment conditions in either initial status or growth rate. Students with a disability classification scored statistically significantly lower on all three grade level measures at intake. Older students grew at a slower rate than younger students on the grade 3 measure. Latino students grew at a slower rate on the third-grade measure. Students with a disability classification displayed a statistically significantly faster growth rate on the eighth-grade measure, although the average growth rate for both treatment conditions was not statistically significantly different from zero for the sixth and eighth-grade measures.

Discussion

This is the first experimental randomized controlled group study to demonstrate significant improvement in the reading comprehension and language abilities of incarcerated students using a comprehensive reading program. Our study accounted for many of the methodological and practical issues previously associated with JC literacy research and instruction. Methodologically, our study addressed the need for more extensive and rigorous studies (Houchins et al., 2008; Shippen et al., 2012; Wexler et al., 2014). Prior to this study, only three studies had used random assignment

Table 9 Treatment effects for AIMS Web Oral Reading Fluency measure by grade level

	Grade 3		Grade 6		Grade 8	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model						
Intercept	182.314784***	6.237738	175.436083***	6.098196	152.494678***	6.896154
Treatment group	1.119862	5.102283	-0.714729	4.961920	4.596694	5.761519
Effect size	0.024		-0.015		0.119	
Age	2.019625	1.069575	1.448556	1.045086	1.740673	0.984795
On medication	-7.152954	5.296111	-3.425028	5.162282	-3.148513	6.065322
Special education status	-26.518575***	5.226251	-24.690620***	5.093001	-14.483963	5.991538
African American	-10.518895	5.554739	-10.822404*	5.395426	-16.432530*	6.379536
Latino	-10.046613	8.778002	-5.193401	8.579248	2.541493	9.219550
Growth Rate Model						

(*Continued*)

Table 9 (Continued).

	Grade 3		Grade 6		Grade 8	
Intercept	0.001317***	0.000293	0.001321***	0.000252	0.001310***	0.000399
Treatment group	-0.000177	0.000240	0.000100	0.000210	0.000070	0.000320
Effect size	-0.062		0.035		0.029	
Age	-0.000082	0.000050	-0.000121*	0.000048	-0.000170*	0.000076
On medication	0.000188	0.000258	-0.000036	0.000221	-0.000010	0.000353
Special education status	-0.000258	0.000247	-0.000300	0.000217	0.000342	0.000332
African American	-0.000566*	0.000263	-0.000547*	0.000227	-0.000345	0.000341
Latino	-0.000760	0.000417	-0.000542	0.000376	-0.000521	0.000534

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 10 Treatment effects for the AIMS Web Spelling Words Spelled Correctly measure by grade level

	Grade 3		Grade 6		Grade 8	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model						
Intercept	13.427671***	0.444971	9.643305***	0.525991	8.481894***	0.686492
Treatment group	-0.135642	0.362502	0.334050	0.428147	0.538108	0.573357
Effect size	-0.045		0.087		0.157	
Age	0.093126	0.077342	0.245728**	0.089821	0.296986**	0.097729
On medication	0.023775	0.377675	-0.173826	0.445760	0.314289	0.609885
Special education status	-2.303225***	0.371139	-2.735398***	0.439093	-1.734121**	0.595105
African American	-0.155476	0.395108	-0.606111	0.465438	-0.333766	0.632426
Latino	-0.532672	0.624820	-0.614748	0.739284	-0.556329	0.909739
Growth Rate Model						

(*Continued*)

Table 10 (Continued).

	Grade 3		Grade 6		Grade 8	
Intercept	0.000108***	0.000037	0.000029	0.000031	-0.000003	0.000058
Treatment group	0.000018	0.000030	-0.000036	0.000026	0.000014	0.000046
Effect size	0.096		-0.152		0.066	
Age	0.000004	0.000006	0.000001	0.000006	-0.000002	0.000011
On medication	-0.000040	0.000033	0.000035	0.000028	0.000006	0.000051
Special education status	0.000029	0.000031	0.000030	0.000027	0.000066	0.000048
African American	0.000029	0.000033	0.000002	0.000028	0.000006	0.000049
Latino	0.000047	0.000052	0.000000	0.000046	0.000064	0.000077

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 11 Treatment effects for the AIMS Web Spelling Correct Letter Sequence measure by grade level

	Grade 3		Grade 6		Grade 8	
Effect	Coefficient	*se*	Coefficient	*se*	Coefficient	*se*
Initial Status Model						
Intercept	103.846567***	1.978468	129.685642***	3.552147	127.943529***	3.737457
Treatment group	-0.961308	1.611462	1.405115	2.890942	3.345630	3.124697
Effect size	-0.072		0.054		0.172	
Age	0.622241	0.344476	1.175532	0.607172	0.902199	0.533498
On medication	-0.882033	1.679614	-2.304764	3.009463	0.360312	3.317913
Special education status	-8.357408***	1.649558	-17.806320***	2.965446	-11.608004***	3.244818
African American	-1.053789	1.756530	-5.148079	3.142935	-1.509685	3.452431
Latino	0.083178	2.777441	-4.747109	4.993498	-1.263288	4.967968
Growth Rate Model						
Intercept	0.000372*	0.000172	0.000017	0.000197	-0.000302	0.000288
Treatment group	0.000098	0.000138	-0.000123	0.000163	-0.000018	0.000226
Effect size	0.119		-0.076		-0.015	
Age	-0.000060*	0.000029	-0.000035	0.000037	0.000040	0.000054
On medication	-0.000042	0.000151	0.000291	0.000173	0.000220	0.000253
Special education status	0.000120	0.000143	0.000158	0.000168	0.000682***	0.000235
African American	0.000012	0.000153	-0.000044	0.000176	-0.000283	0.000241
Latino	-0.000496*	0.000243	0.000047	0.000290	0.000051	0.000376

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

(Calderone et al., 2009; Houchins, et al., 2008; Shippen et al., 2012). We used random assignment to a treatment or control condition over 31-months compared to a maximum of 13-weeks for previous studies thus representing the longest JC literacy study to date (Wexler et al., 2014). We randomly assigned 464 students compared to a maximum of 65 participants in previous studies representing the largest sample size to date (Wexler et al., 2014). Our data analyses used multiple covariates and thus increased the precision of the findings (Allen-DeBoer et al., 2006; Drakeford, 2002; Houchins et al., 2008). Our analyses took into consideration race, disability status, whether or not they were on psychotropic medication(s), their psychological functioning, and instructional minutes. Previous JC reading research (Wexler et al., 2014) rarely accounted for these variables and none precisely accounted for dosage, as we did with instructional minutes. We used multi-level growth modeling with repeated measures, thus accounting for the transitory nature of the JC population and moving beyond pretest/posttest-only designs typical of previous JC studies (Houchins et al., 2010). Multi-component measures were used to assess various literacy aspects (e.g., language, reading comprehension) with both age standardized measures and CBMs. Finally, this is only the third JC literacy study to include a fidelity measure (Houchins et al., 2008; Shippen et al., 2012). We assessed fidelity using a multi-dimensional approach (Durlak & DuPre, 2008; O'Donnell, 2008) for both conditions.

By attending to many of the methodological issues associated with conducting research (U.S. Department of Education, 2014), particularly in JC schools (Houchins et al., 2010; Wexler et al., 2014), our findings have scientific merit. Results indicated that the treatment group outperformed the control group on reading comprehension and language measures. More specifically, significant comprehension differences were found in favor of the treatment group on third and sixth grade Aimsweb maze (comprehension) probes with low to moderate ES (0.284 and 0.408, respectively) and with the WJ-III Passage Comprehension subtest (ES = .22). Our comprehension findings are aligned with previous research findings (U.S. Department of Education, 2010) suggesting the potential positive effect of Read 180®.

Given the dismal literacy abilities of incarcerated students (Harris, Baltodano, Jolivette, & Mulcahy, 2009; Krezmien et al., 2008; Leone & Weinberg, 2010), these findings are noteworthy. Previous research findings have been inconclusive on how to improve the comprehension abilities of incarcerated students (Wexler et al., 2014) and potentially with good reasons. Incarcerated students are disproportionately comprised of (a) students with disabilities (e.g., LD, ED, language disorders) (Quinn, Rutherford, Leone, Osher, & Poirer, 2005; Snow et al., 2015); (b) those who have mental health disorders (Abrams, 2013; Wasserman et al., 2005); (c) students on psychotropic medication(s) (Soller et al., 2006); and (d) students who collectively

have significantly lower literacy abilities for a variety of known and unknown reasons (Gonsoulin, Griller Clark, & Rankin, 2015; Harris et al., 2009; Krezmien et al., 2008). These complexities make it difficult to conduct literacy research in JC. Yet, while our findings are modest, we were able to take into account many of these issues. Thus, future researchers should further explore treatment elements that contribute to improved reading comprehension, taking into consideration the multitude of student and contextual factors.

In addition to positive comprehension findings, significant differences were found in favor of the treatment group on the PPVT (ES = .22). While the ES findings are small, this finding is thought-provoking. Researchers (Bryan, Freer, & Furlong, 2007; Sanger, Creswell, Dworak, & Schultz, 2000; Snow et al., 2015) have indicated that the prevalence rate of language disorders is significantly higher for incarcerated students. These researchers have also suggested that there is a positive relationship between the language abilities of these students and the overrepresentation of students with LD and ED (Quinn et al., 2005). Given the potential comorbid relationship between literacy, language, and disability (Lane, Carter, Pierson, & Glaeser, 2006; Lewis et al., 2015), finding that the treatment had a positive effect on the language abilities, even if small, is important. Future researchers should examine the relationship between the treatment and the language abilities of incarcerated students that might contribute to increased literacy abilities.

One potential explanation for our findings is related to the structure of Read 180®. Instruction was provided using materials that are meant to be of interest to an adolescent. The treatment condition accounted for flexible student groups, allowing for greater use of whole class, small group, individual, and computer-based literacy activities (Loadman et al., 2011; Shippen et al., 2012). Differentiated instruction was possible because students were grouped across three ability levels (i.e., high, medium, low) within classes, allowing teachers to provide students with explicit instruction while other students practiced new skills that had been taught. The instructional grouping structure is important, as JC facilities are increasingly smaller nationally (Abrams, 2013). As facilities increasingly serve smaller number of students (<200 students), there is a need for literacy intervention programs that are flexible enough to meet the diverse literacy needs of students without disturbing the fundamental structure of the class. Increased opportunities for differentiated instruction have the potential to allow teachers to provide more intensive literacy instruction that is associated with improved student literacy outcomes (Solis, Miciak, Vaughn, & Fletcher, 2014). Future researchers should explore how teachers can provide students with targeted differentiated instruction that accommodates student literacy diversity.

While positive comprehension and language outcomes were found, significant differences were not found in the areas of decoding, oral reading

fluency, or spelling. The lack of findings in these areas is concerning. Several instructional factors of Read 180® may have negatively affected student progress in these areas, including (a) the role of explicit and direct instruction, (b) the lack of extended text discussion, and (c) limitations in the usability of data to make informed instructional decisions (Kamil et al., 2008). The structure of Read 180® allows teachers to provide explicit instruction when students are in small groups. However, increasing the quantity and quality of teacher-lead explicit instruction focused on decoding, fluency, spelling, and writing during small group instruction could enhance student learning. Teacher-lead instruction could also allow for extended discussions of text meaning and interpretation. Another potential modification may be to integrate specific discussion protocols that provide students with the opportunity to discuss the text in small groups and ask follow-up questions incorporated into the curriculum.

Finally, providing teachers with the skills necessary to make more informed databased instructional decisions could lead to improved student outcomes. Read 180® provides teachers with built-in formative and summative assessments that could allow teachers to provide students with improved explicit instructional strategies based on individual student needs. Databased decision-making teacher preparation that capitalizes on the extensive amount of Read 180® could increase student achievement.

Limitations of study

Despite the rigor of implementing a randomized trial with repeated measures design, our study has several limitations that might affect the interpretation and generalization of the findings to other JC facilities. This study was conducted in a single facility, which may or may not resemble other facilities in terms of size, student characteristics, teacher quality, length of stay, or other organizational structures. Facility size influences how many personnel (i.e., teachers, paraprofessionals) are available to provide instruction (Houchins et al., 2010). The number of available personnel provided by a facility could influence the degree of instructional differentiation. Additionally, the length of stay could have influenced our findings. The average number of instructional hours for students in the sample was 137 h. Thus, a portion of students did not reach the 140-h assessment checkpoint. Had our study been conducted at a facility where the length of stay was longer or had our assessment points occurred earlier, our findings might have been different. Future studies should address such facility issues to determine the generalizability of our findings across and within states.

Additionally, the characteristics of students in this study may be different from other facilities. The facility used in this study served students

with significant mental health problems. We were unable to determine how mental health issues influenced literacy achievement, since the majority of students had similar mental health diagnoses. There was not enough diagnosis variability to compare student sub-groups. The use of more precise instruments to measure student mental health, behavioral issues, and the use of different types of psychotropic medications, could have influenced our findings. Additionally, our analyses did not take into consideration finding differences by disability classification. While students with disabilities are overrepresented in JC (Harris et al., 2009; McClelland et al., 2004; Puzzanchera & Adams, 2013; Quinn et al., 2005) and in this study, we were unable to disaggregate the data by disability (e.g., LD, ED). However, the major of students had an ED disability classification. Future studies should be sufficiently large enough to allow data to be analyzed by disability classification.

Conclusions

The need for evidence-based JC literacy instruction is substantial. While the study outcomes were modest, given the dismal reading levels of the incarcerated students, they are important. Improving the literacy abilities of incarcerated students has the potential to reduce recidivism and increase school involvement (Blomberg, Bales, Mann, Piquero, & Berk, 2011). If incarcerated students are to be academically successful and earn a high school diploma, greater student literacy growth is required. Moreover, for incarcerated to experience lifelong success and become contributing members of society, accelerated literacy instruction is crucial. The current study provides important data and conclusions that Read 180® has the potential to support effectively incarcerated youth literacy.

References

Abrams, L. (2013). Juvenile justice at a crossroads: Science, evidence, and twenty-first century reform. *Social Service Review*, *87*(4), 725–752. doi:10.1086/674074

Allen-DeBoer, R., Malmgren, K., & Glass, M. (2006). Reading instruction for youth with emotional and behavioral disorders in a juvenile correctional facility. *Behavioral Disorders*, *32*(1), 18–28.

American Psychiatric Association Task Force on DSM-IV. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Blomberg, T. G., Bales, W. D., Mann, K., Piquero, A. R., & Berk, R. A. (2011). Incarceration, education and transition from delinquency. *Journal of Criminal Justice*, *39*, 355–365. doi:10.1016/j.jcrimjus.2011.04.003

Bryan, K., Freer, J., & Furlong, C. (2007). Language and communication difficulties in juvenile offenders. *International Journal of Language & Communication Disorders*, *42*(5), 505–520. doi:10.1080/13682820601053977

Calderone, C., Bennett, S. V., Homan, S., Dedrick, R. F., & Chatfield, A. (2009). Reaching the hard to reach: A comparison of two reading interventions with incarcerated youth. *Middle Grades Research Journal*, *4*, 61–80.

Coulter, G. (2004). Using one-to-one tutoring and proven reading strategies to improve reading performance with adjudicated youth. *Journal of Correctional Education*, *55*, 321–333.

Davis, L. M., Steele, J. L., Bozick, R., Williams, M., Turner, S., Jeremy, N. V., … Steinberg, P. S. (2014). *How effective is correctional education, and where do we go from here? Theresults of a comprehensive evaluation*. Santa Monica, CA: The RAND Corporation.

Drakeford, W. (2002). The impact of an intensive program to increase the literacy skills of youth confined to juvenile corrections. *Journal of Correctional Education*, *53*, 139–144.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (PPVT-4)*. San Antonio, TX: Pearson.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. doi:10.1007/s10464-008-9165-0

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*, 43–53.

Gagnon, J. C., Houchins, D. E., & Murphy, K. M. (2012). Current juvenile corrections professional development practices and future directions. *Teacher Education and Special Education*, *35*(4), 333–344. doi:10.1177/0888406411434602

Gonsoulin, S., Griller Clark, H., & Rankin, V. E. (2015). *Quality education services are critical for youth involved with the juvenile justice and child welfare systems*. Washington, DC: National Evaluation and Technical Assistance Center for Children and Youth Who Are Neglected, Delinquent, or At-Risk (NDTAC).

Harris, P. J., Baltodano, H. M., Jolivette, K., & Mulcahy, C. (2009). Reading achievement of incarcerated youth in three regions. *Journal of Correctional Education*, *60*, 120–145.

Heward, W. L., McCormick, S. H., & Joynes, Y. (1980). Completing job applications: Evaluation of an instructional program for mildly retarded juvenile delinquents. *Behavioral Disorders*, *5*, 223–234.

Hill, N. C., Minifie, D. G., & Minifie, E. L. (1984). Handicapped adolescent delinquents: Educational diagnosis and tutoring. *Remedial and Special Education*, *5*(5), 44–45. doi:10.1177/074193258400500510

Horner, R. M., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165–179.

Houchins, D. E., Jolivette, K., Krezmien, M. P., & Baltodano, H. M. (2008). A multistate study examining the impact of explicit reading instruction with incarcerated students. *Journal of Correctional Education*, *59*(1), 65–85.

Houchins, D. E., Jolivette, K., Shippen, M. E., & Lambert, R. (2010). The advancement of high quality literacy research in juvenile justice: Methodological and practical considerations. *Behavioral Disorders*, *36*(1), 61–69.

Houghton Mifflin Harcourt. (2017). *Welcome Read 180 families*. Boston, MA: Author. Retrieved from http://www.hmhco.com/products/read-180/family/

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

Krezmien, M. P., Mulcahy, C. A., & Leone, P. E. (2008). Detained and committed youth: Examining differences in achievement, mental health needs, and special education status. *Education and Treatment of Children*, *31*, 445–464.

Lane, K. L., Carter, E. W., Pierson, M. R., & Glaeser, B. C. (2006). Academic, social, and behavioral characteristics of high school students with emotional disturbances or learning disabilities. *Journal of Emotional and Behavioral Disorders*, *14*, 108–117.

Leone, P., & Weinberg, L. (2010). *Addressing the unmet educational needs of children and youth in the juvenile justice and child welfare systems*. Washington, DC: Center for Juvenile Justice Reform. Retrieved from http://www.modelsforchange.net/publications/260

Lewis, B. A., Freebairn, L., Tag, J., Ciesla, A. A., Iyengar, S. K., Stein, C. M., & Taylor, H. G. (2015). Adolescent outcomes of children with early speech sound disorders with and without language impairment language disorders, languages, literacy. *American Journal of Speech-Language Pathology*, *24*(2), 150–163. doi:10.1044/2014_AJSLP-14-0075

Loadman, W. E., Moore, R. J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. (2011). *Striving readers year 5 project evaluation report: Ohio: An addendum to the year 4 report*. Columbus, OH: The Ohio State University. Retrieved from http://www2.ed.gov/programs/strivingreaders/springfieldyear5evalrpt.pdf

Malmgren, K., & Leone, P. (2000). Effects of a short-term auxiliary reading program on the reading skills of incarcerated youth. *Education and Treatment of Children*, *23*(3), 239–247.

Mayer, E. E., & Hoffman, R. A. (1982). A comparison of the effects of individualized instruction with group instruction on the academic achievement and self-concept of youthful offenders. *Journal of Correctional Education*, *33*(4), 11–13.

McClelland, G., Elkington, K. M., Teplin, L. S., & Abrams, K. M. A. (2004). Multiple substance use disorders in juvenile detainees. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*, 1215–1224.

Morsy, L., Kieffer, M., & Snow, C. (2010). *Measure for measure: A critical consumers' guide to reading comprehension assessments for adolescents. Final report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy*. New York, NY: Carnegie Corporation.

Murph, D., & McCormick, S. (1985). Evaluation of an instructional program designed to teach minimally literate juvenile delinquents to read road signs. *Education and Treatment of Children*, *8*, 133–151.

New Century Education Corporation. (2010). *New century learning system*. Piscataway, NJ: Author.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. doi:10.2105/AJPH.2013.301361

Platt, J. M., & Beech, M. (1994). The effectiveness of learning strategies in improving performance and increasing the independence of juvenile offenders with learning problems. *Journal of Correctional Education*, *45*(1), 18–29.

Puzzanchera, C., & Adams, B. (2013). *National disproportionate minority contact databook*. Washington, DC: National Center for Juvenile Justice for the Office of Juvenile Justice and Delinquency Prevention. Retrieved from http://www.ojjdp.gov/ojstatbb/dmcdb/

Quinn, M., Rutherford, R., Leone, P., Osher, D., & Poirer, J. (2005). Youth with disabilities in juvenile correctional: A national survey. *Exceptional Children*, *71*, 339–345.

Raudenbush, S. W., & Lui, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*, 387–401.

Sanger, D., Creswell, J. W., Dworak, J., & Schultz, L. (2000). Cultural analysis of communication behaviors among juveniles in a correctional facility. *Journal of Communication Disorders*, *33*, 31–57.

Scarlato, M. C., & Asahara, E. (2004). Effects of "Corrective Reading" in a residential treatment facility for adjudicated youth. *Journal of Direct Instruction*, *4*, 211–217.

Scholastic Inc. (2006). *Scholastic reading inventory: Educator's guide*. New York, NY: Author.

Scholastic Research. (2007). *READ 180 enterprise*. New York, NY: Scholastic Inc.

Scholastic Research. (2010). *System 44*. New York, NY: Scholastic Inc.

Shinn, M. R., & Good, R. H. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, *21*(3), 459–479.

Shippen, M. E., Morton, R. C., Flynt, S. W., Houchins, D. E., & Smitherman, T. (2012). Efficacy of a computer-based program on acquisition of reading skills of incarcerated youth. *Remedial and Special Education*, *33*, 14–22. doi:10.1177/0741932510362512

Sinatra, R. (1984). Use of three writing tasks at an adolescent treatment center. *Perceptual and Motor Skills*, *59*, 355–358. doi:10.2466/pms.1984.59.2.355

Smith, R. R. (2000). How the Lexile framework operates. *Popular Measurement*, *3*(1), 18–19.

Snow, P. C., Sanger, D. D., Caire, L. M., Eadie, P. A., & Dinslage, T. (2015). Improving communication outcomes for young offenders: A proposed response to intervention framework. *International Journal of Language & Communication Disorders*, *50*(1), 1–13.

Solis, M., Miciak, J., Vaughn, S., & Fletcher, J. M. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disability Quarterly*, *37*(4), 218–229.

Soller, M. V., Karnik, N. S., & Steiner, H. (2006). Psychopharmacologic treatment in juvenile offenders. *Child and Adolescent Psychiatric Clinics of North America*, *15*(2), 477–499. doi:10.1016/j.chc.2005.12.003

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2009). *WWC procedures and standards handbook version 2.1*. Washington, DC: Author. Retrieved from http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2010, July). *Read 180 intervention report*. Washington, DC: Author. Retrieved from http://whatworks.ed.gov

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2014). 36th annual report to Congress on the implementation of the Individuals with Disabilities Education Act, 2014. Washington, DC: Author.

Wasserman, G. A., McReynolds, L. S., Ko, S. J., Katz, L. M., & Carpenter, J. R. (2005). Gender differences in psychiatric disorders at juvenile probation intake. *American Journal of Public Health*, *95*, 131–137. doi:10.2105/AJPH.2003.024737

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI)*. San Antonio, TX: Psychological Corporation.

Wexler, J., Pyle, N., Flower, A., Williams, J. L., & Cole, H. (2014). A synthesis of academic interventions for incarcerated adolescents. *Review of Educational Research*, *84*(1), 3–46. doi:10.3102/0034654313499410

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson test of achievement*. Rolling Meadows, IL: Riverside Publishing.

Copyright of Residential Treatment for Children & Youth is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.