

Title: Selecting Fluency Assessments for Adult Learners

Authors: Elena Nightingale, Daphne Greenberg, Lee Branum-Martin

Georgia State University

Published in the Journal of Research and Practice for Adult Literacy, Secondary, and Basic Education (2016), 5, 18-29.

Author Note:

The research reported here is supported by the Institute of Education Sciences, US Department of Education, through Grant R305C120001 Georgia State University. The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education.

Selecting Fluency Assessments for Adult Learners

Abstract

Selecting assessments for adults who struggle with reading can be difficult because few literacy measures used by reading researchers have been normed on this population, leaving uncertainty regarding the validity of these tests for adult learners. This study focused on the performance of 116 adults reading between the 3rd and 8th grade levels on a selection of reading fluency tests and other reading tests. The study examined the convergent and discriminant validity of the measures, taking into consideration the trait and method represented in the assessments in order to analyze how these reading fluency tests (TOSWRF, TOSCRF, WJ Reading Fluency) function when administered to a sample for whom they were not designed. The results suggest that there may be inconsistent patterns in the convergent and discriminant validity of these measures with this group of adult learners. Based on the findings of this study, suggestions are made for assessment selection for this population.

Selecting Fluency Assessments for Adult Learners

Fluency

Reading fluency can be defined as the combination of speed, accuracy, and prosody (Kuhn & Stahl, 2003), and is an area in which adult literacy students struggle (Greenberg et al., 2011; Mellard & Fall, 2012; Sabatini, Sawaki, Shore, & Scarborough, 2010). Fluency is an important skill to possess because it has been shown to be correlated to reading comprehension, the end goal in the development of reading skills (Fuchs, Fuchs, Hosp, & Jenkins, 2001). The link between fluency and reading comprehension skills may be attributed to the fact that fluency reduces the cognitive demand of conscious decoding of each individual word and allows the reader to instead move on to higher-order comprehension processes (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007).

While fluency and its relationship to comprehension has been studied extensively with children (Eason, Sabatini, Goldberg, Bruce, & Cutting, 2013; Kim, Wagner, & Lopez, 2012; Wise et al., 2010), research in this area with adults, particularly struggling adult readers, is limited (readers are encouraged to read the National Academy of Sciences report on Improving Adult Literacy Instruction; NRC, 2011). Some studies with adults have indicated that there may be evidence of a relationship between reading fluency and reading comprehension in adult literacy learners, though possibly weaker than the relationship in younger populations (Greenberg et al., 2011; Mellard, Woods, & Fall, 2011; Sabatini et al., 2010). For example, while test manuals cite correlations between fluency and comprehension for typically developing children at .8 fluency has been found to be correlated with adults at .54 (Sabatini, Sawaki, Shore, & Scarborough, 2010).

Selection of Appropriate Measures

As will be described, there are many different fluency tests on the market, and how to determine the best fluency measure to use can feel like a daunting task. One way to select an appropriate measure is by categorizing measures keeping a careful distinction between traits and methods (Campbell & Fiske, 1959; Maul, 2013). This approach can be useful in teasing out the similarities and differences between measures. “Trait” simply means the skill intended to be measured (e.g., fluency or reading comprehension). “Method” means the mode, test, or process by which the measure is taken. This approach necessitates practitioners and/or researchers to determine which trait one wants to measure and which method of assessment one wants to use.

Assessments may measure the same trait using the same method, the same trait using a different method, a different trait using the same method, or a different trait using a different method. For example, reading comprehension can be measured in speeded versus unspeeded methods. Alternatively, fluency and decoding could both be measured using passages of connected text. The multitrait/multimethod (MTMM) approach defines trait as the construct the test is designed to measure, and in this study, our focus is primarily on fluency (speed and accuracy).

Practitioners and researchers also need to take into account convergent and discriminant validity issues. Convergent validity is the extent to which two measures which should be related in constructs measured are actually related, and is important in our ability to interpret the results of the assessment as reflecting the actual ability it is intended to measure (Trochim, 2006; Messick, 1995). For example, we would expect two measures of reading fluency to have similar results if administered to the same examinee. Convergent validity can be determined by calculating the correlation coefficients between the two tests with a given sample (Campbell &

Fiske, 1959). Similarly, discriminant validity is the extent to which two assessments that are supposedly not related in content are actually unrelated. Discriminant validity can also be evaluated using correlation coefficients, looking for a lower correlation between two tests which measure different constructs. Therefore, we would expect tests measuring different constructs – such as an irregular word reading test and an algebra test – to have a lower correlation than tests measuring the same or similar constructs (such as an irregular word reading test and a general word reading test), and this is called discriminant validity. Among tests with strong validity evidence, one would expect higher correlations between tests that have trait and/or method in common. Because traits must always be measured via a certain method or combination of methods, separating trait from method effects becomes crucial to valid interpretation of test scores. Specifically, one would want to reduce variance based on shared or different methods in order to ensure that what is inferred from the scores is related to examinee ability rather than the methodology of the assessment (Messick, 1995).

Aim of Study

The purposes of this study were to (a) examine the convergent and discriminant validity of a selection of reading fluency tests administered to adult learners as well as to (b) investigate what the correlations among the fluency measures and other literacy measures suggest about their convergent and discriminant validity and the potential influence of measurement methods. This was done by analyzing how the tests described below are similar and dissimilar to each other as reflected in their correlations. The analysis is done with preliminary data from a larger study, as an illustration of this method. If the tests are working as intended with this population, they will have convergent validity with the tests measuring the same constructs (traits) using the

same methods and discriminant validity with tests measuring different constructs using different methods (speeded/not speeded).

Methods and Assessments

Participants

The participants in this study were engaged in a larger federally funded project, (to be acknowledged after peer review is complete), which assessed their underlying reading strengths and weaknesses. As indicated by answers to a demographic questionnaire, of 116 native English participants, 66% were female, 24% were employed, and 81% identified as Black/African American (others reporting as: 2% American Indian/Alaska Native, 9% White/Caucasian, 4% Asian, and 1% Native Hawaiian/Pacific Islander. Their ages ranged from 16-70 years old with a mean age of 39 (SD = 15) and a median age of 37.

Participants were recruited through their classes that targeted individuals reading at the third through eighth grade levels in Metro Atlanta. Research staff described the project to the learners, and those who were interested in participating met with research staff to learn about the study, and go through the consent process.

Assessments

On a mutually agreed date and time, participants were individually assessed in a private room across approximately 3-4 sessions (sessions ranged from 1-3 hours), a large battery of reading and reading related assessments administered in a fixed order. The testers were masters' and doctoral students in the fields of educational psychology, counseling psychology and communication sciences, and they all received intensive six-week training in both test administration and adult literacy sensitivity issues. Testers followed test manual procedures, and hand-recorded participants' responses. Upon participant test completion, all tests and scores were

carefully reviewed for basal/ceiling and scoring mistakes. Any mistakes were either quickly resolved or treated as missing data. Participants were paid ten dollars per hour of participation.

The measures included in this paper reflect a subset of the larger battery of assessments and were selected based on their being most congruent with the reading fluency aims of this study. Specifically, the following fluency assessments were administered: Test of Silent Word Reading Fluency (*TOSWRF*), Test of Silent Contextual Reading Fluency (*TOSCRF*), Woodcock Johnson III – Reading Fluency (*WJ-RF*), and Test of Word Reading Efficiency (*TOWRE*). The following non-fluency assessments were administered: Test of Irregular Word Reading Efficiency (*TIWRE*), Woodcock Johnson III – Letter Word Identification (*WJ III-LWI*), Woodcock Johnson III – Passage Comprehension (*WJ III-PC*), and Woodcock Johnson III - Word Attack (*WJ III-WA*). To assess convergent validity, this study investigated how the fluency measures relate to each other (*TOSWRF*, *TOSCRF*, *WJ-RF*, *TOWRE*). In order to assess discriminant validity, the fluency measures were compared to non-fluency literacy tests (*TIWRE*, *WJ III-LWI*, *WJ III-PC*, *WJ III-WA*).

Fluency Assessments

The Test of Silent Word Reading Fluency (*TOSWRF*; Mather, Hammill, Allen, & Roberts, 2004) is designed to measure silent reading fluency of single words (Mather et al., 2004). It has been normed for examinees aged 6-18 years of age with an average reliability (internal consistency) of .86. The test may be administered to an individual or group, and is 3 minutes in length, preceded by about 1-2 minutes of instruction for a total length of administration of 4-5 minutes. The examinee is presented with lines of words which are printed without spaces, and is asked to draw lines between as many words as possible in 3 minutes.

The Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) is intended to measure the speed with which students can silently recognize the individual words in a series of printed passages that become progressively more difficult in content, vocabulary, and grammar. It has been normed on ages 6-18 years old with an average reliability (internal consistency) of .86. The test may be administered to an individual or group, and is 3 minutes in length with a 2 minute practice form and 1-2 minutes of instruction, for a total length of administration of 6-7 minutes. The examinee is presented with passages in which all the words are printed together without spaces and is asked to draw lines between as many correct words as possible in the context of the passage in 3 minutes. (Hammill et al., 2006).

The Woodcock Johnson III Test of Achievement Reading Fluency subtest (WJ III-RF; Woodcock, McGrew, & Mather, 2001) is designed to assess reading speed by measuring the examinee's ability to silently identify whether a sentence contains correct or incorrect information. The test may be administered to an individual or group, and is normed for examinees ages 2 through 99 years with a reliability (internal consistency) of .90. This speeded test lasts for three minutes, during which the participant is instructed to read each sentence silently and to circle yes or no to identify whether the sentence is correct or incorrect, for as many sentences as they can, within the three minute time limit.

The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 2012) sight word reading subtest is individually administered and assesses the ability to recognize words which must be orthographically decoded as whole units. The test is normed for examinees 6-24 years old, and has a reported reliability (internal consistency) of .87. Administration for this speeded subtest is 45 seconds, in addition to time required for directions. During the test, the

examinee is asked to read aloud from a list of words, while the examiner scores each item as correct or incorrect, from which a final raw score is gathered.

The Test of Word Reading Efficiency (TOWRE; Torgesen et al., 2012) phonemic decoding subtest is individually administered and assesses the examinee's ability to sound out non-words which must be phonemically decoded to pronounce correctly. The test is normed for examinees 6-24 years old and has a reported reliability (internal consistency) of .87.

Administration for this subtest is 45 seconds in addition to the time required for directions.

During the test, the examinee is asked to read aloud from a list of non-words, while the examiner scores each item as correct or incorrect, from which a final raw score is gathered.

As indicated by the above descriptions, the assessments used to measure fluency can be categorized as requiring individuals to read orally or silently, and providing individuals with the words in or out of context (i.e., single words or sentences). Table 1 summarizes how the fluency assessments can be categorized.

Table 1: Fluency Assessments

	Oral/Silent	Real/not real words	In context/out of context
TOWRE: Sight Words	Oral	Real	Out
TOWRE: Phonemic	Oral	Not real	Out
WJ: Reading Fluency	Silent	Real	In
TOSWRF	Silent	Real	Out
TOSCRF	Silent	Real	In

Note: oral/silent indicates tests which are read orally (aloud) or silently. In context and out of context indicates tests with words in sentences or tests with lists of single words not in sentences. TOWRE=Test of Word Reading Efficiency, WJ=Woodcock Johnson, TOSWRF=Test of Silent Word Reading Fluency, TOSCRF=Test of Silent Contextual Reading Fluency

As described, fluency tests are not all alike. Reading fluency is often measured with variations of the words-per-minute method where the number of correctly identified words or

non-words in a set time is used to measure speed and accuracy, which can be done with items in and out of context. For example, fluency assessments may use real words out of context (Test of Word Reading Efficiency – Sight word decoding, TOWRE, Torgesen et al., 1999), non-words out of context (Test of Word Reading Efficiency – Phonemic decoding, TOWRE, Torgesen et al., 1999), and real words in context (Woodcock Johnson III, WJ– Reading Fluency, Woodcock et al., 2001). In addition, reading fluency can be measured orally (e.g., TOWRE, sight word and phonemic decoding, Torgesen et al., 1999) or silently (WJ – Reading Fluency, Woodcock et al., 2001). The Test of Silent Word Reading Fluency (TOSWRF, Mather et al., 2004) and Test of Silent Contextual Reading Fluency (TOSCRF, Hammill et al., 2006) are more recent tests of silent reading fluency. The TOSWRF measures silent word reading fluency out of context, and the TOSCRF measures silent word reading fluency in context. These differences in trait and method may influence what the test is actually measuring, and how one can interpret the results in the context of other measures, which is our focus in this study.

Non-Fluency Assessments.

Irregular Word Reading Efficiency.

The Test of Irregular Word Reading Efficiency (TIWRE; Reynolds & Kamphaus, 2007) assesses irregular word reading efficiency by measuring the examinee’s ability to verbally identify phonetically irregular words from a list. The test has been normed on individuals aged 3 to 94 years old and reports internal consistency for all forms in the mid to high .90s. This test is not a speeded measure, and it involves presenting the examinee with phonetically irregular words and letters, which they identify orally until they identify four words incorrectly, after which administration ceases.

Letter-Word Identification.

The Woodcock Johnson III Test of Achievement subset of Letter-Word Identification (WJ III-LWI; Woodcock et al., 2007) is designed to measure the ability to recognize and orally identify words and letters. This test has been normed on children and adults, ages 2 through 99 years old with internal consistency reliability of .94. It is not speeded and takes about three to five minutes to administer, which is done by presenting the examinee with lists of words, until six words in a row are identified incorrectly, and moving backward from the starting point of the test as needed until six words in a row are identified correctly (in this study, all participants started with item number 33).

Passage Comprehension.

The Woodcock Johnson III Test of Achievement subtest for Passage Comprehension (WJ III-PC; Woodcock et al., 2007) assesses passage comprehension by measuring the examinee's ability to correctly provide the single missing word in a sentence or passage. The measure is normed for ages 2 through 99 years old, is not speeded, and can take between 5 to 20 minutes to administer. The median reliability (internal consistency) reported is .88. Administration involves presenting the examinee with series of sentences with one word left missing and instructing the examinee to read the sentence silently and provide aloud the one word which goes in the blank (in this study, all participants started with the item 14).

Word Attack.

In the Woodcock Johnson III Test of Achievement Word Attack subtest (WJ III-WA; Woodcock et al., 2007), phonemic decoding is assessed by measuring the examinee's ability to pronounce nonsense words. The subtest is normed on individuals from 2 to 99 years of age, at .87 reliability (internal consistency), is not speeded, and takes less than 5 minutes to administer.

The examinee is asked to pronounce pseudo-words orally until six words in a row are pronounced incorrectly, and testing backwards as needed until the six lowest items presented are identified correctly (in this study, all participants start with item number 4).

Results

The performance on these assessments (see Table 2) shows that the participants struggled on all of the assessments, but in particular on assessments which pertain to phonemic decoding abilities, such as the phonemic decoding subtest of the TOWRE. Although they were not strong in any one area of the tests given, the area in which this group performed the highest was on irregular (sight) word reading and the lowest was phonemic decoding. The average grade equivalencies, based on the grade equivalencies used in the manuals for the average total score, for all but two tests were between the 3rd and 5th grade level. Information on missing data is also included below in Table 2, and pair-wise deletion was used to handle missing cases.

Table 2: Descriptive Statistics (raw scores) for literacy tests administered.

	Mean	SD (total)	Min	Max	Grade Equ.	n
TOSWRF	94.55	25.68	23	153	5.2	115
TOSCRF	85.46	28.58	14	155	5.2	114
WJ: RF	44.94	12.50	11	82	5.2	112
TOWRE: Sight	65.68	14.50	29	95	3.5	112
TOWRE: Phonemic	20.76	12.88	1	52	2.2	110
TIWRE	37.98	4.82	23	48	7.1	111
WJ: LWI	55.42	8.10	33	72	5.3	102
WJ: PC	29.51	4.41	18	39	4.5	108
WJ: WA	15.66	7.76	1	31	3.1	104

Note. Varying sample sizes are due to missing data. Fluency tests are listed first above the darkened line, followed by other reading assessments below the darkened line. *TOSWRF*=*Test of Silent Word Reading Fluency*, *TOSCRF*=*Test of Silent Contextual Reading Fluency*, *TOWRE:S*=*Test of Word Reading Efficiency: Sight Word Reading Efficiency*, *TOWRE:P*=*Test of Word Reading Efficiency: Phonemic Decoding Efficiency*, *TIWRE*=*Test of Irregular Word Reading Efficiency*, *WJ-III*=*Woodcock Johnson III*, *RF*=*Reading Fluency*, *PC*=*Passage Comprehension*, *WA*=*Word Attack*, *LWI*=*Letter-Word Identification*; Grade equivalents are only shown to aid in the practical interpretation of the reported means.

This study classified the tests to tease apart potential interactions with trait and method within these assessments (see Table 1). A description of our analysis is provided, followed by a discussion of the results and implications.

1. Same Trait, Same Method (ST/SM): The following three pairs of tests assess similar traits, using similar methods: TOSWRF/TOWRE-S (word identification/speeded), TOSCRF/WJ-RF (sentence fluency/speeded), and TIWRE/WJ-LWI (word identification/not-speeded).

2. Same Trait, Different Method (ST/DM): These seven pairs of tests assess similar traits, but do not use similar methods. TOSWRF/TIWRE (word identification, speeded/not speeded), TOSCRF/WJ-PC (sentence fluency, speeded/not speeded), WJ-RF/WJ-PC (sentence fluency, speeded/not speeded), TOWRE-S/TIWRE (word identification, speeded/not speeded), TOWRE-P/WJ-WA (nonword identification, speeded/not speeded), TOWRE-S/WJ-LWI (word identification, speeded/not speeded), and TOSWRF/WJ-LWI (word identification, speeded/not speeded).

3. Different Trait, Same Method (DT/SM): These 13 pairs of tests do not assess the same trait, but do use similar methods. Specifically, they are all matched by the fact that the pairs are either speeded or non-speeded. TOSWRF/TOSCRF (words out of/in context, speeded), TOSWRF/WJ-RF (words out of/in context, speeded), TOSWRF/TOWRE-P (real words/non-words, speeded), TOSCRF/TOWRE-S (words in/out of context, speeded), TOSCRF/TOWRE-P (real words/non-words, speeded), WJ-RF/TOWRE-S (words in/out of context, speeded), WJ-RF/TOWRE-P (real words/non-words, speeded), TOWRE-S/TOWRE-P (real words/non-words, speeded), TIWRE/WJ-PC (words out of/in context, non-speeded), WJ-WA/TIWRE (non-words/real words, non-speeded), WJ-WA/WJ-PC (non-word identification/passage

comprehension, not speeded), WJ-LWI/WJ-PC (words out of/in context, not speeded), and WJ-LWI/WJ-WA (real words/non-words, not speeded).

4. Different Trait, Different Method (DT/DM): The 13 pairs of tests in this category are neither assessing similar traits nor are they using similar methods: TOSWRF/WJ-PC (word out of/in context, speeded/not speeded), TOSCRF/TIWRE (words in/out of context, speeded/not speeded), WJ-RF/TIWRE (words in/out of context, speeded/not speeded), TOWRE-S/WJ-PC (words out of/in context, speeded/not speeded), TOWRE-P/TIWRE (non-word identification/word identification, speeded/not speeded), TOWRE-P/WJ-PC (non-word identification/passage comprehension, speeded/not speeded), WJ-WA/TOSWRF (non-word identification/word identification, not speeded/speeded), WJ-WA/TOSCRF (non-word identification/passage comprehension, not speeded/speeded), WJ-WA/WJ-RF (non-word/real word, not speeded/speeded), WJ-WA/TOWRE-S (non-word identification/word identification, not speeded/speeded), WJ-LWI/TOSCRF (words out of/in context, not speeded/speeded), WJ-LWI/WJ-RF (words out of/in context, not speeded/speeded), and WJ-LWI/TOWRE-P (word identification/non-word identification, not speeded/speeded).

A summary of the above is found in Table 3, (to make the table visually easy to read, pairs of tests are classified as follows: same trait (ST), same method (SM), different trait (DT), and different method (DM). In addition, the four categories of pairings of trait and method are numbered as: ST/SM (1), ST/DM (2), DT/SM (3), DT/DM (4).

Table 3: *Traits and Methods*

Test	TOSCRF	WJ:RF	TOWRE:S	TOWRE:P	TIWRE	WJ:PC	WJ:WA	WJ:LWI
TOSWRF	DT/SM (3)	DT/SM (3)	ST/SM (1)	DT/SM (3)	ST/DM (2)	DT/DM (4)	DT/DM (4)	ST/DM (2)
TOSCRF	-	ST/SM (1)	DT/SM (3)	DT/SM (3)	DT/DM (4)	ST/DM (2)	DT/DM (4)	DT/DM (4)

WJ:RF	-	-	DT/SM (3)	DT/SM (3)	DT/DM (4)	ST/DM (2)	DT/DM (4)	DT/DM (4)
TOWRE:S	-	-	-	DT/SM (3)	ST/DM (2)	DT/DM (4)	DT/DM (4)	ST/DM (2)
TOWRE:P	-	-	-	-	DT/DM (4)	DT/DM (4)	ST/DM (2)	DT/DM (4)
TIWRE	-	-	-	-	-	DT/SM (3)	DT/SM (3)	ST/SM (1)
WJ:PC	-	-	-	-	-	-	DT/SM (3)	DT/SM (3)
WJ:WA	-	-	-	-	-	-	-	DT/SM (3)

Note: Fluency tests are listed first above the dark line, followed by other reading assessments below the dark line. TOSWRF=Test of Silent Word Reading Fluency, TOSCRF=Test of Silent Contextual Reading Fluency, TOWRE:S=Test of Word Reading Efficiency: Sight Word Reading Efficiency, TOWRE:P=Test of Word Reading Efficiency: Phonemic Decoding Efficiency, TIWRE=Test of Irregular Word Reading Efficiency, WJ-III=Woodcock Johnson III, RF=Reading Fluency, PC=Passage Comprehension, WA=Word Attack, LWI= Letter-Word Identification; DT = Different Trait; ST= Same Trait; SM= Same Method; DM= Different Method, ST/SM (1), ST/DM (2), DT/SM (3), DT/DM (4).

To examine convergent and discriminant validity, a correlation matrix of the correlation coefficients of the total scores for the assessments is shown in Table 4. The correlations seem to show inconsistent patterns of convergent and discriminant validity, with some cases of tests with different traits or methods having higher correlations than tests with the same traits or methods.

More on the specific examples of this is discussed below.

Table 4: Correlations between all measures

Test	TOSCRF	WJ:RF	TOWRE:S	TOWRE:P	TIWRE	WJ:PC	WJ:LWI	WJ:WA
TOSWRF	0.75	0.50	0.47	0.28	0.36	0.33	0.34	0.30
TOSCRF	-	0.67	0.47	0.39	0.47	0.36	0.42	0.41
WJ:RF	-	-	0.59	0.55	0.51	0.33	0.49	0.45
TOWRE:S	-	-	-	0.60	0.45	0.51	0.52	0.39
TOWRE:P	-	-	-	-	0.66	0.51	0.77	0.79
TIWRE	-	-	-	-	-	0.52	0.80	0.69

WJ:PC	-	-	-	-	-	-	0.57	0.36
WJ:LWI	-	-	-	-	-	-	-	0.72

Note. $n = 116$. All correlations are significant at $p < .001$. Fluency tests are listed first above the dark line, followed by other reading assessments below the dark line. *TOSWRF=Test of Silent Word Reading Fluency*, *TOSCRF=Test of Silent Contextual Reading Fluency*, *TOWRE:S=Test of Word Reading Efficiency: Sight Word Reading Efficiency*, *TOWRE:P=Test of Word Reading Efficiency: Phonemic Decoding Efficiency*, *TIWRE=Test of Irregular Word Reading Efficiency*, *WJ-III=Woodcock Johnson III*, *RF=Reading Fluency*, *WA=Word Attack*, *LWI=Letter-Word Identification*, *PC=Passage Comprehension*

The correlations in Table 4 seem to suggest a method effect: shared methods may consistently result in higher correlations where shared traits may not have as strong of an effect. The highest correlation here between fluency measures (thus, they are Same-Trait/Same Method) was between the TOSCRF and WJ Reading Fluency ($r = .67$). Interestingly, this was not the highest correlation of all the tests, which instead was the WJ Letter-Word ID and the TIWRE ($r = .80$), which are not fluency measures, but they do measure the same trait with the same method. But looking at the method effect, examples of fluency tests with the same method and different trait often resulted in high correlations. The highest correlation of this kind is also the highest correlation of all the fluency tests; it is between the very similarly structured TOSWRF and TOSCRF ($r = .75$). Other high correlations indicative of a method effect include WJ Word Attack with both the WJ Letter-Word Id and TIWRE ($r = .72$ and $.69$, respectively).

When examining the results for tests with shared traits, however, it is interesting because one can see two cases of a high correlation: the aforementioned TOSCRF/WJ-RF pairing ($r = .67$), and the WJ Word Attack and the TOWRE Phonemic Decoding subtest ($r = .79$). All other cases were far lower ($r = .33$ to $.52$). Consistently low correlations in this category could indicate that a shared method is more important to concurrent validity than sharing traits being measured.

Surprisingly, there are a few instances of high correlations between tests that are measuring different traits and different methods. The highest correlation of this kind is found between the WJ Letter Word ID and the TOWRE Phonemic Decoding subtest ($r = .77$). Even the lowest correlation of these pairs (TOSWRF and WJ Word Attack, $r = .30$) is not the lowest of the assessments examined.

Discussion

The goal of this study was to analyze the correlations between various reading fluency measures and discuss observed patterns of convergent and discriminant validity while taking into consideration the traits and methods represented. Based on the correlations shown in the results, one can see that correlations ranged from .28 (TOSWRF and TOWRE:P) to .75 (TOSWRF and TOSCRF), meaning that the interrelationships among these assessments range from very low to very high. The patterns of convergent and discriminant validity, however, do not paint a straightforward picture. For example, the study found a few instances of high correlations between tests that are measuring different traits and different methods. The highest correlation of this kind is found between the WJ Letter Word ID and the TOWRE Phonemic Decoding subtest ($r = .77$). Even the lowest correlation of these pairs (TOSWRF and WJ Word Attack, $r = .30$) is not the lowest of the assessments examined. This means that for this set of tests in this sample, some correlations fit as expected from consistent traits and methods, and some correlations did not.

To interpret the trait/method interaction, examining all four categories of correlations here, and the results show that the top four correlations are representative of each of the four relationships. Specifically, a pair representing same method/same trait (SM/ST) is: TIWRE/WJ-LWI - $r = .80$; a pair representing same trait/different method (ST/DM) is: WJ-WA/TOWRE-P -

$r = .79$; a pair representing different trait/same method (DT/SM) is: TOSWRF/TOSCRF - $r = .75$; a pair representing different trait/different method (DT/DM) is: TOWRE-P/WJ-LWI - $r = .77$. Looking at the rest of the correlations, there are some patterns, such as the mostly low correlations in the category with the same trait and different methods (TOSWRF/WJ-LWI - $r = .34$, WJ-RF/WJ-PC - $r = .33$).

As a whole, the correlations are not consistently in line with a straightforward interpretation that these are valid tests with minimal method influence in this population. Tests that had the same method and measured the same trait had a correlation of .47 (the TOSWRF with the TOWRE-S), while several pairs of assessments which shared neither trait nor method had correlations above .60 (eg: TOWRE-P/TIWRE, WJ-LWI/TOWRE-P). This seems to indicate that for this population, these tests do not have the level of convergent and discriminant validity that one would hope to find.

Limitations

Several limitations characterize this study. This sample represents a small, preliminary group of participants from an ongoing project, and as examinees may have chosen to participate in order to contribute to research or receive payment, there may be a selection bias present. Additionally, the reading levels were determined by adult literacy programs' previous administrations of their own assessments – for example, the Test of Adult Basic Education (TABE). In addition, the study only included native English speakers, and Hispanics were not represented in this sample. Larger, more representative samples may give more dependable results. The use of these tests on this population also presents a limitation in that the tests were not designed for struggling adult readers, and therefore some of the information provided for each test, such as internal consistency, may not apply to this sample. Also, responses were not

audiotaped, and it is possible that audiotaped responses might have clarified some of the results of this study. Finally, in addition to the goal of analyzing the assessments/population discussed above, the current study used visual inspection of correlation matrices (Campbell & Fiske, 1959). It should be noted that confirmatory statistical models are available, which can be used to better separate trait from method sources of variability (Nussbeck, Eid, & Lischetzke, 2006; Maul, 2013).

Implications for practitioners

The take-home message for practitioners is that great care should be taken when reading studies about struggling adult readers' performances on reading fluency tests, and/or when selecting reading fluency measures. If a measure is not specifically designed for adults who have difficulties with reading, the results of the test may not be as accurate as they are for the normed population described in the test manual. All standardized tests come with a technical manual. These manuals should be carefully reviewed to evaluate the appropriateness of the intended sample to the norming population, and what types of convergent and discriminant validities are reported. This study shows that when tests are administered to populations for whom they were not designed or normed, they may not be measuring the same trait in the same way and with the same level of accuracy and interpretability as the assessment would on the population on whom it was normed. In addition, convergent and discriminant validities may show different patterns. This information will allow those who are considering tests to make an informed decision when selecting a test, as well as help with understanding of research study results.

Implications for researchers

This investigation builds upon past research which has indicated that tests not designed for adults – and struggling adult learners – may not function the same with them (Greenberg,

Pae, Morris, Calhoun, Nanda, 2009; Nanda, Greenberg, & Morris, 2014), giving further evidence for careful selection and interpretation of assessments for struggling adult learners. The convergent and discriminant validity patterns of this study indicate that the tests do not relate to established measures in expected patterns and to expected degrees for this population. When examining the convergent and discriminant validity of the reading fluency tests analyzed here, in some instances, the pairs of tests fell in line with expected patterns of convergent and discriminant validity while others did not. For example, while tests which assessed the same trait tended to have higher correlations than with tests that did not share the same trait, indicating there is a level of convergent validity, tests that did not measure the same trait still at times had high correlations, particularly when sharing a method. This indicates that method effects may be strong for these tests in this population.

When looking for one test that does fall in line with what one would hope to find in terms of (a) higher correlations with tests with shared traits and/or methods and (b) lower correlations with test with differing traits and/or methods, it is clear that not one test from this list consistently follows this pattern. Every assessment here has at least one example of a test with no shared trait or method correlating higher than a test with a shared trait or method. However, within this population, the fluency test from this sample of assessments with the highest concurrent and discriminant validity that falls in line with what one would expect is the Woodcock Johnson Reading Fluency subtest, followed by the Test of Word Reading Efficiency Sight Word Reading subtest. Of the fluency tests, these two tests had the highest concurrent validity with other fluency measures. Specifically, in the area of fluency assessments, the potentially complex factors of speed and accuracy create the need to take both method and trait into consideration when working to understand these tests with this population. Further research

with larger sample sizes, using confirmatory statistical analyses, is needed before definite conclusions can be reached.

References

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Eason, S., Sabatini, J., Goldberg, L., Bruce, K., & Cutting, L. (2013). Examining the relationship between word reading efficiency and oral reading rate in predicting comprehension among different types of readers. *Scientific Studies of Reading*, *17*, 199-223.
- Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239-256.
- Greenberg, D., Wise, J., Morris, R., Fredrick, L., Rodrigo, V., Nanda, A., & Pae, H. (2011). A randomized control study of instructional approaches for struggling adult readers. *Journal of Research on Educational Effectiveness*, *4*, 101-117.
- Greenberg, D., Pae, H., Morris, R., Calhoun, M. B., Nanda, A. (2009). Measuring adult literacy students' reading skills using the Gray Oral Reading Test. *Annals of Dyslexia*, *59*, 133-149. doi:10.1007/s11881-009-0027-8. PubMed # 19629705.
- Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *Test of Silent Contextual Reading Fluency*. Austin, TX: PRO-ED.
- Kim, Y. S., Wagner, R., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade one to two. *Journal of Experimental Child Psychology*, *113*, 93-111.
- Kuhn, M., & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*, 22-40.
- Mather, N., Hammill, D., Allen, E., & Roberts, R. (2004). *Test of Silent Word Reading Fluency*. Austin, TX: PRO-ED.

- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*, 4, 169. doi:10.3389/fpsyg.2013.00169
- Mellard, D., & Fall, E. (2012). Component model of reading comprehension for adult education participants. *Learning Disability Quarterly*, 35, 10-23.
- Mellard, D. F., Woods, K., & Fall, E. (2011). Assessment and instruction of oral reading fluency among adults with low literacy. *Adult Basic Education and Literacy Journal*, 5(1), 3-14.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Nanda, A. O., Greenberg, D., & Morris, R. (2010). Modeling child-based theoretical reading constructs with struggling adult readers. *Journal of Learning Disabilities*, 43, 139-153.
- Nanda, A., Greenberg, D., Morris, R. (2014). Reliability and validity of the CTOPP Elision and Blending Words subtests for struggling adult readers. *Reading and Writing: An Interdisciplinary Journal*, 27, 1603-1618. doi: 10.1007/s11145-014-9509-0.
- National Research Council [NRC]. (2011). *Improving Adult Literacy Instruction: Options for Practice and Research*. Committee on Learning Sciences: Foundations and Applications to Adolescent and Adult Literacy, Alan M. Lesgold and Melissa Welch-Ross, Editors. Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.
- Nussbeck, F.W., Eid, M., & Lischetzke, T. (2006). Analyzing MTMM data with SEM for ordinal variables applying the WLSMV-estimator: What is the sample size needed for valid

- results? *British Journal of Mathematical and Statistical Psychology*, 59, 195-213. doi: 10.1348/00071100SX67490
- OECD (2013), *OECD Skills Outlook 2013: First results from the survey of adult skills*, OECD Publishing. <http://dx.doi.org/10.1787/9789264204256-en>.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling adult readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11, 289-312.
- Reynolds, C., & Kamphaus, R. W. (2007). *Test of Irregular Word Reading Efficiency*. Lutz, FL: Psychological Assessment Resources, Inc.
- Sabatini, J., Sawaki, Y., Shore, J., & Scarborough, H. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities*, 43, 122-138.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency, 2nd Edition*. Austin, TX: PRO-ED.
- Trochim, W. M. The research Methods knowledge base, 2nd Edition. Internet WWW page, at URL: <<http://www.socialresearchmethods.net/kb/>> (version current as of October 20, 2006).
- Wise, J. C., Sevcik, R. A., Morris, M. D., Lovett, M. V., Wolf, M., Meisinger, B., & Schwanenflugel, P. (2010). The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. *Language, Speech, and Hearing Services in Schools*, 41, 340-348.
- Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock Johnson III Test of Cognitive Abilities*. Rolling Meadows, IL: The Riverside Publishing Company.

