

Standardized Assessment Support for Speech-Language Pathology

Overview:

There are two types of assessment tools that are administered by Speech-Language Pathologists (SLP); standardized and non-standardized. Standardized assessment tools assess a child's speech and language skills by comparing their performance to a large sample of their peers typically representing a variety of regions from across the United States as well as a variety of cultures. Standardized assessments require all test takers to answer the same questions, or a selection of questions from a common bank of questions, in the same way, and those answers are scored in a “standard” or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. Non-standardized assessments tools assess a child's speech and language skills in a more naturalistic and decontextualized manner gathering qualitative data through observation, interviews, checklists/probes, language sampling, dynamic assessment, curriculum-based assessments, etc.

IDEA has several requirements for evaluation relative to the use of standardized and non-standardized assessment tools when conducting an evaluation of a student suspected of a disability (34 CFR § 300.304). First, consideration of eligibility as a student with a disability requires that the presence of a disability is based on having been evaluated in accordance with the evaluation procedures as outlined in IDEA. In addition to the presence of a disability, the data gathered as part of the evaluation must also document that the disability has an adverse educational impact performance (or for preschool aged students, educational impact refers to their ability to “participate in appropriate activities” (34 CFR §300.304(b)(1)(ii)) and a need for specially designed instruction. IDEA further requires that the assessment is administered by trained and knowledgeable personnel in accordance with the instructions outlined by the publisher of the assessment.

IDEA does not require the use of standardized assessment tools but does have several requirements when or if a standardized tool is utilized. First, a single measure or assessment cannot be used as the sole criterion for determining whether a child is a child with a disability. This means that the score from a standardized assessment alone cannot be used to determine eligibility. In fact, IDEA requires three types of data to be gathered using a variety of assessment tools and strategies. This includes gathering relevant functional, developmental, and academic information about the child. This information is then used to both determine whether the child has a disability as well as the content of the child's IEP (Individualized Education Program). The content of the IEP includes an impact statement which should reflect data related to the child's ability to be involved in and progress in the general education curriculum (or for a preschool child, to participate in appropriate activities) in order to in order to meet the educational standards that apply to all students (34 CFR § 300.39(b)(3)(i-ii).

Second, IDEA requires that evaluators use technically sound instruments. Technically sound instruments generally refers to assessments that have been shown through research to be valid and reliable (71 Fed. Reg. at 46642). Validity generally refers to diagnostic accuracy to determine the presence of a disorder or typical development. This is reflected by an assessment tool's sensitivity and specificity data. Reliability

generally refers to consistency of the findings. In short, no matter when the test is administered or who administers the test, the findings should be relatively the same. In addition, Confidence Intervals (CI) and Standard Error of Measurement (SEM) must be considered.

The combination of the CI and SEM indicate the degree of confidence that the child's true score on a test is represented by the actual score the child received." (Betz et al., 2013). The smaller the SEM, the greater the reliability and confidence in the test score. If the standardized assessment tool does not meet IDEA's requirement as a "technically sound instrument" the data collected may be used qualitatively, but the score may not be used to determine eligibility.

Relative to assessment of students who are culturally or linguistically diverse, IDEA requires that any assessments and evaluation materials must not be discriminatory on a racial or cultural basis. Related to standardized assessments, this means that the evaluator and test items must not have implicit or explicit bias. Implicit evaluator bias may be in the form of a lack of familiarity with the student's language, dialect, or culture. For example, if the evaluator is not familiar with a language or dialect that does not include the voiced or voiceless /th/ phonemes, they may count a student's response as incorrect when, in fact, that phoneme is not part of their phonemic inventory. Implicit test item bias may take the form of "modified scoring guidelines" which may lead to under-identification of a true language disability (Hendricks & Adlof, 2017) or over-identification of students with typically developing language for students with a dialect such as African American English (AAE), Spanish-influenced English, Appalachian English, and Chinese-influenced English (Crowley & Bucaj, 2023).

Moreover, if using a standardized assessment to evaluate a culturally or linguistically diverse student suspected of a disability, IDEA requires that the assessment must be provided in the child's native language or other mode of communication and in the form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally, unless it is clearly not feasible to provide or administer. This is important as languages and dialect may differ in syntax, morphology, phonology, semantics, and pragmatics. For example, there are several common languages and dialects where the plurals are marked by using the number and not the addition of the plural -s morpheme at the end of the word (i.e., "There are two cat" vs. "There are two cats"). This is why assessment tools written for speakers of General American English (GAE) should never be translated into another language. If an assessment has been translated,, the score may not be used for determination of eligibility. Standard scores also may not be used if the student is not reflected in the normative sample for the standardized tool (e.g., the student is Russian, but the normative sample only included White, African American, Hispanic, and Asian children in the sample). Even if using a standardized assessment that is written in a language other than GAE, the score can only be used for determination of eligibility if the tool was written to match the student's language ability (e.g., monolingual Spanish tests aren't the same as a bilingual test designed for Spanish-English bilinguals).

The 2024 SC Standards for Evaluation and Eligibility Document (SEED) does not require the use of standardized assessments. Specific to assessment of language, the SEED outlines that in order to be considered a language impairment, three criteria must be met in order to document the presence of a disability (though additional data must be gathered and reviewed to denote adverse educational impact and need for specially designed instruction). Only one of the criteria refers to standardized assessments.

The other criteria refer to the use of non-standardized measures of dynamic assessment, language sampling, and criterion referenced assessment. If a standardized assessment is one of the three criteria being considered, the composite standard score must be considered relative to the cut score for that specific assessment, sensitivity and specificity at that cut score, and confidence intervals.

Finally, it is important to note that standardized assessments may be used qualitatively if they do not meet the requirements as outlined above. Qualitative data can identify specific areas that warrant further assessment using non-standardized assessment tools and strategies. The variety of data collected can then be triangulated outlining converging evidence indicative of a disability.

Key Terms and Definitions:

- **Sensitivity** refers to the degree to which the assessment will accurately identify those who truly have a language disorder as having one (Dollaghan, 2007).
- **Specificity** refers to the degree to which the assessment will accurately identify those who truly do not have a language disorder as typical (Dollaghan, 2007).
- **Normative Sample** refers to the group of individuals whose performance data are used as a reference for evaluating individual test scores. The individual being evaluated should be represented in the normative sample for the test being used.
- **Normalcy of subjects** - Students with a Disability (SWD) refers to normative samples that included specific sub-populations that may alter the overall distribution of scores. When the purpose of the test is to identify the presence of a language disorder, individuals with a known disorder should not be included in the normative group (Pena, Spaulding, & Plante, 2006).
- **Dialectal Consideration** refers to data gathered by the test developers relative to dialects used by those in the normative sample which is often then used to provide optional scoring considerations in the administration manual.
- **Confidence Interval** and **Standard Error of Measurement (SEM)** indicate the degree of confidence that the child's true score on a test is represented by the actual score the child received." (Betz et al., 2013). The smaller the SEM, the greater the reliability and confidence in the test score.
- **Validity** –
 - Content validity refers to how representative the test items are of the content that is being assessed (Paul, 2007).
 - Construct validity assesses how well the test measures what it purports to measure (Paul, 2007).
 - Concurrent validity is the extent to which a test agrees with other valid tests of the same measure (Paul, 2007).
- **Reliability** -
 - Reliability relates to the consistency of the results of a given test, as any assessment has inherent errors. Reliability measures the number of errors or inconsistencies in the test results.
 - Test-retest reliability is a measure used to represent how stable a test score is over time (McCauley & Swisher, 1984).

- Inter-examiner reliability is used to measure the influence of different test scorers or different test administrators on test results (McCauley & Swisher, 1984).
- Inter-item consistency assesses whether “parts of the test are measuring something similar to what is measured by the whole” (Paul, 2007).

List of Various Standardized Language Measures:

Below is a table outlining the information currently available for a variety of standardized assessment tools in the area of language. This is not a list of all standardized assessment tools, but rather a sample of the tools commonly used and/or available at this time (which may or may not reflect the most current version of the test). The data gathered is based on current test manuals and subject to change based on normative updates by publishers. Information in the table should be considered as one piece of information in the selection of a standardized assessments with consideration that IDEA requires that determination of eligibility is not based on a single measure or assessment, but rather a variety of assessment tools and strategies. In addition, the table below does not reflect all psychometric properties for a given tool as this can only be found within the manual. As such, SLP personnel are strongly encouraged to consult the manual for any and all standardized assessment tools used.

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
ALL Assessment of Language and Literacy	Pre-K, K5 and 1 st grade	@ 1.5 SD – 86% @ 2 SD – 54%	@ 1.5 SD – 96% @ 2 SD – 98%	White – 62% Hispanic – 18% African American – 15% Other – 6%	African American English	
BESA Bilingual English-Spanish Assessment	4;0 – 6;11	Two subtests on the BESA, morphosyntax and semantics provides the best diagnostic accuracy and with the addition of story retell sensitivity and specificity are 100% and 93% respectively (Lazewnick, et al., 2018).			17 Spanish dialects and 7 regional dialects for English.	

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
CASL-2 Comprehensive Assessment of Spoken Language	3;0-21;11	@ cut score 70 – 41% @ cut score 75 – 47% @ cut score 80 – 64%	@ cut score 70 – 99% @ cut score 75 – 96% @ cut score 80 – 91%	White – 56.7% Hispanic – 22% African American – 14.4% Asian – 2.8% Asian Pacific Islander – <1% Two or more – 2% American Indian, Eskimo, Aleut – <1% Other – 3.4%	African American English, Southern English	
CAPs Clinical Assessment of Pragmatics	7-18	@ 1.0 SD – 100% @ 1.5 SD – 100% @ 2 SD 90%	@ 1.0 SD – 85% @ 1.5 SD – 90% @ 2 SD 97%	White – 77% Hispanic – 14% African American – 11% Asian – 4% Other – 7%		Students with known disability included in sample (15%)
CCC-2 Children’s Communication Checklist	4;0-16;11	@ 1.0 SD – 89%	@ 1.0 SD – 97%	White – 61.89% Hispanic – 16.96% African American – 15.47% Other – 5.68%	Not reported	Students with known disability included in sample (20%)
CELF:P-3 Clinical Evaluation of Language Fundamentals - Preschool	3;0 – 6;11	@ 1.0 SD – 85% @ 1.5 SD – 89% @ 2 SD 71%	@ 1.0 SD – 82% @ 1.5 SD – 87% @ 2 SD 97%	White – 56% Hispanic – 22% African American – 13% Asian – 2% Other – 7%	African American English, Southern English, Spanish Influenced English, Asian influenced English	Students with known disability included in sample (13%)
CELF:P-2 (Spanish) Clinical Evaluation of Language Fundamentals – Preschool (Spanish)	3;0 – 6;11	@ 1.5 SD - 93%	@ 1.5 SD - 68%			

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
CELF-5 Clinical Evaluation of Language Fundamentals	5;0 – 21;11	@ 1.5 SD – 85% @ 2 SD – 57%	@ 1.5 SD – 99% @ 2 SD – 100%	White – 56.8% Hispanic – 20% African American – 13.8% Asian – 3.6% Other – 5.9%	African American English, Southern English, Spanish influenced English, Asian influenced English	Students with known disability included in sample (20%) “Construct validity is insufficient because the reference standard used to identify the sensitivity and specificity groups for the crucial discriminant accuracy analysis is flawed... Moreover, this test contains very significant racial, cultural, and socioeconomic biases.” (Crowley & Bucaj, 2023)
CTOPP-2 Comprehensive Test of Phonological Processing	4;0- 24;0	Not reported	Not reported	White – 76% Hispanic – 16% African American – 14% Asian Pacific Islander – 2.0% Two or more– 4% Other – 4%	Not reported	Students with known disability included in sample (7%)
DAY-C Developmental Assessment of Young Children - Communication	Birth – 5;11	@ cut score of 90* – 75%	@ cut score of 90* – 84%	White – 77% Hispanic – 18% African American – 15% Asian– 3%		*Cut score and SD reported for Communication section only



Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
DELV Diagnostic Evaluation of Language Variation	4-9	@ 1.0 SD – 95% @ 1.5 SD – 69% @ 2 SD 36%	@ 1.0 SD – 93% @ 1.5 SD – 99% @ 2 SD 100%	White – 63% Hispanic – 14% African American – 16.44% Asian – 4.89% Two or more – 4% Other – 4%	African American English, Appalachian English, Southern English, Cajun English, Spanish influenced English	
DYMOND Dynamic Measures Of Narrative Language and Decoding	5;0-14;5	95%	91%	White – 57.5% African American – 8.3% Hispanic – 25.9% American Indian/Alaska Native – 3.1% Asian – 2.7% Native Hawaiian/ Pacific Islander - <1% Two or more/other – 2.1%		Norm Referenced dynamic assessment of language.
ESB Early Sociocognitive Battery	1;6 – 4; 11	@ 2 SD - 80%	@ 2 SD - 86%	White – 51.5% African American – 10.9% Asian – 3.1% Hispanic – 24.4% Two or more/other – 10.1%	Not applicable for this tool	At a 95% Confidence Interval, Standard Error of Measurement is between 29-35

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
EOWPVT-4 Expressive One Word Picture Vocabulary Test	2;0-80;11	Not reported	Not reported	Caucasian – 63.2% Hispanic – 18% African American – 12.8% Asian American – 3.4% Native American – 1% Other – <1% Not reported – 1.4%	Not reported	Students with known disability included in sample (8.7%)
EVT-3 Expressive Vocabulary Test-3	2-90	@ 1.0 SD – 88%	@ 1.0 SD – 83%	White – 82.5% Hispanic – 25.5% African American – 17% Asian – 5.7% Other – 7.9%		Students with known disability included in sample (3.7%)
LPT – 3 Language Processing Test - 3	5-11	Not reported	Not reported	White (61%); Hispanic (18%); African American (17%); Asian-American and others (4%)		Included students with language learning disorders but not HI, ID, ED, LEP
MAVA Montgomery Assessment of Vocabulary Acquisition	3-12	Receptive @ 1.5 SD – 100% Expressive @ 1.5 SD – 83%	Receptive @ 1.5 SD – 85% Expressive @ 1.5 SD – 100%	White – 63% Hispanic – 15% African American – 17% Other – 6%		Students with known disability included in sample (10%)
NLM Narrative Language Measure (part of the Cubed-3 combination of Narrative Language Measure + Dynamic Decoding Measure)	Normed for PreK-3 rd , but extends to 8 th grade	(Cubed-3) K5 - 85% 1 st grade - 100% 2 nd grade – 86% Third grade – 81%	(Cubed-3) K5 - 85% 1 st grade - 100% 2 nd grade – 71% Third grade – 76%	White – 69% Hispanic – 13% African American – 5% Asian – 1% Native American – 3% Other – 3%	Not reported	Students with known disability included in sample (6%)
OWLS-2 Oral and Written Language Scale	3;0-21;11	Not reported	Not reported	White – 54.9% Hispanic – 18.7% African American – 18.8% Asian – 2.3%	African American English	Not reported

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
PLS-5 Preschool Language Scale	Birth – 7;11	@ cut score of 85 – 91%	@ cut score of 85 – 78%	White – 60.7% Hispanic – 18% African American – 11.6% Asian – 4.0% Other – 5.7%	African American English, Appalachian English, Southern English, Spanish influenced English, Chinese influenced English	Students with known disability included in sample (6.2%) At a 90% confidence interval, the SEM is not appropriate as it falls between a 72 and 87.
PLS-5 (Spanish) Preschool Language Scale (Spanish)	Birth – 7;11	@1.5 SD - 93%	@ 1.5 SD - 71%			
PPVT-4 Peabody Picture Vocabulary Test	2;6-90+	Not reported	Not reported	White – 63.4% Hispanic – 15.4% African American – 15.1% Other – 6.1%	Not reported	Students with known disability included in sample (13.4%)
RESCA-E Receptive, Expressive, & Social Communication Assessment - Elementary	5-12	Not reported Test developer note – “not suitable for diagnosis”	Not reported	White – 77.22% African American – 13.09% Asian American – 3.88% Two or more – 5.09% American Indian/Alaska Native - <1% Native Hawaiian/ Pacific Islander - <1% Not Reported– <1%		

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
ROWPVT Receptive One Word Picture Vocabulary Test	2-80	Not reported	Not reported	Caucasian – 63.2% Hispanic – 18% African American – 12.8% Asian American – 3.4% Native American – 1% Other - <1% Not reported – 1.4%	Not reported	Students with known disability included in sample (8.7%)
REEL-3 Receptive-Expressive Emergent Language Scale - 3	Infants – age three	Not reported	Not reported	White 78%; Black 12%; Other 10%	Not reported	Students with known disability included in sample (7.0%)
SLDT: E (NU) Social Language Development Test - Elementary	6-11	@ cut score 90 – 82%	@ cut score 90 – 86%	White – 73% Hispanic – 13% African American – 9% Asian and other – 5%		
SLDT: A (NU) Social Language Development Test - Adolescent	12-18	@ cut score 90 – 71%	@ cut score 90 – 96%	White – 70% Hispanic – 13% African American – 14% Asian and other – 3%		
SPELT:P-2 Structured Photographic Expressive Language Test - Preschool	3;0 - 5;11	@ cut score 87 – 90%	@ cut score 87– 100%	White – 72.8% Hispanic – 8.6% African American – 12.5% Other – 6.1%	African American English	Students with known disability included in sample (2.5%)
SPELT – III Structured Photographic Expressive Language Test	4;0 – 9;11	@ cut score 95 – 90%	@ cut score 95 – 100%	White – 65.5% Hispanic – 11.2% African American – 16.1% Asian – 5.0% Other – 7.2%	African American English	Students with known disability included in sample (7%)

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
SUGAR Sampling Utterances and Grammatical Analysis Revised	3-7;11	@ 1 SD for MLU used with cut score of 1.25 SC for CPS – 97%	@ 1 SD for MLU used with cut score of 1.25 SC for CPS – 82.96%	White – 71.4% Hispanic – 7.4% African American – 9.3% Asian – 2.9% American Indian/Alaska Native – <1% Native Two or more – 7.8%	African American English, Southern English	
TEEM Test for Examining Expressive Morphology	3;0-7;11	@ 2 SD – 90%	@ 2 SD – 95%	Not reported	Not reported	
TILLS Test of Integrated Language and Literacy	6;0 – 18;11	(sample of cut scores – not full list – see manual) @ age 6;0-7;11 cut score 24 – 84% @ age 8;0-8;11 cut score 34 – 88% @ age 12;0-18;11 cut score 42 – 86%	(sample of cut scores – not full list – see manual) @ age 6;0-7;11 cut score 24 – 84% @ age 8;0-8;11 cut score 34 – 85% @ age 12;0-18;11 cut score 42 – 90%	White – 73% Hispanic – 10% African American – 10% Asian – 5.0% Native American – 1% Other – 1%	African American English, Spanish influenced English, Asian influenced English	
TNL-3 Test of Narrative Language	4;0-15;11	@ cut score 92 – 92%	@ cut score 92 – 92%	White – 78% Hispanic – 22% African American – 14% Asian Pacific Islander – 5.0% Two or more – 2% American Indian, Eskimo, Aleut – <1%	Not reported	Students with known disability included in sample (8%)

Name	Age range	Sensitivity	Specificity	Normative Sample	Dialectal Consideration	Additional Information
TOLD:P-5 Test of Oral Language Development - Primary	4;0-8;11	@ 1.5 SD – 88%	@ 1.5 SD – 88%	European American – 78% African American – 15% Asian Pacific Islander – 4.0% Two or more – 2% American Indian, Eskimo – 1% Other - <1%	Not reported	Students with known disability included in sample (15.1%)
TOLD:I-4 Test of Oral Language Development - Intermediate	8;0-17;11	@ cut score 90 – 77%	@ cut score 90 – 89%	European American – 78% African American – 14% Asian Pacific Islander – 5.0% Two or more – 2% American Indian, Eskimo – 1% Other - <1%	Not reported	Students with known disability included in sample (15.3%)

References

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of test for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*(2), 133-146.

Crowley, C., & Bucaj, E. (2023). Analysis of the CELF-5 for validity, reliability, and bias. *Leaders Project*. Available from: <https://www.leadersproject.org>.

Dollaghan, C. (2007). *The Handbook for Evidence-based Practice in Communication Disorders*. Baltimore, MD: Paul H. Brooks Publishing Co. <https://psycnet.apa.org/record/2007-03655-000>

Hendricks, A.E. & Adlof, S.M. (2017). Language assessment with children who speak nonmainstream dialects: Examining the effects of scoring modifications in norm-referenced assessment. *Language, Speech & Hearing Services in Schools, 48*(3):168-182.

IDEA: Individuals with disabilities education act. (2023). Individuals with Disabilities Education Act. Retrieved April 14, 2023, from <http://idea.ed.gov/>.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*, 34-42.

Paul, R., Norbury, C., & Gosse, C. (2018). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (5th ed.). Maryland Heights, MO: Elsevier/Mosby.