

Bilingual assessment evidence guide

Selecting and interpreting tools for Spanish and English

Choose a defensible use for one assessment result and identify the evidence still needed.

Prepare

Have the child's language history and the exact edition's manual available. Read the profile that matches your referral.

Plan your time

Allow about 10–15 minutes to review one candidate and complete the fit record. This editorial estimate excludes test administration and manual study.

Have ready

This guide, the full test manual or published protocol, referral question, and family or school examples.

Use one focused route

State the question → check population and administration fit → check the relevant accuracy study → select complementary evidence.

Record enough to decide

Edition, score, age and language fit, diagnostic sample, reference standard, accuracy uncertainty, and one next action.

Evidence behind the practice

Publisher facts describe intended use. Independent studies test specific applications. A meta-analysis informs dynamic assessment but does not validate our worksheets. [2–15]

Interpretation limit

No best-to-worst ranking applies to every child. If population fit or score validity is unresolved, do not treat that score as diagnostic evidence.

02 OPEN THE CASE

Start with the child in front of you

The useful question is not simply which test is best. It is what this result can tell you about this child.

Fictional training case

Mateo, age 6, uses Spanish with grandparents and English at school. His teacher reports short explanations. His family describes detailed Spanish stories. The referral asks whether language difficulties interfere with classroom participation.

Clinician to teacher Show me a recent task that was difficult. What did Mateo say, and what support helped?

Get an observable classroom demand.

Clinician to family Cuando Mateo le cuenta algo que pasó, ¿qué le resulta fácil y cuándo necesita ayuda?

Ask about strengths and support in a familiar language.

Build an evidence plan



What I would do next

Collect a Spanish narrative and an English explanation, documenting prompts and topic familiarity. Review language history before selecting a norm-referenced score. The two tasks answer related questions but are not equivalent tests. [1]

Read the profiles relevant to your referral, then complete “Match the tool to the question.” Use the source notes when checking a manual or study; you do not need to read every profile during a session.

Language measures at a glance

Ages are years;months. A population label alone does not establish that a particular score is interpretable.

BESA

Norm referenced | 4;0–6;11. Spanish–English bilingual children; apply manual language-profile rules. Normative **N = 756** in a manual audit. Publisher reports 17 Spanish and 7 English regional dialects; per-variety counts are not provided. [2, 3]

Use for: matched bilingual phonology, semantics and morphosyntax questions. **Check next:** the exact age, language, score and manual accuracy table. One sensitivity/specificity pair cannot describe every subtest and composite.

PLS 5 Spanish

Norm referenced | birth–7;11. Normative **N = 1,150** monolingual Spanish and Spanish-speaking bilingual children in the US and Puerto Rico. It is not intended for primarily English-speaking children with little Spanish. [4, 5]

Use for: early receptive and expressive language when the population fits. **Check next:** prescribed dual-language scoring and the Spanish diagnostic-validation table; the public brochure is incomplete for that purpose. Do not substitute English-edition accuracy.

CELF 4 Spanish US edition

Norm referenced | 5;0–21;11. US Spanish-speaking monolingual and bilingual individuals. CELF 5 Spain has a different population. [6]

Independent study: 656 low-income Spanish–English learners, ages 5;0–7;11, in English-only schools; diagnostic subsample initially 299, finally 28 with language impairment and 265 typical after six exclusions. At cutoff 85: sensitivity approximately 93–94%, specificity 65%. At study-derived 78: 86%, 80%. The abstract/Results differ on 94%/93%. [7]

Clinical move

Study classification used Spanish morphosyntax and language-sample criteria, with English grammar for the impairment group. Keep this reference standard attached to the estimates. The study-derived cutoff is not a universal replacement rule. [7]

Speech and early vocabulary

Keep the language, scoring group and communication question attached to each result.

BAPA

Norm referenced | 3;0–10;11. English-only, bilingual English, and Spanish/bilingual Spanish groups. Publisher normative **N = 438**, Central Texas. Among Spanish speakers, 95% reported Mexican Spanish; 90% of Spanish/bilingual groups attended dual-language programs. Language-group records overlap. [8]

Use for: matched speech-sound questions. **Check next:** exact scoring group and accuracy table. Publisher validation includes the norm group plus 86 additional students; 23 of those additional students were in bilingual/Spanish groups; generalization beyond this context needs care.

GFTA 3 Spanish

Norm referenced | 2;0–21;11. US Spanish edition. The public technical summary contains inconsistent age and geographic information; verify the full norm sample and intended comparison group in the Spanish manual. [9]

Publisher accuracy: Sonidos-en-palabras, score 85: sensitivity .91, specificity .88. Public summary omits accuracy-analysis N, reference standard and confidence intervals. Verify analyzed ages. **Add:** connected speech, intelligibility and dialect review; this Spanish articulation score is not a complete bilingual speech evaluation.

English Spanish Vocabulary Inventory

Caregiver report | toddler research samples. Three studies of **52, 13 and 22** children; caregivers predominantly of Mexican heritage in western US communities. These are research samples, not representative national norms. [10]

Use for: descriptive vocabulary across languages with knowledgeable caregivers. **Check next:** exact study and form. Diagnostic accuracy and norms require further research; no validated DLD cutoff is provided. Monolingual CDI norms cannot be applied to these scores.

A useful gap to record

“The available page does not report this value” is more accurate than assigning a poor-quality rating because a number is missing.

Sampling learning and individualized probes

These methods can answer questions that a standard score leaves open.

SALT bilingual story retell databases

Language sample | approximately 5;0–9;9, K–3; story-specific. Texas/California learners. Counts of 2,070, 1,667 and 930 are story samples, not unique children. Use reference comparisons only with matched story, elicitation, coding and comparison set. No universal diagnostic cutoff. [11]

SLAM

Language sample | preschool through high school materials. Choose the appropriate set and language support. Useful for narrative, explanation and inference. The resource page supplies materials and analysis guidance, not national norms or universal Spanish–English diagnostic accuracy. [12]

MAIN Spanish and English

Structured narrative | principally about 3–10 years. Match the language version and elicitation mode. Useful for production and comprehension across languages. The linked manual is not a US Spanish–English normative database; a study of one version does not validate every version. [13]

Published dynamic word learning protocol

Dynamic assessment | 4–5 years in this study. Kapantzoglou et al.: 28 predominantly Spanish-speaking bilingual children, 15 typical and 13 with primary language impairment. Classification evidence is protocol-specific; broader evidence has methodological limitations. Original activities do not inherit published accuracy. [14, 15]

Original criterion referenced probes

Individualized skill and access. Useful for baseline, target selection and progress monitoring when opportunities and scoring rules are defined. No norm or validation sample. Percent correct describes this task, not a diagnostic threshold.

Match the method to the next decision

Need a teaching target? Examine the exact breakdown. Need learning information? Document instruction and response. Need a reference comparison? Match the published protocol first.

See what specificity means

Keep normative sample size separate from the sample used to estimate diagnostic accuracy.

Two different questions

Sensitivity

Among children with the defined disorder, how many test positive?

Specificity

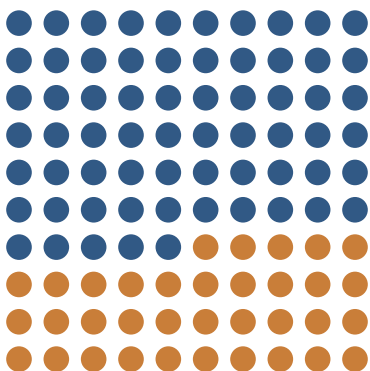
Among children without it, how many test negative?

Neither is the probability that an individual child has the disorder. Both depend on the cutoff, population and reference standard.

Make 65 percent specificity concrete

For illustration only, if specificity were exactly 65% in 100 children without the disorder, 65 would test negative and 35 would test positive. This is arithmetic using the CELF study estimate, not an additional study or a forecast for your caseload. [7]

An illustrative group of 100 children without the disorder



65 test negative

Correct negatives in this illustration

35 test positive

False positives in this illustration

One dot represents one child in this arithmetic example. This is not an observed sample or a prediction for your caseload.

07 WORKED EXAMPLE AND YOUR PLAN

Match the tool to the question

Use the filled example to make your own interpretation plan.

Fit record	Filled example for Mateo	Your child
Question	Explain events clearly at school	
Candidate use	Consider BESA at age 6; confirm language-profile and scoring rules	
Evidence still needed	Manual table, Spanish sample, classroom task and language history	
Decision now	Plan complementary evidence; no diagnosis from the profile alone	

Tool edition and score / manual page / population match

Cutoff / diagnostic sample / reference standard / confidence intervals or not reported

Administration changes and effect on score interpretation

Evidence that agrees or conflicts / next action

If fit is unresolved, gather descriptive evidence and verify the manual before using a standard score. If findings conflict, examine language experience, task demands and access, then collect the missing evidence. US school evaluation uses multiple sources and appropriate language or mode. [1, 16, 17]

Sources for language and speech measures

Product facts and independent research serve different purposes. Consult the exact manual before interpreting a score.

[1] American Speech-Language-Hearing Association. (n.d.). Multilingual service delivery in audiology and speech-language pathology. Practice Portal. <https://www.asha.org/practice-portal/professional-issues/multilingual-service-delivery/>

[2] Brookes Publishing. (n.d.). BESA product information and manual excerpt. Ages 4;0–6;11; Spanish and English subtests developed for their respective languages. <https://brookespublishing.com/product/besa/>

[3] Loveall, S. J., Channell, M. M., Mattie, L. J., & Barkhimer, A. E. (2022). Inclusion of individuals with neurodevelopmental disorders in norm-referenced language assessments. *Frontiers in Psychology*, 13, 929433. Table 1 reports BESA normative N = 756. <https://doi.org/10.3389/fpsyg.2022.929433>

[4] Pearson. (n.d.). Screen or assess emerging communication skills in English or Spanish with the PLS 5 family. Publisher brochure. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/pls5/pls-5-brochure.pdf>

[5] Pearson. (n.d.). Preschool Language Scales Fifth Edition Spanish. Product information. <https://www.pearsonassessments.com/store/en/usd/p/100000401.html>

[6] Pearson. (n.d.). Clinical Evaluation of Language Fundamentals Fourth Edition Spanish. Product information and Spanish edition comparison. <https://www.pearsonassessments.com/store/en/usd/p/100000436.html>

[7] Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals—Fourth Edition, Spanish. *Language, Speech, and Hearing Services in Schools*, 49, 292–305. https://doi.org/10.1044/2017_LSHSS-17-0013

[8] Smarty Ears. (n.d.). Bilingual Articulation and Phonology Assessment. Publisher standardization information. <https://www.smartyearsapps.com/bilingual-articulation-and-phonology-assessment/>

[9] Pearson. (n.d.). Goldman-Fristoe Test of Articulation Third Edition Spanish. Product information and publisher technical summary. <https://www.pearsonassessments.com/store/usd/p/100001920.html>
<https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/gfta-3/GFTA-3-Spanish-Technical-Summary.pdf>

Sources for complementary methods

Original examples and worksheets in this guide are clinical planning aids, not validated tests.

[10] De Anda, S., Cycyk, L. M., Moore, H., Huerta, L., Larson, A. L., & King, M. (2022). Psychometric properties of the English–Spanish Vocabulary Inventory in toddlers with and without early language delay. *Journal of Speech, Language, and Hearing Research*, 65(2), 672–691. https://doi.org/10.1044/2021_JSLHR-21-00240 Full article and supplement.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC9132146/>

[11] SALT Software. (n.d.). Bilingual Spanish English Story Retell reference databases. Database description and elicitation requirements. <https://www.saltsoftware.com/coursefiles/SelfCheck/shared/BilingualSEStoryRetellRDBDoc.pdf>

[12] LEADERSproject. (n.d.). School-Age Language Assessment Measures and guidelines for analysis.

<https://www.leadersproject.org/en/category/slam>

[13] Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN Multilingual Assessment Instrument for Narratives revised. *ZAS Papers in Linguistics*, 63. Spanish revised version 2020.

https://www.leibniz-zas.de/fileadmin/media/Dokumente/ZASPiL/ZASPiL_63/MAIN_Spanish_2020_Revised.pdf

[14] Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology*, 28, 1298–1317.

https://doi.org/10.1044/2019_AJSLP-18-0202

[15] Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools*, 43, 81–96.

[https://doi.org/10.1044/0161-1461\(2011/10-0095\)](https://doi.org/10.1044/0161-1461(2011/10-0095))

[16] U.S. Department of Education. (n.d.). IDEA regulations, 34 CFR §300.304, evaluation procedures.

<https://sites.ed.gov/idea/regs/b/d/300.304>

[17] American Speech-Language-Hearing Association. (n.d.). Collaborating with interpreters, transliterators, and translators. Practice Portal. <https://www.asha.org/practice-portal/professional-issues/collaborating-with-interpreters/>

Sources checked 26 September 2026. Proprietary stimuli and scoring tables are not reproduced. These profiles are a selective desk guide, not an exhaustive test review.