Oral and Written Language Scales*  
(OWLS)

*Requested for review by the Speech-Language Department, this review will focus on the oral language component of the test (Listening Comprehension and Oral Expression subtests)

Publisher/Date:

Purpose:
- Individually-administered, norm-referenced measure of oral language ability in individuals, ages 3-21.

Provides:
- Three scores are derived (from administration of two subtests), including Listening Comprehension, Oral Expression, and Oral Language Composite.

Standardization Issues:
- Standardization included 1,795 participants at 74 sites nationwide. The sample was a reasonable representation of US Census data from 1991.

Reliability and Validity Issues:
- Little information was provided related to predictive validity (comparing OWLS scores to future academic success), though the manual describes relationship between OWLS scores and current academic performance. It should be noted that this latter correlation was demonstrated for children with more severe academic difficulties and that there was little to no significant differences seen in children with mild impairments from the matched students in the standardization sample. The test’s theoretical-basis is solid as it relates to language development processes. Test-retest reliabilities (based on 3 age groups studied) ranged from .58 to .85. Interrater reliability was an impressive .90-.99, but dependent upon well-trained examiners/scorers. In general, the reliability studies were good but not cited as heavily as the test’s validity evidence.

Additional Points:
- The Written Expression component to the test battery is packaged separately. Comparison of the three language areas (Listening Comprehension, Oral Expression, and Written Expression) can provide a more complete comparison across skill areas.
- Task-demands on the test may not be “typical” of those found in the classroom-setting, possibly calling into question the instrument’s “ecological-validity” (which could be said about many individually-administered language tests).

- While no reliability-data exists for this feature, the test does offer a unique analysis of responses beyond “correct” or “incorrect.” Responses can be further classified as “preferred” or “acceptable.” Further, miscues can be indicated by grammar-, semantic-, or pragmatic-error.

- Studies cited in the manual support lower scores for children with special needs across the demographics of LD, speech-language issues, hearing difficulties, and mental handicaps.

- The test’s effectiveness in progress-monitoring may be limited given that standard deviations for each age-level were often larger than the mean difference between different ages, with small differences generally shown after age 9.

- Hartley, et al. (2007), examined the feasibility and utility of the OWLS as a language assessment instrument in assessing autism-spectrum children. Their findings supported the use of the OWLS as a feasible measure for older autism-spectrum children on the basis that 1) it assesses more than a restricted set of language skills, 2) its ability to capture a wide-range of ability levels (spans functioning in ages 3-21), 3) inclusion of situational-inferencing, non-literal and sarcastic language, and pragmatics, 4) ease of administration and the fact that the test materials and structure don’t unnecessarily “invite” challenging behaviors or undue distractibility over the course of administration. Further, their findings in test score performance did differentiate between autism-spectrum students and their typically-developing cohorts.

- The norms are approaching 20 years old.