

Test Review Assignment: TOWL4

Instrument: Test of Written Language 4

Assessment Area: Written Expression

Age/Grade Range: Age 9.0 through 17.11

Author: Hammill, Donald D. & Larsen, Stephen C.

Publisher: Pro-ed Inc.

Publication Date: 2009

Revision Dates: Original TOWL published in 1978 and modified in 1983. Revised TOWL-2 published in 1988. Revised TOWL-3 published in 1996. TOWL-4 published in 2009.

Type of Test: meant for individual administration but has instructions and cautions for small group administration; standardized, norm-referenced

Purpose of Test: Designed to identify students who have problems in writing and to determine the degree of the problem. It assesses the linguistic, conventional and conceptual aspects of writing. It can be used as a diagnostic tool to guide remediation and to qualify students for accommodations, and to monitor progress in writing skills.

Recommended Administrator: The TOWL-4 is a level B test and the examiner's manual recommended that examiners have "some formal training in assessment", such as a college or university course. Examiners also need to have a strong knowledge of English language usage, conventions, grammar and the proper construction of sentences and compositions to be able to score the subtests (recommended college graduate level of proficiency).

Amount/Type of training needed: The TOWL-4 manual states that individuals with the qualifications stated above should be able to competently administer the subtests after familiarizing themselves with the specific administration instructions provided and practicing scoring for the spontaneous writing subtests.

Reliability: The examiner's manual presents evidence to support the reliability of the instrument. Coefficient alpha values are provided for each subtest at different ages and grades. These showed Internal Consistency for most subtests (.72-.96), and for composite scores (.82-.96). Test-Retest reliability was mostly within acceptable range (93% rounded to .80 or above). Alternate Forms of composite scores (Immediate: .94 & .95) showed internal consistency but Alternate Forms Delayed (.73- .94) was not proven to be reliable in the Spontaneous Writing composite score. Interscorer Differences fell within acceptable ranges of (.80-.99). The information and evidence presented shows that the TOWL-4 has reasonable internal consistency and test-retest reliability, with the exception of a few subtest scores in the .71-.75 range, and the Spontaneous Writing composite score. Improvements in reliability over the years are clearly delineated in the manual.

Validity: The TOWL-4 Examiner's Manual Evidence provided support for the content validity by providing a detailed description of the test format and content of the test. A rationale for the specific content of each subtest was provided and a rationale for changing the content from previous versions was included in the Item Analysis. The manual also provides a discussion on the Criterion-Prediction Validity and showed a comparison with 3 other tests of literacy, the WLOS, the ROS and the TORC-4. The results showed Level 1-A criterion of acceptability in two out of three comparison studies, supporting the validity.

Standardization: The TOWL-4 was normed on a sample of 2,205 students in 17 states that was collected between 2006 and 2007. The authors chose four major standardization sites as representative of the regions of the United States. The demographic characteristics of the sample included “geographic region, gender, race, ethnicity, household income, education level of parents, and disabling condition” (Hammill & Larsen, p. 56), and were matched with the 2007 U. S. Census. Data was also further stratified by age.

Format/Organization: The TOWL has four components: an Examiner’s Manual, Record/Story Scoring Forms (used with both form A and B), separate student response booklets for Form A or B, and Picture Cards (a sample and one each for form A and B). There is also a Supplemental Practice Story Scoring Booklet included, to allow examiners to rehearse the scoring process. The test is comprised of seven subtests. Subtests one through five are called “contrived” and include Vocabulary (using a word in a sentence), Spelling (writing dictated sentences with an emphasis on correct spelling), Punctuation (writing dictated sentences with an emphasis on correct punctuation), Logical Sentences (editing illogical sentences) and Sentence Combining (integrating two or more sentences into one grammatically correct sentence). Subtests six and seven are called “spontaneous”, and are based on a story written by the subject in 15 minutes; subtest six, Contextual Conventions, measures orthographic and grammatical conventions. Subtest seven, Story Composition, evaluates the sophistication and interest of the story based on standard literary elements such as plot and character development. These individual scores are then combined into three composite scores, Contrived Writing, Spontaneous Writing and Overall Writing.

Scoring: The TOWL-4 is hand-scored using the Examiner’s Manual. For the contrived subtests, raw scores are calculated by adding together the number correct on both item types (1= correct, 0=incorrect). These all start at the first question and have a ceiling of three, making administration easier. For the spontaneous subtests, examiners score subjectively using a scale between 0 and 3, depending on the question, which are then added to achieve raw scores. The raw scores are converted into percentiles and scaled scores using appendices in the manual, and all scores are recorded on the cover of the Record/Story Scoring Form. Subtest scores can then be combined into composite scores by using the sum of the scaled scores in each composite area and assigning a percentile, a descriptor and a composite index found in the appendices. This can then be analyzed for a difference score to determine statistically or clinically significant differences using tables provided in the manual.

Limitations of Test: Test results are not to be used to diagnose or make specific remediation plans. Subtests such as spelling are not extensive enough to pinpoint strengths and weaknesses, so are to be used as a guideline only. There are also no recommendations for remediation strategies.

Advantages of Test: The TOWL clearly breaks down the subtests into conventional, linguistic and cognitive areas. The test is not timed and therefore is less stressful for all students, and allows for thinking time for students with processing issues and learning disabilities to show they are capable of the cognitive aspects of the test. It is estimated to take 60-90 minutes, and the story is done first, allowing students to display their best writing while still “fresh.” If the test is done by an English teacher familiar with written output problems, or a learning assistance professional, it can yield insights into where a student struggles most and can help them to set goals for further testing and remediation.

Criticisms of Test: Although there is extensive evidence provided for reliability and validity, I felt the author's glossed over some areas that did not quite meet the criteria. For example, they rounded up scores as low as .71 on subtests that did not achieve an alpha of .80. Also, there is no basal, so testing a more advanced student would clearly take far longer than the estimated 60-90 minutes. Also, marking of the sentences and stories have guidelines, but is still subjective.

Overall Rating: Good. The test has evolved and changed to accommodate feedback from examiners and examinees. Although there were no reviews available of the TOWL-4, the authors responded to all the criticisms of the TOWL-3 in creating the new test.

Sources of Review:

Hammill, Donald D. & Larsen, Stephen C. (2009) Test of Written Language Fourth Edition: Examiner's Manual