Test of Language Development-Primary, 3rd Edition
(TOLD-P:3)

Publisher/Date:

Purpose:
- Individually-administered, norm-referenced test for English-speaking children, ages 4-00 through 8-11 years, who show significant delays in language development and to identify specific areas of strength and weakness.

Provides:
- The test consists of 5 core subtests (Picture Vocabulary, Oral Vocabulary, Grammatic Understanding, Sentence Imitation, and Grammatic Completion) which then comprise the overall composite score (Language Quotient). Three supplemental subtests (Word Discrimination, Phonemic Analysis, and Word Articulation) are provided for an expanded clinical analysis of “speech” systems.

Standardization Issues:
- Standardization in 1996 included 1,000 children ages 4-8, with demographics by geographic region, gender, race, ethnicity, urban/rural, parental education, and disability being matched to 1990 US Census data. Individual cell sizes by full year ranged from 107 children (age 4) to 258 (ages 7 & 8). However, the manual’s standard score conversion tables report by 6-month intervals, and it is not specified how many children were used for each table.

Reliability and Validity Issues:
- Internal-consistency reliability estimates were good at both the subtest-level (.80s to low-.90s) and for the composite data (low-.90s). The overall Language Quotient’s internal consistency was a strong .95 or better across all age-levels. Test-retest estimates (over 4-month interval) are quite acceptable (.81-.92) except for the Word Discrimination subtest, and interrater reliability was uniformly high (.99) across the test. Validity studies are cited but don’t appear as strong as the reliability data. In particular, evidence of discriminant- and construct-validity was questioned. Concerns included the fact that the core subtests loaded on a single-factor as opposed to validating the theoretical structure of the test, and, scores from all subtests correlated similarly with both the semantic and the syntactic scales of a comparison measure.
**Additional Points:**

- The authors acknowledge the test’s limited coverage of “spoken language” and suggest areas not covered (i.e.; pragmatics) be addressed through other methods.

- Extreme care should be used with the test’s use in children below the age of 5 ½ due to limited floors. For example, the minimum obtainable Spoken Language Quotient at age 4 is a standard score of 85 (just one standard deviation from the mean). Subtest and composite floors below 5 ½ years do not meet Bracken’s (1987) suggested criterion. In this case, an alternative measure might be considered.

- While in general the standardization sample appears representative, there was a slight overrepresentation of lower-income families.

- Reliability and validity studies were computed separately for gender, minority, and disability groups (learning disability, speech-language disorder, mental retardation, and “other”) as well as for the general population and did not show any meaningful variation across these variables.

- Classical item analyses were used on each item to choose “good” items from a statistical-point, and DIF (Differential Item Functioning) analysis was conducted to identify and eliminate “biased” items.

- TOLD-P:4 (updated; Fourth Edition) was recently published (2008), however, specific independent reviews are not available at the time of this project. Cursory review of changes from the 3rd Edition include updated norms (from 2005 US Census figures), expanded bias studies, and improved psychometric properties (elimination of floor- and ceiling-effects for subtests and composites—a criticism of the 3rd Edition, updated validity studies, and data on the test’s sensitivity and specificity has been collected).