



Published in final edited form as:

J Learn Disabil. 2010 ; 43(2): 108–121. doi:10.1177/0022219409359342.

Reading Component Skills of Learners in Adult Basic Education

Charles MacArthur,
University of Delaware

Timothy R. Konold,
University of Virginia

Joseph J. Glutting, and
University of Delaware

Judith A. Alamprese
Abt Associates Inc

Abstract

The purposes of this study were to investigate the reliability and construct validity of measures of reading component skills with a sample of adult basic education (ABE) learners including both native and non-native English speakers and to describe the performance of those learners on the measures. Investigation of measures of reading components is needed because available measures were neither developed for nor normed on ABE populations or with non-native speakers of English. The study included 486 students, 334 born or educated in the United States (native) and 152 not born nor educated in the US (non-native) but who spoke English well enough to participate in English reading classes. All students had scores on 11 measures covering five constructs: decoding, word recognition, spelling, fluency, and comprehension. Confirmatory factor analysis (CFA) was used to test 3 models: A 2-factor model with print and meaning factors; a 3-factor model that separated out a fluency factor; and a 5-factor model based on the hypothesized constructs. The 5-factor model fit best. In addition, the CFA model fit both native and non-native populations equally well without modification showing that the tests measure the same constructs with the same accuracy for both groups. Group comparisons found no difference between the native and non-native samples on word recognition, but the native sample scored higher on fluency and comprehension and lower on decoding than the non-native sample. Students with self-reported learning disabilities scored lower on all reading components. Differences by age and gender were also analyzed.

Significant numbers of adults in the United States have difficulties with basic literacy. According to the 2003 National Assessment of Adult Literacy (NAAL, Kutner, Greenberg, & Baer, 2005), about 14% or 30 million adults in the United States perform below a basic level of literacy, defined as the skills necessary to perform simple and everyday literacy activities. An additional 2% did not speak sufficient English to participate in the assessment. The figures are worse for racial and ethnic minorities; 24% of Blacks and 44% of Hispanics scored below the basic level. These findings show minimal change from the 1992 National Adult Literacy Survey (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993).

One policy response to this problem has been the funding of adult basic education (ABE) programs by the U.S. Department of Education, which serve over two million adults annually (U.S. Department of Education, 2002, 2006). This population of ABE participants

is diverse in skills, age, race/ethnicity, and background characteristics. Of the 2.5 million adults enrolled in ABE programs during the 2004–2005 program year, 39% had skills at the eighth-grade equivalent level or lower at the time of enrollment, 16% scored above the ninth-grade level in reading or math at entry, and 22% came to ABE programs to improve their English-language skills. These adults ranged in age from 16 to over 60 years old, with 38% under age 25 and more than 80% under age 45. Only 4% were age 60 or older. ABE program participants are varied in terms of race and ethnicity, with Hispanics representing the largest group enrolled in 2004–05 (43%), followed by whites (27%) and African Americans (20%). Adults in ABE programs also vary in their economic backgrounds. Of the adults enrolled during 2004–05, 37% were employed, 10% reported being on public assistance, and 10% received ABE services provided in correctional facilities.

Limited research is available to describe the reading and writing skills of low-literacy adults and to evaluate the effectiveness of reading instruction designed for them. While the NAAL (Kutner et al., 2005) provides considerable information about the functional literacy abilities of adults, little research has been completed that describes the specific reading skills and related cognitive skills of low-literacy adults. A recent longitudinal study provided support for further investigation of the development of adults' reading component skills using instructional strategies derived from K-12 instruction. This research examined 643 low-literate adults (i.e., skills below a seventh-grade equivalent level in reading comprehension) from 130 ABE reading classes in 35 ABE programs (Alamprese, in press). 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 (nine months) and from pre- to follow-up (eighteen 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. A large body of literacy research on school-age children and adolescents has provided a theoretical and practical foundation for the design of instruction and the development of assessment measures to diagnose students' reading problems and monitor their progress (e.g., National Reading Panel, 2000). Two recent reviews of literature on adult literacy have used the research on reading problems in children as a framework for interpreting the limited research on low-literacy adults (Kruidenier, 2002; Venezky, Oney, Sabatini, & Jain, 1998). Both reviews found some similarities between poor adult readers and children with reading disabilities, but they also found some differences in patterns of skills between children and adults and among different groups of adults. Clearly, further research on adult reading is needed to provide a sound basis for assessment and instruction.

Several studies have compared adults with reading problems with children and/or adults with normal reading. Bruck (1990; 1992) compared adults with dyslexia with age-matched and reading-matched controls. She found poor phonemic awareness, word recognition, decoding, and fluency among the adults with dyslexia, similar to findings for children with dyslexia. Carver and Clark (1998) found that students in remedial reading classes in a community college had more uneven patterns of reading skills than normal children and adults. In general, their word recognition and fluency scores were relatively lower than vocabulary and comprehension scores. Greenberg, Ehri, and Perin (1997) Greenberg, Ehri, and Perin (2002) compared adults in ABE programs with reading-matched children in grades three to five. Adults performed worse on phonological tasks, which is consistent with findings for adults with dyslexia. However, the adults performed better on some orthographic tasks, including sight word recognition. Furthermore, analysis of errors on word recognition, decoding, and spelling showed that adults relied more on orthographic processes or visual memory and less on phonological analysis than children. Worthy and Viise (1996) analyzed the spelling errors of ABE learners and reading-matched children and found greater numbers of morphological errors for the adults. The general finding of weak

phonological and decoding skills among low-literacy adults is consistent with results for children with reading problems. However, there is also evidence of variability in reading processes that may be relevant for assessment and instruction of ABE learners.

The research comparing adults and children just cited has focused on native speakers of English. However, ABE programs include ESOL classes and substantial numbers of students in English reading classes who are not native speakers. Research has found differences in the reading skills of ABE students who are native or non-native speakers of English (Davidson & Strucker, 2002; Strucker, 2007). Davidson and Strucker (2002) compared native and non-native speakers in ABE programs. Not surprisingly, the native speakers scored higher on oral receptive vocabulary, as well as silent reading comprehension, although both groups had similar word recognition and decoding skills. A sub-sample was matched on the decoding task (pseudoword reading), and decoding errors were analyzed. The native group made more real-word substitutions, while the non-native group made more phonetically plausible errors. It appears that the non-native group relied more on phonological decoding and the native group on visual memory and vocabulary knowledge. Strucker (2007) reported a cluster analysis of a large sample (n=1042) of adult education students including ESL (English as a Second Language) students. Of the five clusters, two included primarily native speakers and had relatively high vocabulary, and two were primarily non-native with low vocabulary scores. Differences in reading processes between native and non-native speakers are highly relevant for instruction and may also have implications for assessment measures and practices.

The current study focused on assessment of reading components that are frequently assessed as part of reading instruction – for placement, diagnosis, progress monitoring, and outcome evaluation (McKenna & Stahl, 2003). We included measures of five components: decoding, word recognition, spelling, fluency, and comprehension. Three of these components focus on individual words – word recognition, decoding, and spelling. They are generally highly correlated, but are, nonetheless, separate theoretical constructs. All three skills draw on phonological, orthographic, and morphological knowledge but to varying degrees (Richards, et al., 2006; Venezky, 1970, 1999). Decoding draws most heavily on phonological processes, especially at lower levels (Share & Stanovich, 1995; Vellutino, Tunmer, Jacard, & Chen, 2007). It is typically assessed with pseudoword reading to prevent interference from sight word recognition. Word recognition draws on phonological processes but also on orthographic processes for recognition of common sight words and word parts; morphological knowledge becomes more important with increasing difficulty (Ehri & McCormick, 1998). Spelling draws on the same base of phonological, orthographic, and morphological knowledge as word recognition, but it is harder because it requires production rather than recognition and because there many alternate spellings of sounds (Templeton & Bear, 1992; Ehri, 2000; Berninger, Abbott, Abbott, Graham, & Richards, 2002). Thus, it requires more visual memory.

It may be of practical importance to assess these three components separately despite relatively high correlations because individual students or groups of students may have different patterns of skills. For example, many individuals with dyslexia who have developed adequate accuracy in reading continue to have substantial difficulties with spelling (Lefly & Pennington, 1991). The research on adults mentioned above (Greenburg, et al., 1997; 2002) found that low-literacy adults in ABE programs, in comparison to children, tended to have weaker decoding skills in relation to their sight word recognition. In addition, native and non-native speakers in ABE programs appear to rely differently on decoding and sight word recognition skills (Davidson & Strucker, 2002). Thus, it may be important diagnostically to differentiate skills in these three word-level components.

In recent years, reading researchers have paid more attention to the importance of fluency in contributing to comprehension. Theoretically, fluency is a result of relative automaticity in word recognition that frees working memory to attend to comprehension (Samuels, 2004). Some scholars (Fuchs, Fuchs, & Hosp, 2001) have argued that oral reading fluency (ORF), measured as correct words per minute, is a good measure of general reading competence. Fuchs et al., 2001) cited a correlation of .91 between ORF and comprehension for a group of students with reading disabilities. Correlations are lower but still substantial for average readers; for example, Cutting and Scarborough (2006) found correlations of .75 between ORF and three reading comprehension measures for average readers in grades 1–10. Oral reading fluency, of course, requires accurate word recognition, but ORF explains additional variance in comprehension beyond word recognition. Cutting and Scarborough (2006) found that word reading and oral language comprehension explained much of the variance in reading comprehension, but ORF predicted an additional 1 to 6% of variance across grades. A study with a sample of students with reading disabilities (Aaron, Joshi, Gooden, & Bentum, 2008) found that after including the effects of decoding and listening comprehension, fluency predicted an addition 2–11% of variance in reading comprehension. Some work on reading fluency has focused on adults. Bristow & Leslie (1988) measured accuracy, rate, and comprehension of short texts of varying difficulty by low-literacy adults. They found that both accuracy and rate on easier passages predicted comprehension of harder passages. Sabatini (2002) studied a sample of young adults from ABE programs and college with a wide range of reading ability. He found pervasive differences in both accuracy and rate of reading as well as in comprehension. Research with adults with a history of dyslexia found that they continued to read very slowly despite the fact that they had compensated for their problems and could read accurately with good comprehension (Lefly & Pennington, 1991). Fluency is an important component of reading that is related to comprehension and that continues to be a problem for adult learners.

Comprehension is the end result and purpose of reading. According to the simple view of reading (Hoover & Gough, 1990), in addition to accurate word reading, general language skills determine comprehension. Vocabulary, knowledge about texts, prior content knowledge, and other factors also influence reading comprehension (Pressley, 2000). Our study is limited to a single comprehension construct, measured as a combination of vocabulary and passage comprehension, both assessed through reading. This limitation of the study is due to the fact that the data set comes from an instructional study focused primarily on the development of word skills.

The overall purpose of the current study was to investigate the reliability and construct validity of measures of reading component skills with a sample of ABE learners including both native and non-native speakers of English. Adult education programs use broad assessments of reading comprehension that were specifically developed for use with low-literacy adults, such as the Tests of Adult Basic Education (CTB/McGraw-Hill, 1994) and the Comprehensive Adult Student Assessment System (2005). These measures are used to place students in appropriate level classes and to document the outcomes of instruction for accountability purposes. However, measures of multiple components of reading skill are needed to diagnose students' strengths and weaknesses in reading in order to design effective instruction and to monitor progress more closely. The measures of word-level skills and fluency that have been used in research on adult learners and that are available for use by practitioners were neither developed nor normed for use with low-literacy adults. Thus, although they may have solid reliability and validity for typical populations of children, adolescents, and, in some cases, adults, they may not be valid with adults functioning considerably below norms for their age.

The study addressed the following three research questions:

1. Are the measures reliable with an ABE population? Are they reliable both with native and non-native born adults that participate in ABE classes?
2. Does the hypothesized model with five reading components fit the data? Do the five constructs capture more variance than alternate models with fewer constructs? Do the measures assess the same constructs equally well for native and non-native born groups?
3. How do native and non-native born adults compare on the five components? How does performance on the constructs vary by other background variables?

Method

Participants

The study included 486 adult learners from 23 ABE programs, representing 12 states. Programs were selected based on the following criteria: a) provided class-based reading instruction to English-speaking adults whose reading level in comprehension was at the Low Intermediate level (roughly between the fourth- and seventh-grade equivalence level, as defined by the U.S. Department of Education (2007)), b) met certain basic standards for learner recruitment and assessment, program management, and support services, and c) had instructors who were trained or experienced in teaching reading. These programs and adults participated in an instructional study. The data used in this analysis were collected as pretest data.

The majority of adults were female (67%). Average age was 35 (SD = 14, range 16 to 71). The racial and ethnic breakdown was as follows: White 32%, Hispanic 28%, Black 22%, Asian 12%, and other 6%. All participants were sufficiently fluent in English to participate in English reading classes. The majority (69%) had been born in the U.S. or educated in the U.S. from the primary grades (native); the remaining 31% were born and educated outside of the U.S. (non-native). We used place of birth to represent whether participants were native speakers of English because it was deemed more reliable than self-reports of primary language as an indicator of native English proficiency. We included in the native group individuals who received their education in the U.S. beginning in the primary grades because previous research indicated that individuals who immigrated to the U.S. before the age of 12 performed more like native-born residents than like immigrants who came after the age of 12 (Davidson & Strucker, 2002). A majority (66%) spoke English at home. A minority spoke Spanish at home (10%), while 24% spoke some other language. Education ranged widely; 7% had less than a sixth-grade education; 49% had completed some secondary education; 14% had a high school diploma or GED; and 30% had some education but not in the U.S. Just less than half (43%) were currently employed although another 47% had been employed previously; 8% had never worked; and 2% were retired. Of the learners who were born or educated in the U.S., 48% self-reported that they had had a learning disability when they were young. Almost none of the non-U.S. born learners reported a learning disability.

Measures

Eleven measures of reading component 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 three through nine with 3,800 students per grade representative of a range of geographic areas, SES, and school district

size. 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 subtest (WRAT3-R) assesses ability to read words in isolation. The WRAT3 Spelling subtest (WRAT3-S) assesses ability to spell individual words from dictation. The test was normed on a nationally representative sample stratified for age, region, gender, ethnicity, and SES. 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 given. 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, non-words that follow the phonological, orthographic, and morphological patterns of English. The test was normed on a nationally representative sample of individuals from age 2 to 90. 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 subtest (TOWRE-SWE) presents words of increasing difficulty and tests how many a person can read in 45 seconds. The Phonemic Decoding Efficiency subtest (TOWRE-PDE) has the same format but uses pseudowords. The TOWRE was normed on a nationally representative sample ranging in age from 6 to 24 years old.

The Letter-Sound Survey (LSS) was developed by our project (Venezky, 2003) to assess decoding. It consists of 26 pseudowords of one or two syllables that represent common phonological, orthographic, and morphological patterns. It is untimed.

The Passage Reading test (PR) is a measure of oral reading fluency that uses a passage developed for the fluency test in the NAAL. 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 scored for correct words per minute.

A Developmental Spelling test (DS) was developed by our project. It consists of 20 words of increasing difficulty that represent common phonological, orthographic, and morphological patterns in English. For the present purposes, words were scored correct or incorrect.

In summary, the 11 measures include three tests of decoding (WJR-WA, TOWRE-PDE, LSS), two tests of untimed word recognition (WJR-LW, WRAT3-R), two tests of spelling (WRAT3-S, DS), two tests of fluency (TOWRE-SWE, PR), and two tests of comprehension (NWM and NRC).

In addition, a Learner Background Interview was administered to gather information on demographics, education, employment, health and disabilities, goals for participating in the program, and literacy activities at home and work.

Procedures

The administration of the eleven reading tests and learner background interview was conducted by 40 individuals from the study's 23 ABE programs. These test administrators were ABE professional staff that: 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.

All test administrators participated in a three-day training session conducted by one of the senior researchers prior to collecting data, and follow-up training was conducted via telephone and email. During the three-day training session, test administrators learned about the study's design, data collection procedures, and the content and design of the eleven reading measures and learner background interview; they also received a manual with detailed instructions for administration of the test battery and interview (Author, 2003). Training participants administered the eleven tests twice during the training session, were observed by the senior researcher, and received feedback on their test administration procedures. After the training session and prior to the beginning of data collection, the testers audiotaped their administration of the battery of eleven tests and conducted the learner interview with two individuals who were not students in the study's classes. The senior researcher listened to the audiotapes of the tests administered by the testers to determine the accuracy of test administration and provided feedback to the testers on their administration procedures. Note that throughout the study, the seven tests that required oral responses were audiotaped for later scoring and determination of interrater reliability.

The test administrators also were responsible for scoring the tests that involved establishing a basal and ceiling (WJR-LW, WJR-WA, WRAT3-R, and WRAT3-S) and the LSS. Training in scoring these tests, including procedures for determining basal and ceiling levels, was included in the training sessions. Training included scoring audiotaped responses from sample learners, or written responses for the spelling test (WRAT3-S), with corrective feedback.

The remaining seven 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-SDE, PR) were scored from the audiotapes. They also scored a sample of tests scored by each test administrator to determine reliability.

The study adopted the scoring guidelines for native English speakers and non-native speakers of English that were developed by Strucker (2004) and used in his study of 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. The senior researcher established reliability with the first cohort of data collectors including a lead scorer who then became the standard for test scoring with subsequent scorers. The training process included two sessions in which staff scored tests from audiotapes, compared scores, and resolved disagreements through re-playing the contested items and discussing them. At the end of the second session, additional tests were scored and percent of agreement was calculated. Each staff scorer scored twelve test batteries that were independently scored by the senior researcher or the lead scorer. Ranges of interrater reliability across scorers were as follows: WJR-LW: .93-.97; WJR-WA: .82-.88; WRAT3-R: .95-.97; TOWRE-SDE: .95-.98; TOWRE-PDE: .81-.88; LSS: .77-.99; PR: .95-.99. Reliability for test administrators was calculated by rescoring a sample of half of the tests given by each administrator. 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.

Analyses

The study analyzed a set of 11 measures of reading components: three tests of decoding, two of word recognition, two of spelling, two of fluency, and two of comprehension. Scores for the 11 measures were converted to sample based standard scores (i.e., z-scores). Instrument based normative standard scores could not be used for three reasons. First, some of the measures did not have available norms because they were developed for the project. Second, some of the norm-referenced measures did not have norms for adults or for the range of ages of adults in the study. Third, even when adult norms were available, our sample generally scored below the norm range or in the lowest end of the range. Thus, raw scores were converted to z-scores to provide a common metric.

Our general analytic approach was as follows: We began by examining the internal consistency reliability of the measures for the total population and separately for the native and non-native groups. We then used a series of hierarchically-ordered confirmatory factor analyses (CFA) to investigate alternate models of reading components. We tested a two-factor model comprised of a print factor that included the three word-level components and fluency, and a meaning factor comprised of vocabulary and comprehension. Next we separated out fluency to create a three-factor model. Finally, we tested a five-factor model based on our hypothesized constructs. Following identification of the best fitting model (i.e., two-, three-, or five-factor model), multi-group CFA (MGCFA) analyses were conducted to investigate whether the identified reading dimensions were similarly measured for the native and non-native groups. It is important to note that our primary interest in MGCFA was to determine whether the factor loadings linking the subtests to their respective factors were statistically indistinguishable across groups. Although it is also possible to test the invariance of parameters involving variances and covariances, these tests are viewed as overly restrictive (Keith et al., 1995). Moreover, there is often little to be gained from tests of these parameters as their values may fluctuate from group to group even when the factors are being similarly measured (MacCallum & Tucker, 1991; Marsh, 1993).

A graphic representation of the five factor model is illustrated in Figure 1. The observed reading measures are enclosed in boxes to differentiate them from the directly unobservable factors and uniqueness terms associated with each observed variable. Each observed variable was modeled to be directly influenced by a single factor as illustrated through the use of single headed arrows. The curved double-headed arrows reflect the fact that factor correlations were estimated, as well as the covariance between two uniqueness terms. Parameterization of the model included fixing one factor loading to one in order to set the scale of the latent variables.

Numerous measures of model fit exist for evaluating the quality of measurement models, most developed under a somewhat different theoretical framework focusing on different components of fit (Browne & Cudeck, 1993; Hu & Bentler, 1995). For this reason, it is generally recommended that multiple measures be considered to highlight different aspects of fit (Tanaka, 1993). Use of the chi-square statistic was limited to testing differences (χ^2_D) between competing models. As a stand-alone measure of fit, chi-square is known to reject trivially mis-specified models estimated on large sample sizes (Hu & Bentler, 1995; Kaplan, 1990; Kline, 2005). The chi-square ratio (χ^2/df) was, however, used to evaluate stand-alone models. This index tends to be less sensitive to sample size, and values less than 3 (or in some instances 5) are often taken to indicate acceptable models (Kline, 2005). Three additional measures of fit were considered in evaluating model quality. These included the Bentler-Bonett normed fit index (NFI), Tucker-Lewis index (TLI), and the comparative fit index (CFI). These three measures generally range between 0 and 1.0. Traditionally, values of .90 or greater have been taken as evidence of good fitting models (Bentler & Bonett, 1980). However, more recent research suggests that better fitting models produce values

around .95 (Hu & Bentler, 1999). All models were estimated with the Analysis of Moment Structures (AMOS; Arbuckle & Wothke, 1999) program.

CFA and MGCFA procedures permitted an investigation into whether the subtests measure the same factors with same degree of accuracy across groups. They did not, however, address whether groups differ on the identified latent dimensions. Differences between native status, self-reported learning disability, gender, and age groups were examined on the identified factors through procedures of multivariate analysis of variance (MANOVA).

Results

Overall Performance

Descriptive data on the 11 tests are presented in Table 1. The table includes grade equivalent scores to provide a rough measure of the general reading levels of the students in comparison to normal readers, and measures of skewness that attest to the relative normality of the scores. Correlations among the 11 tests are provided in Table 2. Multivariate outlier analysis was conducted on the 11 measures through the use of Mahalanobis distance. Results revealed only 7 outliers ($\chi^2 > .001\chi^2_{p=11} = 31.3$) that were retained throughout the analyses described below.

Internal Consistency Reliability

Cronbach's alpha was calculated for the total sample and separately for the native and non-native groups (see Table 3). Internal consistency could not be calculated for the speed-dependent TOWRE tests because students complete different numbers of items. For the Passage Reading test, the individual item scores used were the individual words in the passage. For the total sample, internal consistencies across the tests ranged from .86 to .96. For the native adults, results ranged from .88 to .98. Internal consistencies were slightly lower for the non-native adults, ranging from .77 to .88. The internal consistencies were judged to be adequate for all tests for both groups.

It is worth noting that the lowest two internal consistency scores for the non-native group were on the two decoding measures. The lowest interrater reliabilities were also on the measures of decoding. These measures required pronunciation of pseudowords, where interference from language differences and pronunciation accents might be expected to play the greatest role.

Confirmatory Factor Analysis

Three hypothesized factor structures were investigated with the eleven subtests through methods of confirmatory factor analysis (CFA). A summary of each model is provided below. In all instances, slight modifications were made to the originally hypothesized model. Estimates of model fit for each model are provided in Table 4.

Two-Factor Model—We tested a two-factor model with a print factor (comprised of WJR-LW, WRAT3-R, WJR-WA, LSS, WRAT3-S, DS, PR, TOWRE-PDE, and TOWRE-SWE) and a meaning factor (consisting of NWM and NRC). Three nested two-factor solutions were examined, see Table 4. The first model (Model 2A) specified a simple structure in which each subtest was modeled to load on one of the two hypothesized factors as described above. All measures of fit suggested that the model did not provide a good representation of the data. The second model (Model 2B) allowed for correlated residual terms between WRAT3-S and DS that were suggested by modification indices. Although allowing these correlated errors to be estimated (Model 2B) resulted in a statistically better model fit than Model 2A, $\chi^2_D(1) = 130.26$, $p < .05$; stand alone measures of fit for this

model continued to suggest poor fit. The third and final two-factor model (Model 2C) was also informed by modification indices that suggested material model improvement through the estimation of an additional set of correlated error terms linking PR and TOWRE-SWE. Support for this model was mixed. Relaxation of the uniqueness covariance resulted in a statistically better fit than Model 2B, $\chi^2_{\text{D}}(1) = 147.19$, $p < .05$. However, although the model would be defensible under historical standards for gauging model fit (e.g., NFI, TLI, CFI $> .90$), it fails the test of more contemporary standards for gauging good fitting models (e.g., NFI, TLI, CFI $< .95$), see Table 4. In addition, the χ^2/df ratio = 9.02 exceeded acceptable thresholds.

Three Factor Model—Next we separated out fluency to create a three-factor model. The fluency factor included both speeded TOWRE tests (i.e., TOWRE-PDE and TOWRE-SWE) and Passage Reading (PR) that were previously specified to load on the print factor when only two factors were considered. Here again, three nested three-factor solutions were examined. All measures of model fit (NFI, TLI, CFI, and χ^2/df) for the first model were deemed unacceptable. Two correlated error terms were identified through inspection of modification indices, and estimated in turn. The first model relaxation was applied to the estimation of the uniqueness covariance between WRAT3-S and DS (i.e., Model 3B). This model was found to fit statistically better than Model 3A, $\chi^2_{\text{D}}(1) = 125.36$, $p < .05$, and two of the three stand-alone measures of fit *approached* defensible values, see Table 4. The second model relaxation was applied to the estimation of the correlated residuals associated with LSS and TOWRE-PDE (i.e., Model 3C). This model was found to fit statistically better than Model 3B, $\chi^2_{\text{D}}(1) = 85.43$, $p < .05$. Here again, stand-alone measures of fit approached defensible values, but remained below contemporary standards. A comparison between the best fitting two factor model (Model 2C) and the best fitting three factor model (Model 3C) indicated that the three factor solution fit statistically better than the two factor model, $\chi^2_{\text{D}}(2) = 61.76$, $p < .05$.

Five Factor Model—Last, we tested a five-factor model based on our hypothesized constructs, see Figure 1. Here, two nested five-factor solutions were examined. Results of the uncorrelated residual model (Model 5A) are shown in Table 4. All measures of fit exceeded those obtained from both the two (Model 2A) and three factor (Model 3A) solutions, and two of the three measures of fit (i.e., NFI and CFI) exceeded recently recommended thresholds for good fit ($> .95$). Although this form of the model could likely be deemed defensible, modifications were sought to determine whether further substantively meaningful model relaxations could improve fit. Inspection of modification indices suggested estimation of the TOWRE-PDE and TOWRE-SWE residual covariance. This model (Model 5B) resulted in a statistically better fit than that provided by Model 5A, $\chi^2_{\text{D}}(1) = 65.14$, $p < .05$, and all stand-alone measures of fit (NFI, TLI, CFI, and χ^2/df) were suggestive of good fit, see Table 4. A comparison between the best fitting three factor model (Model 3C) and the best fitting five factor model (Model 5B) indicated that the five factor solution fit statistically better than the three factor model, $\chi^2_{\text{D}}(6) = 182.78$, $p < .05$. Standardized model parameter estimates (i.e., factor loadings, factor correlations, squared multiple correlations, and correlated errors) are shown in Figure 1. All factor loadings and correlations were large and statistically significant. In addition, the squared multiple correlations (shown above the observed variable boxes in Figure 1) were appreciable and indicate that the factors are accounting for meaningful portions of observed score variance. This best fitting five-factor measurement model was next evaluated to determine whether the measurement of the underlying five constructs was the same for native and non-native speaking groups through methods of multi-group confirmatory factor analysis.

Native and Non-native Multi-group Analyses

Multi-group confirmatory factor analyses were conducted by first evaluating whether the general form of this configuration (i.e., the pattern of free and fixed loadings) was viable for the two groups. In general, this model provides a baseline against which more restrictive hypotheses can be tested that result from imposing additional cross group equality constraints on different aspects of the model. Results of the general form model (Model 5C) were favorable and suggest that the “five” factor model as illustrated in Figure 1 works well for these two groups, see Table 4. A second multi-group model (Model 5D) was specified that constrained the factor loadings linking the factors to their respective subtests to be equal between the two groups. These additional model constraints failed to result in a statistically significant decline in model fit in comparison to the general form model (Model 5C) in which these values were allowed to be freely estimated for each of the two groups, $\chi^2_D(6) = 8.05, p > .05$. As a result, the examined subtests can be said to be measuring the same factors, with the same degree of accuracy, for both groups. Estimated correlations among the five factors from Model 5D are provided separately for the native and non-native groups in Table 5.

Comparisons by Demographic Groups

Next we investigated whether performance on the five reading components varied by demographic variables. We compared native and non-native adults because previous research suggested strongly that they had different patterns of skills. We also compared adults with and without self-reported learning disabilities (LD) because previous research has suggested that the performance of low-literacy adults has some similarities with younger students with reading disability. Gender was included as a factor because it is often related to literacy performance. Age was similarly of interest because performance on many cognitive skills is known to change with age during adulthood. Skills related to memory and rapid performance are thought to be more susceptible to declines with age than skills based on knowledge. In addition, age may be related to performance because older people went to school in a different time period when educational opportunities may have been different. Our measures were not based on standard scores normed on adults, so age was not already factored into the scores.

Preliminary analyses explored the relationships among these demographic factors. We found that the proportion of women did not differ between the native (65%) and non-native (70%) groups ($p > .3$). Surprisingly, the proportions of women and men reporting a LD did not differ either; in the native group, 47% of women and 49% of men reported a LD. Learning disabilities (LD) were almost never (2%) reported among the non-native group. Consequently, our analyses of LD were limited to the native group.

Age did differ significantly between the native and non-native groups. The native group had a higher proportion of young adults between the ages of 16 and 25, as well as a lower mean age (33.5 versus 38.0), $F(1, 484) = 11.9, p = .001$. Preliminary analyses found significant negative correlations ($p < .01$) between age and scores on three of five components (decoding, $-.25$; spelling, $-.13$; fluency, $-.23$). To include age as a factor, we used a median split. (We also tried dividing the sample into quartiles and found the same results.)

A 3-factor (birthplace by gender by age) MANOVA was conducted on the five identified reading factors, followed by univariate analyses. The analysis failed to meet the requirement of equality of covariance matrices. As a result, alpha levels of .01 (vs. .05) served as thresholds for gauging statistical significance to control for the sometimes liberal effect that violation of this assumption has on the multivariate F statistic. For analyses of individual comparisons, effect sizes (ES) are reported as adjusted z-score differences, which are the

differences in standard deviation units. These give some idea of the magnitude of differences.

Statistically significant effects were found for age, $F(5, 466) = 2.4, p = .002$, birthplace; $F(5, 466) = 20.1, p < .001$; and the interaction of birthplace and gender, $F(5, 466) = 5.5, p < .001$. The main effect of gender was not statistically significant, $F(5, 466) = 2.2, p = .055$.

Follow-up univariate tests were conducted for the five reading components. For birthplace, native and non-native adults did not differ on word recognition, but non-native adults scored higher on decoding, $F(1, 470) = 7.2, p = .008, ES = .31$; and lower on comprehension, $F(1, 470) = 31.6, p < .001, ES = .67$; and fluency, $F(1, 470) = 8.6, p = .004, ES = .34$, than native adults. These findings are consistent with previous research. For gender, women scored higher on fluency, $F(1, 470) = 8.3, p = .004, ES = .34$. No significant effects were found for the interaction between birthplace and gender. For age, younger adults scored significantly higher on decoding, $F(1, 478) = 14.1, p < .001$, and fluency, $F(1, 478) = 9.5, p = .002$.

A one-factor (LD) MANOVA with age as covariate was conducted including just the native adults. A significant main effect was found for LD, $F(5, 326) = 4.3, p = .001$. Follow-up univariate analyses found significant differences on all five reading components with adults with LD scoring lower: decoding, $F(1, 330) = 16.1, p < .001, ES = .45$, word recognition, $F(1, 330) = 13.0, p < .001, ES = .42$, spelling, $F(1, 330) = 15.9, p < .001, ES = .44$, comprehension, $F(1, 330) = 11.1, p = .001, ES = .38$, and fluency, $F(1, 330) = 6.3, p = .012, ES = .29$.

Discussion

The overall purpose of the current study was to investigate the reliability and construct validity of measures of reading component skills with a sample of ABE learners including both native and non-native speakers of English. This investigation is important because the measures of word-level skills and fluency that have been used in research on adult learners and that are available for use by practitioners were neither developed nor normed for use with low-literacy adults. Measures of decoding, word recognition, spelling, fluency, and comprehension were investigated.

The internal consistencies for all measures were judged to be adequate though they were somewhat higher for the native group than the non-native group. The lowest internal consistency results for non-native learners and the lowest interrater reliabilities were on measures of decoding, which required pronunciation of pseudowords. These measures may be sensitive to differences in letter-sound associations across languages or dialects and may be most subject to errors in scoring because of pronunciation issues.

Confirmatory factor analysis (CFA) was used to test 3 models: A 2-factor model with print and meaning factors; a 3-factor model that separated out a fluency factor; and a 5-factor model based on the hypothesized constructs. The 5-factor model fit best. In addition, the CFA model fit both native and non-native populations equally well without modification, showing that the tests measure the same constructs with the same accuracy for both groups. One implication is that it is potentially important to assess all five constructs to develop an accurate description of the reading skills of individuals or groups.

To further test the validity of the five constructs, we investigated whether performance on the five reading components varied by demographic variables including birthplace (native or non-native), gender, age, and LD. Native and non-native adults did not differ on word recognition, but non-native adults scored higher on decoding and lower on comprehension and fluency than native adults. The difference in comprehension (which included

vocabulary) is expected and consistent with the limited previous research. Davidson & Strucker (2002) compared native and non-native English speakers in ABE programs and reported that non-native adults scored lower on vocabulary and comprehension measures. The differences we found in decoding and fluency are also consistent with Davidson and Strucker's results. Although they did not find differences in overall performance on decoding or word recognition, a detailed analysis of errors in word recognition found that native adults made twice as many real-word substitutions as non-native adults, while non-native adults made nearly three times as many phonetically plausible errors. Davidson and Strucker's findings suggest that the non-native adults relied on their phonological decoding skills, which is consistent with the current study's finding of better decoding of pseudowords, whereas the native adults relied on sight word recognition. Better sight word recognition would be consistent with the current study's finding of better fluency for the native group.

In this study, women scored higher than men on fluency but not on other reading components. Previous research has not investigated gender differences in reading performance in the ABE population, but research on the reading skills of the overall adult population (Kutner et al., 2005) has found higher reading comprehension for women than men. Younger adults scored significantly higher than older adults on decoding and fluency. Research on the overall adult population using a general literacy measure has found lower performance on reading among older adults (Kutner et al., 2005). Among the native adults in the current study, 48% self-reported having had a LD while in school; their performance was lower on all five reading components than other adults. This finding is consistent with expectations. Overall, differences by demographic variables were consistent with theoretical expectations and previous findings, lending additional support to the validity of the measured constructs.

The study has a number of limitations. First, the study was limited to measures of five reading components and included only one measure of vocabulary and one of reading comprehension. Furthermore, it did not include measures of important related constructs such as oral vocabulary and language comprehension and phonemic awareness. Second, the sample was limited to learners at the low intermediate level of ABE. Although it was drawn from 23 ABE programs in 12 states, it was not intended as a representative sample of ABE learners. For example, compared to the national population of ABE programs, the sample included a smaller proportion of Hispanic adults and a larger proportion of non-native adults from non-Spanish-speaking countries.

Despite the limitations, the study provides support for the reliability and construct validity of measures of reading components in the ABE population. Although the correlations among the three word-level components -- decoding, word recognition, and spelling -- were relatively high, the CFA demonstrated better fit when modeling them as separate components. Furthermore, the CFA model fit equally well for both native and non-native populations. The importance of measuring all three word-level components is further demonstrated by the finding that the native and non-native groups differed on decoding skills but not word recognition.

The findings have practical implications as well as theoretical ones. They support the importance of measuring multiple reading components in planning instruction. Such assessment is common in planning instruction for struggling school-age readers (McKenna & Stahl, 2003; Walpole & McKenna, 2007) and is being recommended in adult reading instructional programs (Curtis, Bercovitz, Martin, & Meyer, 2007). Different patterns of performance on decoding, word recognition, spelling, and fluency have practical importance for instruction.

In considering the implications for practice and future research, it is important to note the considerable effort and care spent in training test scorers. For the measures of decoding and word recognition, we found it was important to modify the pronunciation guidelines in the standard scoring manuals to adjust for differences in language background and dialect. We used guidelines developed by Strucker (2004) for his study of reading development of adult learners. Without those guidelines, the scorers found it difficult reliably to make scoring decisions. Test administrators in the field, even those with some experience in testing, might not get reliable results for adults from diverse language backgrounds if they just used the standard scoring guides.

Further research is needed to examine the reliability and validity of reading component measures for use in adult education. The current findings provide general support for the use of these measures that were developed for younger populations. However, measures developed for diverse populations of low-literacy adults might well provide more accurate estimates of their reading skills, and norms for such populations would be helpful in describing their skills and assessing growth.

Acknowledgments

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

References

- Aaron PG, Joshi RM, Gooden R, Bentum KE. Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*. 2008; 41(1):67–84. [PubMed: 18274504]
- Alamprese, JA.; Stickney, EM. Building a knowledge base for teaching adult decoding: Training and reference manual for data collectors. Bethesda, MD: Abt Associates Inc; 2003.
- Alamprese, JA. 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; in press
- Arbuckle, JL.; Wothke, W. *Amos 4.0 user's guide*. Chicago: Small Waters; 1999.
- Bentler PM, Bonett DG. Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*. 1980; 88:588–606.
- Berninger VW, Abbott RD, Abbott SP, Graham S, Richards T. Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities*. 2002; 35:39–56. [PubMed: 15490899]
- Bristow PS, Leslie L. Indicators of reading difficulty: Discrimination between instructional- and frustration-range performance of functionally illiterate adults. *Reading Research Quarterly*. 1988; 23(2):200–218.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993. p. 136-162.
- Bruck M. Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*. 1990; 26:439–454.
- Bruck M. Persistence of dyslexics' phonological deficits. *Developmental Psychology*. 1992; 28:874–886.
- Carver RP, Clark SW. Investigating reading disabilities using the reading diagnostic system. *Journal of Learning Disabilities*. 1998; 31(5):453–471. 481. [PubMed: 9763775]
- Chall J. Patterns of adult reading. *Learning Disabilities*. 1994; 5(1):29–33.
- CTB/McGraw-Hill. *TABE: Tests of Adult Basic Education*. Monterey, CA: CTB/McGraw-Hill; 1994.
- Comprehensive Adult Student Assessment System. *CASAS technical manual*. San Diego, CA: Author; 2005.

- Curtis, MB.; Bercovitz, L.; Martin, L.; Meyer, J. *STudent Achievement in Reading (STAR) tool kit*. Fairfax, VA: DTI Associates, Inc; 2007.
- Cutting LE, Scarborough HS. Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10(3):277–299.
- 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(3):299–316.
- Ehri LC. Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*. 2000; 20:19–36.
- 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(2): 135–163.
- Fuchs LS, Fuchs D, Hosp MK. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*. 2001; 5(3):239–256.
- 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(2):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(3):221–243.
- Hanna, G.; Schell, LM.; Schreiner, R. *The Nelson Reading Skills Test*. Itasca, IL: Riverside; 1977.
- Hoover WA, Gough PB. The simple view of reading. *Reading and Writing*. 1990; 2:127–160.
- Hu, L.; Bentler, PM. Evaluating model fit. In: Hoyle, RH., editor. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage; 1995. p. 76-99.
- Kaplan D. Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*. 1990; 25:137–155.
- Keith TZ, Fugate MH, DeGraff M, Dimond CM, Shadrach EA, Stevens ML. Using multi-sample confirmatory factor analysis to test for construct bias: An example using the K-ABC. *Journal of Psychoeducational Assessment*. 1995; 13:347–364.
- Kirsch, IS.; Jungeblut, A.; Jenkins, L.; Kolstad, A. *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington: U.S. Government Printing Office; 1993.
- Kline, RB. *Principles and practice of structural equation modeling*. 2. New York: Guilford; 2005.
- Kruidenier, J. *Research-based principles for adult basic education reading instruction*. Portsmouth, NH: RMC Research Corporation; 2002.
- Kutner, M.; Greenberg, E.; Baer, J. *National Assessment of Adult Literacy (NAAL): A first look at the literacy of America's adults in the 21st century (Report No. NCES 2006–470)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education; 2005.
- Lefly DL, Pennington BF. Spelling errors and reading fluency in compensated adult dyslexics. *Annals of Dyslexia*. 1991; 41:143–162.
- MacCallum RC, Tucker LR. Representing sources of error in common factor analysis: Implications for theory and practice. *Psychological Bulletin*. 1991; 109:501–511.
- Marsh HW. The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*. 1993; 30:841–860.
- McKenna, MC.; Stahl, SA. *Assessment for reading instruction*. New York: Guilford; 2003.
- National Reading Panel. *Report of the National Reading Panel: Teaching children to read – Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development; 2000.
- 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. 3. Mahwah, NJ: Erlbaum; 2000. p. 545-562.
- Richards T, Aylward E, Raskind W, Abbott R, Field K, Parsons A, Richards A, Nagy W, Eckert M, Leonard C, Berninger V. Converging evidence for triple word form theory in children with dyslexia. *Developmental Neuropsychology*. 2006; 30:547–589. [PubMed: 16925475]

- Sabatini J. Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*. 2002; 6(3):267–298.
- Samuels, SJ. Toward a theory of automatic information processing in reading, revisited. In: Ruddell, RB.; Unrau, NJ., editors. *Theoretical models and processes of reading*. 5. Newark, DE: International Reading Association; 2004.
- Share DL, Stanovich KE. Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology*. 1995; 1(1):1–58.
- 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.
- Templeton, S.; Bear, D. *Development of orthographic knowledge and the foundations of literacy: A memorial Festschrift for Edmund Henderson*. Mahwah, NJ: Lawrence Erlbaum; 1992.
- 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. *FY2000 State-Administered Adult Education Program: Data and Statistics*. Washington, DC: Author; 2002.
- U.S. Department of Education, Division of Adult Education and Literacy. *State Administered Adult Education Program: Program Year 2004–2005 Enrollment*. Washington, DC: Author; 2006.
- U.S. Department of Education, Office of Vocational and Adult Education. *Adult Education Annual Report to Congress Year 2004–05*. Washington, DC: Author; 2007.
- U.S. Department of Education, Office of Vocational and Adult Education. *Implementation Guidelines: Measures and Methods for the National Reporting System for Adult Education*. Washington, DC: Author; June. 2007
- Vellutino FR, Tunmer WE, Jaccard JJ, Chen R. Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*. 2007; 11(1): 3–32.
- Venezky, RL. Unpublished test. 2003. Letter-sound survey.
- Venezky, RL.; Oney, B.; Sabatini, J.; Jain, R. *Technical Report*. Bethesda, MD: Abt Associates; 1998. Teaching adults to read and write: A research synthesis.
- Walpole, S.; McKenna, MC. *Differentiated reading instruction*. New York: Guilford Press; 2007.
- Wilkinson, GS. *Wide Range Achievement Test – Revision 3*. Wilmington, DE: Jastak Associates, Inc; 1993.
- 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.

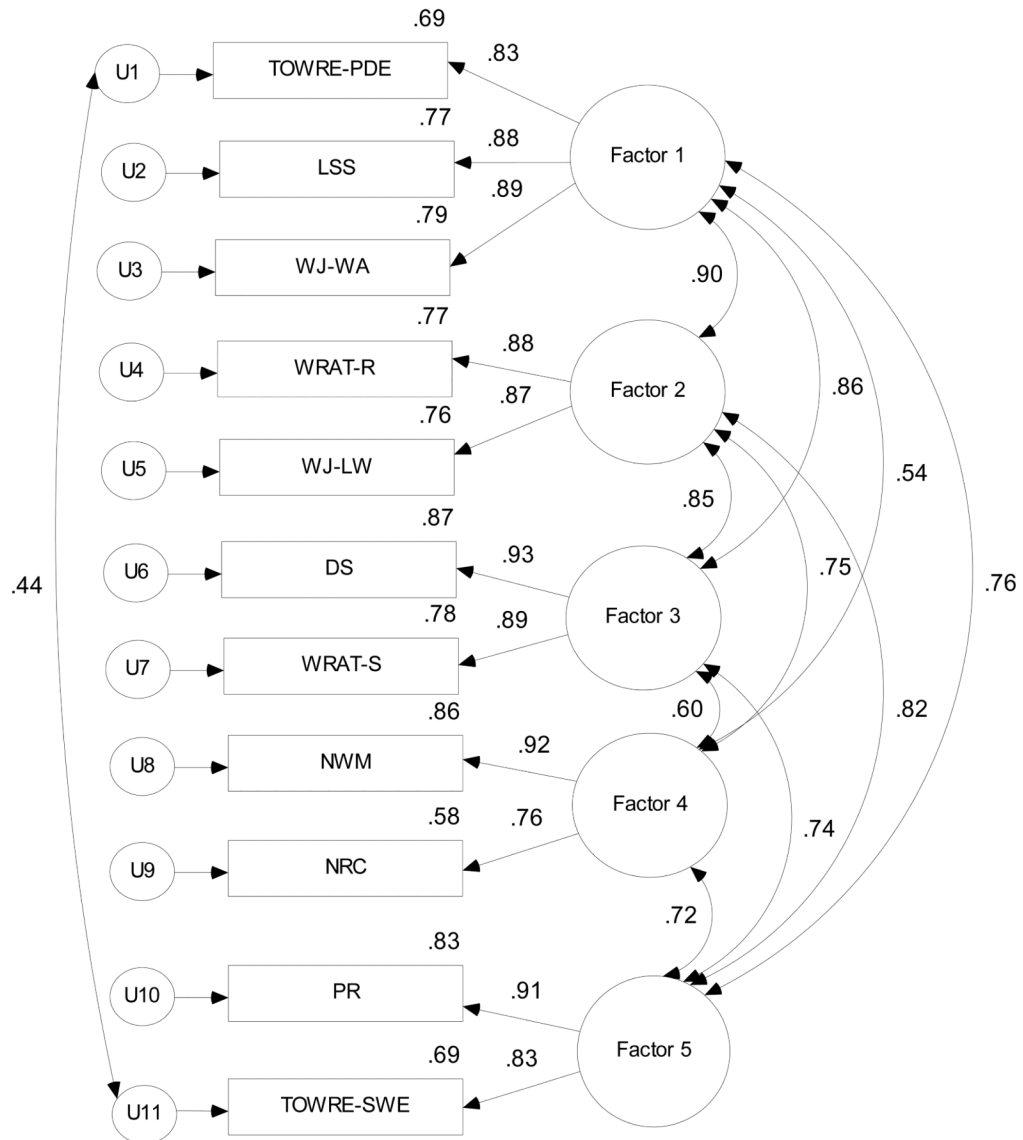


Figure 1. Five-Factor Correlated Model with Complete Sample Standardized Estimates
 Note: TOWRE-PDE = *Test of Word Reading Efficiency, Phonemic Decoding Efficiency*; LSS = *Letter-Sound Survey*; WJR-WA = *Woodcock–Johnson Tests of Achievement, Revised, Word Attack*; WRAT3-R = *Wide Range Achievement Test–Revision 3, Reading*; WJR-LW = *Woodcock–Johnson Tests of Achievement, Revised, Letter-Word Identification*; DS = *developmental spelling*; WRAT3-S = *Wide Range Achievement Test–Revision 3, Spelling*; NWM = *Nelson Word Meaning*; PR = *Passage Reading Test*; TOWRE-SWE = *Test of Word Reading Efficiency, Sight Word Efficiency*; U = *unique variances*.

Table 1

Reading Component Descriptive Measures

Test	Mean	SD	Skewness	Mean Grade Equivalent
Woodcock Johnson Word Attack ^a	488.5	15.6	-.23	3.3
TOWRE Phonemic Decoding	26.4	11.7	.27	3.7
Letter Sound Survey ^b	13.1	5.8	-.26	NA
WRAT Reading ^a	500.8	12.7	-.34	4.5
Woodcock Johnson Letter Word ^a	497.6	22.4	-.21	5.1
TOWRE Sight Word	61.8	14.7	-.01	4.0
WRAT Spelling ^a	498.8	11.8	-.29	4.0
Developmental Spelling ^b	6.9	4.9	.32	NA
Passage Reading – Correct words/min. ^b	111.2	39.7	.15	NA
Nelson Reading Comprehension ^b	20.9	6.6	.27	5.5
Nelson Word Meaning ^b	19.7	6.5	.35	5.1

^a Absolute or W-scores.

^b Raw scores.

Table 2

Correlations among Reading Component Scores

	Observed Variable Correlations									
	WA	PD	LSS	R	LW	SW	S	DS	PR	RC
WJR Word Attack (WA)	1									
TOWRE Phonemic Decoding (PD)	.726	1								
Letter-Sound Survey (LSS)	.788	.747	1							
WRAT Reading (R)	.747	.658	.708	1						
WJR Letter Word (LW)	.679	.539	.620	.772	1					
TOWRE Sight Word (SW)	.558	.683	.511	.563	.538	1				
WRAT Spelling (S)	.674	.656	.666	.664	.607	.555	1			
Developmental Spelling (DS)	.716	.677	.705	.707	.614	.571	.833	1		
Passage Reading (PR)	.609	.651	.570	.628	.629	.782	.625	.650	1	
Nelson Reading Comp (RC)	.356	.238	.283	.451	.425	.455	.360	.411	.516	1
Nelson Word Meaning (WM)	.502	.371	.443	.594	.583	.530	.476	.564	.618	.706

Table 3

Internal Consistency (Cronbach's alpha) for Native and Non-Native Adults

Test	Native	Non-native	Total sample
Woodcock Johnson-R Word Attack (WJR-WA)	.910	.768	.883
Letter Sound Survey (LLS)	.878	.791	.860
WRAT Reading (WRAT3-R)	.894	.805	.872
Woodcock Johnson-R Letter Word (WJR-LW)	.900	.794	.877
WRAT3 Spelling (WRAT3-S)	.905	.878	.897
Developmental Spelling (DS)	.902	.800	.879
Passage Reading – Correct words/min. (PR)	.975	.880	.964
Nelson Word Meaning (NWM)	.888	.816	.875
Nelson Reading Comprehension (NRC)	.902	.850	.896

Table 4

CFA and MGCFA Measures of Model Fit

	χ^2	Df	χ^2/b	NFI	TLI	CFI	χ^2/df
Two Factor Model							
Model 2A	647.08	43		.86	.87	.87	15.05
Model 2B	516.82	42	130.26*	.89	.86	.90	12.31
Model 2C	369.63	41	147.19*	.92	.90	.93	9.02
Three Factor Model							
Model 3A	525.17	41		.89	.86	.89	12.81
Model 3B	399.81	40	125.36*	.91	.89	.92	9.99
Model 3C	314.38	39	85.43*	.93	.91	.94	8.06
Five Factor Model							
Model 5A	196.74	34		.96	.94	.96	5.79
Model 5B	131.60	33	65.14*	.97	.96	.98	3.99
Model 5C (MG)	162.87	66		.97	.96	.98	2.47
Model 5D (MG)	170.92	72	8.05	.96	.97	.98	2.37

Note.

* $p < .05$. χ^2 = chi-square, Df = degrees of freedom, NFI = normed fit index, TLI = Tucker-Lewis index, CFI = comparative fit index.

Table 5

Model 5 Estimated Factor Correlations

Correlations for Native (Upper Diagonal) and Non-Native (Lower Diagonal) Groups					
	F1	F2	F3	F4	F5
Factor 1		.900	.887	.610	.820
Factor 2	.947		.875	.773	.845
Factor 3	.762	.739		.642	.804
Factor 4	.619	.810	.618		.690
Factor 5	.837	.842	.668	.774	

J Estimates obtained from Model 5D in which cross-group equality constraints were imposed on factor loadings.