

Angela Orne

EDUC 314

Prof. Deterding

September 25, 2011

Formal Assessment Review

Test of Early Written Language

I. Test

- a. Title: Test of Early Written language
- b. Author: Wayne P. Hresko, Shelley R. Herron, Pamela K. Peak, Deanna L. Hicks
- c. Publisher: Pro Ed
- d. Copyright: 1996

II. Description:

- a. General Purpose: The authors of this test listed four purposes for the TEWL-3. The first purpose being, to identify those students who are significantly below average in writing and need academic intervention. The second purpose being, to identify weaknesses in writing that can be remediated. The third purpose being, to document or monitor progress in writing results from instruction and remediation. The fourth purpose being, to provide researchers with a standardized tool for studying writing in young children.
- b. Materials provided/needed: Examiner's Manual, Form A and Form B Administration/Record Booklets, Form A and Form B student workbooks, three color picture cards, pencil
- c. Alternate forms: There is a form A and a form B to this test

III. Administration:

- a. Age ranges: The TEWL is for age ranges 4-0 to 11-11.
- b. Administration time: There is no set time limit for the TEWL, aside from the 15 to 30 minutes for the contextual writing subtest. On average the TEWL takes about 50 minutes to administer.

- c. Scoring time: There is no specific scoring time listed in the examiner's manual. However, from experience, the scoring for this test is easy and not too time consuming.
- d. Types of scores reported: The raw score, percentile Rank, index score, and descriptive term is found for both the basic writing subtest and the contextual writing subtest. Then there is a difference score found between the two tests. Also, there is a sum of index scores, percentile rank, composite index, and descriptive term found for the overall writing score.
- e. Basal and ceiling levels: The Basal for this test is 5 questions in a row correct. The ceiling for this test is 5 questions in a row incorrect.
- f. Standard error of measurement: The standard error of measurement for the basic writing subtest is 3 and the standard error of measurement for the contextual writing subtest is 4. For the overall writing, the standard error of measurement is 3.
- g. Confidence intervals: The confidence interval for this test is 68%.

IV. Reliability:

- a. Coefficient Alpha: All of the coefficient alphas exceed or round to .90 and are very similar across groups. The coefficients for the basic writing subtest are .95 for Form A and .96 for form B. The coefficients for the contextual writing subtests are .91 for both Form A and Form B. The coefficients for the overall writing is .96 for both Form A and Form B. This test is used to help with measuring standard error of measurement.
- b. Alternate Forms (Immediate Administration): Multiple forms of the test are given during one testing session to estimate error due to content sampling between test forms. To find this reliability both forms A and B were correlated using the test performance in the entire normative sample. All coefficients are .90 or above with the overall writing composite score being .95. This proves that Form A and Form B are equivalent forms with little content sampling error.
- c. Test-retest reliability: This type of reliability is used to measure a student's test performance over time and to estimate time sampling error in this test. In this approach you have to administer the test and then re-administer it a week or two

later. The amount of similarity that is proven between the test scores indicates the reliability level of the test. Both forms of this test were administered to a sample of 89 students with two weeks in between the administrations. The average coefficients ranged between 88-94 for the different grades, subtests, and forms.

- d. Alternate Forms (Delayed administration): This type of procedure is considered to be highly reliable because it measures both the factor of time sampling as well as content sampling. Each student in this study was administered both tests and then re-administered both tests two weeks later. The coefficients for the combined sample for age-based norms was .95 and the coefficients for grade based norms was .94.

V. Validity:

- a. Author's intent: The author's intent for the TEWL-3 was to test the written language ability of early learners, ages 4-11.
- b. Does test measure what the author purports to measure? : Through the various validity tests that were conducted I think that it is obvious that the TEWL-3 measures what the authors set out to measure when putting the test together and creating the test questions.
- c. Content-Description Validity: The authors of this test provided three demonstrations of content-description validity for the TEWL-3 Subtests. The first being, a rationale for each subtest's content is discussed, the second being the results of conventional item analysis, and the third being the results of differential item functioning analyses are reviewed.
- d. Criterion-Prediction Validity: The TEWL- 3 was compared to three tests that measure similar reading and writing skills as the TEWL-3 does. The TEWL-3 was compared to the Wechsler Individual Achievement Test-2, The Woodcock-Johnson III: Tests of Achievement, and the 6 +1 Writing Trait Rubric. The coefficient data found for comparing the TEWL-3 to these three tests range from 63-74. This data shows that there is a positive correlation between the TEWL-3 and the other three tests. Thus, proving that the TEWL measures what it is suppose to measure in reading and writing as shown by criterion-prediction validity.

- e. Construct-Identification Validity: This validity is used to examine the degree to which the underlying traits of the TEWL-3 can be identified and the extent to which these traits reflect what the test is based on. There was a three step procedure used to demonstrate this validity for the TEWL-3. First, several constructs about test performance were identified. Second, hypotheses were identified based on these identified constructs. Third, the hypotheses were verified by logical or empirical methods. The different constructs being hypothesized were relationship to age and grade, relationship of TEWL-3 to TEWL-2, relationships among the subtests, relationship to intelligence, difference among groups, and item validity. Overall, all of the constructs that were tested by logical or empirical methods came back with coefficients at least above .60 making the TEWL-3 a highly valid test.

VI. Norming Procedures:

- a. Sampling Procedures: There was not a specific norming procedure listed in the examiner's manual.
- b. Size of sample: The sample size of the TEWL-3 was 2,085 people from 30 states and British Columbia. This sample was done in two different units with one sample being in the fall of 2007 with 1,033 students and the second sample being in the spring of 2009 with 1,053 students. The sample conducted in the spring had a little bit larger of a male population as opposed to the fall sample which was an even population for males and females. Demographic characteristics in this norming sample ranged from geographic region, gender, ethnicity, Hispanic, household income, educational attainment of parents, and exceptionality status.
- c. Was item analysis conducted and indexes reported? To the best of my knowledge the examiner's manual does not address item analysis.
- d. Date of norms: Fall of 2007-Spring of 2009

VII. Classroom Uses:

- a. As suggested by authors: The authors of the TEWL-3 do not give many suggestions for classroom uses. The authors of this test stress that the test does not diagnose an individual disability. The authors did say however, that the TEWL-3

would “contribute to the selection of long-term educational goals” but not day to day goals (74).

- b. Your opinion of appropriate uses: I could see this test being used in the middle elementary grades such as for 2nd-4th graders. This test could help determining if the student has a processing deficiency or problem expressing themselves in words on paper. Sometimes students have a serious problem not only figuring out what they want to say but also figuring out how to put it on paper and this test might help dissect a students specific problem to provide them better educational help. However, I do not really see too many classroom appropriate uses for this test below the second grade level. I administered this test to a kindergartener and felt as if it didn’t really establish anything with him because he was too young.

| <u>Desirable Features</u> | <u>Undesirable Features</u> |
|---|--|
| <ul style="list-style-type: none"> - Easy to score - Appropriate age range - Excellent reliability and validity coefficients - Two separate easily readable manuals for both examiner and student - Simplicity of the test administration time | <ul style="list-style-type: none"> - Scores once calculated seem to be a bit more ambiguous - Contextual writing should have a firm age range on it, I do not think it is appropriate for any student under 7 - Not an appropriate amount of variety in question type |