# Measuring Reading Comprehension in Groups:

# A Review of the Test of Silent Contextual Reading Fluency

Elizabeth A. Allen Laura Rogers

PRO-ED, Inc.

#### Overview

The *Test of Silent Contextual Reading Fluency* (TOSCRF; Hammill, Wiederholt, & Allen, 2006) is a theoretically sound, research-based method of assessing the silent reading ability of school-aged students in a quick, accurate, and cost-efficient way. The test was normed on 1,898 students ranging in age from 7-0 to 18-11. The test consists of four equivalent forms (A, B, C, and D) and provides raw scores, standard scores, percentiles, and age and grade equivalents. It can be administered individually or in group in approximately 10 minutes and can be used by classroom teachers, special education teachers, reading specialists, school psychologists, speech pathologists, or any other professional who has some training in standardized test administration.

The test measures the speed with which students can recognize the individual words in a series of printed passages that become progressively more difficult in their content, vocabulary, and grammar. The passages are adapted from the passages in the *Gray Oral Reading Tests—Fourth Edition* (GORT-4; Wiederholt & Bryant, 2001) and the *Gray Silent Reading Tests* (GSRT; Wiederholt & Blalock, 2000). The easy passages use preprimer- and first-grade-level words and simple grammar; the difficult passages use adult-level words and complex grammar (embedded phrases, sequenced adjectives, affixes, etc.). Each passage is printed in uppercase without punctuation or spaces between words. Students are given 3 minutes to draw a line between as many words as possible.

# Background

This format belongs to a family of similar word puzzle formats and has a long history in psychology, education, and popular culture. In devising his Structure of Intellect model, J.P. Guilford (1959; Guilford & Hoepfner, 1971) used several word search measures to help establish his Convergent Production of Symbolic Transformations factor. One of these had an untimed format similar to the timed format used in the TOSCRF. Guilford's measures were Sentence Gestalt I and Four Letter Words. In Sentence Gestalt I, the examinee is asked to separate the words in a printed string, as in this example

#### LIKESOMANYOFTHEPURSUITSWHICHHAVEBECOMEPARTAND

In Four Letter Words, the examinee is asked to find and circle four-letter words in lines of letters. For example, he or she is asked to find wind, rock, and howl in the following row of letters.

#### AMGEWINDTEYKQCIROCKWZEHOWLP

Using Guilford's Structure of Intellect model as a guide, Meeker and Meeker (1975) built the *Structure of Intellect Learning Abilities Test*, a battery of tests to evaluate a wide variety of cognitive abilities. A

second edition was published 10 years later (Meeker, Meeker, & Roid, 1985). These authors designed four sets of word search items, the sum of which represents Guilford's Convergent Production of Symbolic Transformations factor. All four sets measure some type of word identification ability under timed conditions ranging from 2 to 5 minutes. Interestingly, three of the four sets used the contextual word-strings-without-spaces format. The format of the first set is very similar to the one used in the TOSCRF. The examiner says the following: "You see a long line of words here (point). All the words have been run together in one long line. You find as many words as you can. Each time you find a word draw a line after it..." (p. 43). After 3 minutes, the examinee is told to stop. A sample line follows:

#### SAYINGASSUMESONEWAYTOKEEPPEOPOLEFROMJUMPING

Meeker et al. (1985) claimed that the item sets designed to measure the Convergent Production of Symbolic Transformations factor assess "speed of word recognition" (p. 76). Support for their claim is found in the validity section of their test manual.

In designing the TOSCRF, we selected sentences from the GORT-4 and GSRT because the stories used in them are nationally normed on a large number of school-aged students throughout the United States. The syntactic structure of the sentences is carefully controlled, and age- and grade-appropriate vocabulary is used. The sentences were modified, where necessary, to ensure that no passage ended with a line containing only one word. As a result, the test requires not only sight word recognition but also basic reading comprehension skills. Unlike in the *Test of Silent Word Reading Fluency* (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004), no effort was made to eliminate words that had little words inside them or to eliminate two words forming a third word when run together.

# Scoring Key—Form A

Forr	n A	Line Word Count	Total Word Count	For	rm A	Count	Count
1. [ 2. [	udcan Iokþown		2 4	13.	THEYOUNGMANWASDELIGHTEDTOOBTAINHISDEPUTY Sheriffsbadgeashisfirstassignmenthehadbeen appointedtoescortaprisonentoftheauthoritiesat thecountyseatandthesiteoftheimpendingtrial	9 9 9 10	232 241 250 260
3.	JUMPUPRUNFAST	4	8	14.	THEREARESUNDRYDEFINITIONSOFUAZZALLOFTHEMVAGUE	10	270
4.	seefatherweßing	4	12		ТНЕ IRVAGUENESSISEEMSIMPERATIVEHOWEVERIFTHEY АRETOACCOMMODATETHECUSTOMOFJAZZTOAPPROPRIATE	7 9	277 286
5.	a Yellow Bird with Blue wings Sat On Mothers Pretty Hat it was The Missing Bird	10 6	22 28		EVERYTHINGINSIGHT	3	289
6.	тне[F0X]GAVE[THE]STORҢАВІТ ОF[THE]SOUP IN A SHALLOW DISHBUT[THE]STORҢCOULD[NOT[EAT]FROM[THE]DISH	13 10	41 51	15.	ASTHEEMPERORSAGENTHEEXPLOREDTHEPAHOONNEHS OFTHEKINGDOMHEKEPTANOTEBOOKFILLEDWITH FASCINATINGACCOUNTSOFHISEXPERIENCESAS ECONOMICANDPOLITICALREPORTSFORTHEEMPERORSUSE	9 9 8 8	298 307 315 323
7.	THEBOYSWENTINTOASTOREANDLOOKEDATBOOKSABOUT ANIMALSONEBOYSAWABOOKABOUTHORSES	П 8	62 70	16.	THEADROITMAN WASAFFABLEIN DE BATEAN DOFTHAT	10	333 342
8.	тне sтопк served тне soup in тwo таll sottles with тнім narrow necks she stuck неr long sill into тне воттle and ate everv Drop	10 10 5	80 90 95		BROUGHTINGOPINIONSWITHIIMBUTINATHERIAFIDELITYTO ENLIGHTENMENTANDINSTRUCTIONHEHADISOISUBTLE AWAYOFINTERROGATINGTHEIDTHERIMINISTERSANDUNDER THEIGUISOISSINIJATINGHISODJECTIONS	10 7 9 7	352 359 368 375
9.	SOMEDAYSJAFTERWARDTHEISTORK/INVITEDTHEIFOXTO DINNERISHEIHADIMADEISOMEDELICIOUSISOUPIANDTHEIFOX WASIVERYHUNGRY	9 10 3	104 114 117		ТОНРАОРОЗЕДЕЛІСТІТНАТІНЕЛІХТІLLЕДНІ SPERSPECTIVE INTHOSEFROMWHOMHEPRETENDEDTOLEARN	9	384 392
D.	Α S¦T Η ΕΊC LOC K¦T ICKE D¦OFFONE[MINUTE]AFTER ANOTHER ΕΑCΗ OF THE GIRLS TRIED TO PIN A RIBBON ON THE LIONS ΤΑΙL THE V PUT RIBBONS ON THE LIONS PAWS HEAD AND BODY WHICH GAVE HIM A FRIENDLY LOOK	9 12 10 7	126 138 148 155	17.	CERTAINLYSHEHAQLATELYBEENCOGNIZANTJOHTHEJSUAVE YOUNGMANSJUNNECESSARYCHIVALRYBUTJSHEHAD CONSTRUEDJTASJANJABERRATIONJOHJUDGMENTJOHTASTE ASJONEJPROOFAMONGJOTHERSTHATJHEHADNOTJALWAYS MOVEDJHTHEBESTCIRCLESTHATJWITHALLHISJASSIDUOUS CONCERNED AND EOTHERSCOULTBEMENTSJOETBUE	9 7 9 10 10 7	401 408 417 427 437 444
1.	THISMANWASONEOFTHEFIRSTEUROPEANTRAVELERSTO VISITTHEDISTANTREGIONOFCHINAHELEARNEDSEVERAL LANGUAGESSPOKENTHEREANDSERVEDITSGREATRULERS FORMANYYEARS	10 9 8 3	165 174 182 185		GENTILITYWEREWANTINGBUTUNTILINOWSHEHADNEVER SURMISEDITTOMEANANYTINGHISMANNERMUSTHAVE BEENEQUIVOCALORSHECOULDNOTHAVEBEENSOBEGUILED	9 9 10	453 462 472
2.	а\$ 4 Youth не Траvе Led From One Farm To Another Picking Crop\$ A\$ TheY Ripened Since Hi\$ Family Had Nopermanent Hom = Had Attended 4 Multitude OF DiFFerent Schools BY The Time He Reached The seventh Grade	10 9 8 8 3	195 204 212 220 223				

#### Scoring

Scoring is quick and easy. Before scoring, make sure that every student has attempted to identify the words on the Practice Form and that no rows were skipped on the test. If the student successfully completed the example words and demonstrated understanding of the task on the Practice Form, begin scoring with the last row that the student attempted. Score backwards until you reach the point at which the student has correctly identified all of the words in one passage or until all words have been scored. A word is correctly identified when a line is clearly drawn separating the word from another word. Disregard any lines drawn before the first word or after the last word on a row. The student receives credit for all the words from Passage 1 to the completely correct passage and is awarded 1 point for each word correctly identified after that correct passage.

Characteristics	Percentage of Sample	Percentage of Population	Characteristics	Percentage of Sample	Percentage of Population
Geographic Area <sup>a</sup> Northeast	18	18	Exceptionality Status <sup>a</sup> Learning disorder	8	5
Midwest South	26 38	23 35	Hearing impairment Articulation disorder	<1	2
Gender	19	25	Blind/partially sighted	<1	<1
Male Female	52 48	51 49	Physical impairment	~	1
Ethnicity <sup>b</sup> White	81	81	Age	-	
Black/African American Asian/Pacific Islander	12 4	13 4	7 (n = 145) 8 (n = 165) 9 (n = 151)	8 9 8	NA NA
Eskimo/Aleut	<1	1	10 (n = 125) 11 (n = 185)	7 10	NA
Hispanic <sup>b</sup> Yes	11	13	12 (n = 258) 13 (n = 183)	14 10	NA NA
No Family Income (in dollars) <sup>c</sup>	89	87	14 (n = 192) 15 (n = 143)	10 8	NA NA
Under 15,000 15,000–24,999	7	10 12	16 (n = 119) 17 (n = 124)	6 7	NA NA
25,000-34,999 35,000-49,999	12 18	12 16	18 (n = 108)	0	INA
50,000-74,999 75,000 and over	23 34	22 30			
Educational Attainment of Parents <sup>6</sup>					
Less than bachelor's degree Bachelor's degree	70	74			
Master's, professional, doctoral degrees	19	9			

#### **Demographic Characteristics**

The TOSCRF normative sample was stratified with regard to geographic area, gender, race and ethnicity, family income, educational attainment of parents, exceptionality status, and age.

	Types of Reliability Coefficients							
TOSCRF Coefficients	Alternate Form (Immediate)	Test–Retest	Alternate Form (Delayed)ª	Scorer				
r		86		99				
r <sub>BB</sub>		88		99				
r <sub>cc</sub>	<u> </u>	85	<u> (1 - 1 - 1</u> - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	99				
r <sub>DD</sub>		88		99				
r <sub>AB</sub>	84		85	_				
r <sub>AC</sub>	82	<del>8</del>	85					
r <sub>AD</sub>	85	<u></u>	83					
r <sub>BC</sub>	86		85					
r <sub>BD</sub>	85		87	3				
r <sub>CD</sub>	86		84					
Sources of Test Error <sup>b</sup>	Content Sampling	Time Sampling	Time Sampling, Content Sampling	Interscorer Differences				

#### Reliability

The TOSCRF's overall reliability was compared to Anastasi and Urbina's (1997) types of reliability coefficients and sources of test error. The TOSCRF evidences a high degree of reliability across all four types of reliability. The magnitude of the reported coefficients strongly suggests that the TOSCRF possesses little test error and that test users can have confidence in its results.

Corrected (and Uncorrected) Coefficients Representing the Relationship Between TOSCRF Forms and Types of Reading								
Criterion Reading Tests	Types of		TOSCI	Average				
	Reading	А	В	C	D	Forms <sup>a</sup>	Magnitude <sup>b</sup>	
GORT-4 Total Score	Gen	.75 (.52)	.67 (.47)	.61 (.45)	.65 (.47)	.67	LARGE	
Stanford 9 Total Reading	Gen	.76 (.54)	.65 (.45)	.64 (.47)	.66 (.47)	.68	LARGE	
TOSWRF Total Score	WI, F	.81 (.71)	.84 (.75)	.89 (.79)	.85 (.77)	.85	VERY LARGE	
Reading Efficiency	WI, F	.64 (.80)	.75 (.86)	.73 (.85)	.68 (.84)	.70	VERY LARGE	
WJ–III Broad Reading	Gen	.63 (.55)	.73 (.65)	.74 (.66)	.63 (.59)	.69	LARGE	
Global Reading	Mixed	.69 (.69)	.71 (.72)	.72 (.73)	.69 (.71)	.70	VERY LARGE	

Note. Gen = general reading; WI = word identification; F = fluency; GORT-4 = Gray Oral Reading Tests-Fourth Edition (Wiederholt & Bryant, 2001); Stanford 9 = Stanford Achievement Test Series-Ninth Edition (Harcourt Brace Educational Measurement, 1996); TOSWRF = Test of Silent Word Reading Fluency (Mather et al., 2004); TOWRE = Test of Word Reading Efficiency (Torgesen et al., 1999); WJ-III = Wood-cock-Johnson III (Woodcock et al., 2001).

All coefficients are significant at the p < .0001 level.

<sup>a</sup>Average of the corrected coefficients.

<sup>b</sup>Magnitude of the averaged corrected coefficients; based on Hopkins's (2002) criteria for interpreting correlation coefficients.

#### **Criterion-Prediction Validity**

Data were collected on four samples of students from Arizona, Florida, New York, and Texas. Sample sizes ranged from 42 students to 243 students. In all, data were collected on five criterion reading measures: GORT-4, *Stanford Achievement Test Series—Ninth Edition* (Stanford 9; Harcourt Brace Educational Measurement, 1996), *Test of Silent Word Reading Fluency* (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004), *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner, & Rashotte, 1999), and *Woodcock-Johnson III* (WJ III; Woodcock, McGrew, & Mather, 2001). Once the data were collected, the TOSCRF were correlated with those of the criterion tests. The coefficients were corrected to account for any effects of range that might artificially repress or inflate the correlation coefficients. Similarly, the coefficients were attenuated to control for any lack of reliability in the criterion reading measures. Finally, student performance on each individual criterion measure was recoded into a new metavariable called Global Reading. This variable includes all TOSWRF, TOWRE Total Reading Efficiency, GORT-4 Total Score, Stanford 9 Total Reading, and W-J III Broad Reading scores.

Hopkins (2002) suggested a Likert-scale approach in determining the magnitude of correlation coefficients. He suggested that coefficients between 0.00 and 0.09 are very small, coefficients between 0.10 and 0.29 are small, coefficients between 0.30 and 0.49 are moderate, coefficients between 0.50 and 0.69 are large, coefficients between 0.70 and 0.99 are very large. The results indicate that the TOSCRF possesses a large to very large relationship with the criterion reading measures.

				Standard S for TC	core Means (a SCRF and Crit	and Standard terion Reading	Deviations) g Tests			
Criterion Reading Tests/ TOSCRF Comparisons	Sample(s)	N	Type of Sample	Criterion Form A	Criterion Form B	Criterion Form C	Criterion Form D	Criterion Average	Descriptive Rating	scriptive Rating t
GORT-4 Total Score TOSCRF Total Reading	2	119	LD	81 (14) 87 (10)	81 (14) 86 (10)	81 (14) 87 (11)	81 (14) 86 (11)	81 (14) 87 (11)	Below Avg. Below Avg.	5.23*
Stanford 9 Total Reading TOSCRF Total Reading	2	103	LD	89 (14) 87 (10)	89 (14) 87 (10)	89 (14) 88 (11)	89 (14) 87 (11)	89 (14) 87 (11)	Below Avg. Below Avg.	-1.36 (ns
TOSWRF Silent Word Reading Fluency TOSCRF Total Reading	3	243	Regular Ed Deaf/HH, ADHD, LD	99 (15) 97 (13)	99 (15) 98 (13)	99 (15) 98 (13)	99 (15) 98 (14)	99 (15) 98 (13)	Average Average	—1.31 (ns)
TOWRE Total Word Reading Efficiency TOSCRF Total Reading	4	42	Regular Ed, LD	99 (21) 98 (18)	99 (21) 97 (17)	99 (21) 98 (17)	99 (21) 98 (19)	99 (21) 98 (18)	Average Average	-0.05 (ns
WJ–III Broad Reading TOSCRF Total Reading	1	134	ADHD, LD	105 (16) 98 (12)	105 (16) 99 (12)	105 (16) 99 (12)	105 (16) 99 (13)	105 (16) 99 (12)	Average Average	-4.89*
Global Reading TOSCRF Total Reading	1-4	641	Mixed	95 (17) 94 (13)	95 (17) 94 (14)	95 (17) 95 (13)	95 (17) 95 (14)	95 (17) 95 (14)	Average Average	-1.33 (ns)

Note. LD = learning disabilities; ADHD = attention-deficit/hyperactivity disorder; Deaf/HH = deaf or hard of hearing; GORT-4 = Gray Oral Reading Tests-Fourth Edition (Wiederholt & Bryant, 2001); TOSWRF = Test of Silent Word Reading Fluency[Mather et al., 2004); Stanford 9 = Stanford Achievement Test Series-Ninth Edition (Harcourt Brace Educational Measurement, 1996); TOWRE = Test of Word Reading Efficiency (Torgesen et al., 1999); WJ–III = Woodcock-Johnson III (Woodcock et al., 2001). \*p < .05.

It is not enough to determine a strong relationship between the TOSCRF and other criterion reading measures, one must also examine if these tests provide similar scores. To do this, the differences between the standard score means of the TOSCRF and each individual criterion reading tests were also evaluated. This study indicated that in only two comparisons were the means significantly different (p < .05) while all of the means of the criterion tests were within one *SEM* (i.e., 6 points) of the corresponding TOSCRF means, and the descriptive ratings of the TOSCRF and criterion test means were identical.

#### Predicting Reading Efficiency from the TOSCRF Form A Score

	Stan	ford 9 Total Reading	Score	
TOSCRF Composite	Poor	Good	Total	
Poor	44 <sup>a</sup>	19 <sup>b</sup>	63	
Good	13°	27 <sup>d</sup>	40	
Total	57	46	103	

*Note.* N = 103. Percent agreement [(44 + 27)/(44 + 27 + 13 + 19)] = .69; Positive Predictive Value [44/(44 + 19)] = .70; Sensitivity [44/(44 + 13)] = .77; Specificity [27/(27 + 19)] = .59.

<sup>a</sup>True positives.

<sup>b</sup>False positives (overreferrals).

<sup>c</sup>False negatives (underreferrals).

<sup>d</sup>True negatives.

Level I-A: Sensitivity and specificity  $\geq$  .70

Level I-B: Sensitivity and positive predictive value  $\geq$  .70

Level II: Sensitivity, specificity, and positive predictive value  $\geq$  .70

Level III: Sensitivity, specificity, and positive predictive value  $\geq$  .75

To examine the effectiveness of the TOSCRF in predicting reading ability, a series of positive predictive outcome analyses were conducted using measures of word identification (with or without speed), fluency, comprehension, and general reading as the criterion. These analyses, which are described by Bernstein and Weiler (2000), Gredler (1997), and Mausner and Kramer (1985), involve calculating three statistics: sensitivity index, specificity index, and positive predictive value. In the current context, sensitivity index reflects the ability of a test to correctly identify individuals who have reading problems. The specificity index reflects the ability of a test to correctly identify individuals who do not have reading problems. The positive predictive value reflects the proportion of individuals who truly have a reading problem among all those whom the screening measure identifies as having a reading problem. We used a cutoff of the 25<sup>th</sup> percentile to establish dichotomous groups of Below Average (i.e., "Poor") and Average and Above Average (i.e., "Good").

Authorities vary regarding how large a test's sensitivity index, specificity index, and positive predictive value should be to justify its use for screening purposes (see Carran & Scott, 1992; Gredler, 2000; Jansky, 1978; Kingslake, 1983; Wood, Flowers, Meyer, & Hill, 2002). In an effort to clarify the criteria for acceptability, we have developed a system for designating the level of acceptability in these types of analyses. These are described in the table above. All researchers agree that at least two indexes be .70 or greater and that the sensitivity index be one of them, but the ambiguity lies in which of the other

indexes is most important. Although Level III should be strived for, Level 1 is probably acceptable for screening purposes.

			Predictiv			
Criterion Reading Tests	N	Level of Acceptability	Sensitivity Index	Specificity Index	PPV	Percent Agreement
GORT-4 Total Score	119	I-B	.71	.59	.84	.68
Stanford 9 Total Reading	103	I-B	.77	.59	.70	.69
TOSWRF Total Score	275	II	.78	.87	.73	.84
Efficiency	42	111	.81	.85	.76	.83
WJ III Broad Reading	130	I-A	.76	.83	.46	.82
Global Reading	641	II	.76	.80	.72	.78

Note. PPV = positive predictive value; GORT-4 = Gray Oral Reading Tests-Fourth Edition (Wiederholt & Bryant, 2001): Stanford 9 = Stanford Achievement Test Series-Ninth Edition (Harcourt Brace Educational Measurement, 1996); TOSWRF = Test of Silent Word Reading Fluency (Mather et al., 2004); TOWRE = Test of Word Reading Efficiency (Torgesen et al., 1999); WJ III = Woodcock-Johnson III (Woodcock et al., 2001); Level of Acceptability: I-A = sensitivity and specificity  $\geq$ .70, I-B = sensitivity and positive predictive value  $\geq$ .70, II = sensitivity, specificity, and positive predictive value  $\geq$ .70, III = sensitivity, specificity, and positive predictive value  $\geq$ .70, III = sensitivity, specificity, and positive predictive value  $\geq$ .70, III = sensitivity, specificity, and positive predictive value  $\geq$ .70, III = sensitivity, specificity, and positive predictive value  $\geq$ .70, III = sensitivity, specificity, and positive predictive value  $\geq$ .75.

The level of acceptability and indexes of sensitivity, indexes of specificity, and positive predictive values for the TOSCRF composite when predicting scores on the criterion reading measures are presented in the table above. Additionally, percent of agreement rates are presented in the column to the far right. The contents of the table show that the TOSCRF met some level of acceptability in all six comparisons studied. Most importantly, the TOSCRF met the Level II criteria for acceptability when predicting the Global Reading metavariable. The TOSCRF only missed Level III criteria by 3% on the PPV.

The results of these studies suggest that the TOSCRF is a reasonably accurate predictor of reading. Therefore, the TOSCRF can be used confidently as a measure for detecting (i.e., screening or identifying) students experiencing reading difficulty. The degree of accuracy in predicting poor readers is impressive, especially considering that the TOSCRF can be given to entire classes of students in only 10 minutes.

				TOSCRF Valu	les		
Measures of School Achievement	Sample <sup>a</sup>	Form A	Form B	Form C	Form D	Avg. of the Corrected Coefficients	Magnitude <sup>b</sup>
Stanford 9: Total Math $(N = 103)$	2	.64 (.36)	.64 (.39)	.69 (.45)	.67 (.41)	.66	LARGE
Stanford 9:Vocabulary $(N = 103)$	2	.66 (.40)	.55 (.34)	.51 (.33)	.53 (.33)	.56	LARGE
WJ III: Calculation (N = 140)	1	.45 (.29)	.57 (.38)	.59 (.40)	.53 (.37)	.54	LARGE
WJ III: Spelling $(N = 138)$	1	.69 (.58)	.83 (.70)	.76 (.65)	.75 (.67)	.76	VERY LARG
WJ III: Academic Skills (N = 62)	1	.66 (.52)	.69 (.54)	.71 (.57)	.67 (.55)	.68	LARGE

Note. Stanford 9 = Stanford Achievement Test Series-Ninth Edition (Harcourt Brace Educational Measurement, 1996); WJ III = Woodcock-Johnson III (Woodcock et al., 2001).

All coefficients are significant at the p < .001 level.

<sup>a</sup>Sample characteristics are described in Table 6.1.

<sup>b</sup>Magnitude of the averaged corrected coefficients; based on Hopkins's (2002) criteria for interpreting correlation coefficients.

# **Construct-Identification Validity**

Because reading problems are common for students who fail in scholastic subjects, the authors of most tests of reading report correlations with academic achievement to be large (e.g., Hammill, Mather, & Roberts, 2001; Newcomer, 2001; Wiederholt & Bryant, 2001). Because of this, the TOSCRF should also correlate significantly with measures of school achievement and to a large degree. To investigate this relationship, the TOSCRF was correlated to the Calculation test, Spelling test, and Academic Skills cluster of the WJ III and to the Vocabulary and Total Math scores of the Stanford 9. The coefficients were corrected for restricted range and for any lack of reliability in the achievement measure. The average of the corrected coefficients are classified as either large or very large by Hopkins's (2002) criteria and provide strong evidence of the test's construct-identification criteria.

#### Corrected (and Uncorrected) Correlations Between TOSCRF and WISC

Measure of Intelligence	Form A	Form B	Form C	Form D	Average of the Corrected Coefficients	Magnitude <sup>a</sup>
WISC: Verbal Scale WISC: Performance Scale WISC: Full Scale	.45 (.37) .39 (.27) .53 (.39)	.44 (.37) .41 (.30) .54 (.40)	.36 (.30) .37 (.27) .47 (.34)	.38 (.33) .35 (.26) .46 (.35)	.41 .38 .50	MODERATE MODERATE LARGE

Note. WISC = Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991) and Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003).

All coefficients are significant at the p < .001 level.

<sup>a</sup>Magnitude of the averaged corrected coefficients; based on Hopkins's (2002) criteria for interpreting correlation coefficients.

Most individuals would assume that reading and intelligence (or aptitude) are related to some degree. To determine the relationship between the TOSCRF and intelligence, the test was correlated to the *Wechsler Intelligence Scale for Children—Third Edition* (Wechsler, 1991) and the *Wechsler Intelligence Scale for Children—Fourth Edition* (Wechsler, 2003). The data are based on 215 students from Texas. The resulting correlation coefficients are reported above. Using Hopkins's (2002) classification for interpreting correlation coefficients, one may determine that all of these coefficients are indicative of a moderate to large relationship and that this study supports the construct-identification validity of the TOSCRF.

			TOSCRF	Values	
Subgroup	Average of All Forms	Form A	Form B	Form C	Form D
Normative sample ( $N = 1,898$ )	100 (15)	100 (15)	100 (15)	100 (15)	100 (15
Female ( $n = 906$ )	103 (15)	103 (15)	103 (15)	103 (15)	103 (15
Male ( $n = 991$ )	98 (15)	98 (14)	98 (15)	98 (14)	98 (15
Asian/Pacific Islander ( $n = 80$ )	110 (17)	109 (18)	110 (17)	110 (16)	109 (17
European American ( $n = 1,530$ )	102 (14)	101 (14)	101 (14)	102 (14)	102 (14
Hispanic ( $n = 214$ )	96 (14)	95 (14)	96 (15)	97 (14)	96 (14
American Indian/Eskimo/ Aleut ( $n = 14$ )	94 (13)	93 (14)	94 (13)	95 (13)	94 (13
African American ( $n = 231$ )	94 (14)	94 (13)	94 (14)	93 (13)	94 (15
Gifted and talented ( $n = 70$ )	113 (13)	113 (14)	112 (13)	113 (13)	112 (12
Attention-deficit/hyperactivity disorder ( $n = 134$ )	100 (13)	99 (13)	100 (14)	100 (13)	100 (13
Learning disabilities ( $n = 154$ )	93 (13)	93 (13)	93 (13)	93 (12)	94 (14
Deaf/hard of hearing ( $n = 49$ )	93 (13)	90 (11)	93 (14)	94 (13)	93 (15
Poor readers ( $n = 316$ )	86 (11)	86 (10)	86 (11)	86 (11)	86 (12

Another way of establishing a test's validity is to study the performance of different groups of people on the test. Each group's results should make sense, given what is known about the relationship of the test's content to the group. Thus, in the case of the TOSCRF, a test of reading fluency, one would expect that individuals with disabilities that negatively affect reading ability would do less well than people without such disabilities. One would certainly anticipate that students who are diagnosed as having reading disorders would do poorly on the test compared with other students.

The mean standard scores for the total sample used to norm the TOSCRF as well as for demographic subgroups are listed in the table above. The students in the samples were tested by the same professionals who participated in the norming of the TOSCRF. In all, we have data on three "mainstream" subgroups (males, females, and European Americans), four "minority" subgroups (African American, Hispanic, Asian/Pacific Islander, or American Indian/Eskimo/Aleut), and five "exceptionality" subgroups (students identified as being gifted and talented, being deaf or hard of hearing, having attention-deficit/hyperactivity disorder, having learning diabilities, or being poor readers). The poor reader group comprised students who scored below average (i.e., < 90) on any of one of the criterion measures studied previously.

The mean standard scores are very supportive of the construct-identification validity of the TOSCRF. All of the TOSCRF scores made by the seven mainstream and minority subgroups are within expectations. The scores made by the exceptionality subgroups are about what one would predict. Based on our knowledge of reading problems among disability subgroups, we would predict that the "poor reader" subgroup would have the lowest mean standard score and that the rank order from lowest to highest for the other subgroups would be as follows: deaf or hard of hearing, learning disabilities, attention-deficit/hyperactivity disorder, and gifted and talented. The anticipated order is the observed order.

# Summary

We have presented a considerable amount of evidence to show that the TOSCRF is both a reliable and valid measure of general reading and reading fluency and can be used with confidence. We recognize, however, that our work is only the initial stage in studying the test's validity. We will continue to provide additional studies in the years to come and encourage other researchers to study the test using different samples, statistical procedures, and related measures.

# References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

- Bernstein, J. H., & Weiler, M. D. (2000). "Pediatric neuropsychological assessment" examined. In G.
  Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (3<sup>rd</sup> ed.) (pp. 263-300).
  Kidlington, Oxford: Pergamon.
- Carran, D. T., & Scott, K. G. (1992). Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in Early Childhood Special Education*, 12(2), 196-211.
- Gredler, G. R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools,* 34(2), 99-106.
- Gredler, G. R. (2000). Early childhood screening for developmental and educational problems. In B. A. Bracken (Ed.) *The psychoeducational assessment of preschool children* (pp. 399-411). Boston: Allyn & Bacon.
- Guilford, J. P. (1959). Three faces of intellect. American Psychologist, 14, 469-470.
- Guilford, J. P., & Hoepfner, R. (1971). The analysis of intelligence. New York: McGraw-Hill.
- Hammill, D., Mather, N., & Roberts, R. (2001). *Illinois Test of Psycholinguistic Abilities—Third Edition*. Austin, TX: PRO-ED.
- Harcourt Educational Measurement (1996). *Stanford Achievement Test Series, Ninth Edition*. San Antonio: Author.
- Hopkins, W. G. (2002, June 23). A scale of magnitudes for the effect statistics. In *A new view of statistics*. Retrieved from http://www.sportsci.org/resource/stats/effectmag.html
- Jansky, J. J. (1978). A critical review of some developmental and predictor precursors of reading disabilities. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 331-347). New York: Oxford University Press.
- Kingslake, B. J. (1983). The predictive (in)accuracy of on-entry to school screening procedures when used to anticipate learning difficulties. *British Journal of Special Education, 10,* 24-26.
- Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of Silent Word Reading Fluency*. Austin , TX: PRO-ED.

- Mausner, J. S., & Kramer, S. K. (1985). Screening in the detection of disease. In *Mausner & Bahn* epidemiology—An introductory text (2<sup>nd</sup> ed.; pp. 214-238). Philadelphia: W. B. Saunders Company.
- Meeker, M., & Meeker, R. (1975). *Structure of Intellect Learning Abilities Test*. El Segundo, CA: SOI Institute.
- Meeker, M., Meeker, R., & Roid, G. H. (1985). *Structure of Intellect Learning Abilities Test—1985 Edition*. Los Angeles: Western Psychological Services.
- Newcomer, P. L. (2001). *Diagnostic Achievement Battery—Third Edition*. Austin, TX: PRO-ED.
- Torgesen, J. K., Wagner, R., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio: The Psychological Corporation.
- Wiederholt, J. L., & Bryant, B. R. (2001). Gray Oral Reading Tests—Fourth Edition. Austin, TX: PRO-ED.
- Wiederholt, J. L., & Blalock, G. (2000). Gray Silent Reading Tests. Austin, TX: PRO-ED.
- Wood, F., Flowers, L., Meyer, M., & Hill, D. (2002, November). How to evaluate and compare screening tests: Principles of science and good sense. Paper presented at the meeting of the International Dyslexia Association, Atlanta, GA.
- Woodcock, R. W. (1998). Woodcock Reading Mastery Test—Revised, Normative Update. Circle Pines, Minnesota: American Guidance Services.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.