Published in final edited form as:

J Res Educ Eff. 2011 ; 4(2): 154–172. doi:10.1080/19345747.2011.555294.

Effects of a Structured Decoding Curriculum on Adult Literacy Learners' Reading Development

Judith A. Alamprese,
Abt Associates Inc.

Charles A. MacArthur,
University of Delaware

Cristofer Price, and
Abt Associates Inc.

Deborah Knight
Atlanta Speech School

Abstract

This article reports the results from a randomized control field trial that investigated the impact of an enhanced decoding and spelling curriculum on the development of adult basic education (ABE) learners' reading skills. Sixteen ABE programs that offered class-based instruction to Low-Intermediate level learners were randomly assigned to either the treatment group or the control group. Reading instructors in the 8 treatment programs taught decoding and spelling using the study-developed curriculum, *Making Sense of Decoding and Spelling (MSDS)*, and instructors in the 8 control programs used their existing reading instruction. A comparison group of 7 ABE programs whose instructors used K-3 structured curricula adapted for use with ABE learners were included for supplemental analyses. Seventy-one reading classes, 34 instructors, and 349 adult learners with pre- and posttests participated in the study. The study found a small but significant effect on one measure of decoding skills, which was the proximal target of the curriculum. No overall significant effects were found for word recognition, spelling, fluency, or comprehension. Pretest to posttest gains for word recognition were small to moderate, but not significantly better than the control classes. Adult learners who were born and educated outside of the U.S. made larger gains on 7 of the 11 reading measures than learners who were born and educated within the U.S. However, participation in the treatment curriculum was more beneficial for learners who were born and educated in the U.S. in developing their word recognition skills.

Adult Basic Education (ABE) programs funded by the U.S. Department of Education under the Adult Education and Family Literacy Act of 1998 are estimated to serve about 2.4M adults annually. Almost one third (31%) of the adults enrolled in ABE programs during the 2008–2009 program year entered instruction with reading comprehension skills at the fourth-to- ninth-grade level. (U.S. Department of Education, 2010). Categorized as either High or Low Intermediate-level learners on the U.S. Department of Education's National Reporting System (NRS) for Adult Education, these adults are able to read texts on familiar subjects with a clear underlying structure and can use context to determine meaning (U.S. Department of Education, 2007). Generally, adults at or below the Intermediate level lack the range of reading skills that they need to compete for family-sustaining jobs and are less

likely to be offered opportunities for advancement or have access to further educational training from employers (Tamassia, Lennon, Yamamoto, & Kirsch, 2007).

ABE programs have been hampered by the lack of evidence-based research on effective reading instruction that they can use to guide their services, particularly regarding approaches for teaching Intermediate-level learners enrolled in adult basic education and adult secondary education services. Syntheses of reading research have pointed to the lack of experimental studies on adult reading and the need for adult education instructors to rely on children's reading research for strategies to address adults' reading difficulties (Kruidenier, 2002).

This article discusses the results from an experimental study, focused on adults at the Low-Intermediate level (i.e., reading comprehension at roughly the fourth-to-seventh-grade level), which tested efficient and effective methods for teaching adults decoding and spelling. The study was motivated in part from findings from previous research that investigated the association between reading instruction and reading skill development in ABE programs. This research examined 643 low-literacy adults (i.e., below the seventh-grade level in reading comprehension) from 130 ABE reading classes in 35 ABE programs (Alamprese, 2009). The study found that learners made significant gains on six standardized reading tests used to assess their word recognition, decoding, vocabulary, comprehension, and spelling skills from pre- to posttest (9 months) and from pretest to follow-up (18 months). Learners in classes that emphasized phonics instruction and used a published scope and sequence had larger gains on decoding than did learners in the study's other classes.

One challenge the programs in the previous research encountered in implementing published phonics-based reading curricula was the time required for the curricula that did not align with ABE learners' attendance patterns. These published curricula are based on a K-12 schedule in which literacy is taught daily over nine months. Learners' average hours of attendance during the study period was 124 hours over about 22 weeks, which is the equivalent of two instructional sessions (e.g., fall and winter) (Alamprese, 2009). National data on ABE programs indicate that on average learners participate for less than 100 hours per program year (Tamassia et al., 2007). This lack of alignment between the time requirements for K12-adapted reading curricula and learners' patterns of participation in ABE motivated the current study's investigators to develop and test a curriculum that could be implemented in the average timeframe that adult learners participate in ABE.

Reading Skills of Adult Literacy Learners

The limited research on adults with literacy problems has found evidence of variability in reading skills and related cognitive processes. Comparing adults in ABE programs with reading-matched children, Greenberg, Ehri, and Perin (1997; 2002) found that the adults performed worse on phonological tasks but better on sight word recognition. Further analysis of word recognition and decoding errors showed that adults relied more on orthographic processes and less on phonological analysis than children. Binder and Borecki (2009) compared ABE learners to normal college readers on a homophone reading task and found that the ABE learners were less efficient in using phonological information and relied more on context. Consistent with this finding of weak phonological skills, three recent large studies of ABE learners (Mellard, Fall, & Woods, 2010; MacArthur, Konold, Glutting, & Alamprese, 2010; Sabatini, Sawaki, Shore, & Scarborough, 2010) found relatively lower performance on pseudoword decoding tests than on word recognition. Other research has documented problems with fluency (Sabatini, 2002) and spelling (Worthy & Viise, 1996.)

ABE learners' English proficiency has been found to be a factor in the variability of their performance on reading assessments in ABE reading classes. Substantial numbers of

learners in ABE reading classes are non-native English speakers who have sufficient oral English to participate in these classes and who often bypass English-as-a-Second Language (ESL) classes that are offered in adult literacy programs (Alamprese, 2009). A cluster analysis of adult education learners from ABE and ESL classes (Strucker, Yamamoto, & Kirsch, 2007) found five clusters, of which two were primarily native speakers with higher vocabulary than word recognition and two were primarily non-native speakers with lower vocabulary. The fifth group was mixed and demonstrated low performance in all reading components. The results from a study of the reading errors of native and non-native speakers of English in ABE indicated that non-native speakers scored lower on vocabulary compared to native speakers with equivalent word recognition skills. (Davidson & Strucker, 2002). Research on ABE reading classes (Alamprese, 2009) found similar results in which learners who were neither born nor educated in the U.S. performed better at pretest on word recognition and decoding assessments but less well on vocabulary and comprehension than learners who were born and educated in the U.S.

Overall, ABE learners have substantial difficulty with all reading skills including decoding, word recognition, fluency, vocabulary, and comprehension. There is some evidence that decoding skills may be particularly weak, and learners' English proficiency is a factor related to their development of reading skills.

Purpose of the Study

The purpose of this study was to develop and test the impact of a structured decoding curriculum, *Making Sense of Decoding and Spelling (MSDS)*, on the reading skills of adult literacy learners. It was intended to be used as one component of a comprehensive adult reading course in combination with instruction in vocabulary and comprehension. The curriculum was based on a morphological analysis of English orthography (Venezky, 1970; 1999) and was designed to be efficient because of the limited time for instruction typically available in adult basic education classes. This study focused on adults at the Low-Intermediate level (approximately fourth to seventh grade reading comprehension level) and was an experimental study to test efficient and effective methods for teaching adults decoding and spelling.

The study addressed the following research questions:

- What are the effects of an enhanced decoding and spelling curriculum on the reading skills of ABE learners?
- How are ABE learners' background characteristics, including their place of birth and education, and their attendance in reading classes related to the improvement of their reading skills?

Curriculum Design

Making Sense of Decoding and Spelling was designed to teach adult learners to decode and spell words more accurately and fluently. It begins with a review of basic alphabetic decoding skills and then teaches the most common and useful patterns of English words, and their applications in decoding, spelling, and fluent reading. The design was based on a theoretical framework and on design studies conducted in ABE classes. Discussion of the design studies is beyond the scope of this article, but some of the conclusions are mentioned below.

The core theoretical framework for the curriculum is Venezky's work (1970; 1999) on orthography, based on the understanding that a parsimonious analysis of English orthography requires analysis of morphemes and phonemes and their relationships. This

theoretical framework is consistent with stage or phase theories of spelling development (Ehri & McCormick, 1998; Templeton & Morris, 2000). From an early logographic or pre-alphabetic stage, children move to an early alphabetic phase in which they learn grapheme-phoneme relationships that enable them to read and spell highly regular words. Later, they learn about the orthographic patterns required because English has more phonemes than letters, and still later, they learn how morphemes influence spelling, particularly in multisyllabic words. Orthographic patterns and morphemes were selected for inclusion in the scope and sequence based on efficiency and productivity, i.e., the most common and useful patterns were taught. In addition, a few rules with high applicability were included.

One conclusion from the design studies was that adults were interested in understanding how English spelling and pronunciation work and that such metalinguistic information seemed to help them remember and apply what they learned. Thus, the curriculum includes explicit information about phonology, orthography, and morphology and includes occasional interesting items of etymology. The first lesson introduces the curriculum as the study of how the English language works and teaches the concepts of phonemes and syllables with a few exercises. Linguistic terms such as phoneme, suffix, and prefix are used freely, and notes on the English language are interspersed throughout the lessons. These notes raise motivation and help distinguish the program from what some learners recall from primary school reading classes. This metalinguistic principle influenced the name of the curriculum; we wanted learners to *make sense of decoding and spelling*.

The curriculum also includes a comprehensive metacognitive strategy for decoding multisyllabic words. A large body of research demonstrates the effectiveness of teaching strategies for reading, writing, and other academic tasks (Graham, 2006; Pressley, 2000). Strategies help learners to develop independence in applying knowledge to meaningful tasks. In the curriculum, the strategy is intended to support learners in using their new decoding knowledge while reading and writing. The strategy is modeled and practiced in the context of reading brief passages, and instructors encourage learners to use it during other class reading activities. The curriculum emphasizes the importance of flexibility in applying the strategy. No curriculum can teach all the morphophonemic patterns used in skilled reading. A flexible, strategic approach to decoding can get learners to attend to word structure so that they can extend their knowledge through reading (Juel & Minden-Cupp, 2000).

Fluent reading tends to lag behind development of accurate decoding and requires practice (Kuhn & Stahl, 2003). The design studies revealed that timing adults' reading led them to sacrifice accuracy for speed. Thus, the curriculum includes untimed repeated reading practice. Each lesson concludes with a brief *smooth reading* passage of 50–75 words that learners read repeatedly in pairs. Prior to this repeated reading, the passage is used to model and practice the multisyllabic decoding strategy.

Application of new skills to meaningful reading and writing is important at all ages and, perhaps, especially with adult learners (Beder, 2007; Wagner & Venezky, 1999). The smooth reading passage at the end of each lesson is one attempt to encourage such application. Each lesson begins with a brief text that is informative and contains words with the patterns taught in the lesson. As noted above, the metacognitive strategy is intended to support application. Instructors also are to encourage learners to apply the strategies during other reading activities.

Spelling draws on much the same knowledge base as decoding (Ehri, 2000), and it is integrated with decoding instruction in many instructional approaches, from invented spelling in whole language approaches (Clarke, 1988), to word study methods (Bear,

Invernizzi, Templeton, & Johnston, 2004), to structured remedial approaches (Wilson, 1996). Spelling requires attention to all the letters and patterns in words that may enhance the development of clear mental representations. In addition, interviews in the design studies indicated that spelling problems were highly salient to adults and that they were motivated to learn to spell better. Thus, spelling is integrated with the curriculum in practice exercises and in progress monitoring assessments.

Curriculum development also required decisions related to instructional delivery and staff development. First, instruction was designed primarily as whole group instruction. There are some paired fluency activities, and instructors support learners in individual application of the content and strategy during the rest of the class time, but the core instruction is delivered to groups. Second, progress monitoring assessments were included with each lesson, and a review lesson was included about every five lessons. Instructors could use the assessments to decide whether to re-teach and extend a lesson with the entire group or to provide additional work for individuals. Finally, based on the advice of instructors from the design studies, the lessons were scripted. The lesson plans included all the instructional steps and examples and presentation materials needed, and a student booklet included all materials needed by the learners. Instructors were expected to use the script as a guide in delivering the curriculum using their own words.

Method

Design

The study was a randomized control field trial with random assignment at the program level to treatment and control groups. Sixteen ABE programs that offered class-based reading instruction to adult learners at the Low-Intermediate level were recruited for the study. Eight programs were randomly assigned to the treatment group and eight to the control group. In the treatment group, reading instructors were trained to use the study curriculum to teach decoding and spelling, but used their own lessons for vocabulary and comprehension instruction. In the control group reading instructors continued their existing reading instruction. The study also involved a comparison group of seven ABE programs whose instructors used commercially produced K-3 structured decoding curricula adapted for use with adult learners. The data from the comparison programs were used in the study's supplemental analyses.

Sample

The 23 adult literacy programs included in the study were located in 12 states. All eligible reading classes (71), instructors (34), and adult learners were included in the sample. Five hundred and sixty-one learners were pretested, and there were 349 learners with both pre and posttests. All programs met three criteria: (a) provided class-based instruction to English-speaking adults at the intermediate level, (b) had a basic level of operations in learner recruitment, learner assessment, program management, program improvement, and support services (Alamprese, 1993), and (c) had instructors who were trained or experienced in teaching reading and whose instruction followed a discernable scope and sequence. All instructors in the study taught the components of reading and used some form of lesson plans. In each program, all reading classes that served the study's target population and whose instructors met the study's criteria were selected for the study. Half of the programs in the study had more than one class participate. To meet the sample requirements for the study, data were collected from programs over a three-year period. Each program in the treatment and control groups had two cohorts of classes, one for each of two years of the study, and the programs in the comparison group had three cohorts of classes with data collected over three years.

Learner characteristics—In each class in the study, all learners were recruited to participate in the data collection. Participation in the study was voluntary and 99 percent of the learners in the classes agreed to participate. Learners in the pre-post sample of 349 learners ranged in age from 16 to 76 years, with an average age of 37 years. The majority of learners were female (66%). The racial and ethnic distribution was as follows: White, 35%; Hispanic, 24%; Black, 20%; Asian, 15%; and other, 6%. All participants were sufficiently fluent in English to participate in English reading classes. The majority (65%) had been born or educated in the United States since primary grades (native); the remaining 35% were born and educated outside of the United States (non-native). Among those that were born and educated outside of the United States, about one third were Hispanic, about one third were Asian, and the remaining was either Black (18%) or White (12%). Place of birth was used to represent whether participants were native speakers of English because it was considered more reliable than self-reports of primary language as an indicator of native English proficiency. Included in the native group were individuals who received their education in the United States beginning in the primary grades because previous research indicated that ABE participants who migrated to the United States before the age of 12 performed more like native-born residents than like immigrants who came after the age of 12 (Davidson & Strucker, 2002). The education levels of learners varied, with 8% having less than a sixth-grade education, 44% having completed between 7 to 12 years of school, 15% having a high school diploma or General Educational Development (GED), and 33% having some education but not in the United States. Almost half (46%) were employed, although another 46% had been employed previously; 5% had never worked; and 3% were retired. More than half (63%) had an income below the poverty threshold of \$12,000, and almost one-third (31%) of the learners reported having a learning problem or disability at pretest.

No significant differences between the treatment and control group learners were found on any of the above demographic characteristics. However, relative to treatment and control group members, the comparison group learners were older and more likely to be Hispanic.

Instructor characteristics—Thirty-five instructors participated in the study; all but one was female. Close to two-thirds (63%) held Master's degrees; 31% had Bachelor's degrees; and 2 instructors (from the control and comparison groups) had completed less than a Bachelor's degree. This level of educational attainment is similar to that reported in prior studies of adult education instructors (Alamprese, Tao, & Price, 2003; Smith & Hofer, 2003). Over half (61%) of the instructors had an academic specialty in education, with 29% specializing in reading. Of these instructors, 41% were in the treatment group and 27% were in the control group. Half (53%) of the instructors were full-time instructors, which is similar to the data (52%) reported in a nationally representative study of ABE programs (Tamassia et al., 2007). The instructors were experienced, with 82% having taught reading more than five years and 63% having taught adult education for more than five years. Furthermore, 44% of the instructors had taught the targeted study reading class for more than five years. All but one instructor had participated in formal reading training. There were no significant differences between treatment and control instructors on any of these characteristics.

Classes—For each cohort of classes, instruction lasted approximately eight months or about 30 weeks. However, the time between pretest and posttest and the hours of instruction varied considerably for individual learners due to their attendance. Data on hours of instruction are reported in Results. The classes in the study met from one to five days per week. About half (52%) met twice per week, and about a quarter (27%) met in the evenings. Control classes met more frequently than treatment classes. Of control classes, 24% met twice a week and 76% met more often; of treatment classes, 84% met twice a week and 16%

met more often. Of comparison classes, 44% met twice, 48% met more often, and 8% met just once a week.

Control group instruction—The control group continued their existing reading instruction that included teaching reading components but not following a published scope and sequence. These instructors varied in their approaches to organizing reading instruction, such as using reading pretest results to identify the reading skills to teach or selecting chapters from published reading workbooks as a guide for instruction. While most control instructors taught some decoding in their classes, they emphasized spelling, vocabulary, and comprehension rather than decoding. Generally the control teachers' instruction was less systematic than the curriculum used by the treatment instructors. They varied the sequence of the reading skills that they taught from class to class and some adapted their lessons based on their perceived needs of the learners during a particular class.

Measures

Learner measures—Eleven measures of reading skills were administered. The *Nelson Reading Test* (Hanna, Schell, & Schreiner, 1977) was administered to classroom groups and yielded scores for vocabulary and comprehension. The *Nelson Word Meaning* (NWM) test assesses vocabulary with items that present a term in a sentence and a choice of meanings. The *Nelson Reading Comprehension* (NRC) test presents short passages followed by multiple-choice questions. The Nelson was standardized for Grades 3 through 9. Internal consistency reliability ranged from .81 to .93 on vocabulary and comprehension.

The remaining tests were administered individually, and the oral reading tests were audiotaped. The Reading and Spelling subtests of the *Wide Range Achievement Test-Revision 3* (WRAT3; Wilkinson, 1993) were used. The WRAT3 Reading (WRAT3-R) subtest assesses ability to read words in isolation. The WRAT3 Spelling (WRAT3-S) subtest assesses ability to spell individual words from dictation. Internal consistency ranged from .85 to .95 and test-retest correlations were .98 and .96 for reading and spelling.

Two subtests of the *Woodcock-Johnson Tests of Achievement Revised* (Woodcock & Johnson, 1989) were administered. The Letter-Word Identification (WJR-LW) subtest assesses ability to read words in isolation (and to identify letters at the lowest levels). The Word Attack (WJR-WA) subtest requires pronunciation of pseudowords, nonwords that follow the phonological, orthographic, and morphological patterns of English. Internal consistency ranged from .87 to .95 across age groups.

Two subtests of the *Test of Word Reading Efficiency* (TOWRE; Torgeson, Wagner, & Rashotte, 1999) were given. The Sight Word Efficiency (TOWRE-SWE) subtest presents words of increasing difficulty and tests how many words a person can read in 45 seconds. The Phonemic Decoding Efficiency (TOWRE-PDE) subtest has the same format but uses pseudowords. Internal consistency ranged from .93 to .94 on sight word and phonemic decoding.

The *Letter-Sound Survey* (LSS) was developed by our study (Venezky, 2003) to assess decoding. It consists of 26 pseudowords of one or two syllables that represent common phonological, orthographic, and morphological patterns. The item set designed did not include words that were part of the treatment curriculum. Words were scored as correct or incorrect. Internal consistency (*Cronbach's alpha*) for our sample at pretest was .86.

The *Passage Reading Test* (PR) is a measure of Oral Reading Fluency (ORF) that uses a passage developed for the fluency test in the National Assessment of Adult Literacy (NAAL) (see Bauer, Kutner, & Sabatini, 2009). The passage is 161 words long and is

written at a fourth-grade level according to the Flesch-Kincaid index. Adults were told to read the passage at a comfortable speed and to skip words that they could not figure out. The reading was audiotaped and timed, and then scored for correct words per minute. Internal consistency for our pretest sample was .96.

A developmental spelling (DS) test also was developed by our study. It consists of 20 words of increasing difficulty that represent common phonological, orthographic, and morphological patterns in English. Words that were in the curriculum were intentionally avoided; commercially available tests reviewed had more words in common with the curriculum. For the present purposes, words were scored correct or incorrect. Internal consistency for our pretest sample was .89.

In addition, a learner background interview was administered to gather information on learners' demographics, education, employment, health and disabilities, goals for participating in the program, and literacy activities at home and work.

Instructor measures—An Instructor Background Characteristics Form was used to collect data on instructors': (a) demographic characteristics, (b) experience and credentials, and (c) participation in reading professional development and state leadership activities.

Measures of instruction and fidelity of treatment—Instructors' teaching activities were measured through the use of a Class Observation Form that documented: (a) the time each lesson segment began, (b) the information taught during each lesson segment, (c) instructor and learner interactions during the lesson segment, and (d) materials used during the lesson segment. An instructor interview protocol also was used to collect information about instructors' approach to teaching reading instruction, the instructional activities for the lesson that was observed, and instructors' use of computers and homework.

The fidelity of treatment for the experimental curriculum was measured by treatment instructors' completion of a Teacher Feedback form for each lesson in the curriculum and by the class observation and instructor interview instruments described above. The Teacher Feedback form was customized for each lesson with the list of the segments that were conducted during the lesson. Each lesson's form was designed with a grid in which instructors were asked to check whether they had: (a) taught the segment according to the script, (b) modified the segment, or (c) not taught the segment. Instructors also indicated whether they or the learners had difficulty with each segment in the lesson. Data on number of lessons completed and adherence to the script were calculated from the form. As an additional check, the Teacher Feedback Forms for the observed classes were compared to the Class Observation Forms for the treatment classes observed.

Procedures

Learner data collection—The 11 reading tests and the learner background interview were administered by 40 individuals from the study's 23 ABE programs. These test administrators were ABE professional staff members who (a) had experience in administering reading tests, (b) had worked with low-literacy adult learners, and (c) were not scheduled to teach any of the ABE reading classes in the study. Prior to collecting data, all test administrators participated in a three-day training session conducted by one of the senior researchers, and follow-up training was conducted via telephone and email. Note that throughout the study the 7 tests that required oral responses were audiotaped for later scoring and determination of interrater reliability.

The test administrators were responsible for scoring the tests that involved establishing basal and ceiling levels (WJR-LW, WJR-WA, WRAT3-R, and WRAT3-S) and the LSS. The

remaining 7 tests were scored by the research project staff. The subtests of the *Nelson Reading Test* (NWM and NRC) were scored using answer keys provided by the test developer. The spelling test (DS) was scored from the written responses. Three tests involving oral responses (TOWRE-PDE, TOWRE-SWE, and PR) were scored from audiotapes. The research project staff also scored a sample of tests scored by each test administrator to determine reliability. The study adopted the scoring guidelines for native speakers and non-native speakers of English that were developed by Strucker (2004) and used in his study of the reading development of ABE learners. These guidelines take into account regional variations in speech, dialects, and foreign accents.

Eight research project staff with backgrounds in test administration and reading were trained by the senior researcher to score the oral reading tests. The senior researcher established reliability with the first cohort of data collectors and with a lead scorer from the project staff, who then became the standard for test scoring with subsequent staff scorers. Each staff scorer scored 12 test batteries that were independently scored by the senior researcher or the lead scorer. Ranges of inter-rater reliability between the lead scorer and the staff scorers were as follows: WJR-LW, .94 to .96; WJR-WA, .88 to .90; WRAT3-R, .90 to .94; TOWRESWE, .94 to .98; TOWRE-PDE, .82 to .88; LSS, .88 to .99; and PR, .95 to .99. Reliability for test administrators was calculated by rescoring a sample of half of the tests given by each test administrator. Staff test scorers whose reliability on any test was below .90 were identified, and the tests that had been scored by these individuals were rescored by the lead scorer.

Instructor data collection procedures—Four senior members of the study's research team conducted the class observations and the face-to-face interviews with the instructors. The first class observation and instructor interview were conducted during the first three months of a program's participation in the study, and the second observation and interview were conducted during the second year of data collection. Instructors' Background Characteristics Forms were completed by instructors at the time of the observation visit. Inter-observer reliability was established among the observers through their documentation of video-taped adult reading classes from prior research (Alamprese, Tao, & Price, 2003). The observers established a documentation reliability of 92 percent.

Training of instructors—Treatment instructors participated in two, 2-day training workshops to prepare them to teach the study curriculum. The first workshop was held during the summer prior to the first year that the study classes were taught. Throughout the first year of the classes, program staff were available via email and telephone to provide technical assistance to the instructors on their use of the curriculum. A refresher workshop was held during the summer prior to the second year of the study classes to discuss the instructors' experiences using the curriculum and address any implementation questions. Technical assistance also was available to the treatment instructors during their second year of study classes. Information on treatment instructor fidelity is reported later.

Analysis

Analysis approach—Five types of analyses were conducted. Descriptive analyses included: (a) learner characteristics and baseline reading skill levels, (b) classes and instructors, and (c) the amount of change in reading skill outcomes from baseline to follow-up assessments. Baseline balance testing was undertaken to determine whether learners in the treatment, control, and comparison groups were comparable on demographic measures and baseline reading skill levels. An experimental impact analysis and comparison of treatment and control group outcomes to outcomes of learners in the comparison group were conducted. Subgroup analysis was performed to determine if impacts varied for particular

subgroups. Non-experimental exploratory analysis of predictors of gain was conducted and included analyses to identify the learner characteristics that are associated with increased gains in reading skills, and analyses to determine whether increased class attendance was associated with greater gains.

Analysis of change in reading skill outcomes—The analysis approach described below was used to assess pretest to posttest change for each of the reading skill outcome measures. Gain scores were constructed by subtracting each learner's pretest scale score from his/her posttest scale score. To produce a standardized score, the gain score was divided by the pooled pre-test standard deviation.

The mean gain for the entire sample or for a subgroup (e.g., native learners) was calculated as the mean of the standardized gains. We tested the null hypothesis that learners did not improve in their test scores, versus the alternative that there was change in their average test scores using a one-sample t-test of the standardized gain scores. This is equivalent to a paired t-test. We note that our analyses of the distribution of gain scores indicated no major challenges to the distributional assumptions associated with the use of the t-test methodology.

Models for estimation of treatment impacts on learner skill levels—Impacts were estimated in two-level hierarchical linear models where learners (level-1) were nested in programs (level-2). Since there was no sampling of classes (all eligible classes were included in the sample), there was no need for a third level to represent classes in the impact models (Schochet, 2008). The models simultaneously produced estimates of experimental impact (i.e., treatment group contrasted to the control group), and estimates of outcome differences between the comparison group and the treatment and control groups.

To increase precision of the estimates of treatment and control differences in outcomes, and to increase precision and reduce possible baseline differences in learner characteristics between comparison group and treatment and control group learners, baseline covariates were included in the impact model. All impact models controlled for pretest scores and an indicator for whether learners were non-native (vs. native) learners.

A list of baseline learner-level characteristics thought to be correlated with outcomes was identified a priori. This list was informed by prior research in modeling predictors of reading assessment outcomes in adult learners (Alamprese, Tao, & Price, 2003). Decision rules were used to determine which variables from the a priori list should be included in the impact model for each outcome. Pretest scores and an indicator for whether students were non-native (vs. native) learners were included in all impact models for reasons of face validity. The remaining variables included age, and indicators for race/ethnicity, current employment at baseline, gender, and indicators for several types of health issues or disabilities at baseline. Budz-Jorgensen (2007) and Maldonado and Greenland (1993) have shown that using a p-value <0.20 criterion for deciding whether to include or drop a covariate from a model is a good method for identifying and retaining variables that either control for confounding or increase precision. In this method, an initial model includes all covariates and, starting with the covariate with the largest p-value, the covariate is dropped and the model is refitted to the data. The process iterates until only those covariates with p-values less than 0.20 are retained in the model. The utility of this method for improving precision also has been shown in Price (2008).

We show treatment impacts and mean differences between comparison and control, and between comparison and treatment groups in standardized effect size units. The impacts and

mean differences obtained from the model described above are converted to standardized effect size units by dividing the estimates by the standard deviation of the pretest scores.

Results

Preliminary Analyses

Sample sizes and attrition—Baseline reading tests and learner background interviews were administered to 561 students in 76 classes, nested in 23 programs. Posttest data were obtained for 349 learners in 71 classes, nested in 23 programs. Posttest scores were obtained from 62% of learners that were tested at baseline for an attrition rate of 38%. There were no significant differences between the treatment and control groups in the learner-level or class-level attrition. The proportion of learners that were posttested was higher in the comparison group than in the treatment or control groups.

Attrition analysis found no significant difference on baseline reading test scores between the pre-posttested group ($n=349$) and the attritted (non-posttested) group ($n=212$). However, the attritted group was younger, less likely to be non-native, more likely to be Hispanic and less likely to be White, more likely to have less than a high school education, more likely to report that they had never been employed, more likely to have a physical handicap when growing up, more likely to have had a drug or alcohol problem when growing up, and more likely to be single.

Reading skill levels—Means and standard deviations and grade equivalents of baseline and follow-up reading assessment scores for the full sample of pre-and-posttested learners are shown in Table 1. There were no significant differences between treatment and control group scores on any of the baseline reading tests. The comparison group had significantly lower scores than the treatment group on the TOWRE-SWE and the Passage Reading Test, but did not differ significantly from the treatment or control groups on any of the other measures.

Learner attendance—The study programs provided the daily attendance of learners in the study as well as the number of hours and weeks that the study class was in session during each year of data collection. Learners' attendance was calculated for the number of hours of reading instruction that learners received between their pre- and posttest, since some study classes included instruction other than reading (e.g., math) and the date of the pre- and posttests varied among learners. The mean number of hours of reading instruction that learners received between pre- and posttest was 57. There was variation but no significant differences in mean hours among the groups: treatment, 50 hours; control, 60 hours; comparison, 65 hours. The study data on attendance align with data from a national representative study of ABE programs which found that almost half (40%) of ABE programs reported that learners received 30 to 50 hours of instruction between the administration of the pretest and the posttest during a program year (Tamassia et al., 2007).

We also examined the amount of reading instruction that learners received out of the total hours of reading instruction available if they had attended all classes between the pre- and posttests. Learners accessed about half (54%) of the reading instruction that was available, with small but not significant differences among the groups: treatment, 55%; control, 51%; and comparison, 57%. In the exploratory analyses that we conducted, attendance did not predict differences between treatment and control groups.

Fidelity of treatment—There was some variation in the number of study lessons that was taught in the classes, the number of hours that the study curriculum was offered, and in the treatment instructors' fidelity to the scripted lessons. The median percentage of lessons

taught by class was 92% with a range from 28% to 100%. In 44% of the classes, all of the study lessons were taught. In 56% of the classes, more than 90% of the lessons were taught. In all classes except for one, at least 60% of the lessons were taught. The mean total hours devoted to the study lessons by class was 27 hours with a range from 12 to 51. The mean of 27 hours represents 54% of the reading instruction that treatment learners received. While we estimated that the study curriculum would take approximately 35 minutes to teach out of a one-hour reading lesson, on average the treatment instructors spent more than 35 minutes teaching *MSDS*, particularly in the first year of the study as they were becoming familiar with the lessons. The remainder of each hour was spent primarily on vocabulary and comprehension instruction.

Fidelity scores for instructors' use of the study curriculum were calculated from the Instructor Feedback forms. Scores ranged from 0.5 (almost none of the lesson segments taught or taught as scripted) to a possible score of 3.0 (all segments taught and all taught as scripted), with a mean score of 2.18.

Impact on Reading

Gains on the 11 reading measures for the treatment and control groups are presented in Table 2. The treatment group made significantly greater gains than the control group on the Woodcock-Johnson-R Word Attack test with a small effect size of 0.19. Although differences were in the expected direction for all of the decoding, spelling, and word recognition measures, none of them were statistically significant and all were small (Cohen, 1988).

Independent of group, there were significant differences in the reading gains of native and non-native learners (See Table 3). Non-native learners made significantly greater gains on both word recognition measures (WJR-LW and WRAT3-R), two decoding measures (WJR-WA, TOWRE-PDE, and LSS), the experimenter-designed spelling measure (DS), and reading comprehension (NRC). Native learners made significantly larger gains on vocabulary (NWM). The results for word recognition and decoding are similar to those found in a previous study that examined the association between reading instruction and reading skill development for low-level ABE learners (Alamprese, 2009).

To understand whether treatment effects were different for native and non-native learners, tests for interactions between treatment and native status were conducted for all 11 outcome measures. Significant interactions were found for the Woodcock-Johnson-R Letter-Word Identification test and the WRAT3 Word Reading test ($p < .05$). Impacts were estimated separately for each of the two groups (See Table 4). In the native group, the estimated impact of treatment on Word Reading was positive (0.24) but not significantly different than zero ($p > 0.05$). In the non-native group, the treatment impact was negative (-0.21) but not significantly different than zero ($p > 0.05$). Similar results were obtained for Letter-Word Identification (Table 4). Given the smaller size of these subgroups, these analyses were underpowered. Gains on the 11 reading measures for the comparison group and effects compared to the treatment and control groups are presented in Table 5. None of the differences were statistically significant.

Discussion

The overall purpose of this study was to evaluate the impact of a structured decoding curriculum on the reading skills of adult literacy learners. The curriculum was based on a morphological analysis of English orthography (Venezky, 1970; 1999) and was designed to be efficient because of the limited time for instruction typically available in adult basic education classes. It taught basic patterns for decoding and spelling along with a

metacognitive strategy for decoding multisyllabic words intended to be applied during reading activities beyond the curriculum. The study found a small but significant effect on one measure of decoding skills (WJR-WA), which was the proximal target of the curriculum. However, no overall significant effects were found for word recognition, spelling, fluency, or comprehension. Pretest to posttest gains for a second decoding measure (LSS) and word recognition (WTR-LW) and WRAT3-R were small to moderate, but not significantly better than the control classes. Given that the study was a field study with instruction implemented by regular adult education reading instructors given normal resources, even small positive effects may be educationally significant. Very little research is available on methods for teaching reading to adult basic education learners.

One explanation for the limited positive results is that learners' participation in instruction was modest. On average, adults in the treatment and control groups received 50 and 60 hours of reading instruction, respectively, which was approximately 55% and 51% of the reading instruction that was available in the classes. In the treatment classes, approximately half of this time (mean = 27 hours) was devoted to the enhanced decoding curriculum. For the treatment instruction that was provided, instructors' fidelity in teaching *MSDS* was adequate. The median percentage of lessons taught across classes was 92%. All instructors except one taught at least 60% of the lessons. Fidelity to the lesson scripts, self-reported, was reasonably good. While the treatment curriculum generally was implemented as intended, learners' varied attendance in *MSDS* classes resulted in many learners receiving only a portion of the treatment. It is interesting to note that national data reveal a decrease in attendance in adult education for the target population of this study during the period of the study's data collection (2004–2006). The most recent year for which national ABE data are available (2008–2009) indicate that attendance levels for ABE Intermediate-level learners have returned to their 2002–2003 level of approximately 100 hours in a program year (U.S. Department of Education, 2010).

Effects differed for native and non-native adults (based on whether they were born and educated from the primary grades in the U.S.). A significant interaction between treatment and native status was found for both measures of word recognition, indicating that the curriculum was more effective for the native than for the non-native adults. Analysis of the gains for all study participants regardless of group assignment found that the non-native adults gained more than the native adults on 7 of the 11 measures, including both word recognition tests and two decoding tests. Thus, it is interesting to speculate on why the treatment had a relatively larger effect on word recognition for the native adults. This interaction effect was found only for word recognition, not for decoding. One possible explanation is that the emphasis of the curriculum on morphological as well as phonological patterns in words interacts with vocabulary knowledge to increase word recognition. That is, learners with relatively greater knowledge of English vocabulary are better able to use the patterns they learn to figure out real English words. Further research would be needed to test this possibility. The sample size for the non-native learners was too small for confident interpretation of results.

This research study is one of a limited number of randomized control trials that have tested the use of a curriculum in operating ABE programs. Despite barriers such as learners' limited participation, the successful implementation of the study's data collection activities demonstrates the feasibility of conducting rigorous research in ABE programs. The ABE programs and instructors implemented *MSDS* sufficiently well to produce some positive effects on learning. Research in operating ABE programs using regular instructors is important for building a research base for the adult education field. Rigorous research on new interventions is needed to answer questions about the length of an intervention that can be well implemented in an ABE program, the time required for ABE instructors to become

proficient in the use of a new curriculum, and whether instructors' degree of proficiency is related to learners' skill development. The recent data on the average attendance of the target ABE population offer hope for future research in ABE programs in which instructional interventions can be fully tested.

Acknowledgments

This research was supported by a grant to the University of Delaware and Abt Associates Inc. jointly funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R01HD43798), the National Institute for Literacy, and the Office of Vocational and Adult Education of the U.S. Department of Education.

References

- Alamprese, JA.; Tao, F.; Price, C. Study of reading instruction for low-level learners in adult basic education, v. 1. Bethesda, MD: Abt Associates Inc.; 2003.
- Alamprese, J. Research Triangle Institute, Alternative designs for evaluating workplace literacy programs. Research Triangle Park, NC: Research Triangle Institute; 1993. Key components of workplace literacy projects and definition of project "models."
- Alamprese, J. Developing learners' reading skills in adult basic education programs. In: Reder, S.; Bynner, J., editors. Tracking adult literacy and numeracy skills: Findings from longitudinal research. New York, NY: Routledge; 2009. p. 107-131.
- Baer, J.; Kutner, M.; Sabatini, J. Basic reading skills and the literacy of America's least literate adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) supplemental studies, (NCES 2009-481). Washington, DC: U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics; 2009.
- Bear, DR.; Invernizzi, M.; Templeton, S.; Johnston, F. Words their way: Word study for phonics, vocabulary, and spelling instruction. 3rd edition. Upper Saddle River, NJ: Merrill; 2004.
- Beder, H. Quality instruction in adult literacy education. In: Belzer, A., editor. Toward defining and improving quality in adult basic education. Mahwah, NJ: Erlbaum; 2007. p. 87-106.
- Binder K, Borecki C. The use of phonological, orthographic, and contextual information during reading: a comparison of adults who are learning to read and skilled adult readers. Reading and Writing. 2009; 21:843-858.
- Budtz-Jorgensen E, Keiding N, Grandjean EM, Weihe P. Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure. Annals of Epidemiology. 2007; 17:27-35. [PubMed: 17027287]
- Clarke LK. Invented versus traditional spelling in first graders' writings: Effects on learning to spell and read. Research in the Teaching of English. 1988; 22:281-309.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd Ed.. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- Davidson RK, Strucker J. Patterns of word-recognition error among adult basic education native and nonnative speakers of English. Scientific Studies of Reading. 2002; 6:299-316.
- Ehri LC, McCormick S. Phases of word learning: Implications for instruction with delayed and disabled readers. Reading and Writing Quarterly: Overcoming Learning Difficulties. 1998; 14:135-163.
- Ehri LC. Learning to read and learning to spell: Two sides of a coin. Topics in Language Disorders. 2000; 20:19-36.
- Graham, S. Strategy instruction and the teaching of writing: A meta-analysis. In: MacArthur, CA.; Graham, S.; Fitzgerald, J., editors. Handbook of Writing Research. New York: Guilford; 2006. p. 187-207.
- Greenburg D, Ehri L, Perin D. Are word-reading processes the same or different in adult literacy students and third--fifth graders matched for reading level? Journal of Educational Psychology. 1997; 89:262-275.
- Greenburg D, Ehri L, Perin D. Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? Scientific Studies of Reading. 2002; 6:221-243.

- Hanna, G.; Schell, LM.; Schreiner, R. The Nelson Reading Skills Test. Itasca, IL: Riverside; 1977.
- Juel C, Minden-Cupp C. Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*. 2000; 35:458–492.
- Kruidenier, J. Research-based principles for adult basic education: Reading instruction. Washington, DC: National Institute for Literacy; 2002.
- Kuhn MR, Stahl SA. Fluency: A review of developmental and remedial practice. *Journal of Educational Psychology*. 2003; 95:3–21.
- MacArthur CA, Konold TR, Glutting JJ, Alamprese JA. Reading component skills of learners in adult basic education. *Journal of Learning Disabilities*. 2010; 43:108–121. [PubMed: 20179306]
- Maldonado G, Greenland S. Stimulation of confounder-selection strategies. *American Journal of Epidemiology*. 1993; 138(11):923–936. [PubMed: 8256780]
- Mellard DF, Fall E, Woods KL. A path analysis of reading comprehension for adults with low literacy. *Journal of Learning Disabilities*. 2010:154–165. [PubMed: 20179309]
- Pressley, M. What should comprehension instruction be the instruction of?. In: Kamil, ML.; Mosenthal, PB.; Pearson, PD.; Barr, R., editors. *Handbook of Reading Research*. Vol. Vol. 3. Mahwah, NJ: Erlbaum; 2000. p. 545-562.
- Price, C.; Goodson, B.; Stewart, G. Technical report, v. II: Infant environmental exposures and neuropsychological outcomes at ages 7 to 10 years. Atlanta, GA: Centers for Disease Control and Prevention; 2008.
- Sabatini J. Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*. 2002; 6(3):267–298.
- Sabatini JP, Sawaki Y, Shore J, Scarborough H. Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities*. 2010; 43:122–138. [PubMed: 20179307]
- Schochet PZ. Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*. 2008; 33:62–87.
- Smith, C.; Hofer, J. The characteristics and concerns of adult basic education teachers. Cambridge, MA: National Center for the Study of Adult Learning and Literacy; 2003.
- Strucker, J. TOWRE scoring guidelines. Cambridge, MA: National Center for the Study of Adult Learning and Literacy; 2004.
- Strucker, J.; Yamamoto, K.; Kirsch, I. The relationship of the component skills of reading to IALS performance: Tipping points and five classes of adult literacy learners. Cambridge, MA: National Center for the Study of Adult Learning and Literacy (NCSALL); 2007 March.
- Tamassia, T.; Lennon, M.; Yamamoto, K.; Kirsch, I. Adult education in America: A first look at results from the adult education program and learner surveys. Princeton, NJ: Educational Testing Service; 2007.
- Templeton, S.; Morris, D. Spelling. In: Kamil, ML.; Mosenthal, PB.; Pearson, PD.; Barr, R., editors. *Handbook of Reading Research*. Vol. Vol. 3. Mahwah, NJ: Erlbaum; 2000. p. 525-544.
- Torgeson, JK.; Wagner, RK.; Rashotte, CA. Test of Word Reading Efficiency (TOWRE). Austin, TX: ProEd; 1999.
- U.S. Department of Education, Division of Adult Education and Literacy. Implementation guidelines: Measures and methods for the National Reporting System for Adult Education. Washington, DC: 2007.
- U.S. Department of Education, Division of Adult Education and Literacy. State administered adult education program: Program year 2008–2009 enrollment. Washington, DC: 2010.
- Venezky, RL. The structure of English orthography. The Netherlands: Mouton & Co; 1970.
- Venezky, RL. The American way of spelling: The structure and origins of American English orthography. New York, NY: Guilford; 1999.
- Wagner DA, Venezky RL. Adult literacy: The next generation. *Educational Researcher*. 1999; 28(1): 21–29.
- Wilkinson, GS. Wide Range Achievement Test-Revision 3. Wilmington, DE: Jastak Associates, Inc.; 1993.
- Wilson, BA. Wilson reading system. Millbury, MA: Wilson Language Training Corporation; 1996.

- Woodcock, R.; Johnson, MB. Woodcock-Johnson Tests of Achievement-Revised. Itasca, IL: Riverside; 1989.
- Worthy J, Viise NM. Morphological, phonological, and orthographic differences between the spelling of normally achieving children and basic literacy adults. *Reading and Writing*. 1996; 8:139–154.

Table 1

Mean Test Scores and Grade Equivalents at Baseline and Follow-up for Pre-post Sample

Test	Pretest Overall Score Mean (SD)	Pretest Grade Equivalent	Posttest Overall Score Mean (SD)	Posttest Grade Equivalent
Decoding				
WJ-R Word Attack	487.95 (15.8) ^a	3.0–3.3	492.14 (16.3) ^a	3.5–3.8
Letter–Sound Survey	13.01 (6.0) ^b	--	14.41 (6.4) ^b	--
TOWRE Phonemic Decoding	97.32 (15.9) ^c	3.6	97.48 (16.4) ^c	3.6
Spelling				
WRAT3 Word Spelling	498.47 (11.7) ^d	4	499.89 (12.1) ^d	4
Study Spelling Test	6.83 (4.9) ^b	--	7.4783 (5.2) ^b	--
Word Recognition				
WRAT3 Word Reading	500.39 (12.8) ^d	4	504.75 (14.1) ^d	5–6
WJ-R Letter-Word Identification	497.44 (23.0) ^a	4.7–5.1	499.49 (20.7) ^a	5.1–5.4
Comprehension and Vocabulary				
Nelson Word Meaning	38.41 (15.7) ^e	5.4	41.18 (19.6) ^e	5.8
Nelson Reading Comprehension	36.23 (16.4) ^e	4.6	38.24 (16.9) ^e	4.8
Fluency				
TOWRE Sight Word Efficiency	101.69 (14.7) ^c	4.0	101.54 (15.4) ^c	4.0
Passage Reading Test (Words Per Minute)	106.64 (37.7) ^b	--	106.92 (36.1) ^b	--

^aW score;^bRaw score;^cStandard score (scores were standardized to norms for 9 year-olds;^dAbsolute score;^eStandard score (scores were standardized relative to norms for spring of 6th grade)

Table 2

Overall Impact on Reading Gains

Test	Treatment Group Mean Gain ^a (n=163)	Control Group Mean Gain ^a (n=98)	Impact ^b (T – C)	P-value
Decoding				
WJ-R Word Attack	0.35	0.16	0.19 *	0.047
Letter-Sound Survey	0.29	0.26	0.03	0.828
TOWRE Phonemic Decoding	0.05	-0.01	0.06	0.573
Spelling				
WRAT-3 Word Spelling	0.15	0.04	0.11	0.098
Study Spelling Test	0.16	0.09	0.07	0.256
Word Recognition				
WRAT-3 Word Reading	0.43	0.32	0.11	0.371
WJ Letter-Word Identification	0.11	0.09	0.02	0.912
Comprehension and Vocabulary				
Nelson Word Meaning	0.08	0.31	-0.23	0.072
Nelson Reading Comprehension	0.13	0.17	-0.04	0.817
Fluency				
TOWRE Sight Word Efficiency	0.07	-0.03	0.10	0.360
Passage Reading Test	0.05	0.06	-0.01	0.945

*
p < .05

^a Mean gains are the model-adjusted pre-test to post-test gains expressed in effect size units. All models included as covariates the pre-test score and an indicator for whether the learner was born and educated outside of the United States.

^b Impacts are expressed in standardized effect size units.

Table 3

Gains for Native and Non-Native Learners for Total Sample

Test	Native (N=226) Effect Size	Non-Native (N=123) Effect Size
Decoding		
WJ-R Word Attack	0.24	0.33 *
Letter –Sound Survey	0.17	0.37 *
TOWRE Phonemic Decoding	0.02	–0.05
Spelling		
WRAT3 Word Spelling	0.09	0.14
Study Spelling Test	0.06	0.23 **
Word Recognition		
WRAT3 Word Reading	0.21	0.58 **
WJ-R Letter-Word Identification	0.05	0.17 *
Comprehension and Vocabulary		
Nelson Word Meaning	0.27	0.01 *
Nelson Reading Comprehension	0.03	0.28 **
Fluency		
TOWRE Sight Word Efficiency	–0.02	0.05
Passage Reading Test	0.0008	0.09 *

*
 $p > .05$ **
 $p < .01$

Table 4

Interactions between Treatment and Native/Non-Native

Test	Treatment Group Mean Gain ^a	Control Group Mean Gain ^a	Impact ^b (T – C)	P-value
Native	(n=112)	(n=68)		
Word Recognition				
WRAT-3 Word Reading	0.34	0.10	0.24	0.068
WJ Letter-Word Identification	0.11	-0.03	0.14	0.221
Non-Native	(n=51)	(n=30)		
Word Recognition				
WRAT-3 Word Reading	0.55	0.76	-0.21	0.244
WJ Letter-Word Identification	0.11	0.30	-0.19	0.241

^a Mean gains are the model-adjusted pre-test to post-test gains expressed in effect size units (i.e. the dependent measures (reading assessment gain scores) were standardized prior to analysis by dividing each score by the pooled pre-test standard deviation of the reading assessment score). All models included the pre-test score as a covariate.

^b Impacts are expressed in standardized effect size units.

Table 5**Comparison Group Reading Gains**

Test	Comparison Group Mean Gain ^a (n=88)	Difference ^b Comparison minus Treatment	Difference ^c Comparison minus Control
Decoding			
WJ-R Word Attack	0.25	-0.10	0.09
Letter-Sound Survey	0.10	-0.19	-0.16
TOWRE Phonemic Decoding	-0.09	-0.14	-0.08
Spelling			
WRAT-3 Word Spelling	0.12	-0.03	0.08
Study Spelling Test	0.05	-0.11	-0.04
Word Recognition			
WRAT-3 Word Reading	0.21	-0.22	-0.11
WJ Letter-Word Identification	-0.09	-0.20	-0.18
Comprehension and Vocabulary			
Nelson Word Meaning	0.22	0.14	-0.09
Nelson Reading Comprehension	0.07	-0.06	-0.10
Fluency			
TOWRE Sight Word Efficiency	-0.08	-0.15	-0.05
Passage Reading Test	-0.03	-0.08	-0.09

^a Mean gains are the model-adjusted pre-test to post-test gains expressed in effect size units (i.e. the dependent measures (reading assessment gain scores) were standardized prior to analysis by dividing each score by the pooled pre-test standard deviation of the reading assessment score). All models included as covariates the pre-test score and an indicator for whether the learner was born and educated outside of the United States.

^b The differences (comparison group mean gain minus treatment group mean gain) are expressed in standardized effect size units. None of the differences were statistically significant.

^c The differences (comparison group mean gain minus control group mean gain) are expressed in standardized effect size units. None of the differences were statistically significant.