Using the TOSWRF-2 to Identify Students with Specific Learning Disabilities

Elizabeth A. Allen, Melissa E. Morey, and Donald D. Hammill

The *Test of Silent Word Reading Fluency, Second Edition* (TOSWRF-2; Mather, Hammill, Allen & Roberts, 2014) is a theoretically sound, research-based method of assessing the silent reading ability of school-age students in a quick, accurate, and cost-efficient way. The TOSWRF-2 has six components: this manual, the supplemental Administration and Scoring Instructions, and four equivalent forms (A, B, C, and D). The test provides raw scores, standard scores, percentiles, and age/grade equivalents. It was normed on a representative sample of 2,429 students ranging from 6 years 3 months to 24 years 11 months in age. These students resided in 35 states. The TOSWRF-2 can be used by classroom teachers, special education teachers, reading specialists, school psychologists, speech pathologists, or any other persons who have some training in standardized test administration.

Students are presented with 220 unrelated printed words, ordered from pre-primer to adult-level, with no spaces between the words. For example:

<p style="text-align:center">DIMHOWFIGBLUE</p>

Students are given 3 minutes to draw a line between the boundaries of as many words as possible (e.g., dim/how/fig/blue/). While most uses require only the administration of a single form, you may administer two of the alternate forms concurrently for increased reliability. A single form can be administered in less than 10 minutes and two forms can be administered in less than 15 minutes, including the time necessary for explaining the directions and completing the practice items.

The TOSWRF-2 is primarily a measure of word identification, word comprehension, and reading speed (also known as silent reading fluency). Because its scores reflect competence in so many aspects of reading, its results can also be taken as a valid estimate of general reading ability and can be used to identify poor readers with confidence.

The TOSWRF-2 has strong alternate form (immediate and delayed) and test-retest reliability (correlation coefficients ranging from .84 to .91). Correlations with criterion tests across 21 studies averaged .76 (Very Large). Expanded studies of diagnostic accuracy as it relates to its sensitivity (median = .75), specificity (median = .75), classification accuracy (median = .76), and receiver operating characteristic/area under the curve (ROC/AUC; median = .85).

The present study examined the performance of a sample of students between the ages of 6 and 24 years who were diagnosed by local school personnel as having learning disabilities (82 unspecified; 124 reading; 22 reading and writing; 25 reading, writing, math) on the TOSWRF-2. Sample demographics are presented in Table 1. This study also examined the TOSWRF-2's ability to differentiate students diagnosed with LD from all other students in the normative sample.

Differences Among Groups

The first part of this study examined the mean standard score performance of the LD groups on the TOSWRF-2. The mean index scores for all of the clinical samples in the TOSWRF-2 normative study are listed in Table 2. The reported means for these subgroups are as expected; they are consistent with our experience and with

similar data reported in manuals of other tests of reading (e.g*., Gray Oral Reading Tests—Fifth Edition* [Wiederholt & Bryant, 2012]; *Test of Word Reading Efficiency—Second Edition* [Torgesen, Wagner, & Rashotte, 2012]; *Test of Silent Contextual Reading Fluency—Second Edition* [Hammill, Wiederholt, & Allen, 2014]). Based on our knowledge of reading problems among subgroups with disabilities and non-native English speakers, we would predict that students with a learning disability diagnosis affecting reading would have the lowest mean standard scores and that the remaining subgroups would range from low average to below average. The results were as expected, providing strong evidence of the TOSWRF-2's validity in identifying students with learning disabilities.

Diagnostic Accuracy

Diagnostic accuracy refers to the precision with which a test differentiates individuals with a disorder from those without a disorder. Researchers such as Betz, Eickhoff, and Sullivan (2013) and Dollaghan (2004) have suggested that this is "the most important criterion for evaluating a diagnostic measure" (Dollaghan, 2004, p.395). Methods for establishing diagnostic accuracy involve the computation of a test's sensitivity and specificity indexes. In the current context, the sensitivity index reflects the ability of a test to correctly identify students who have a reading impairment. The specificity index refers to the ability of a test to correctly identify examinees who do not have a reading impairment.

The results for sensitivity and specificity are reported as proportions (i.e., percentages). The size of the proportions necessary to be considered acceptable varies depending on the purpose of the analysis (e.g., when screening for cancer, a relatively high number of false positives is tolerable in order to ensure that the number of true positives identified is high). Educational researchers vary in their opinions about how large a test's sensitivity and specificity indexes should be. Wood, Flowers, Meyer, and Hill (2002) recommend that the sensitivity and specificity indexes should be at least .70. Jansky (1978), Gredler (2000), and Kingslake (1983) prefer .75 for both indexes. Carran and Scott (1992) and Plante and Vance (1994) recommend a more rigorous standard of .80 or higher. Jenkins and others (Jenkins, 2003; Jenkins, Hudson, & Johnson, 2007; Johnson, Jenkins, Petscher, & Catts, 2009) recommend that sensitivities be high—perhaps as high as .90—and that specificity levels be relatively high as well.

The receiver operating characteristic/area under the curve (ROC/AUC) "is a measure of the overall performance of a diagnostic test and is interpreted as the average value of sensitivity for all possible values of specificity" (Park, Goo, & Jo, 2004, p. 13). ROC/AUC values range from 0 (representing zero predictive ability) to 1 (representing perfect predictive ability). Zhou, Obuchowski, and Obuchowski (2002) recommend that screening measures designed to distinguish between students with satisfactory and unsatisfactory reading ability should have ROC/AUC values that are close to 1. Compton, Fuchs, L. Fuchs, and Bryant (2006) suggest that ROC/AUCs of .90 and above are considered excellent; .80—.89 are good; .70—.79 are fair; and .70 or below are poor.

In the second part of this study, we conducted a series of diagnostic accuracy analyses to examine the TOSWRF-2's ability to differentiate LD students from all other students in the normative sample. We examined the TOSWRF-2's diagnostic ability at three different cut points—standard scores of 90, 92, and 94, which represent the 25th, 30th 35th percentile ranks, respectively. The results of these analyses are reported in Table 3. These findings suggest that the TOSWRF-2 is a valid and reliable predictor of students at risk for reading failure.

Discussion

The TOSWRF-2 is an efficient index of overall reading ability that could be used in a RTI framework to accurately screen students and identify those with learning disabilities. Professionals recognize the critical role fluency plays in reading proficiency (Kuhn & Stahl, 2003; National Reading Panel, 2000; Wolf, 2001). The TOSWRF-2 incorporates essential abilities like word identification and word meaning (vocabulary). Mastery of these abilities enable the student to automatize words, which is necessary for the development of fluent (i.e., accurate and speedy) silent reading.

While the validation of a measure is always ongoing, the data presented here indicate that the TOSWRF-2 is a promising school-wide, repeated measure of reading fluency that is suitable for an RTI framework, as well as in clinical settings. Its wide-range format allows for quick, economical use of single forms across multiple grade levels and its four equivalent forms are especially useful for the kind of repeated measures advocated by the National Reading Panel report (2000) and as part of the popular Response to Intervention model of service delivery (Jimerson, Burns, VanDerHeyden, 2007).

References

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors Influencing the Selection of Standardized Tests for the Diagnosis of Specific Language Impairment. Language, speech and hearing services in schools, 44, 133-146.

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,* 196–211.

Committee on Children With Disabilities, American Academy of Pediatrics. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics, 108,* 192–196.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention. *Journal of Educational Psychology, 98,* 394–409.

Dollaghan, C. (2004). Evidence-based practice in communication disorders: What do we know, and when do we know it? Journal of Communication Disorders, 37, 391-400.

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, MA: Allyn & Bacon.

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2014). *Test of silent contextual reading fluency* (2nd Ed.). Austin, TX: PRO-ED.

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.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO. Retrieved April 3, 2006, from http://www.nrcld.org/symposium2003/jenkins/index

Jenkins, J. R., Hudson, R. G., & Johnson, E. S. (2007). Screening for service in an RTI framework: Candidate measures. *School Psychology Review, 36*, 582-599.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. (2007). *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24,* 174-185.

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.

Kuhn, M.R., *& Stahl, S.* (2003). Fluency: A review of developmental and remedial strategies. *The Journal of Educational Psychology. 95*, 1–19.

Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2014). *Test of silent word reading fluency* (2nd Ed.). Austin, TX: PRO-ED.

National Reading Panel. (2000). *Report of the national reading panel: Teaching children to read*. Washington, DC: National Institutes of Health.

Park, S. H., Goo, J. M., & Jo, C. H. (2004). Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean Journal of Radiology, 5(1),* 11–18.

Plante, E. & Vance, R. (1994). Diagnostic accuracy of two tests of preschool language. American Journal of Speech-language Pathology, 4, 70-76.

Torgesen, J. K., Wagner, R., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray oral reading tests* (5th ed.). Austin, TX: PRO-ED.

Wolf, M. (Ed.) (2001). Dyslexia, Fluency, and the Brain. Timonium, MD: York Press.

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.

Zhou, X. H., Obuchowski, N. A., & Obuchowski, D. M. (2002). Statistical methods in diagnostic medicine. New York, NY: Wiley.

Table 1
Demographic Characteristics of the Learning Disability and Normative
Sample

Sample characteristic	LD Sample (N)	Normative Sample (N)
Total number of participants	206	2,369
Age range	6-24	6-24
Gender		
Male	123	1,144
Female	83	1,293
Race		
White	169	1,898
Black/ African American	20	237
American Indian/Eskimo/Aleut	1	18
Asian/Pacific Islander	6	131
Two or more	10	153
Hispanic		
Yes	41	472
No	165	1,965
LD Status		
Unspecified	82	0
Reading	124	0
Disability Status		
Intellectual disability	5	18
Attention deficit/hyperactivity disorder	54	115
Articulation disorder	9	24
Language impairment	16	37
Deaf/hard of hearing	6	91
Emotional/behavioral disturbance	13	44
Visual impairment	2	4
Autism spectrum disorder	8	33
Developmental delay	12	15
Traumatic brain injury	6	11
Other	18	73

Table 2

Standard Score Means (and Standard Deviations) for Clinical Subgroups

Subgroup	N	M (SD)
Gifted and talented	168	115 (16)
Attention-deficit/hyperactivity disorder	157	92 (16)
Emotional or behavioral disturbance	37	88 (21)
Deaf/hard of hearing	93	88 (17)
Autism spectrum disorder	37	87 (20)
Learning disability		
Unspecified	82	82 (15)
Reading only	56	87 (13)
Reading and writing	22	85 (12)
Reading, writing, and math	25	80 (13)
English as a second language	67	87 (12)
Language impaired	44	86 (16)

Table 3 Diagnostic Accuracy and ROC/AUC Curve Analyses for TOSWRF-2

Criterion	Cut Score	Sensitivity Index	Specificity Index	ROC/AUC	Classification Accuracy	True Positives	False Positives	True Negatives	False Negatives
LD-Unspecified (n = 82)	90	.73	.78		.78	60	540	1,953	22
	92	.77	.75	.82	.75	63	630	1,863	19
	94	.79	.70		.71	65	736	1,757	17
LD-Reading (n = 124)	90	.64	.79		.78	79	521	1,930	45
	92	.73	.75	.81	.75	90	603	1,848	34
	94	.78	.71		.72	97	704	1,747	27
LD-Reading and Writing (n = 22)	90	.55	.77		.77	12	588	1,965	10
	92	.68	.73	.80	.73	15	678	1,875	7
	94	.82	.69		.69	18	783	1,770	4
LD-Reading, Writing, Math (n = 25)	90	.72	.77		.77	18	582	1,986	7
	92	.84	.74	.85	.74	21	672	1,878	4
	94	.88	.69		.70	22	779	1,171	3

Note. ROC/AUC = receiver operating characteristic/area under the curve.