



CEL F5
Clinical Evaluation of Language Fundamentals

Elisabeth H. Wiig, PhD, Eleanor Semel, EdD & Wayne A. Secord, PhD

Evidence of **Reliability**

The reliability of a test refers to its accuracy, consistency, and stability of test scores across situations (Anastasi & Urbina, 1997; Sattler, 2008). More specifically, reliability refers to the consistency of scores obtained with the theoretical concept of repeatedly testing the same student on the same test under identical conditions (including no changes to the student). Although this can never be done, various estimates of reliability are obtained in practice.

The difference between a student’s hypothetical true score and the student’s obtained score is called measurement error. Measurement error consists of both systematic and random errors. A reliable test will have relatively small random measurement error and provide consistent scores within and across administrations. The reliability of a test score and systematic error should always be considered in the interpretation of obtained test scores and differences between a student’s test scores on multiple occasions. The reliability of CELF®-5 was evaluated using internal consistency, test-retest stability, and inter-scoring reliability.

Evidence of Internal Consistency

One type of estimated reliability is internal consistency. Internal consistency reliability measures how consistently the items in the domain tested (e.g., a single test or a group of tests) measure one construct. Internal consistency reliability coefficients are used to describe the homogeneity of the items in a test.

The internal consistency of CELF-5 test and composite scores was examined using the split-half method. The split-half reliability coefficient is the correlation between the total scores of the two half-tests, corrected by the Spearman-Brown formula for the full test (Crocker & Algina, 1986; Li, Rosenthal, & Ruben, 1996). The composite score internal consistency reliability coefficients were calculated with the formula

Reliability of Index Scores

The average reliability coefficients are excellent ($r_{xx} \geq .90$) for all of the composite scores (Core Language Score, and the Receptive Language Index, Expressive Language Index, Language Content Index, Language Structure Index, and Language Memory Index scores), ranging from .95 to .96.

recommended by Guilford (1954) and Nunnally and Bernstein (1994).

As the data in the table below indicate, the average reliability

CELF-5 Test	Average Reliability Coefficients (across target ages)
Sentence Comprehension	.87
Linguistic Concepts	.91
Word Structure	.89
Word Classes	.90
Following Directions	.91
Formulated Sentences	.86
Recalling Sentences	.94
Understanding Spoken Paragraphs	.85
Word Definitions	.89
Sentence Assembly	.93
Semantic Relationships	.89
Pragmatics Profile	.98
Reading Comprehension	.87
Structured Writing	.75

Acceptable
 Good
 Excellent

CELF-5 Test	Average Reliability Coefficients (across target ages)
Core Language Score	.96
Receptive Language Index	.95
Expressive Language Index	.95
Language Content Index	.95
Language Structure Index	.96
Language Memory Index	.95

Acceptable
 Good
 Excellent

Reliability coefficients by age are reported in the CELF-5 Technical Manual.

Evidence of Reliability for Clinical Groups

Evidence of internal consistency reliability for clinical groups was obtained from a sample of 301 students in three groups: language disorder (LD), autism spectrum disorder (ASD), and learning disability in the area of reading and/or writing (LDR/W).

The table below provides internal consistency reliability coefficients of tests for each of the three clinical groups who took the tests (i.e., because the tests were not appropriate

to their ages, students in the LDR/W group did not take the SC, LC, or WS tests). This table shows that the test reliability coefficients of most of the clinical groups are either higher than or similar to those coefficients reported for the normative sample, which suggests that CELF-5 is equally reliable for measuring the language skills of students from the general population or students with clinical diagnoses.

Complete information about the studies can be found in the CELF-5 Technical Manual.

Clinical Group Test	Language Disorder (n=166)	Learning Disability (Reading & Writing) (n=66)	Autism Spectrum Disorder (n=66)	Average r_{xx}
Sentence Comprehension	.94		.96	.95
Linguistic Concepts	.96		.98	.97
Word Structure	.93		.94	.94
Word Classes	.96	.92	.97	.95
Following Directions	.96	.90	.98	.96
Formulated Sentences	.97	.89	.96	.95
Recalling Sentences	.98	.92	.97	.96
Understanding Spoken Paragraphs	.81	.75	.91	.84
Word Definitions	.87	.91	.95	.92
Sentence Assembly	.92	.94	.97	.95
Semantic Relationships	.88	.89	.96	.92
Pragmatics Profile	.99	.99	.99	.99
Reading Comprehension	.93	.86	.93	.91

Acceptable
 Good
 Excellent

Evidence of Test-Retest Stability

Another way of estimating the reliability of a test is to examine its test-retest stability. Test-retest stability is the correlation between the test and retest scores and is a direct measure of test stability for repeated testing. To examine retest stability, the student is given the same test twice, each time under conditions that are as similar as possible. The student will not perform exactly the same way in each of the two test sessions. The time interval between the test and retest is as short as possible, to minimize changes in the individual, yet long enough for any practice or memory effects to dissipate.

The CELF-5 test-retest reliability was evaluated in a study in which the test was administered to a group of students on two separate occasions and then the scores were compared. The sample used to assess the stability of CELF-5 scores over time included 137 examinees in three age bands (ages 5:0–6:11, 8:0–9:11, and 12:0–16:11) selected from the standardization sample. The sample included 62 females and 75 males. In the sample 62.8% was white; 14.6% was African American; 13.9% was Hispanic; 2.9% were Asian; and 5.8% were students of other races/ethnic origins. The parent/caregiver education level of the sample was as follows: 15.3% had 9–11 years of school; 21.9%, 12 years; 32.8%, 13–15 years; and 29.9%, 16 or more years. The age of the participants ranged from 5:0 through 16:11, with a mean age of 9.0 years.

After being tested as part of the standardization study, these students repeated the test within a range of 7 to 46 days (mean of 19 days), with both tests administered by the same examiner. The test-retest reliability was estimated using Pearson's product-moment correlation coefficient for the three age bands. As the data indicate, the CELF-5 test scores possess adequate stability across time for all three age groups tested. The average corrected stability coefficients for all ages for Recalling Sentences is excellent (.90); the average corrected stability coefficients for Linguistic Concepts, Word Structure, Word Classes, Understanding Spoken Paragraphs, Word Definitions, Sentence Assembly, and Pragmatics Profile, are good (in the .80s); those of the other tests are adequate (in the .70s). In addition, the average corrected stability coefficients of the composite scores are all excellent or good, ranging from .83 to .90.

This test-retest study was conducted to evaluate stability of test scores. The shortest test-retest interval that will not result in significant practice effects on CELF-5 has not been determined; however, one or more of the following three criteria should be met before CELF-5 is administered again:

1. Retesting should be conducted after the student no longer remembers the test items and/or his or her responses when tested previously.
2. Retesting should be conducted when the examiner thinks the child has made progress since the previous test administration; otherwise, there is no reason to retest.
3. Retesting can be conducted when the student's age at testing requires the next-age norms table to score.

Evidence of Inter-Scorer Agreement

Most CELF-5 tests are scored objectively; however, the Word Structure, Formulated Sentences, Word Definitions, and Structured Writing tests require familiarity with different scoring criteria that require clinical judgment, and qualitative

and quantitative judgments about student responses. Familiarity with varieties of American English is also important, so the clinician can identify the variety of American English spoken by the student being tested. Because there is room for interpretation, it is necessary to evaluate the extent to which these interpretations are consistent from one scorer to another. Scoring rules were developed for the Word Structure, Formulated Sentences, Word Definitions, and Structured Writing tests, and scorers were trained in applying the rules before the standardization protocols (i.e., record forms) were scored. The training materials for the Formulated Sentences and Structured Writing tests are provided in chapter 3, Table 3.2 and in Appendix H in the CELF-5 Examiner's Manual. The Formulated Sentences, Word Definitions, Word Structure, Structured Writing, and all other CELF-5 tests were scored by a team of seven trained scorers under the supervision of the test developers. To ensure accuracy of scoring before analysis of the test data, two different scorers from the team were randomly selected to score each protocol independently. The scores were compared, and a third independent scorer resolved any differences.

Inter-scorer reliabilities were .99 for Word Structure; .95 for Formulated Sentences; .91 for Word Definitions; and .96 for Structured Writing. These results demonstrate that although these tests require more judgment, they can be scored reliably, even by individuals with no experience with the CELF-5 scoring rules prior to scoring training.

Additional information about internal consistency, standard error of measurement, test-retest stability, and inter-scorer agreement can be found in the CELF-5 Technical Manual.

**For more information about CELF-5, please visit
PearsonClinical.ca or call 1-866-335-8418.**

1-866-335-8418 |  PsychCorp | PearsonClinical.ca

Copyright © 2013 Pearson Education, Inc. or its affiliate(s). All rights reserved. CELF, Always Learning, Pearson, design for Psi, and PsychCorp are trademarks, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s)