

Test Review

Journal of Psychoeducational Assessment
2015, Vol. 33(5) 495–500
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
jpa.sagepub.com



Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.

Reviewed by: Marian C. Coret & Adam W. McCrimmon, University of Calgary, Alberta, Canada
DOI: 10.1177/0734282914557616

Test Description

The *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5; Wiig, Semel, & Secord, 2013)* is a recently updated battery of tests designed to assess, diagnose, and measure changes in language and communication in individuals 5 to 21 years of age. Designed to identify language strengths and weaknesses, determine service eligibility, provide intervention strategies, and measure intervention efficacy, the CELF-5 assesses both oral and written language as well as non-verbal communication skills. As a Level B measure, the CELF-5 can be administered by speech and language pathologists, school psychologists, special educators, and qualified diagnosticians with a master's degree or certification in standardized testing. The core tests can be completed within 30 to 45 min, whereas the full battery may take 90 to 120 min.

The measure includes an Examiner's Manual, Technical Manual, two Stimulus Books, 25 copies of Record Forms 1 and 2, 10 copies of Reading and Writing Supplements 1 and 2, 50 Observational Rating Scales, and 30 Q-global Score Reports. The Examiner's Manual outlines the recommended assessment process and provides instructions for administration, scoring, and interpretation of subtests as well as suggestions for intervention and further assessment. The Technical Manual provides details of the purpose, design, development, standardization, and psychometric properties of each subtest.

Measure Development

The CELF-5 was developed to improve usability and update test content and scope to align with curricula and assessment trends. Based on literature reviews, clinician feedback, and pilot research, the authors removed and revised items and subtests from the previous edition (*Clinical Evaluation of Language Fundamentals—Fourth Edition [CELF-4]; Semel, Wiig, & Secord, 2003*), added new items and tests, and refined administrative procedures. Age-specific start rules were more widely incorporated, discontinue rules were reduced, and administration and scoring procedures were simplified. The CELF-5 has newly introduced measures for written language, reading comprehension, and social communication, and also includes the ability to track performance over time. In addition, a recommended administration process was developed to allow for efficient use of individual and groups of subtests.

Description, Administration, and Scoring

The CELF-5 is comprised of 16 age-specific subtests, 12 of which differentially combine to create the Core Language Score (CLS), Receptive Language Index (RLI), Expressive Language

Table 1. Subtests Contributing to Core and Index Scores by Age.

	Ages 5-8	Ages 9-12	Ages 13-21
CLS: Measures overall language performance	Sentence Comprehension Word Structure Formulated Sentences Recalling Sentences	Word Classes Formulated Sentences Recalling Sentences Semantic Relationships	Formulated Sentences Recalling Sentences Understanding Spoken Paragraphs Semantic Relationships
RLI :Measures listening and auditory comprehension skills	Word Classes Following Directions Semantic Relationships	Word Classes Understanding Spoken Paragraphs Semantic Relationships	Word Classes Understanding Spoken Paragraphs Semantic Relationships
ELI: Measures expressive language skills	Word Structure Formulated Sentences Recalling Sentences	Formulated Sentences Recalling Sentences Sentence Assembly	Word Classes Understanding Spoken Paragraphs Sentence Assembly
LCI: Measures semantic development	Linguistic Concepts Word Classes Following Directions	Word Classes Understanding Spoken Paragraphs Word Definitions	Word Classes Understanding Spoken Paragraphs Word Definitions
LSI: Measures the ability to interpret and produce sentence structures	Sentence Comprehension Word Structure Formulated Sentences Recalling Sentences	NA	NA
LMI: Measures memory for language tasks	NA	Following Directions Formulated Sentences Recalling Sentences	Following Directions Formulated Sentences Recalling Sentences

Note. CLS = Core Language Score; RLI = Receptive Language Index; ELI = Expressive Language Index; LCI = Language Content Index; LSI = Language Structure Index; NA = not applicable; LMI = Language Memory Index.

Index (ELI), Language Content Index (LCI), Language Structure Index (LSI), and Language Memory Index (LMI). The CLS is a measure of overall language performance, whereas the index scores provide more information about specific language skills. A visual depiction of the factor structure and subtest composition for each age group appears in Table 1.

The CELF-5 authors recommend the initial use of the Observational Rating Scale to identify areas of concern, followed by administration of relevant tests to determine the existence and nature of a language deficit. This process allows for a flexible and comprehensive assessment as subtests can be administered individually to assess specific language skills or as a battery for a comprehensive language evaluation.

Each test is age-specific and includes differential start points, sample items, and specific reversal and discontinue rules. All test items are scored as either 0 (incorrect) or 1 (correct), except for Formulated Sentences (0-, 1-, or 2-point basis) and Recalling Sentences (0-, 1-, 2-, or 3-point basis). Subtest raw scores are converted to scaled scores, and confidence intervals and percentile ranks can be determined (except on the Pragmatics Activities Checklist which uses criterion referenced-scores). Index scores are converted to standard scores and compared to determine any discrepancies. Age equivalents and growth scale scores can also be determined for most subtests.

Subtest Description

The Observational Rating Scale. The Observational Rating Scale consists of 40 items that measure language and communication skills across environments for ages 5 to 21. Listening, speaking,

reading, and writing behaviors are rated by parent/caregivers, teachers, and, if appropriate, the individual being assessed.

Sentence Comprehension. Sentence Comprehension consists of 26 items that evaluate 5- to 8-year-olds' understanding of grammatical rules at the sentence level. Examinees are required to identify one picture from a set of four that best represents an orally presented description.

Linguistic Concepts. Linguistic Concepts consists of 25 items that examine 5- to 8-year-olds' ability to interpret and follow verbally presented instructions with logical operations. Children are required to identify objects corresponding to the examiner's verbal description.

Word Structure. Word Structure is a cloze procedure consisting of 33 items designed to measure 5- to 8-year-olds' ability to correctly use pronouns and apply English morphological rules. Children are required to provide the ending to an incomplete sentence presented by the examiner with reference to one to two illustrations provided.

Word Classes. Word Classes consists of 40 items that measure 5- to 8-year-olds' ability to understand connections between words related by semantic features, function, place, or time. Initial items involve three to four pictured objects whereas more challenging items involve four orally presented words, with examinees verbally identifying the two that are most similar.

Following Directions. Following Directions consists of 33 items that measure 5- to 21-year-olds' ability to interpret, recall, and follow verbally presented directions. Examinees are shown 4 to 18 pictured objects and are required to point to objects in the order described by the examiner.

Formulated Sentences. Formulated Sentences is comprised of 24 items that examine 5- to 21-year-olds' ability to orally produce grammatically and semantically correct sentences of increasing length and difficulty using one to two stimulus words provided by the examiner and with reference to a presented illustration.

Recalling Sentences. Recalling Sentences consists of 26 items that test 5- to 21-year-olds' ability to attend to, recall, and reproduce sentences of increasing length and difficulty verbally presented by the examiner.

Understanding Spoken Paragraphs. Understanding Spoken Paragraphs is composed of 20 items that measure 5- to 21-year-olds' ability to attend to a short story and interpret overall themes, order of events, details, and formulate inferences and predictions.

Word Definitions. Word Definitions consists of 21 items that evaluate 9- to 21-year-olds' ability to define words based on semantic features.

Sentence Assembly. Sentence Assembly is composed of 20 items that measure 9- to 21-year-olds' ability to assemble words and/or word groups into grammatically and semantically correct sentences.

Semantic Relationships. Semantic Relationships is comprised of 20 items that measure 9- to 21-year-olds' ability to understand sentences based on comparative, spatial, temporal, and sequential information. Examinees select two multiple choice answers, from a set of four, in response to an orally presented question by the examiner.

Pragmatics Profile. The Pragmatics Profile is comprised of 50 behavioral statements that form a behavioral checklist that measures 5- to 21-year-olds' verbal and non-verbal pragmatic language skills.

Reading Comprehension. Reading Comprehension consists of 16 to 19 items that evaluate 8- to 21-year-olds' ability to understand information presented in written paragraphs. Examinees independently read a written passage and are subsequently asked 8 to 10 questions regarding themes, order of events, details, inferences, and predictions.

Structured Writing. Structured Writing is comprised of two age-specific narratives of which 8- to 21-year-old examinees read one complete sentence, complete a subsequent incomplete sentence, and write one to four additional sentences within the theme of the narrative.

Pragmatics Activity Checklist. The Pragmatics Activity Checklist consists of 32 statements that measure functional verbal and non-verbal communication skills. The examiner engages in authentic social interaction with the examinee and documents the occurrence of particular non-verbal and verbal behaviors.

Technical Adequacy

Test Standardization

Standardization occurred in the United States from March to December 2012 on a sample of 3,250 English-speaking 5- to 21-year-olds representative of the U.S. population in 47 states. Specifically, 200 children of each age year between 5:0-5:11 and 12:0-12:11, 150 children in each year ages 12:0-12:11 to 16:0-16:11, and 180 individuals from ages 17:0-17:11 to 21:0-21:11 participated. The sample was stratified according to age, race/ethnicity (White, Hispanic, African American, Asian, and Other), geographical region (West, Midwest, Northeast, and South), and parent/caregiver education level (less than a high school diploma, high school diploma, some college or technical school, and 4 or more years of college). The Technical Manual states that 5% of the sample reported having an attention disorder; 1% having a learning disability; 1% having intellectual disability, pervasive developmental disorder, Down syndrome, or developmental delay; and less than 1% each of emotional disturbance, cerebral palsy, color blindness, central auditory processing disorder, visual impairment, autism, or other diagnoses. Approximately 7% of the sample was diagnosed with speech and/or language disorders, 4% with articulation or phonological disorder, and <1% with fluency/voice disorder.

Reliability

Internal consistency. Internal consistency was measured utilizing the split-half method with the Spearman–Brown correction formula for the full test. The average subtest reliability coefficients ranged from acceptable (.77) to excellent (.99) for the younger (5-8) sample, while the reliability coefficients for the indexes were excellent and ranged from .93 to .97. Regarding the older (9-21) sample, the average subtest reliability coefficients were acceptable (.60) to excellent (.99), while the reliability coefficients for the indexes were excellent and ranged from .92 to .97.

Internal consistency was also calculated for individuals from three special populations: language disorders, autism spectrum disorder, and reading and/or writing learning disability. Coefficients ranging from acceptable (.75) to excellent (.99) were reported for the subtests; however, index coefficients were not reported.

Test–retest stability. Test–retest stability was obtained via Pearson's product–moment correlation by administering the measure twice within a 7- to 46-day interval to 137 participants in three age bands (5:0-6:11, 8:0-9:11, and 12:0-16:11) that were representative of the overall normative sample.

Overall results for the younger sample indicate acceptable (.68) to excellent (.92) subtest stability and good (.84-.89) composite stability. Similarly, results for the middle age group indicated adequate (.77) to good (.89) subtest stability and good (.87) to excellent (.92) composite stability. Finally, results for the oldest age group indicated poor (.56) to excellent (.93) subtest stability and good (.86) to excellent (.91) composite stability.

Interrater reliability. The majority of subtests are objectively scored (i.e., correct or incorrect) and so were not analyzed for interrater reliability. However, Word Structure, Formulated Sentences, Word Definitions, and Structured Writing necessitate qualitative judgments regarding scoring of examinee responses. Interrater reliability on these subtests was evaluated by a pair of trained scorers from a team of seven who were randomly selected to score each protocol separately. Scores were then compared, with a third scorer assisting resolution of any discrepancies. Overall interrater reliability for these subtests was excellent and ranged from .91 (Formulated Sentences) to .99 (Word Structure).

Validity

The CELF-5 is designed to be a good indicator of language ability in children, adolescents, and young adults. Confirmation of this was obtained through examination of test content, internal structure, correlations with other related measures, and special group studies.

Internal structure. Good to strong interrelationships among all subtests, as well as composites, support the validity of the CELF-5. Specifically, intercorrelations ranged from .19 to .65 for the subtests and from .72 to .97 for the indexes. The range of intercorrelations for the subtests is not unexpected as the specific aspect of language functioning measured by different subtests can vary, resulting in reduced overlap among tasks. The factor structure of the measure was investigated by confirmatory factor analysis using three age bands (5:0-8:11, 9:0-12:22, and 13:0-21:11). Results indicated support for either of two different models for each age band—a second-order unitary factor consisting of the CLS and two first-order factors consisting of the RLI and ELI, LCI and LSI, or LCI and LMI, depending on the age band. Specific details are reported in the Technical Manual.

Concurrent validity. The relations among CELF-5 scores and scores on other measures of language development informed the measure's concurrent validity. Specifically, correlations between CELF-5 and CELF-4 subtests were adequate (.64) to good (.88), whereas correlations between the indexes were good (.82) to excellent (.92). Additional comparison with the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4; Dunn & Dunn, 2007) indicated adequate (.75) to excellent (.95) correlations with CELF-5 subtests and adequate (.68) to good (.80) correlations with CELF-5 indexes. Similarly, comparisons with the Expressive Vocabulary Test—Second Edition (EVT-2; Williams, 2007) indicated adequate (.71) to excellent (.98) correlations with CELF-5 subtests and adequate (.65-.78) correlations with CELF-5 indexes.

Special group studies. Test-criterion relationships are also provided based on special group studies. A sample of 67 children aged 5:0-15:11 and identified as presenting with a form of language disability were recruited for a specific clinical study. The test developers compared the performance of this sample to a sample without a language disorder matched on age, gender, race/ethnicity, and parent/caregiver education level. Results indicated significant score differences between these samples at the .01 level for all subtests and indexes, providing evidence of the measure's ability to identify examinees with a language disorder.

Commentary and Recommendations

Considerable changes were made to the CELF-5 from the previous edition. Although the latest revision has a number of improvements, some limitations are noted.

Strengths

The CELF-5 can be used within educational, clinical, and research settings and is standardized on a large and representative population. The CELF-5 has revised content and scope, improved administration and scoring procedures, and new processes for evaluating written language and pragmatics. In addition, the new assessment process allows for flexible and efficient use of individual and groups of subtests.

Limitations

Despite its strengths, the CELF-5 also has some limitations. Although most reported indicators of reliability are acceptable to excellent, there is evidence of low test stability in some instances. Q-global scoring software provides web-based scores and reports at an additional cost per report not included in the standard kit. Also, performance on pragmatic subtests emphasizes North American sociocultural behaviors, which may have implications for use with other cultures.

Conclusion

Since the development of the CELF-4, it is clear that the test developers have made significant improvements in user-friendliness, content, and flexibility of the measure, which largely outweigh the minor limitations. Overall, the CELF-5 is a useful and dynamic tool for assessing language weaknesses and strengths.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Bloomington, MN: NCS Pearson.
- Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4)*. San Antonio, TX: NCS Pearson.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.
- Williams, K. T. (2007). *EVT-2: Expressive Vocabulary Test* (2nd ed.). Circle Pines, MN: Pearson Assessments.