

Evidence of test validity refers to the degree to which specific data, research, or theory support that a test measures the concepts it purports to measure and is applicable to the intended population (AERA, APA, & NCME, 2014). There are multiple sources of information required in the process of test validation. Although different sources of evidence may represent different aspects of validity, these sources do not represent distinct types of validity.

The validity of a test is demonstrated by providing evidence to support the test’s interpretations and uses. Evidence of validity provided in the CELF®-5 Metalinguistics Technical Manual includes evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relationships with the prior version of the test and with CELF-5, and evidence based on special group studies, including students who speak English as a second language and students diagnosed as having a language disorder or autism spectrum disorder.

Sensitivity and specificity are diagnostic validity statistics that describe how a test performs. Sensitivity tells us the probability that someone who has the condition will test positive for it. Specificity tells us the probability that someone who does not have the condition will test negative. Sensitivity and specificity provide overall summary statistics of how well the test can classify, although this overall summary can be misleading for specific base rates. For example, a test might have both high sensitivity and specificity, yet still have a large false positive rate.

The table in the next column provides the classification table for children with a language disorder based on cut scores of 1, 1.5, and 2 SDs below the mean for the Total Metalinguistics Index, the Meta-Pragmatics Index,

and the Meta-Semantics Index scores. The table also reports diagnostic validity statistics and adjusted PPPs based on different base rates. The results indicate fair to excellent sensitivity and specificity at -1, -1.5, and -2 SD.

Total Metalinguistics, Meta-Pragmatics, or Meta-Semantics Index	Standard Cut Score	Sensitivity	Specificity
-1 SD	85	.96	.78
-1.5 SD	77	.74	.93
-2 SD	70	.31	.96

For detailed information about evidence of validity for CELF-5 Metalinguistics and this study, see the CELF-5 Metalinguistics Technical Manual.

So, what do these values mean?

At -1 SD, you will have excellent sensitivity—you will be able to confidently identify students having difficulty with higher level language skills. You may also over-identify some typically developing students (22%) whose skills are in the low-average range.

At -1.5 SD, you may miss up to 26% of students having difficulty using higher level language skills, such as students with mild language difficulties, but you will correctly identify 93% as not having a disorder.

At -2 SD, the criterion is set so low that you will only identify the students with severe to profound impairments, and you will not identify students with mild or moderate difficulties with higher level language skills.

Always remember that a complete assessment process requires that you use a variety of tools to evaluate a student’s skills across multiple contexts. Diagnostic decisions should never be made on the basis of performance on a standardised tool alone.

For more information about CELF-5 Metalinguistics, please visit [Pearsonclinical.co.uk/meta](https://pearsonclinical.co.uk/meta)

Pearsonclinical.co.uk