

Submitted to: *Assessments for Effective Interventions*

**Waterford Assessment of Core Skills: A Computerized Adaptive Reading
Test for Pre-kindergarten through 2nd Grade**

Haya Shamir, Ph.D.

Waterford Research Institute, Salt Lake City, UT

Erin Phinney Johnson, Ph.D.

Waterford Research Institute, Salt Lake City, UT

Author Contact:

J c { c " U j co kt

Waterford Research Institute

55 West 900 South

Salt Lake City, UT 84101

jc{cuj.co.kt@waterford.org

(801) 349-2453

FAX: (801) 363-1508

**Waterford Assessment of Core Skills: A Computerized Adaptive Reading Test for
Pre-kindergarten through 2nd Grade**

Abstract

Assessing students' emerging literacy skills is crucial for identifying areas where a child may be falling behind and can lead directly to increased reading success. The Waterford Assessment of Core Skills (WACS), a new computerized adaptive test of early literacy for students in prekindergarten through 2nd grade, addresses this need with young school-aged children. Initial content validity for WACS was established against state and national standards for the 11 subtests. All items were then calibrated for IRT to determine item difficulty. To establish concurrent validity and predictive validity student performance on WACS was compared to performance on five commonly-used standardized tests also measuring early reading skills; all correlations between tests are highly significant. Additional analyses indicate that WACS is internally coherent and has strong test-retest reliability.

Keywords: early education, literacy, computerized adaptive test, validity

Waterford Assessment of Core Skills: A Computerized Adaptive Reading Test for Pre-kindergarten through 2nd Grade

Introduction

While acquiring basic literacy skills has long been considered among the most important elements of early childhood education, many recent efforts to improve reading instruction in U.S. primary schools have not met with encouraging results, especially among lower-performing students (Viteritti, 2004, p. 69; Guthrie & Springer, 2004; Cohen, Raudenbush, & Ball, 2003). In 2005, the National Assessment of Educational Progress (NAEP) reported that more than one-third of American 4th-grade students performed at the lowest level (Below Basic) on the NAEP reading skills test, a measure of reading comprehension (National Center for Education Statistics NCES, 2008). It is clear that a large percentage of students continue to struggle with basic reading skills during their first years in school, and that these difficulties can result in deficits that remain, or grow, in the later grades (Cunningham and Stanovich, 1997; Whitehurst, 2003). Some researchers have noted that large differences in reading technique and achievement begin to emerge as early as first grade (Stanovich, 2000) and, unfortunately, students behind during the first years of school tend to learn at a slower rate than students who begin ahead; often, this results in a so-called “Matthew effect” for reading skills, in which the academically “rich” become richer and the “poor” become poorer (Walberg, 2003).

The Waterford Assessment of Core Skills (WACS) is a newly available computerized adaptive test (CAT) of pre-literacy and early literacy skills for kindergarten

through 2nd-grade students. The assessment is comprised of 11 separate subtests, including measures of letter recognition, letter sound and initial sound recognition, blending, segmenting, real and non-word reading, reading comprehension, listening comprehension, and vocabulary. Results from WACS are available immediately after a student completes the assessment, giving educators real-time data for instructional planning and evaluation. Because the assessment contains many items and tailors itself to each student's abilities, it can be reliably administered multiple times per year. Just as importantly, administering WACS is a simple, streamlined process for educators. The test requires minimal training and minimal time from teachers.

Constructs Assessed by WACS

According to some researchers, it is the cumulative, sequential nature of reading skills that accounts for much of the difficulty lower-performing students experience when trying to catch up to their peers (Bast & Reitsma, 1998; Snow et al, 1998). These observations have led to the so-called "causal" model of early reading skills: early-appearing abilities like phonological awareness are thought to facilitate decoding, which in turn facilitates word recognition, which in conjunction with listening comprehension determines reading comprehension (Stanovich, 2000). Based on this model, WACS tests several concepts, some of which build upon one another.

Phonological and phonemic awareness

Phonological and phonemic awareness refer to the ability to parse the sounds of language into increasingly smaller units. These abilities often appear earliest and are

considered among the most important for a child's transition from speaker to reader. Phonological awareness, the more rudimentary skill, has been defined as the general appreciation of the sounds of speech as distinct from their meaning (Snow et al, 1998). Because it has proven to be very important for the development of a number of later reading skills—as well as widely predictive of successful reading comprehension later in life—phonological awareness is strongly emphasized in both educational literature and the development of new instructional techniques (National Reading Panel, 2000). Phonemic awareness, defined as an understanding that words can be divided into a sequence of sounds (phonemes), is a further refinement of earlier phonological awareness skills (see Snow, 1998).

While basic phonological awareness has been described as more of a school-readiness skill, many teachers and educational programs begin explicit instruction in phonemic awareness when a child is in kindergarten or 1st grade. Because the development of phonological and phonemic awareness skills is tied causally to later pre-literacy skills, falling behind due to inadequate or inappropriate instruction may have a considerable effect on a child's progress in the subsequent months and years (Snow et al, 1998). The Initial Sound, Blending, and Segmenting subtests in WACS are designed to assess these early phonological skills.

The alphabetic principle and decoding

Phonological skills lead either directly or indirectly toward an understanding of the alphabetic principle, which Stanovich (2000) points to as a “discovery” that letters on the page correspond to or map out word sounds. The skill known as decoding, then, refers to the

process of identifying the sounds of each of a word's graphemic (written) components. A child is said to have decoded a word when he or she can pronounce—though not necessarily comprehend—all the sounds in that word. When children are just beginning to read, decoding can sometimes be so slow that even familiar word meanings will not be understood (Stanovich, 1986).

As with phonological and phonemic awareness skills, students who fall behind while learning basic decoding skills often have difficulty catching up. Children who experience difficulty “breaking the spelling-to-sound code” (Stanovich, 2000) naturally read more slowly and begin to be exposed to less text than other peers; these difficulties combined with resulting motivational problems contribute to increasingly lower achievement (Snow et al, 1998). Subtests addressing decoding skills in WACS include Letter Recognition, Letter Sound, and Real and Non-word Reading.

Reading comprehension

Phonological and decoding skills together constitute what has been called a “map” for comprehending written language. It is by these abilities, rather than by comprehension abilities per se, that children are most often constrained by in the early grades. According to some research, nearly all children in the early grades are capable of understanding the content of what they are asked to read, but deficiencies or slowness with more basic reading skills can sometimes prevent them from doing so (Gough, Hoover, & Peterson, 1996). The NAEP results cited above clearly demonstrate that comprehension, if not as often a problem for students when they begin school, can present considerable difficulties of its own in subsequent years (see NCES, 2008; 2010). Other research has pointed to specific

comprehension-related difficulties not thought to be related to earlier skills (Oakhill & Yuill, 1996). In order to help educators distinguish whether or not a problem may be related to comprehension difficulties, WACS includes subtests for both Reading Comprehension and Listening Comprehension.

Vocabulary

Research into vocabulary development has demonstrated that, as with other aspects of early literacy, vocabulary knowledge represents a “malleable factor”; in other words, a child’s vocabulary can be directly and positively affected by early educational programs (see Marzano, 2004; Coyne, McCoach, & Kapp, 2007). There is substantial evidence that large individual differences in vocabulary size develop during early childhood (e.g., Biemiller & Slonim, 2001) and that vocabulary sizes during the primary grades are at least predictive of later academic achievement (Scarborough, 2001; Cunningham & Stanovich, 1997; Storch & Whitehurst, 2002). Vocabulary deficits are also thought to contribute directly to reading comprehension difficulties (Beck & McKeown, 1991; Stahl & Nagy, 2006; Wagner, Muse, & Tannenbaum, 2007). WACS includes a subtest specifically designed to detect vocabulary deficits.

Instrument

The Waterford Assessment of Core Skills was created to be engaging for young children. Throughout the test, students are guided by a groundhog named Wyatt, a “town sheriff” who asks for their help in completing a number of tasks they have to do together. At the end of each section and at the end of the assessment, students receive a non-judgmental

reward screen that serves as a short mental break. Then, after students have completed the entire assessment, Wyatt presents them with a deputy badge as a reward for finishing the test.

Content

Content validity was established for the items described above by thoroughly researching recommended content for this age group. Content experts first investigated the most important skills for pre-kindergarten through 2nd grade students and established guidelines for writing items based on published research. All comprehension passages and questions were then written by professional writers, reviewed by content experts, and edited by writing experts. Reading comprehension passages were also Lexile certified for their grade levels. Upon completion, items from all of the subtests were reviewed by additional content experts and sent to Marilyn Jager Adams, an external content expert, for review. After IRT testing, analysis was conducted to ensure that item difficulty, as determined statistically by IRT analysis, correlated with the item difficulty as determined by the content experts.

In addition to creating items based on researched concepts, a valid reading test should also cover standards accepted by the states for reading and language development. Thus, state standards were examined and correlated with WACS skills and items. With the exception of Iowa (which did not list standards below grade three), WACS assessed skills (from a minimum of three to a maximum of eleven) were also listed as state education standards for pre-kindergarten through grade two.

Item

All test items are presented both visually and aurally, though none require the child to speak the answer. For all skills, the computer introduces the question, including the correct answer and the distracters. Students can use the mouse to roll over the question or the answer options to hear the instructions again. However, the actual layouts of the individual assessments differ. For Letter Sound, Real Words, Sight Words, and Nonwords, the letter/word appears on the screen with three speakers underneath. Each speaker says a different letter/word name. The student must click on the correct speaker to match the word or letter that is on the screen. Vocabulary differs slightly from this arrangement with a sentence missing one word written at the top of the screen. The sentence is read to the child and the child must pick a word from the speakers that best completes the sentence. In the case of Letter Recognition, Initial Sound and Blending there is a single speaker or picture at the top of the screen and three pictures or letters at the bottom of the screen (see Figure 1 for example). The speaker/picture emits a sound or a series of sounds and the student must select the picture below that begins with that same sound or that matches the series of sounds, or select the letter that matches the letter name from the speaker.

For Reading and Listening Comprehension the student is given a passage to read or listen to. When finished, the child is presented with a question about the passage followed by three possible answers. In Listening Comprehension the questions and answers are presented aurally. Each reading or listening passage includes four questions of varying difficulty. Each child receives three passages depending on skill level. The final task, Segmenting, differs from all of the other tasks. Here, a picture is presented to the student and he/she must move a

series of blocks representing the sounds into the correct order for the word associated with the picture.

Adaptive Sequencing

Based on Item Response Theory (IRT), the computer-adaptive sequencing in WACS avoids floor and ceiling effects by choosing appropriate questions based on prior student responses. Rather than simply identifying that a student is “at-risk” or “at benchmark” on a predetermined set of assessments, WACS reports grade-level performance on a nationally-normed scale for each of its 11 sub-tests. As a child answers questions correctly within a content area, the questions’ difficulty level increases; if a child fails to answer questions correctly, the child will receive less difficult questions.

Importantly, all children do not receive all assessments; the assessments with which a student is tested depend on grade level and prior performance. Limiting the number of assessments completed by each child reduces the time required for testing and provides educators with a more child-appropriate, “pinpointed” report. If students from advanced grades fail initial grade-level testing, for example, WACS will automatically test these students on more basic skills as well in order to determine where more instruction and practice is needed. On the other hand, advanced students in earlier grades may receive more advanced assessments if they prove to have mastered the more basic skills.

Reports

One of the greatest benefits of computerized testing in the elementary grades is their ability to report scores immediately after the test is completed. WACS has been designed to

not only provide instant reports on individual test takers but also to compile data for use at the class, school, and district levels. Individual reports indicate the test taker's "grade" level for each of the completed assessments. A student's raw scores are supplemented with detailed information about which abilities each subtest assesses, in which areas he or she can be said to have demonstrated "competence," and ways that any problem areas can be addressed in the home or classroom. Each report also includes results from the student's previous WACS assessments, allowing educators and parents to evaluate changes and improvements over time. Individual scores can then be compared to class, school, and district averages. WACS provides specialized reports based on demographic information, NCE, grade level, raw scores, and proficiency groups; results can be represented either graphically or numerically (as tables). To provide educators with the tools for further analysis, WACS can also export student or group scores to an external spreadsheet program.

Methods

Phase I: Item Response Theory

Sample and Procedure

About 8,700 students in Utah, Idaho, Nevada, California, New York, Texas, North Carolina, and Florida completed the first round of testing with WACS. Two trained test administrators were present at each school to ensure valid data collection. Students completed a random sample of questions from each assessment, all questions representing varied expected difficulty levels. The sample of students from twenty six schools was

representative of US socio-economic status, ethnicity, geographic location, and type of school, based on information obtained from the US 2002 census.

Phase II: Validation

Sample

About 2,000 students in Utah, Idaho, New York, Texas, and Florida completed the validation testing with WACS (Table 1). Two trained test administrators were present at each school to ensure valid data collection. Students completed the entire WACS test in one or two 20 minute sessions, depending on student grade level and proficiency. The sample of students was selected from seven of the original 26 schools participating in IRT data collection.

Table 1. Number of students completing comparison test.

Assessment		Number of Students		
		Kindergarten	First	Second
DIBELS	Fall	206	142	137
	Spring	135	140	139
IRI	Fall	120	61	126
	Spring	115	136	79
ITBS	Fall	--	66	69
	Spring	--	91	102
TPRI	Fall	170	155	190
	Spring	--	94	103
SAT 10	Fall	--	--	--
	Spring	--	132	142

Instruments

WACS. As described earlier, the Waterford Assessment of Core Skills includes 11 subskills which may or may not be seen by a given child, depending on ability and grade level. Thus, each child's final score for the purposes of these analyses is based on an average of the tasks that child completed, excluding any tasks passed at the ceiling in order to avoid artificial deflation of the final score. In general, most Kindergarten scores include Blending, Initial Sound, Letter Recognition, Letter Sound, Vocabulary, and Listening Comprehension. Most first and second grade combinations include Real Words, Nonwords, Sight Words, Reading Comprehension, and Vocabulary.

DIBELS. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. They are designed to be short (one minute) fluency measures used to regularly monitor the development of early literacy and early reading skills. DIBELS are comprised of five measures to function as indicators of phonemic awareness, alphabetic principle, accuracy and fluency with connected text.

TPRI. The Texas Primary Reading Indicator is a valid assessment that provides a picture of a student's reading/language arts development. Designed to be used with students in Kindergarten through third grade, specific reading concepts are assessed in the inventory dependent on the grade level and the time of year the inventory is administered. It includes assessments associated with print awareness, phonemic awareness, graphophonemic knowledge, oral reading ability, listening and reading comprehension.

IRI. The Idaho Reading Indicator (IRI) is a benchmark reading test that is administered three times per school year to Idaho public school students in grades K-3. Depending on grade level, the IRI assesses student skills in letter naming fluency, letter sound fluency, phoneme segmentation fluency, nonsense word fluency, as well as accuracy and fluency with connected text.

ITBS. The Iowa Tests of Basic Skills (ITBS) are standardized tests administered nationwide to students in grades K-8. For the purposes of this validation study, only the Reading Comprehension test is administered. The Reading Comprehension test includes using print, context, and picture cues to identify unfamiliar words; completing sentences that tell about a picture by choosing a word for filling in a blank; and answering multiple-choice questions after reading a brief story.

SAT 10. The Stanford Achievement Test Series (SAT 10) is a standardized assessment used to measure academic knowledge of K-12th grade students and has several components, like the ITBS. For the purposes of this validation study, only the Reading Comprehension subtest is examined. At appropriate levels, the Reading subtest measures phonemic awareness, decoding, phonics, vocabulary, and comprehension.

Procedure

The effectiveness of a test in predicting performance on a related task can be measured by assessing performance on two tests at the same point (concurrent validity) or at two different time points (predictive validity). To assess concurrent validity WACS was administered to students nationwide in fall of 2008 and again in spring of 2009. Student

performance was compared to performance on DIBELS, the IRI, the ITBS, and the TPRI.

Test data from the SAT 10 was only collected in the spring.

Data for predictive validity were also collected in the spring, when WACS was administered to the same students who had previously taken the assessment in the fall. Those students also completed a second round of testing for DIBELS, IRI, or TPRI, making it possible to examine predictive validity as well as a second assessment of concurrent validity.

Reliability Determination

Reliability refers to the consistency of a measure; tests that have adequate reliability will yield more or less the same scores across periods of time and across different examiners. Because WACS is administered on the computer there is no error generated from different examiners. However, error may still be introduced into the resulting final scores through lack of attention to the task at hand, faulty headphones, and disinterest. Because of these concerns, it is important to examine test-retest correlations with a small gap between testing dates as well as the internal consistency of the test.

Computerized adaptive tests differ on measures of test-retest reliability since an individual does not see the exact same test at each time point. The resulting correlation coefficient is regarded as a conservative estimate since content sampling adds an extra degree of error beyond individual performance. CATs also differ on measures of internal consistency. Traditional methods, split-half reliability and Cronbach's Alpha, are statistically inaccurate when applied to a CAT tailored to achievement. Instead, the marginal reliability coefficient provides a better measure of internal consistency by combining measurement

error estimated at multiple points on the scale. The resulting coefficient is almost identical to Cronbach's alpha.

Results

Phase I: Item Response Theory

Sufficient amounts of data were collected to ensure that each of the 2,680 items was seen and answered by a minimum of 200 students per item. With that minimum met, item calibration, based on student responses, could be performed using the Rasch model for item response theory in WINSTEPS. The initial analysis revealed 127 misfit students that were subsequently removed from the dataset and the analysis was rerun. The resulting item calibration revealed 262 items that could be considered misfit based on their outfit and infit MNSQ and point-biserial correlation. These items were excluded from the resulting dataset. Subsequent differential item functioning (DIF) analysis exploring discrimination for gender, geography, and school and revealed 21 items that had a gender bias. These items were removed for content review. With all these items removed, the final RMSE model fit is .17.

Phase II: Validation

Concurrent validity

The fall concurrent validity analyses included comparisons of WACS with DIBELS, IRI, ITBS, and TPRI, while spring testing added a comparison with the SAT 10 and some additional ITBS participants. Overall, correlations between relevant WACS assessments and

the associated paper and pencil test are highly significant (Table 2), even with the low number of participants taking the ITBS (see Table 1).

The DIBELS Beginning Kindergarten assessment, consisting of Letter Naming Fluency and Initial Sound Fluency, was found to significantly correlate with WACS Kindergarten Skills in both fall ($r = .74, p < .001$) and spring ($r = .69, p < .001$). Similarly, DIBELS Beginning First Grade assessment, consisting of Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency, significantly correlated with WACS 1st grade Skills at both time points ($r = .72, p < .001$; $r = .68, p < .001$). DIBELS Second grade assessment, consisting of Nonsense Word Fluency and Oral Reading Fluency, also correlated with WACS 2nd Grade Skills at both time points ($r = .61, p < .001$; $r = .59, p < .001$).

Patterns or correlations for the IRI are similar to those seen with DIBELS. The IRI includes only one test for Kindergartners, the Letter Naming Fluency test, and this task correlated significantly with WACS Kindergarten Skills at both time points ($r = .57, p < .001$; $r = .53, p < .001$). First grade IRI tasks, Letter Naming Fluency and Nonword Fluency, were also highly correlated with 1st-grade WACS Skills ($r = .74, p < .001$; $r = .73, p < .001$). Second graders taking the IRI receive only the RCMB, a reading fluency task; this task also significantly correlated with WACS 2nd grade Skills ($r = .58, p < .001$; $r = .59, p < .001$).

The TPRI is designed for students to receive additional assessments based on previous performance. Only Letter Sound, Blending, Letter Name, and Comprehension are given to all Kindergartners. This combination was found to significantly correlate with the WACS Kindergarten Skills in the fall ($r = .52, p < .001$), but spring data has been excluded due to 90% of the kindergartners finishing at the ceiling of the TPRI combination. The TPRI

combination given to all first graders includes Letter Sound, Word Reading, Words Per Minute Rate, and Comprehension Questions. This combination correlated significantly with the WACS 1st Grade Skills in both the fall and the spring ($r = .64, p < .001$; $r = .58, p < .001$). Finally, the TPRI combination given to all second graders, including Word Reading, Words Per Minute Rate, and Comprehension, correlated significantly with WACS 2nd Grade Skills at both time points ($r = .5, p < .001$; $r = .45, p < .001$).

The ITBS includes a number of assessments. For our purposes, comparisons are only made to the Reading subtest for 1st and 2nd graders. Also, the ITBS is unique in these analyses because fall analyses were completed with students from a different school than those used for the spring analyses. Despite the change in students, ITBS Reading significantly correlates with WACS 1st grade Skills ($r = .7, p < .001$; $r = .74, p < .001$) as well as with WACS 2nd grade Skills ($r = .41, p < .001$; $r = .67, p < .001$) to a similar degree in both spring and fall.

The SAT 10, only taken by students during the spring, also includes a number of assessments but comparisons are only made to the Reading subtest for 1st and 2nd graders. This test was found to correlate significantly with the WACS Skills assessment for both 1st and 2nd grade ($r = .755, p < .001$; $r = .65, p < .001$)

Table 2. Concurrent Validity Correlations

		WACS Assessments		
		Kindergarten Skills	1st Grade Skills	2nd Grade Skills
DIBELS Kindergarten	Fall	$r = .72, p < .001$		
	Spring	$r = .68, p < .001$		

DIBELS 1st Grade	Fall		$r = .73, p < .001$
	Spring		$r = .69, p < .001$
DIBELS 2nd Grade	Fall		$r = .61, p < .001$
	Spring		$r = .62, p < .001$
IRI, Kindergarten	Fall	$r = .63, p < .001$	
	Spring	$r = .47, p < .01$	
IRI, 1st Grade	Fall		$r = .72, p < .001$
	Spring		$r = .73, p < .001$
IRI, 2nd Grade	Fall		$r = .62, p < .001$
	Spring		$r = .62, p < .001$
TPRI, Kindergarten	Fall	$r = .47, p < .001$	
	Spring	--	
TPRI, 1st Grade	Fall		$r = .64, p < .001$
	Spring		$r = .60, p < .001$
TPRI, 2nd Grade	Fall		$r = .51, p < .001$
	Spring		$r = .48, p < .001$
ITBS 1st Grade	Fall		$r = .72, p < .001$
	Spring		$r = .74, p < .001$
ITBS 2nd Grade	Fall		$r = .41, p < .001$
	Spring		$r = .67, p < .001$
SAT 10, 1st Grade	Fall		--
	Spring		$r = .76, p < .001$
SAT 10, 2nd Grade	Fall		--
	Spring		$r = .65, p < .001$

Predictive validity

For the most part, schools that participated in WACS testing in the fall also participated in the spring, making it possible to assess the predictability of WACS scores from one test to the next. Table 3 reveals how well fall scores predict spring scores for each of the three grade levels. All three beta weights all similar in magnitude, suggesting that WACS is a stable measurement tool for the intended age group.

Table 3. Predicting spring WACS scores with fall WACS scores

	Standardized Beta	N
Fall_K→ Spring_K	$\beta = .62^{***}$	456
Fall_1 st → Spring_1 st	$\beta = .67^{***}$	561
Fall_2 nd → Spring_2 nd	$\beta = .66^{***}$	521

Additionally, the majority of students completed WACS in the fall and also took a spring standardized test, making it possible to explore predictive validity in addition to concurrent validity. Table 4 presents the results from those regressions, revealing similar numbers as those seen in Table 3, indicating that WACS predictive value is as strong for other tests as it is for itself.

Table 4. Predictive Validity Standardized Beta.

	Fall WACS Assessment	
	Standardized Beta	N
DIBELS, Spring Kindergarten	$\beta = .61^{***}$	133
DIBELS, Spring 1 st Grade	$\beta = .68^{***}$	135
DIBELS, Spring 2 nd Grade	$\beta = .78^{***}$	266
IRI, Spring Kindergarten	$\beta = .47^{***}$	109
IRI, Spring 1 st & 2 nd Grade	$\beta = .70^{***}$	203
TPRI, Spring 1 st Grade	$\beta = .60^{***}$	109
TPRI, Spring 2 nd Grade	$\beta = .60^{***}$	203
ITBS Spring Kindergarten	$\beta = .62^{***}$	70
ITBS Spring 1 st Grade	$\beta = .62^{***}$	82
ITBS Spring 2 nd Grade	$\beta = .62^{***}$	93
SAT 10 Spring 1 st & 2 nd Grade	$\beta = .7^{***}$	223

Construct validity

Construct validity is typically measured with factor analysis or principle component analysis. Because the data was acquired with a computer adaptive test, the large amount of missing data makes a typical factor analysis less useful. Instead, a Rasch Principal Components Analysis was completed as well as a modified factor analysis.

In general, when over 60% of the variance is explained by a single factor, a test is considered to have only one underlying factor. For WACS, 63.5% of the variance is explained by a single factor with only 4.5% of the variance unexplained. An additional modified factor analysis, a principal components analysis with varimax rotation, produces similar results, with the first factor explaining 60.5% of the variance and the next highest factor only explaining 9.7% of the variance. All assessments load strongly on the first factor (all weights above .63) and only Letter Recognition and Letter Sound have weights above .4

on the second factor. In addition, the scree plot indicates a dramatic drop from the first (eigen value of 6.7) to the second factor (eigen value of 1). Finally, another test of the internal coherence of WACS overall is to examine correlations between subtests. Resulting correlations indicate significant relationships among all of the WACS subtests, ranging from $r = .38$ (between Letter Recognition and Listening Comprehension) to $r = .74$ (between Letter Sound and Initial Sound), supporting the conclusion that all subtests can be grouped together as a unidimensional test.

Reliability

Test-retest correlations were completed when students took the second WACS test in April 2009. Students at the Waterford School in grades K, 1st, and 2nd ($n = 188$) completed the WACS test in April, and then again three weeks later in May. For each time point a combination score was created from all the WACS subtests completed by each student. The correlation between the two time points was highly significant ($r = .90, p < .001$).

Internal reliability was measured with the marginal reliability coefficient, examining internal test consistency. Reliability for WACS, is very strong (item reliability of $r = .98$), indicating that the test reliably places student scores in a correct order.

Discussion

Educators can trust data from a new reading assessment only to the degree that it demonstrates an ability to provide accurate and reliable results. The Waterford Assessment

of Core Skills (WACS), a newly available computerized adaptive test (CAT) of pre-literacy and early literacy skills for prekindergarten through 2nd-grade students, has undergone an extensive and rigorous battery of testing. The initial creation of the test required content validation for the included subtests and questions, item difficulty calibration, and construct validation. Because the use of a CAT for such a young age group is a fairly new approach, WACS underwent a thorough battery of validation testing beyond the initial content validation and item calibration, including examinations of concurrent validity, predictive validity, test-retest reliability, and internal consistency reliability.

The creation of WACS required a thorough exploration of the important elements in early reading skills, both as viewed by researchers and by state and national standards of education. Based on these standards, WACS tests phonological awareness, phonics, comprehension, and vocabulary for pre- and early-readers. The 11 subskills included in WACS correlate with all published state standards for kindergarten through second grade. Additionally, all questions were tested in schools across the nation in order to calculate item difficulties, the basis for adaptive nature for WACS.

Following item calibration, WACS was administered to additional groups of students nationwide in order to compare student scores on WACS to scores on other standardized reading tests. These validity analyses indicate that WACS correlates well with the Iowa Test of Basic Skills (ITBS), Stanford Achievement Test (SAT 10), Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Idaho Reading Indicator (IRI), and Texas Primary Reading Indicator (TPRI). Resulting reliability analyses also indicated that the test is internally coherent and has very strong test-retest reliability.

Over time it has become evident that educators are often constrained by the difficulty of assessing pre-reading and early reading skills among young children. This difficulty arises in large part due to the shortcomings of the established testing methods. Paper-based tests that do not require one-on-one administration (like ITBS) are often limited in scope and hampered by floor and ceiling effects. These tests also risk introducing a significant amount of variance among young students, often because children have difficulty completing the test or paying attention. To help address these difficulties, a number of the reading assessments available for the K-2 age group (like DIBELS) require testing to be administered one-on-one with a proctor or teacher. This process, however, is both expensive and slow; consequently, it is often difficult to administer regularly. One-on-one testing also introduces a significant degree of error variance resulting from differing techniques among teachers and test administrators.

Because it takes the important step of moving early reading assessment away from the paper-based model, WACS is capable of addressing many of the challenges faced by existing tests: it allows whole classes to be tested together, reduces the time burden on teachers and administrators, provides immediate and accurate reports, and assesses a large number of content areas in less time. Because the test is computer-based, it is able to combine the adaptability of one-on-one reading tests with a more consistent and impartial administration.

Currently, very few standardized assessments are capable of being used in the pre-K through 2nd grade educational group. Due to numerous challenges with this age group—both administrative and developmental—even tests that can be used with young children are often not used enough. The presentation of WACS has been designed specifically for young

children; its animated characters and quick transitions help students stay engaged without becoming distracted. While testing children at this age will always present numerous challenges, many problems with older tests can be addressed by proper design and innovation. Because identifying struggling students early on increases their chances of becoming successful readers, innovation and research into early literacy testing remains a crucial component of evaluation research. Measures like WACS aim to leverage newer technologies in the interest of deepening and broadening the usefulness of early assessments; the goal is to make reading instruction at this level both easier and more effective.

Future Directions

Despite a well-researched development and thorough validation, WACS could still expand in a few key areas in order to broaden the scope and usefulness of the test. First, WACS was designed to test students who are preschool-aged, but validation analyses could not be completed with the preschool age group at the time of this study due to lack of standardized testing in preschool. Future work should examine what information can be gained from preschoolers, regardless of the ability to determine concurrent validation. Second, WACS is intended to be used with kindergarten through 2nd grade, but students of all grade levels can benefit from a valid reading test with immediate reporting. As such, we hope to extend the reading comprehension and vocabulary assessments through elementary and middle school. Third, WACS would benefit from expanded validation testing in order to include nationwide norms for income, language, ethnicity, and learning disability status. This new information will allow the test to become more useful for schools working with diverse student populations, and will help in monitoring students who are most in danger of

falling behind. By focusing on student assessment in the early grades, we can ensure that difficulties in reading are discovered early and can be addressed appropriately to diminish the achievement gap.

References

- Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew Effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*(6), 1373-99.
- Beck, I. & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Kamil, P. B. Mosenthal, and P. D. Pearson, (Eds.) Handbook of Reading Research, Vol. 2, (789-814). New York, NY: Longman.
- Biemiller, A., & Slonim, N. (2001). Estimating Root Word Vocabulary Growth in Normative and Advantaged Populations: Evidence for a Common Sequence of Vocabulary Acquisition *Journal of Educational Psychology, 93*, 498-520.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119.
- Coyne, M.D., McCoach, D.B. & Kapp, S. (2007). “Vocabulary Intervention for Kindergarten Students: Comparing Extended Instruction to Embedded Instruction and Incidental Exposure”. *Reading Disability Quarterly 30*, 74-88.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934-945.
- Good, R. H., & Kaminski, R. A. (2003). Dynamic indicators of basic early literacy skills, 6th edition. Eugene: Institute for the Development of Educational Achievement, University of Oregon

- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a Simple View of Reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ: L. Erlbaum Associates.
- Guthrie, J. W., & Springer, M. G. (2004). *A Nation at Risk* revisited: Did “wrong” reasoning result in “right” results? At what cost? *Peabody Journal of Education*, 79 (1), 7-35.
- Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., Lewis, J. C., Quails, A. L., Mengeling, M. A., & Shannon, G. P. (2003). *The Iowa Tests, guide to research and development, Forms A and B*. Itasca, IL: Riverside Publishing.
- Marzano, R. (2004). “The Developing Vision of Vocabulary Instruction”. In: *Vocabulary Instruction: Research to Practice*, New York: Guildford, 100-117.
- National Center for Education Statistics (NCES). (2008). *NAEP overview*. Retrieved October 23, 2008 from <http://www.nces.ed.gov/nationsreportcard/about/>
- (2010). *The Nation’s Report Card: Reading 2009: Executive Summary*. Retrieved from <http://nces.ed.gov/nationsreportcard/pubs/main2009/2010458.asp#section2>
- National Reading Panel (2000). *Teaching Children To Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Retrieved from National Reading Panel website: http://www.nichd.nih.gov/publications/nrp/upload/smallbook_pdf.pdf
- Scarborough, H. (2001). *Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice*. In S. B. Neuman and D. Dickinson

- (Eds.) *Handbook of early literacy research* (pp. 97-110). New York, NY: Guilford Press.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stahl, S. A. & Nagy, W. E. (2006). *Teaching Word Meanings*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Stanovich, K. E. (1986). Matthew Effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407.
- Stanovich, K.E. (2000). Progress in understanding reading: Scientific foundations and new frontiers. New York: Guilford Press.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38, 934-947.
- Viteritti, J. P. (2004). From excellence to equity: Observations on politics, history, and policy. *Peabody Journal of Education*, 79(1), 64-86.
- Wagner, R. K., Muse, A. E., & Tannenbaum, K. R. (2007). *Vocabulary Acquisition: Implications for Reading Comprehension*. New York, NY: Guilford Press.
- Walberg, H. J. (2003). *Improving educational productivity*. Publication Series No. 1. Philadelphia: Laboratory for School Success. Retrieved November 5, 2008 from <http://www.temple.edu/lss/pdf/publications/pubs2003-1.pdf>
- Whitehurst, G. J. (2003). September 10, 2003 interview with David Boulton. Retrieved October 16, 2008 from <http://www.childrenofthecode.org/interviews/whitehurst.htm>