

Determining the Diagnostic Accuracy of Tests of Children's Language

Tammie Spaulding, Ph.D., CCC-SLP
University of Connecticut

Luanne Andersson, Ph.D., CCC-SLP
Ithaca College

Overview of this Session

- **Background information**
- **Reporting of diagnostic accuracy in test manuals**
- **Influences on diagnostic accuracy**
- **Practice determining accuracy of selected tests**
- **Wrap up**

What is Diagnostic Accuracy?

Diagnostic accuracy refers to the **accuracy with which scores obtained on a test classify children** as having language impairment (LI) or having typically developing language (TLD).

Why is Diagnostic Accuracy Important?

Clinicians use test results as part of the diagnostic process

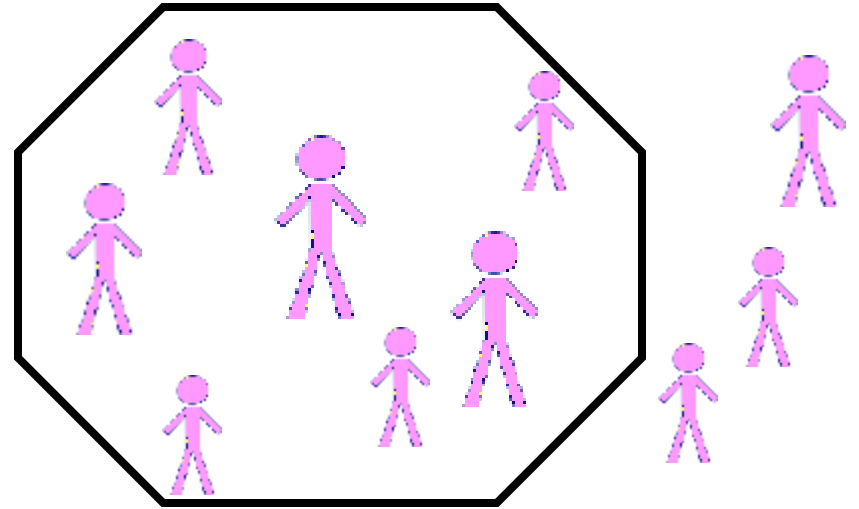
- We want our diagnoses to be accurate

Clinicians use tests to determine eligibility for service

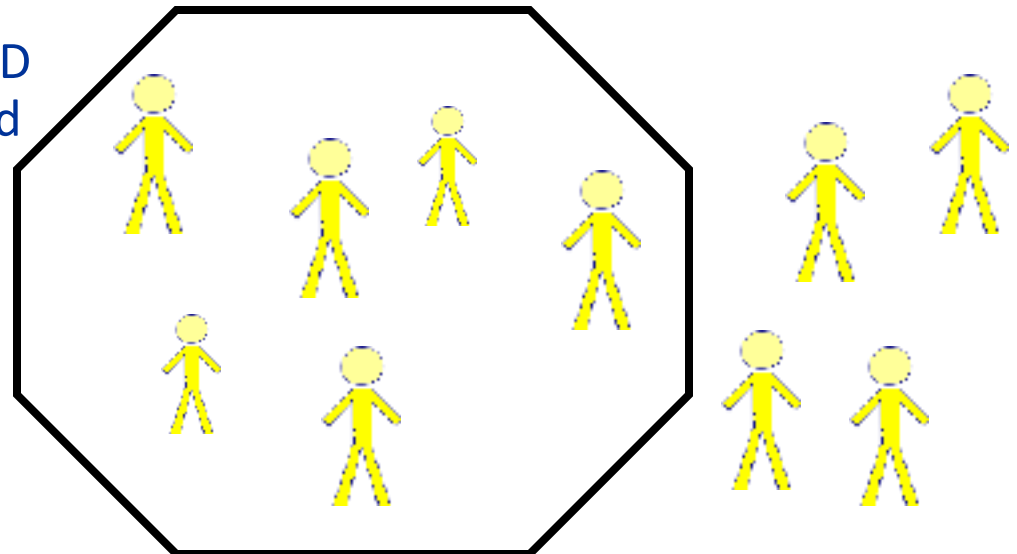
- We want to provide services to children with LI
- We do not want to provide services to children with TLD

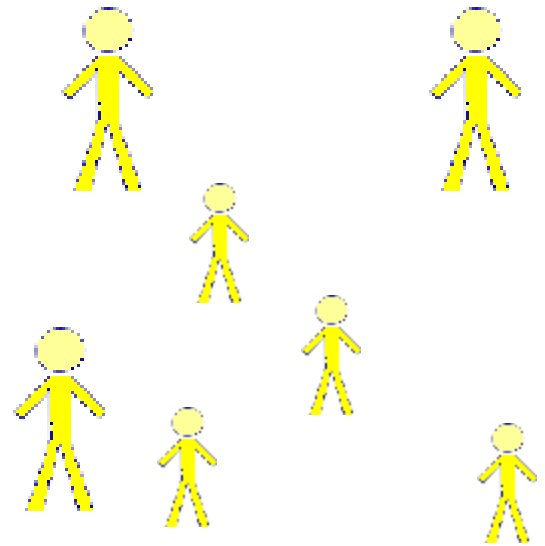
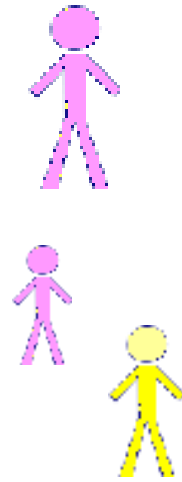
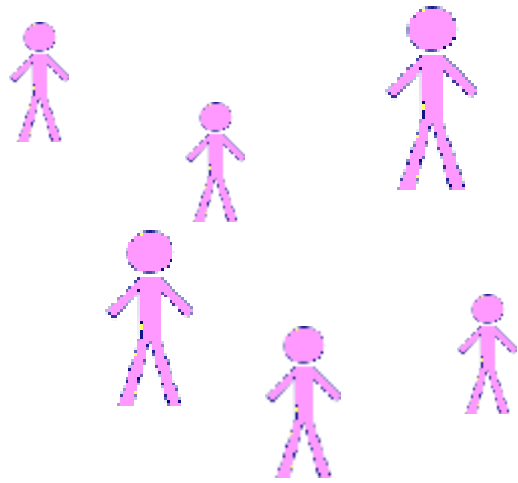
Diagnostic Accuracy as Reported in Test Manuals

- **Impaired diagnosed as Impaired**
 - **Sensitivity**-proportion of LI children correctly diagnosed as LI by the test



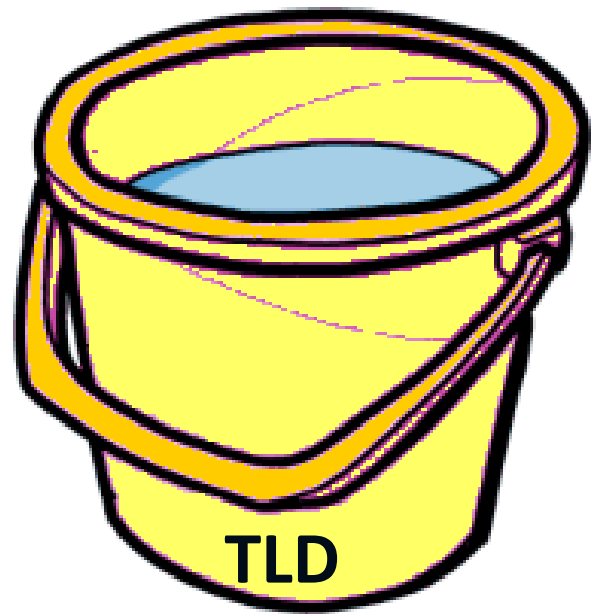
- **Normal diagnosed as Normal**
 - **Specificity**-proportion of TLD children correctly diagnosed as TLD





LI

SENSITIVITY

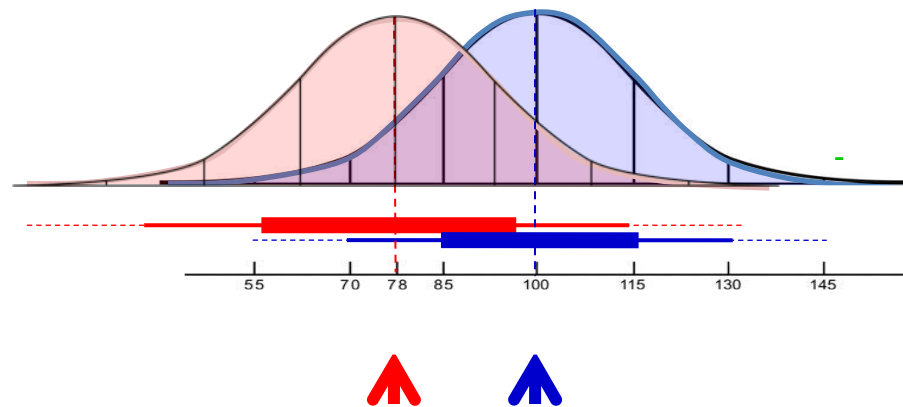


TLD

SPECIFICITY

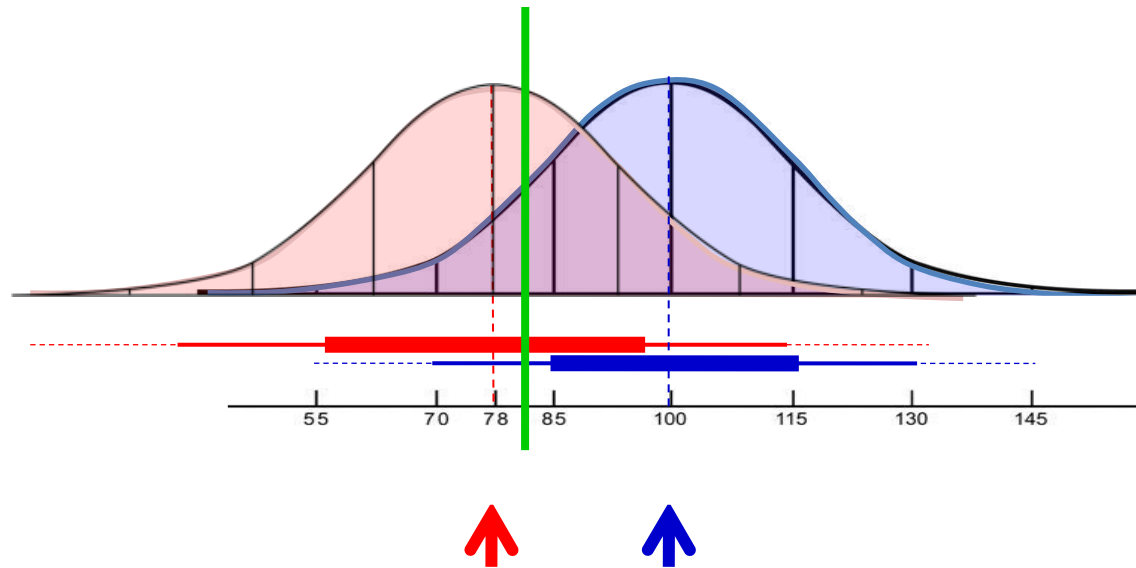
Two Influences on Sensitivity and Specificity

1. The **distribution (mean and sd)** of
 - a. the children with LI
 - b. the children with TLD



Two Influences on Sensitivity and Specificity

2. The cut score



Refresher: The Normal Distribution

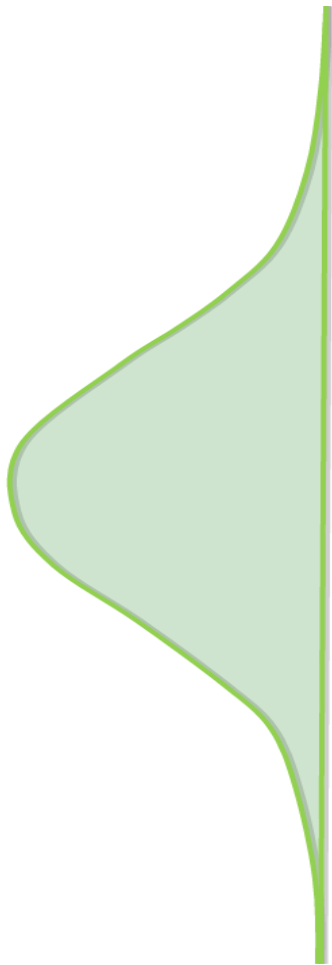
Test scores for the normative sample create a distribution.

(normal distribution = normal curve = bell curve)



Assumption: We are referring to tests with standard scores having a mean of 100 (with a standard deviation of 15).

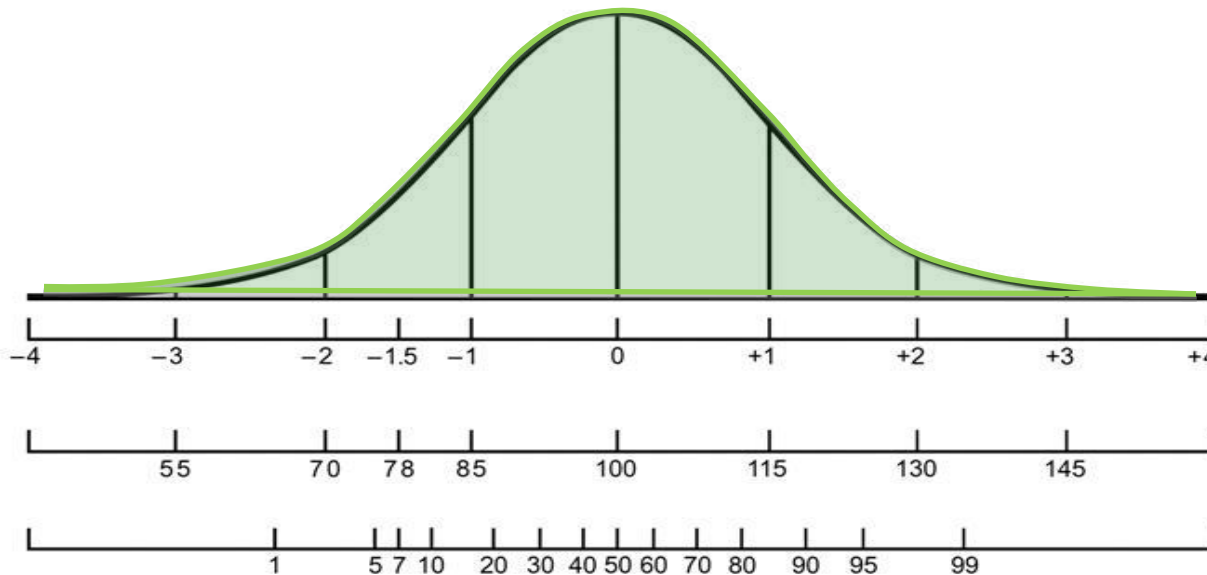
Refresher: Relationships among scores based on the normal distribution



SD	Standard Score (SD 15)	% rank
+ 3	145	> 99
+ 2	130	98
+ 1	115	84
0	100	50
- 1	85	16
- 1.5	78	7
- 2	70	2
- 3	55	< 1

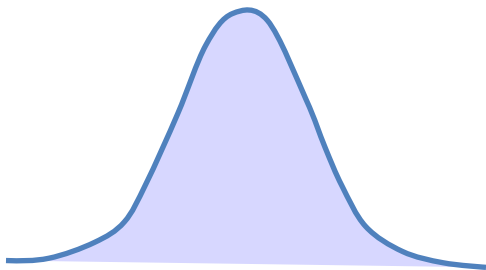
Refresher: The Normal Distribution

In a nutshell, the lower the standard score, the fewer the children who received a score at or below that score.

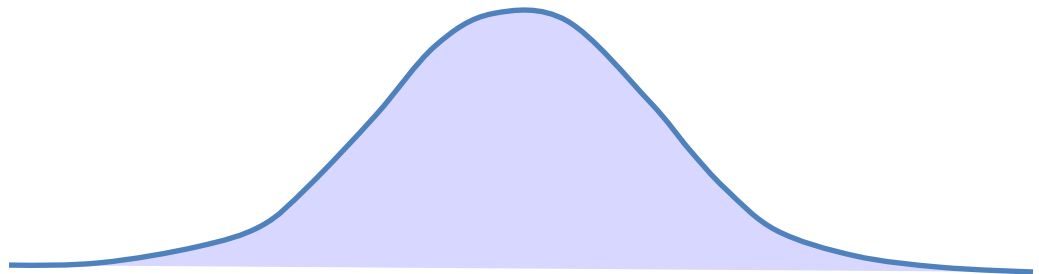


Refresher: The Normal Distribution

Not all distributions are normal.

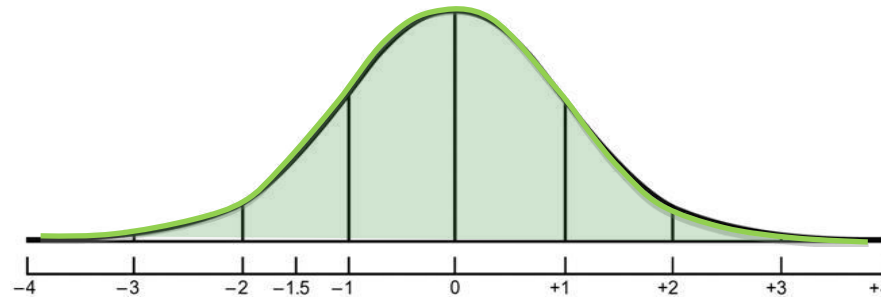


**Distributions with
a small SD
are narrower.**

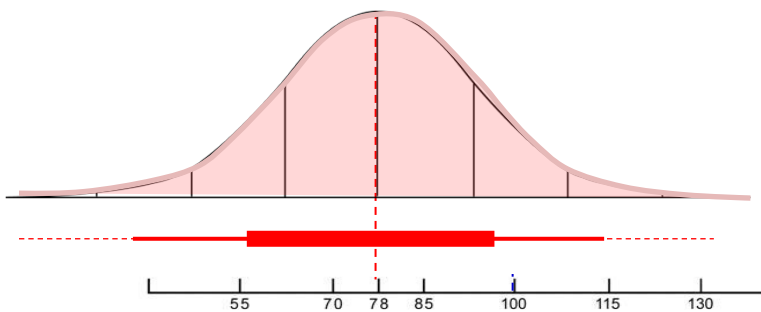


**Distributions with a large SD
are wider.**

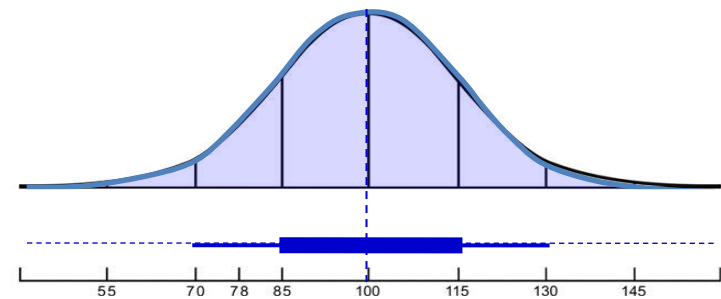
Three Distributions Important in Determining Diagnostic Accuracy



Normative Sample:
Results used to develop a test's conversion table

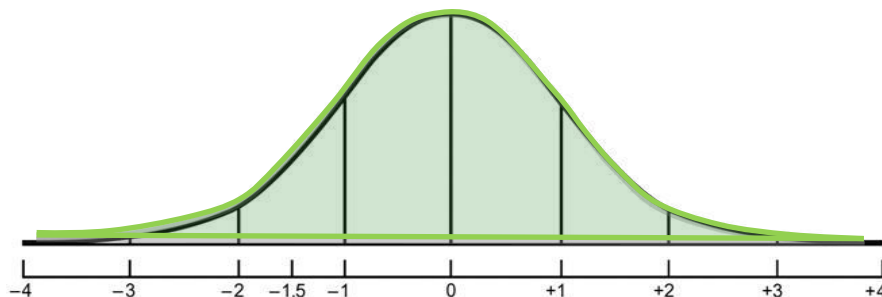


Children with LI

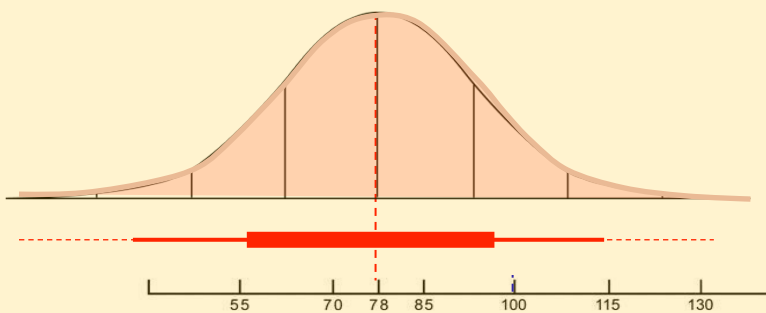


Children with TLD

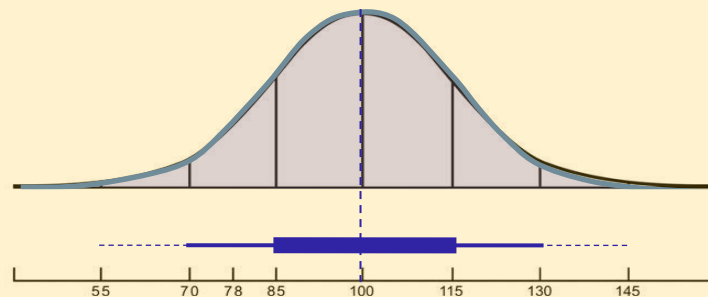
Three Distributions Important in Determining Diagnostic Accuracy



Normative Sample



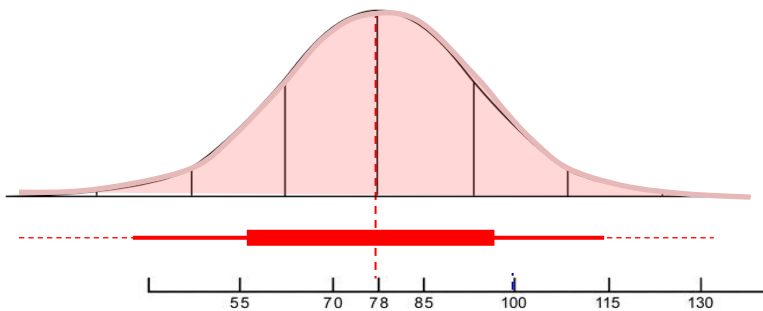
Children with LI



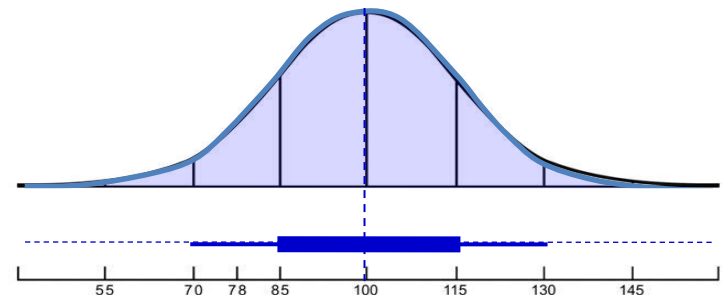
Children with TLD

Influence on Sensitivity and Specificity # 1

- The relationship between the distributions of the children with LI and the children with TLD.



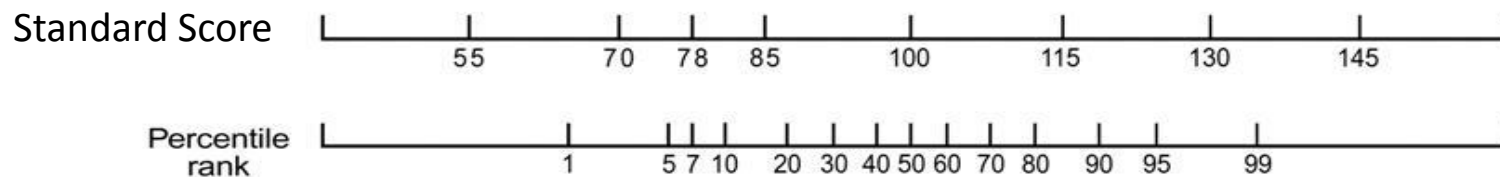
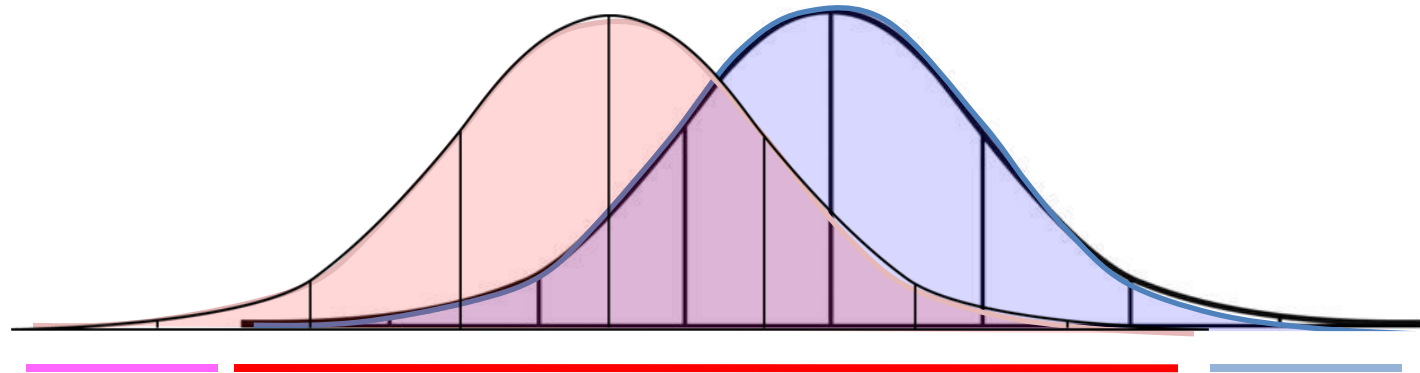
Children with LI



Children with TLD

Relationship between the Distributions

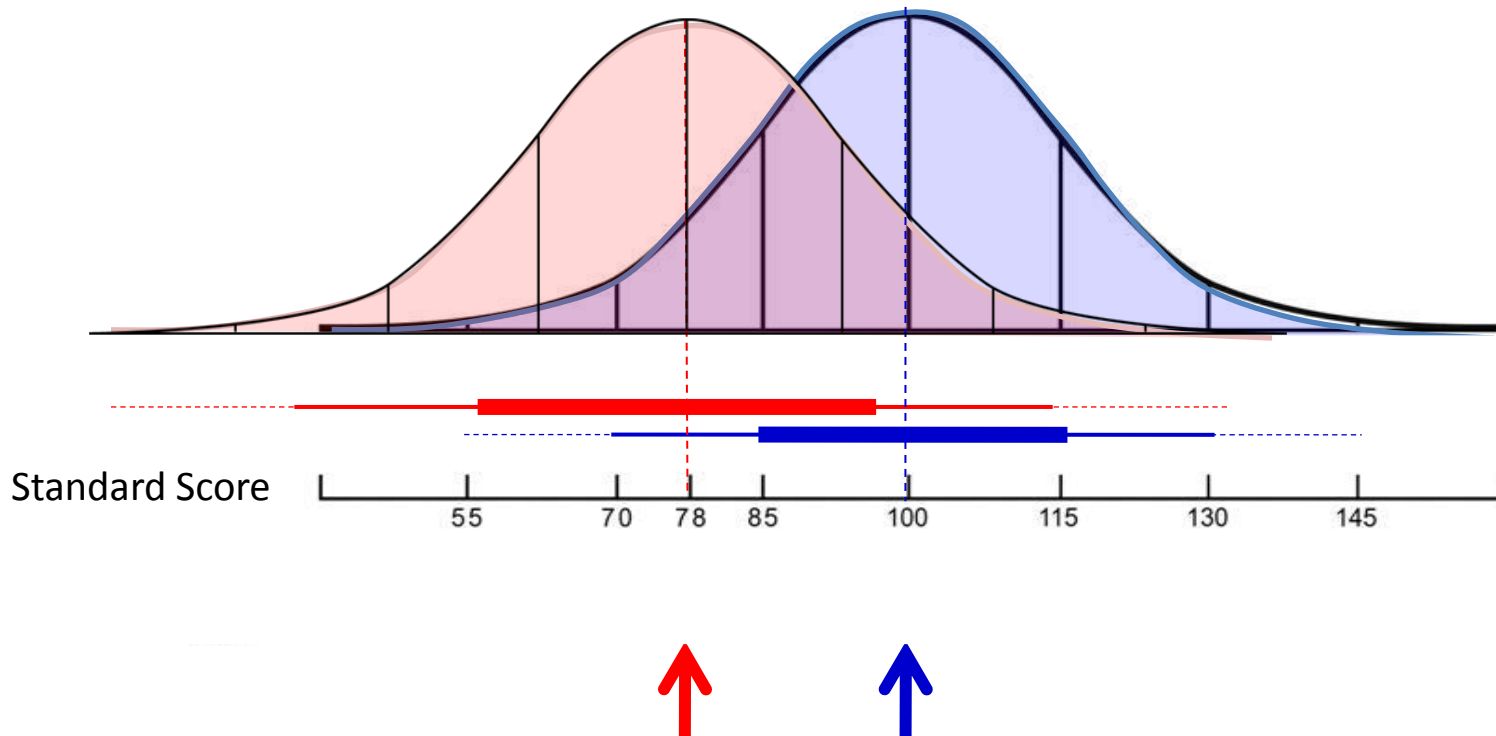
When the distributions overlap, any standard scores that are achieved by both groups of children (by children with LI as well as by children with TLD) are problematic.



The Relationship between the Two Distributions

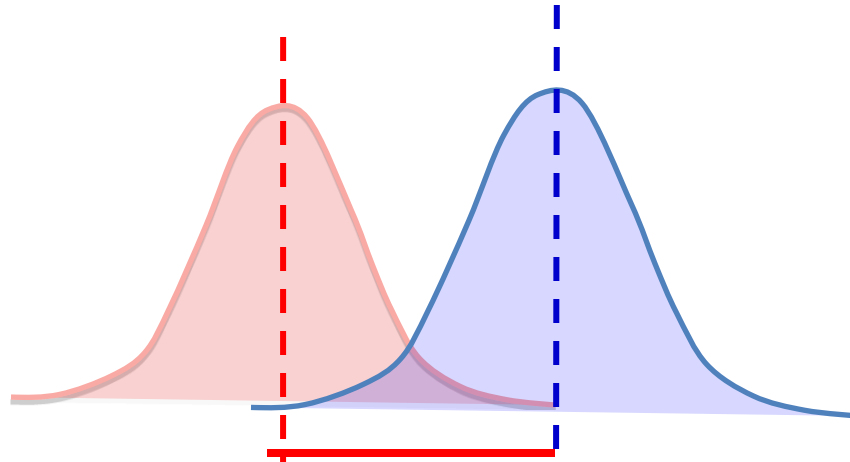
The relationship between the distributions is affected by

- the amount of **variation** in each distribution
- the **distance between the means** of the two groups

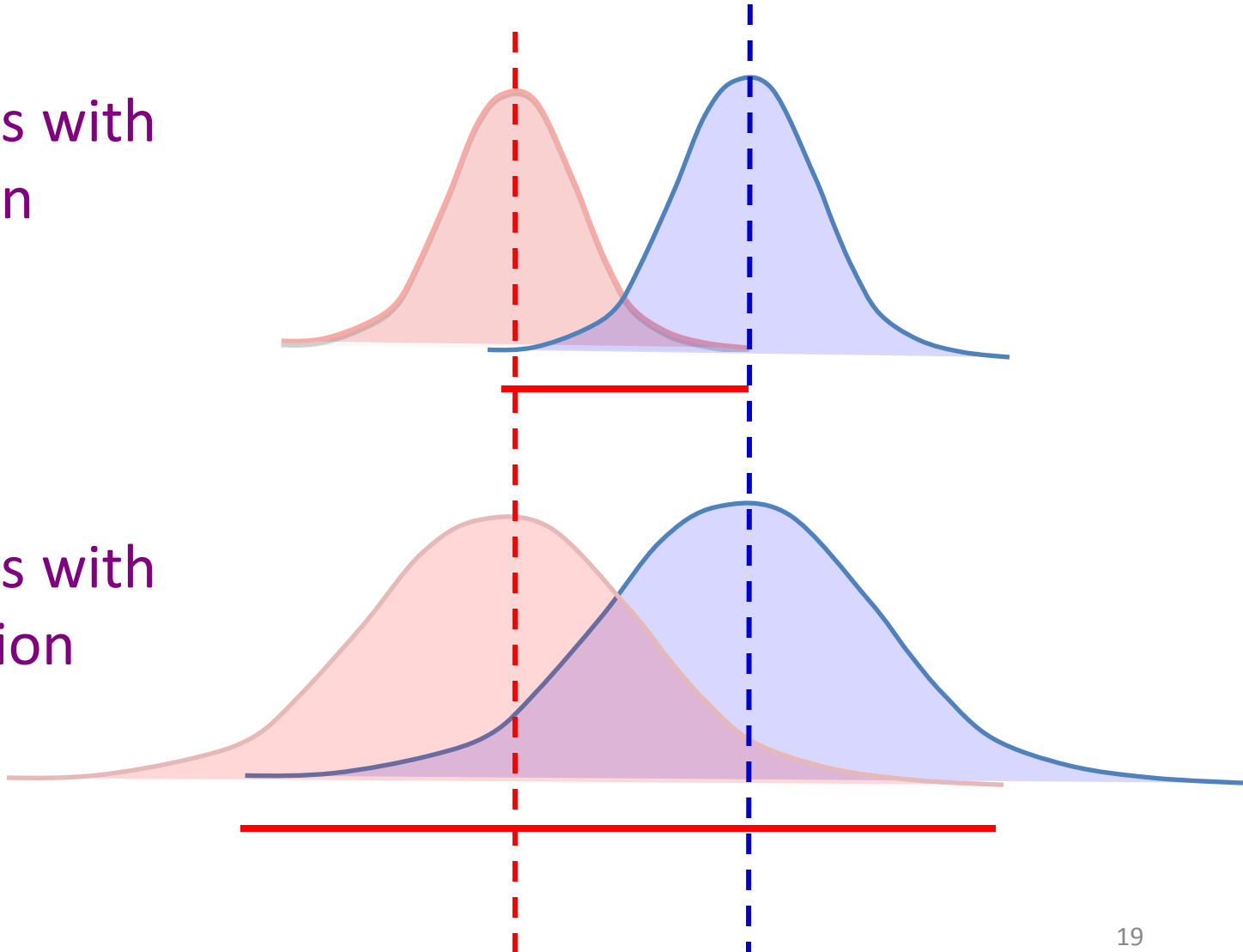


Variation in the Distributions

Distributions with
less variation
($SD < 15$)

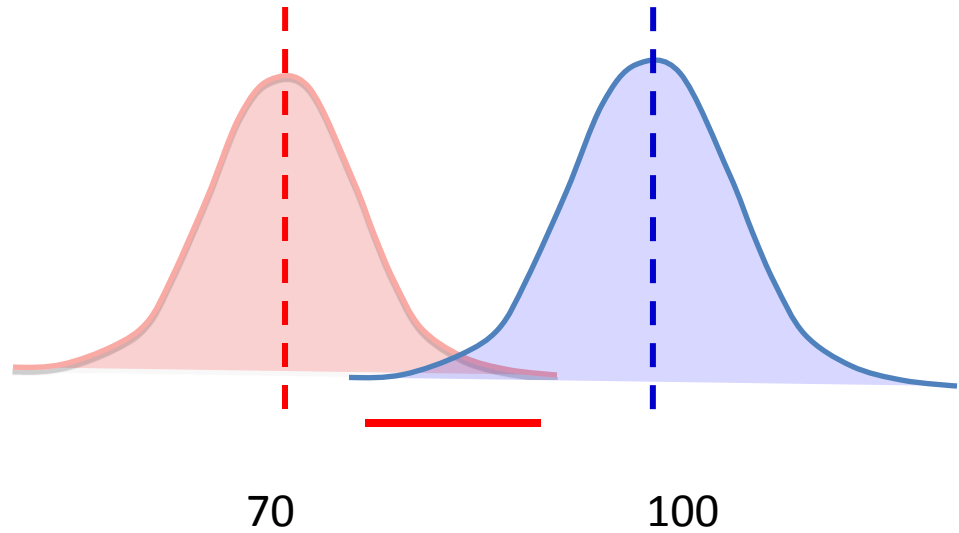


Distributions with
more variation
($SD > 15$)

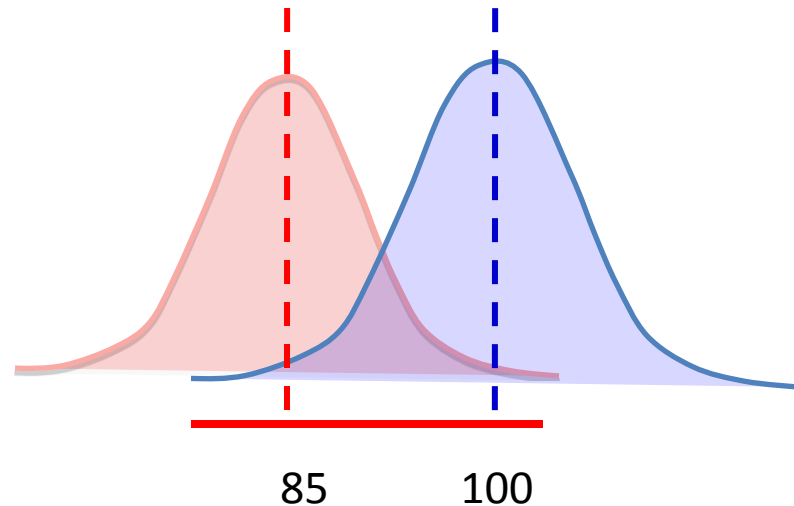


Distance between Group Means

Large distance
between means



Small distance
between means



Relationship between the Distributions

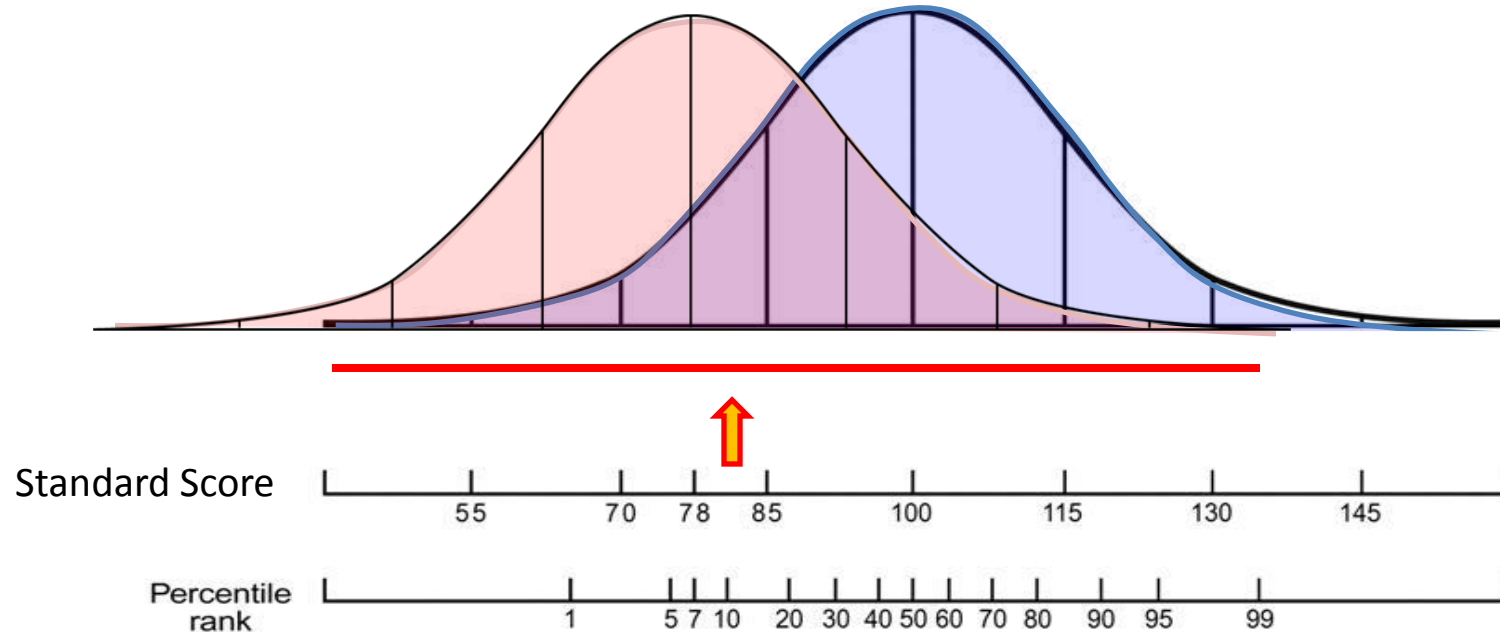
Overlapping distributions are problematic.

If a child receives a standard score of 80 -

Does the child have language impairment?

Or

Does the child have typically developing language?



Relationship between the Distributions

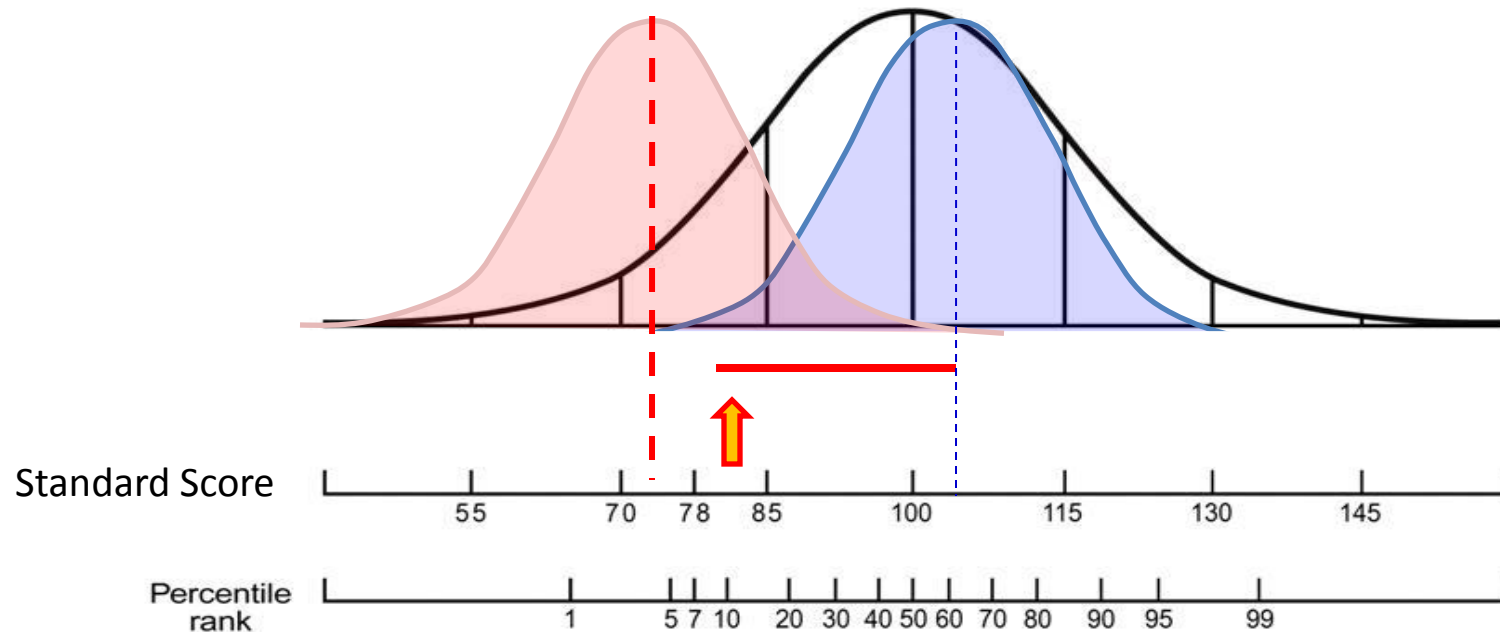
Overlapping distributions are problematic.

If a child receives a standard score of 80 -

Does the child have language impairment?

Or

Does the child have typically developing language?



Influence on Sensitivity and Specificity # 2

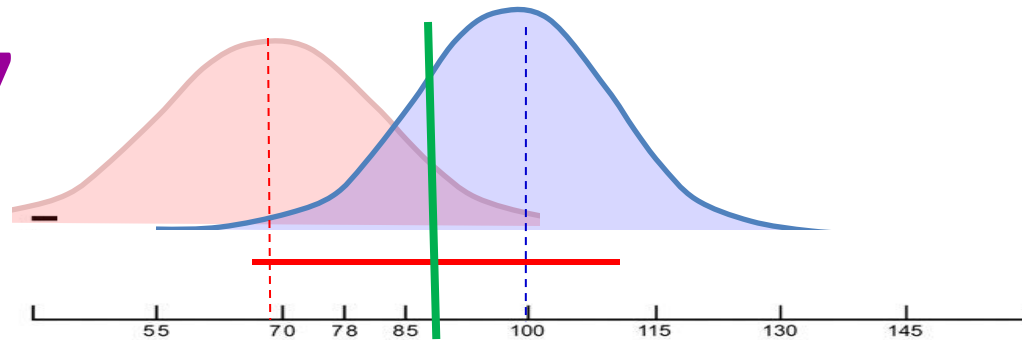
The cut score

- The location of the cut score impacts the child's diagnosis.

Influence # 2: The cut score

Example 1:

Cut score = 87

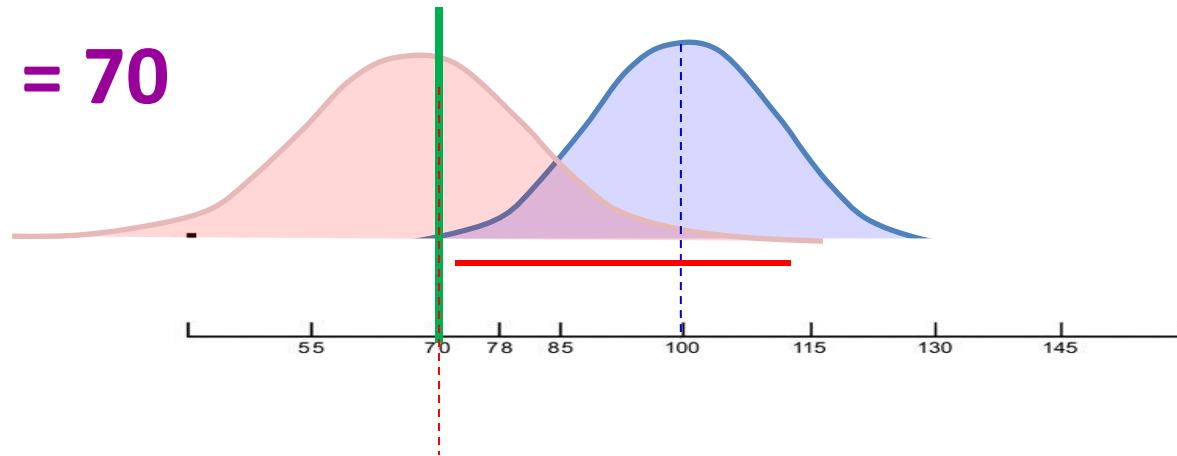


A child who receives a standard score of 87 or less will be identified as having LI.

Influence # 2: The cut score

Example 2:

Cut score = 70



A child with LI who receives a standard score of 70 or lower will be identified as having LI.

Reporting of Information Related to Diagnostic Accuracy in Test Manuals

Information Provided in Test Manuals

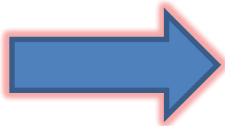
- 1. The diagnostic accuracy table**
- 2. Sensitivity/Specificity**

Four Tests that Include Sensitivity and Specificity

- **PLS-4** (Preschool Language Scale – 4)
- **CELF-4** (Clinical Evaluation of Language Fundamentals – 4)
- **CELF-P2** (Clinical Evaluation of Language Fundamentals – Preschool)
- **DELV-NR** (Diagnostic Evaluation of Language Variation – Norm Referenced)

Diagnostic Accuracy in the PLS-4

Image removed due to copyright restrictions



Diagnostic Accuracy Table

		Existing Diagnosis	
		Language Impairment (LI)	Typically Developing Language (TLD)
TEST	Language Impairment (LI)	A	B
	Typically Developing Language (TLD)	C	D

Example: Diagnostic Accuracy of the PLS-4

(Zimmerman, Steiner, and Pond, 2002)

Children ages
3:0 to 5:11
(Table 7.13, p. 214)

		Existing Diagnosis	
		LI	TLD
PLS-4 (total test score)	LI	60	9
	TLD	15	66
	Total	75	75

Cut Score of 85 (found on page 212 in the PLS-4 manual).

Example: Diagnostic Accuracy of the PLS-4

(Zimmerman, Steiner, and Pond, 2002)

Children ages
3:0 to 5:11
(Table 7.13, p. 214)

		Existing Diagnosis	
		LI	TLD
PLS-4 (total test score)	LI	60	9
	TLD	15	66

Calculating Sensitivity

Example: Diagnostic Accuracy of the PLS-4

(Zimmerman, Steiner, and Pond, 2002)

Children ages
3:0 to 5:11
(Table 7.13, p. 214)

		Existing Diagnosis	
		LI	TLD
PLS-4 (total test score)	LI	60	9
	TLD	15	66

Calculating Sensitivity

PLS-4 Sensitivity

		Existing Diagnosis	
		LI	TLD
<u>TEST</u>	LI	60	
	TLD	15	

Calculating Sensitivity: Determine the percentage of children with an existing diagnosis of LI who also received a diagnosis of LI on the test under investigation.

$$\text{Sensitivity} = A/(A+C)$$

PLS-4 Sensitivity

		Existing Diagnosis	
		LI	TLD
PLS-4 <small>Total test</small>	LI	60	9
	TLD	15	66

• Total Test Sensitivity .80 [60/(60 + 15)]

80% of the children with an existing diagnosis of LI were also identified as having LI using the PLS-4.

Cut Score of 85 (found on page 212 in the PLS-4 manual).

Example: Diagnostic Accuracy of the PLS-4

(Zimmerman, Steiner, and Pond, 2002)

Children ages
3:0 to 5:11
(Table 7.13, p. 214)

		Existing Diagnosis	
		LI	TLD
PLS-4 (total test score)	LI	60	9
	TLD	15	66

Calculating Specificity

Example: Diagnostic Accuracy of the PLS-4

(Zimmerman, Steiner, and Pond, 2002)

Children ages
3:0 to 5:11
(Table 7.13, p. 214)

		Existing Diagnosis	
		LI	TLD
PLS-4 (total test score)	LI	60	9
	TLD	15	66

Calculating Specificity

PLS-4 Specificity

		Existing Diagnosis	
		LI	TLD
<u>TEST</u>	LI		9
	TLD		66

Calculating Specificity: Determine the percentage of children with an existing diagnosis of TLD who also received a diagnosis of TLD on the test under investigation.

$$\text{Specificity} = D / (D + B)$$

PLS-4 Specificity

		Existing Diagnosis	
		LI	TLD
PLS-4 Total test	LI	60	9
	TLD	15	66

• Total Test Specificity .88 [66/ (66 + 9)]

88% of the children with an existing diagnosis of TLD were also identified as having TLD using the PLS-4.

Cut Score of 85 (found on page 212 in the test manual).

Deciding what Level of Diagnostic Accuracy is Acceptable

- **Clinical importance** of test's ability to discriminate
- Are the **social values and social consequences** of mis-identification serious?
 - TLD misdiagnosed as LI
 - LI misdiagnosed as TLD
- Accurate discrimination should have relatively **stringent guidelines**

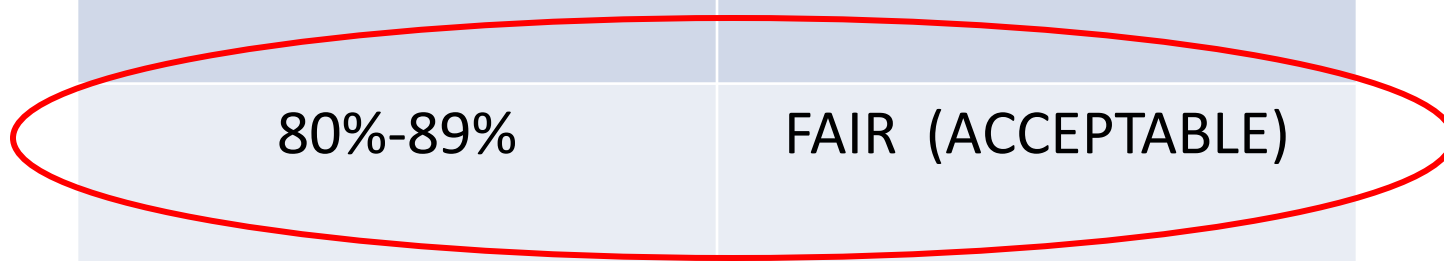
Recommended levels of Sensitivity and Specificity

(Plante & Vance, 2004)

SENSITIVITY & SPECIFICITY	INTERPRETATION OF DIAGNOSTIC ACCURACY
≥90%	GOOD
80%-89%	FAIR (ACCEPTABLE)
<80%	UNACCEPTABLE

Results for the PLS-4

SENSITIVITY & SPECIFICITY	INTERPRETATION OF DIAGNOSTIC ACCURACY
≥90%	GOOD
80%-89%	FAIR (ACCEPTABLE)
<80%	UNACCEPTABLE



PRACTICE: Locating Sensitivity and Specificity Information

Use the PLS-4

- Find the sensitivity and specificity for children ages **4:0 to 4:11**.
- How does it compare to the results for all ages collapsed (3:0 to 5:11)?

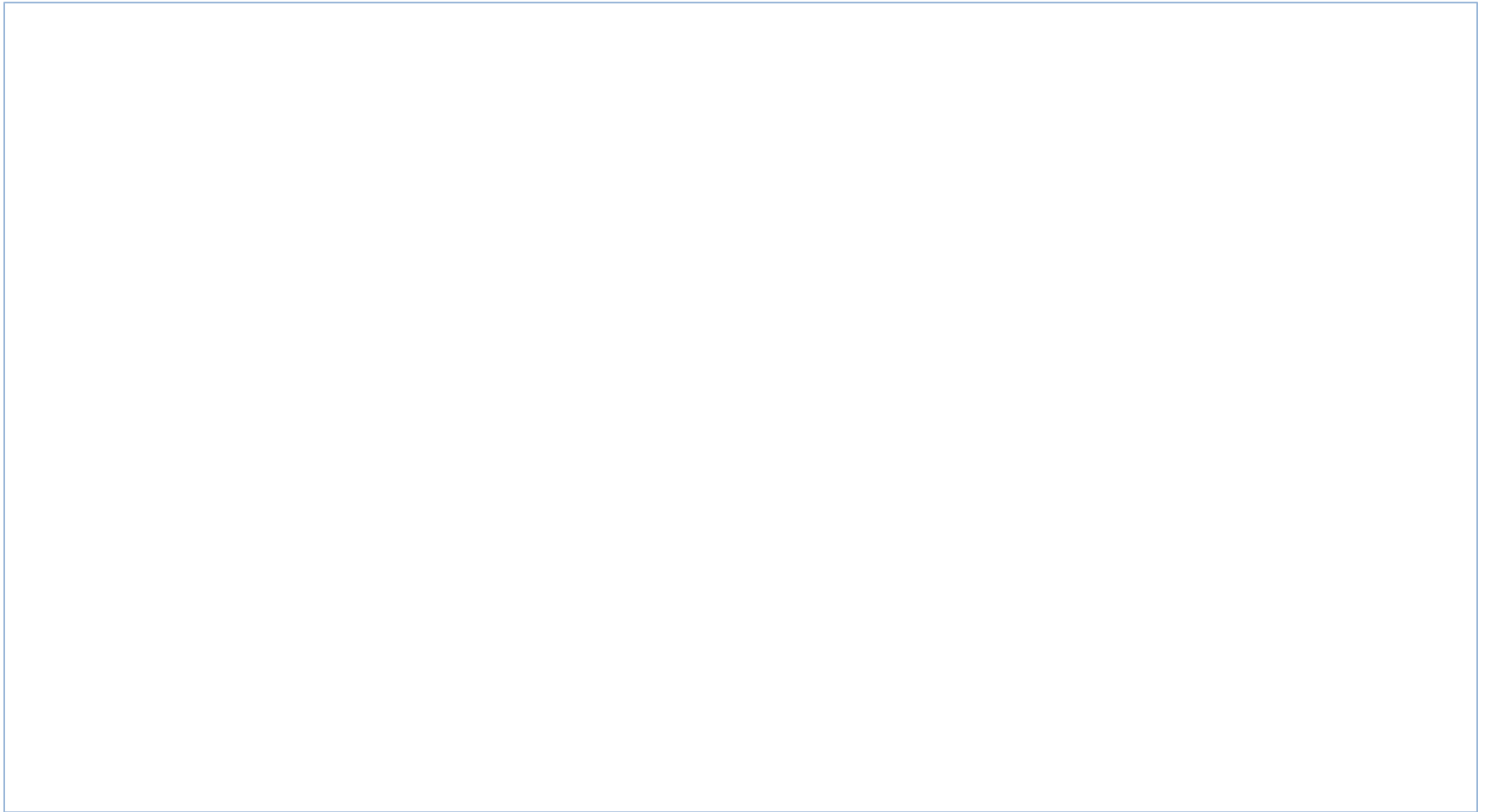
PLS-4 page 213 Image removed due to copyright restrictions



PLS-4 Review

- Consistent -1.0SD cut-off
- Different sensitivity and specificity based on age of child (3.0-3.11; 4.0-4.11; 5.0-5.11)
- Different sensitivity and specificity based on whether you are interested in
 - » Total Language Score
 - » Auditory Comprehension Score
 - » Expressive Communication Score

Changes in the Cut Score Impact Sensitivity and Specificity

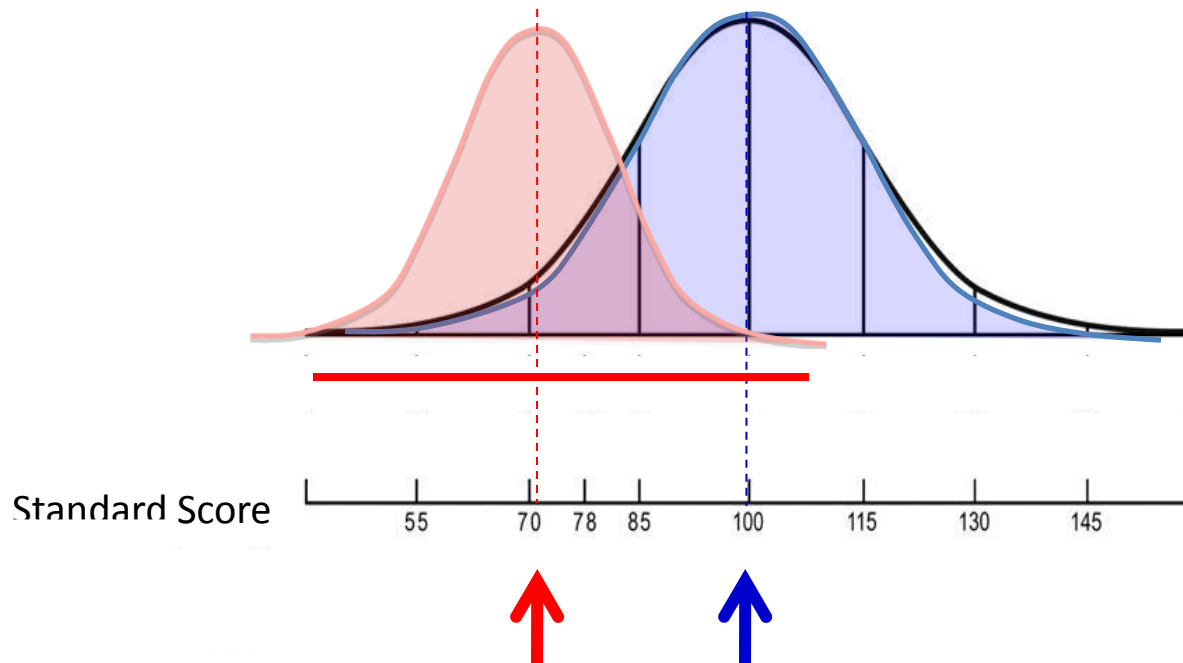


Influence of the distributions

DELV-NR

The children with LI have mean of 72.7 with an SD of 9.9.

The children with TLD have a mean of 100.1 with an SD of 12.4.



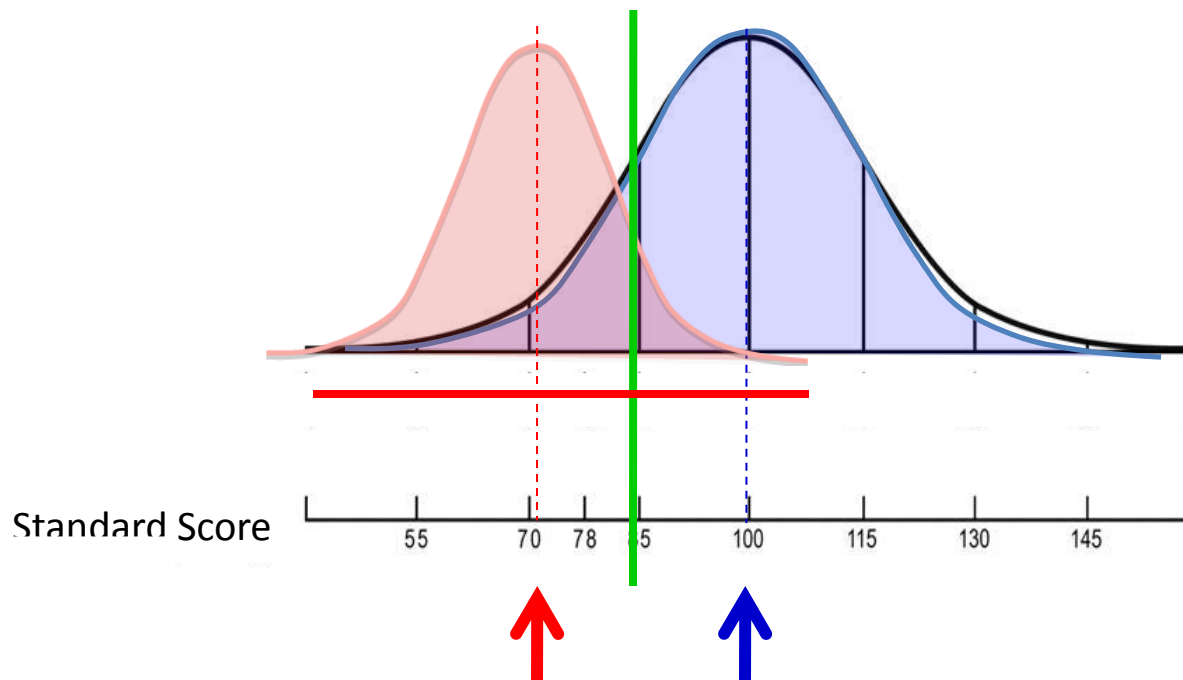
Influence of the distributions

DELV-NR

The children with LI have **mean of 72.7** with an **SD of 9.9**.

The children with TLD have a **mean of 100.1** with an **SD of 12.4**.

Cut Score = - 1 SD



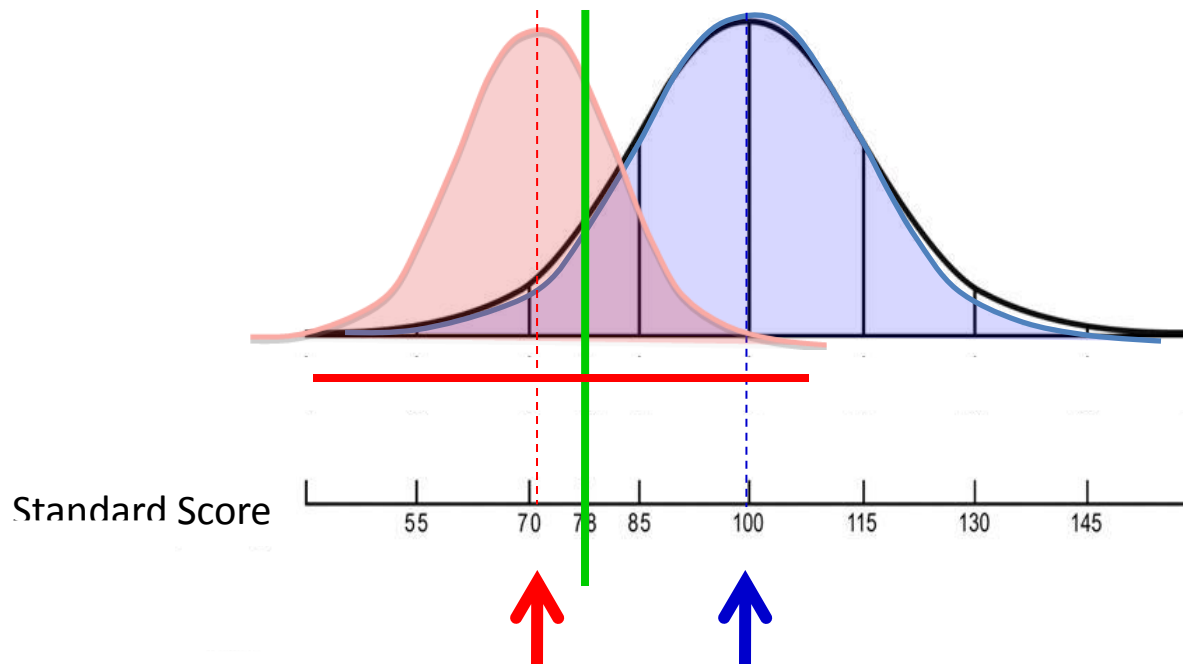
Influence of the distributions

DELV-NR

The children with LI have **mean of 72.7** with an **SD of 9.9**.

The children with TLD have a **mean of 100.1** with an **SD of 12.4**.

Cut score = - 1.5 SD



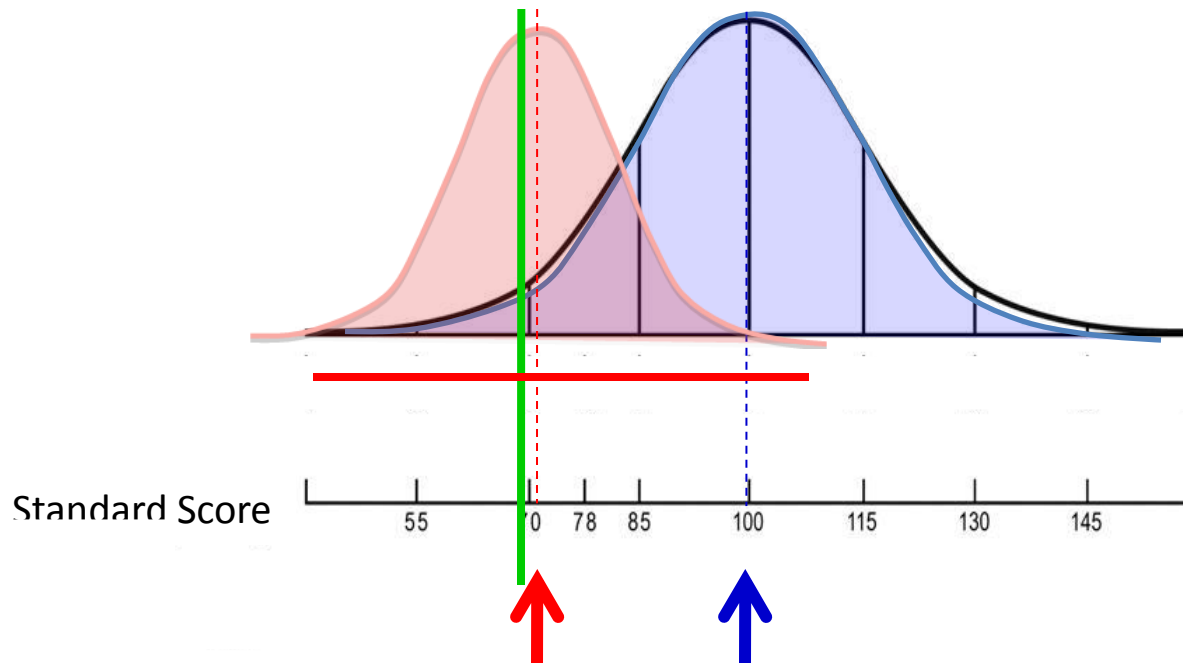
Influence of the distributions

DELV-NR

The children with LI have **mean of 72.7** with an **SD of 9.9**.

The children with TLD have a **mean of 100.1** with an **SD of 12.4**.

Cut score = - 2 SD



Influence of the Distributions Combined with Changes in the Cut Score: DELV-NR

- What happened to diagnostic accuracy as the cut score moved farther from the mean?

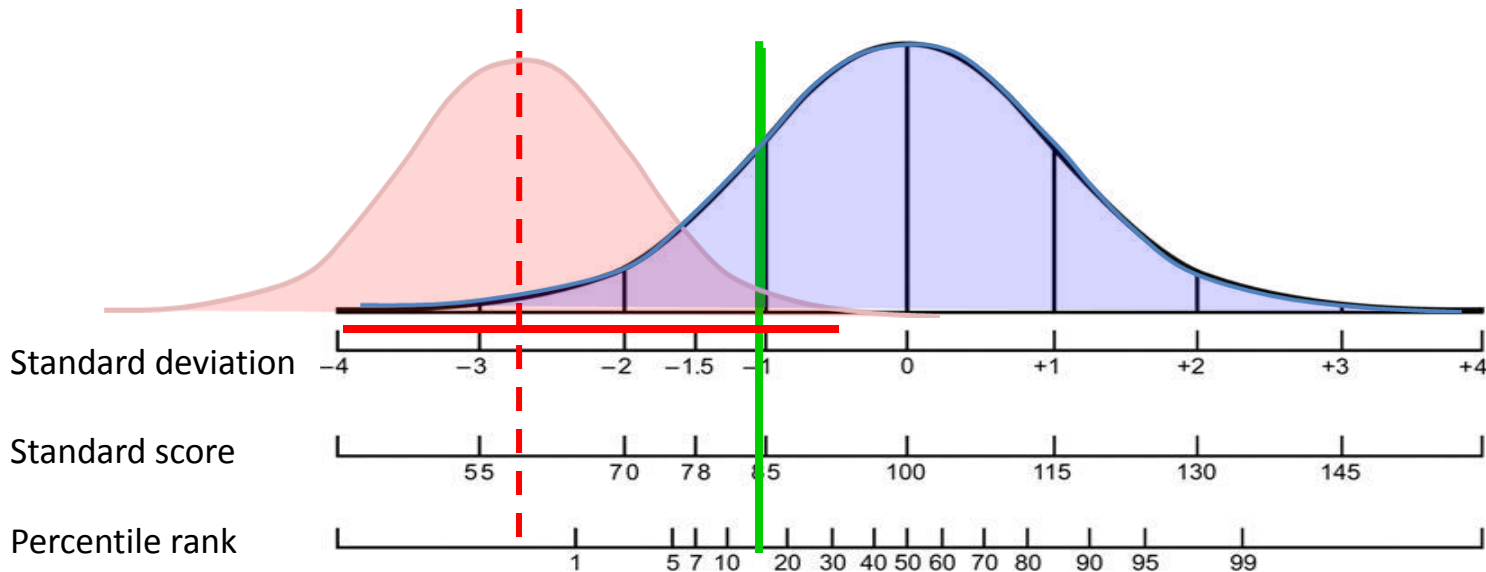
	- 1 SD	- 1.5 SD	- 2 SD
Sensitivity decreased	.95	.69	.36
Specificity increased	.93	.99	1.0

(Table 7.11, p. 140)

PRACTICE: The influence of the cut score

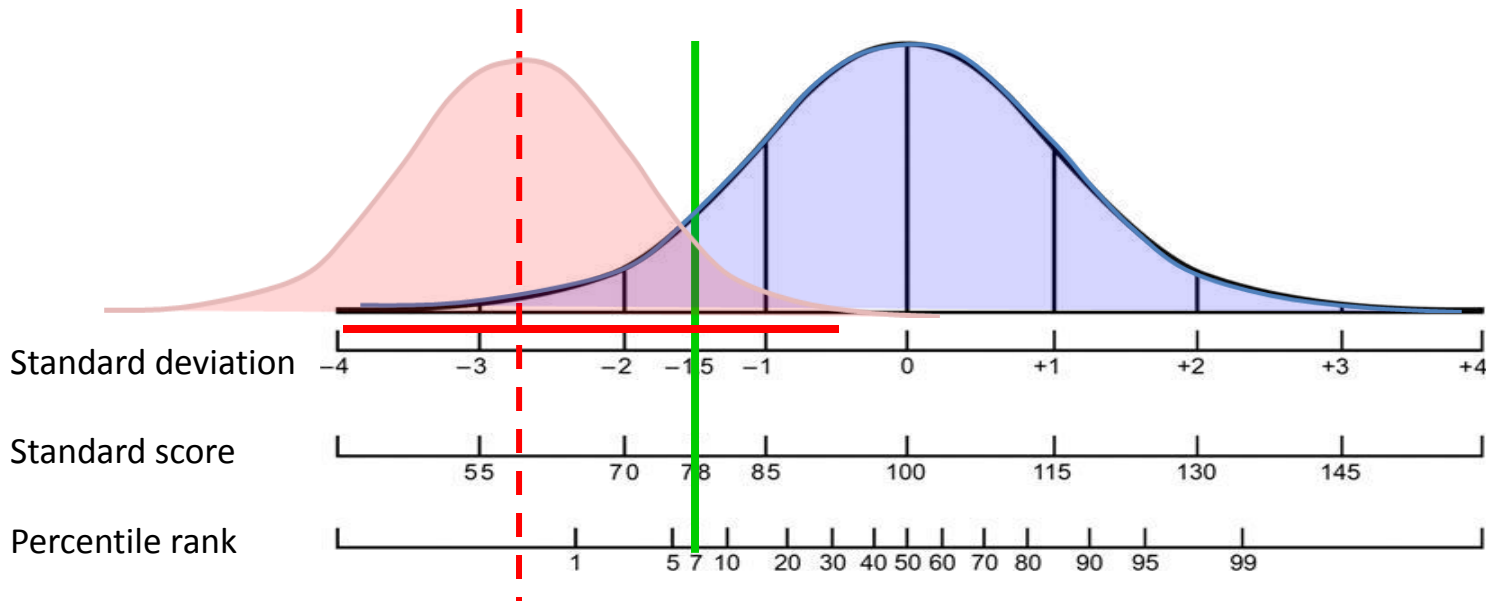
- Use the **CELF-4**
- Find the classification table (Table 8.18)
- As the cut score moves farther from the mean
 - what happens to the sensitivity?
 - what happens to the specificity?

The influence of the Cut Score: CELF-4



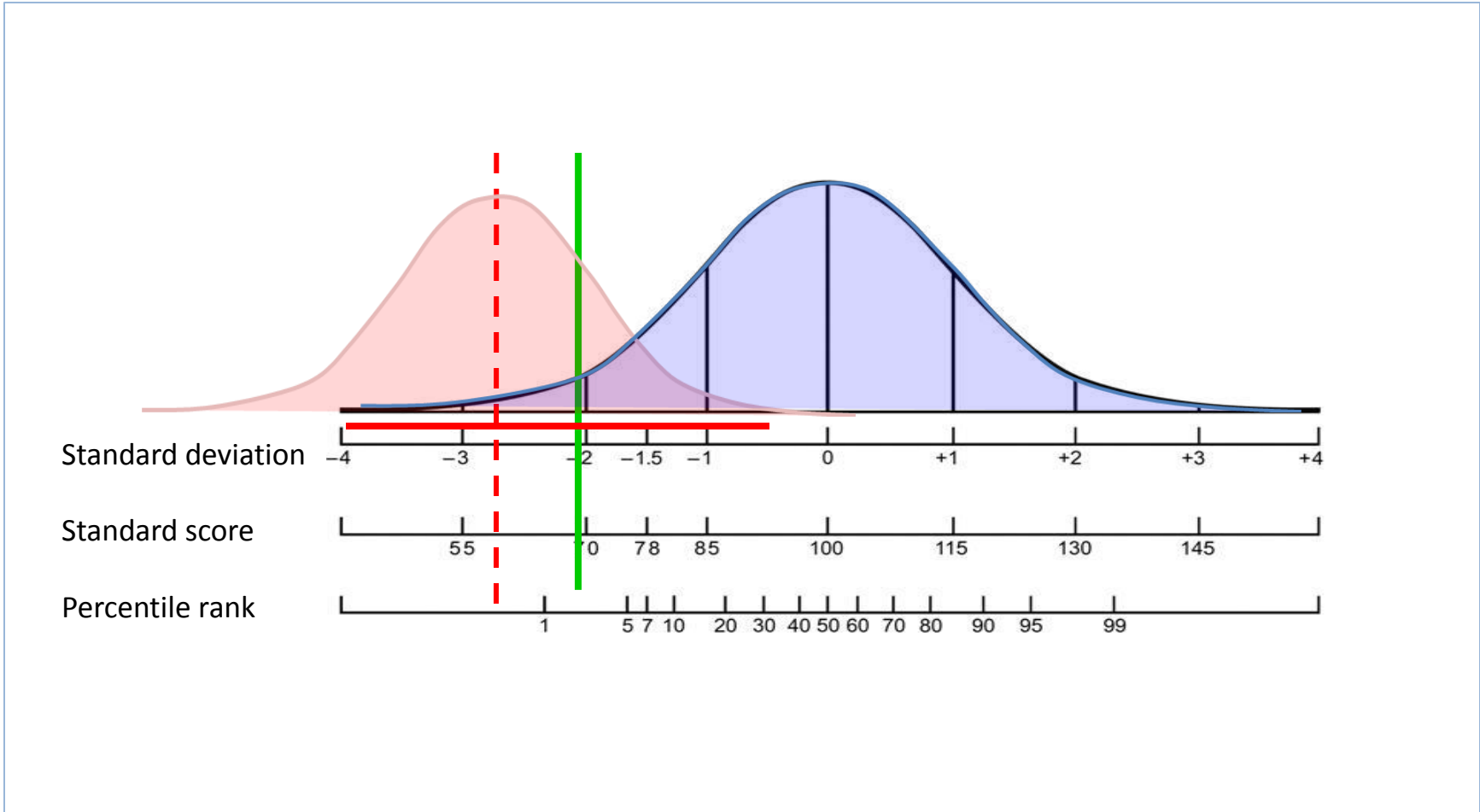
- Cut score of -1 SD

The influence of the Cut Score: CELF-4



- Cut score of -1.5 SD

The influence of the Cut Score: CELF-4



Page 272 CELF-4 image removed due to
copyright restrictions



Results for the CELF-4

SENSITIVITY & SPECIFICITY	INTERPRETATION OF DIAGNOSTIC ACCURACY
≥90%	GOOD
80%-89%	FAIR (ACCEPTABLE)
<80%	UNACCEPTABLE

CELF-4 Review

- Provides sensitivity and specificity info for all ages combined
- Provides sensitivity and specificity info for different cut-off scores (-1.0SD, -1.5SD, -2.0SD)
- Sensitivity decreases as the cut score moves farther from the mean

What should you do if sensitivity and specificity are not provided in the test manual?

- **Best approach: Read research** that provides diagnostic accuracy information

Found in Research:

TEST	SENSITIVITY	SPECIFICITY	CUT-OFF	CUT-OFF SD	AUTHORS
SPELT-II	90%	90%	51	-3.25 <i>SD</i>	Plante & Vance, 1994
SPELT-III	90%	95%	95	-.33 <i>SD</i>	Perona, Plante & Vance, 2005
SPELT-P	83%	95%	79	-1.40 <i>SD</i>	Plante & Vance, 1995
SPELT-P2	90.6%	100%	87	-.87 <i>SD</i>	Greenslade, Plante, & Vance, 2009

Found in Research (continued)

TEST	SENSITIVITY	SPECIFICITY	CUT-OFF	CUT-OFF SD	STUDY AUTHORS
<i>PPVT-III</i>	74%	71%	104	+ .23 <i>SD</i>	Gray, Plante, Vance, & Henrichsen, 1999
<i>ROWPVT</i>	77%	77%	97	- .20 <i>SD</i>	
<i>EOWPVT-R</i>	71%	71%	96	- .27 <i>SD</i>	
<i>EVT</i>	71%	68%	97	- .20 <i>SD</i>	

Found in Research (continued)

TEST	SENSITIVITY	SPECIFICITY	CUT-OFF	CUT-OFF SD	STUDY AUTHORS
<i>THE RENFREW BUS STORY</i>	84%	78%	Info Score: 87 Length Score: 99	Info Score: $-.87 SD$ Length Score: $-.10 SD$	Pankratz, Plante, Vance & Insalaco, 2007
<i>TEEM PEST</i>	90% 90%	95% 95%	75 60	$-1.69 SD$ $-2.67 SD$	Merrell & Plante, 1997

Test Review: Sensitivity & Specificity Summary

(Spaulding, Plante, & Farinella, 2006)

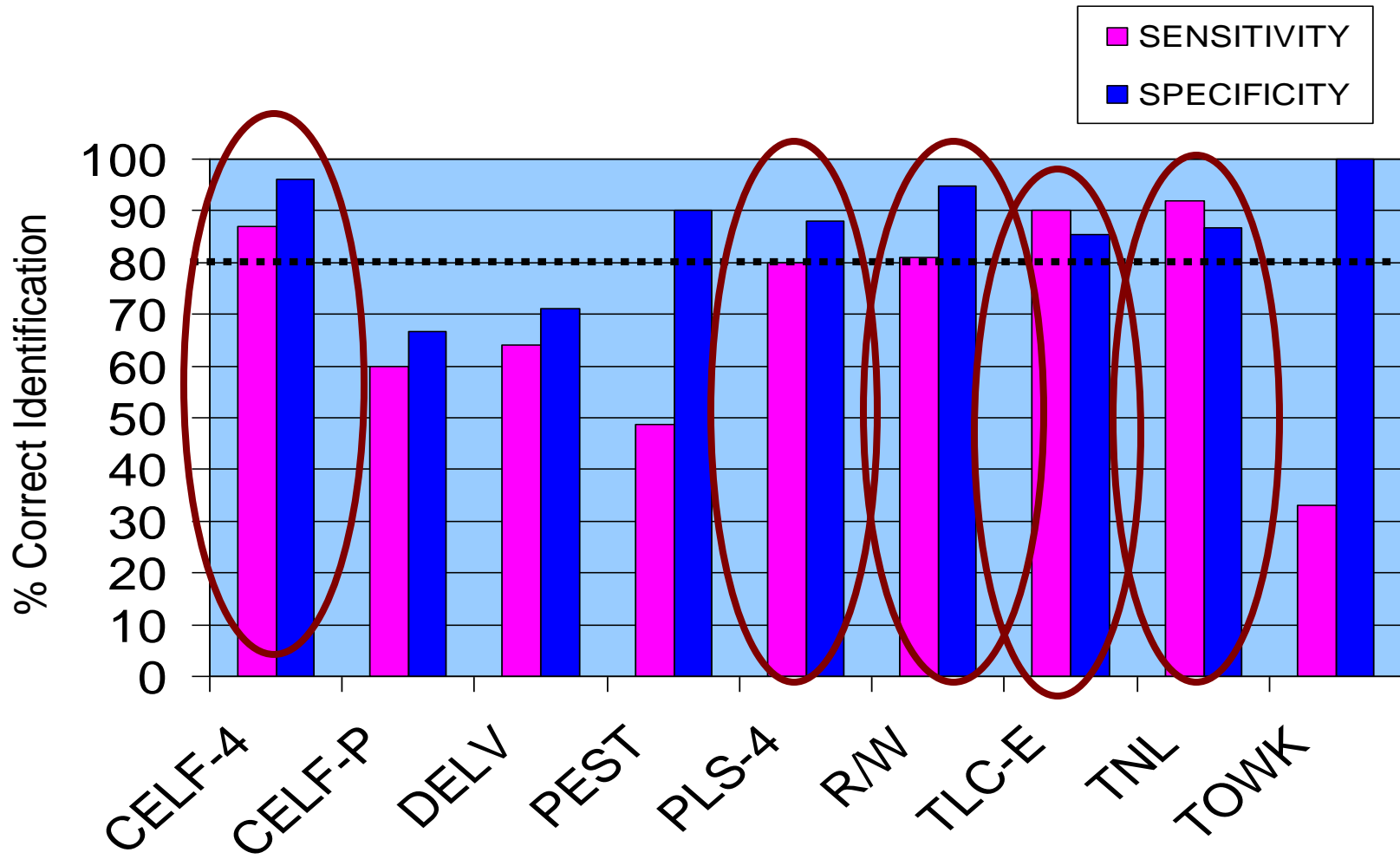
- *Do the test manuals report information on sensitivity and specificity?*

- Only **9** of the 43 tests reported

The cut-off scores that they reported ranged from -1.0SD to -2.0SD

Sensitivity & Specificity of Tests

(Spaulding, Farinella, & Plante, 2006)



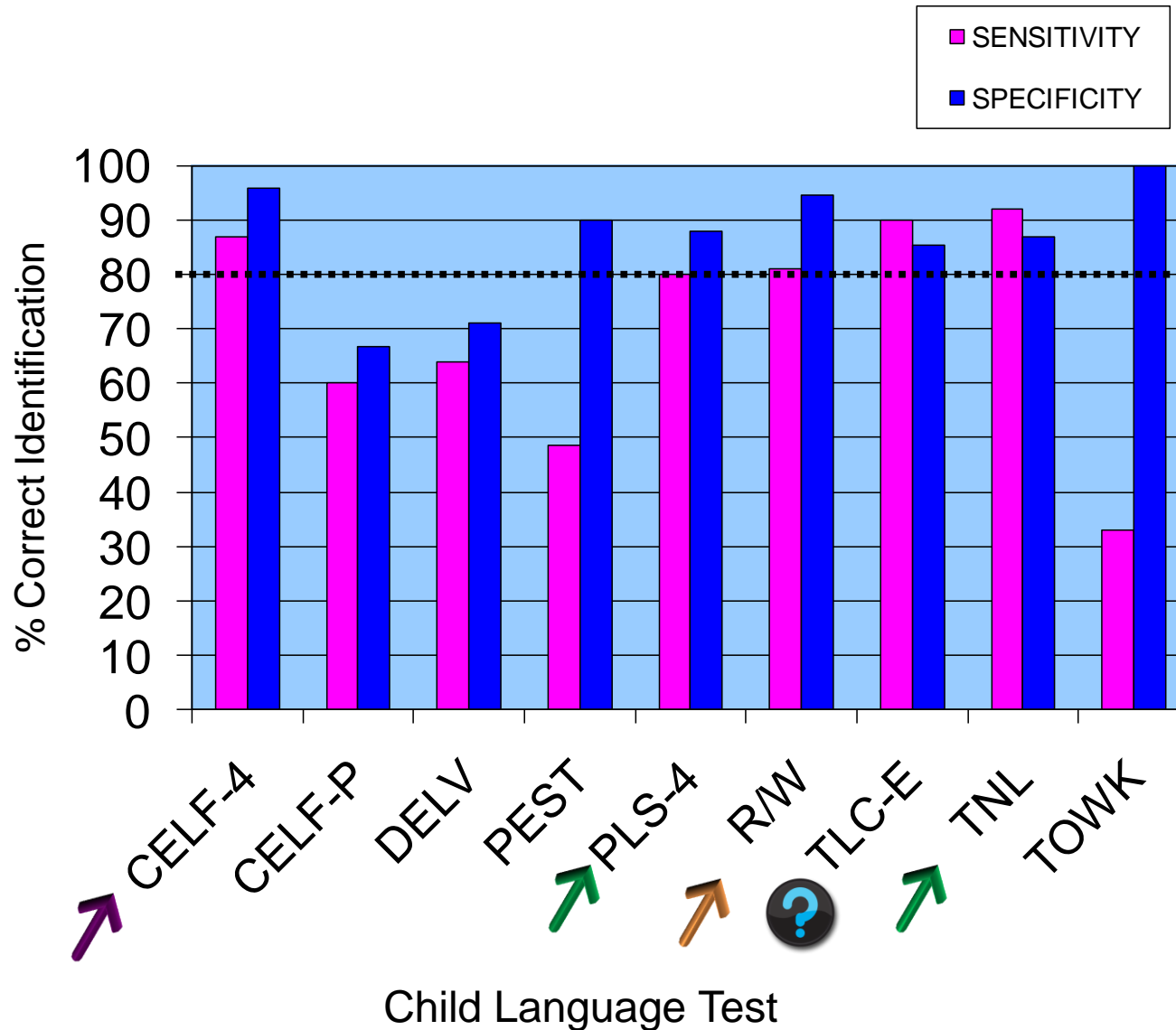
Child Language Test

Sensitivity & Specificity of Tests

Cut-off Scores

(Spaulding, Farinella, & Plante, 2006)

- = -2.0 SD
- = -1.0 SD
- = by age



Update: Sensitivity/Specificity In Manuals...2009

Three years later

TEST	SENSITIVITY	SPECIFICITY	CUT-OFF	CUT-OFF SD	POPULATION
<i>Bracken-Expressive*</i>	83%	86%	85	-1.0 SD	Language Impaired
<i>Receptive</i>	71%	83%	85	-1.0 SD	
<i>DELV-NR**</i>	95%	93%	85	-1.0 SD	Language Disordered
<i>CELF-P2*</i>	85%	82%	85	-1.0 SD	Not Specified

***Fair** level of diagnostic accuracy (criterion: Plante & Vance, 1994)

****Good** level of diagnostic accuracy (criterion: Plante & Vance, 1994)

Update: Sensitivity/Specificity In Manuals...2009

Three years later (continued)

TEST	SENSITIVITY	SPECIFICITY	CUT-OFF	CUT-OFF SD	POPULATION
MAVA**	100% 100%	97% 100%	85	-1.0 SD -1.0 SD	Special Ed with Vocab Deficits
TWFD**	90.7%	97.7%		PT 1 SCORE OF 31%	Word Finding Disordered

***Fair** level of diagnostic accuracy (criterion: Plante & Vance, 1994)

****Good** level of diagnostic accuracy (criterion: Plante & Vance, 1994)

What should you do if sensitivity and specificity are not provided in the test manual?

- What if there isn't any research?

- You might be tempted to use a cut score that seems reasonable.

Using + 1 SD above the mean for children with LI as the Cut Score

- A cut score of + 1 SD above the mean for children with LI will have sensitivity of approximately .84 (correctly identifying approximately 84% of children with LI).

Examples of Standard Score Means for Children with LI

- Tests in which the mean standard score is directly stated in the manual

TEST	Mean	SD	+ 1 SD
PPVT-4 (ages 8 – 12 yrs)	89.7	13.2	102.9
REEL-3	86	unknown	-----
EOWPVT-3	84.4	16.52	100.92
TOLD-I:4	84	unknown	-----
EVT-2 (ages 3 – 7 yrs)	88.1	12.3	100.4

- Using a cut score of 100 to 103 would be hard to justify! (specificity may be too high!)

What does this mean for you, as an SLP?

- You will need to read the test manuals!
- You will need to teach others about appropriate cut scores for tests of language development and why the cut scores are not the same across tests!
- You will need to advocate for change!

What this Means.....

- If we are using standardized tests for diagnostic purposes, we should be using **test specific cut-off scores** to assist in the diagnosis
- We should be choosing tests with high levels of **sensitivity** and **specificity**
- This is possible---so let's demand that from test developers

References

- Andersson, L., & Begley, E. (November, 2006). Opening the Assessment Cupboard: What are school-based clinicians doing today? Poster session presented at the Annual Convention of the American Speech-Language Hearing Association.
- Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30*, 196-206.
- Greenslade, K.J., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test-Preschool; Second Edition. *Language, Speech, and Hearing Services in Schools, 40*, 150-160.
- Pankratz, M.E., Plante, E., Vance, R., & Insalaco, D.M. (2007). The diagnostic and predictive validity of The Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools, 38*, 390-399.
- Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the Structured Photographic Expressive Language Test: Third Edition (SPELT-3). *Language, Speech, and Hearing Services in Schools, 36*, 103–115.
- Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*, 15–24.
- Plante, E., & Vance, R. (1995). Diagnostic accuracy of two tests of preschool language. *American Journal of Speech-Language Pathology, 4*, 70–76.
- Spaulding, T.J., Plante, E., & Farinella, K. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37*, 61-72.

Tests

- Bowers, L., Huisingsh, R., LoGiudice, C., & Orman, J. (2005). The WORD Test 2: Adolescent. East Moline, IL: Linguisystems.
- Bowers, L., Huisingsh, R., & LoGiudice, C. (2005). Test of Problem Solving 3: Elementary. East Moline, IL: Linguisystems.
- Brownell, R. (2000). Expressive One-Word Picture Vocabulary Test (3rd ed.). Novato, CA: Academic Therapy Publications.
- Bzoch, K., League, R., & Brown, V. (2003) Receptive-Expressive Emergent Language Test (3rd ed). Austin, TX: Pro-Ed.
- Carrow-Woolfolk, E. (1999). Comprehensive Assessment of Spoken Language. Circle Pines, MN; American Guidance Service
- Dawson, J., Stout, C., Eyer, J., Tattersall, P., Fonkalsrud, J., & Croley, K. (2005). Structured Photographic Expressive Language Test-Preschool (2nd ed.). DeKalb, IL: Janelle.
- Dunn, L. & Dunn, D. (2007). Peabody Picture Vocabulary Test (4th ed.). San Antonio, TX: Psychological Corporation.
- Hammill, D. & Newcomer, P. (2008). Test of Language Development: Intermediate (4th ed.). Austin, TX: Pro-Ed.
- Semel, E., Wiig, E., & Secord, W. (2003). Clinical Evaluation of Language Fundamentals (4th ed). San Antonio, TX: Psychological Corporation.
- Seymour, H., Roeper, T., & de Villiers, J. (2003). Diagnostic Evaluation of Language Variation: Norm-Referenced. San Antonio, TX: Psychological Corporation.
- Williams, K. (2007). Expressive Vocabulary Test (2nd ed.) Minneapolis, MN: NCS Pearson.
- Zimmerman, I., Steiner, V., & Pond, R. (2002). Preschool Language Scale (4th ed.). San Antonio, TX: Psychological Corporation.

TAKE HOME GUIDE

•This HOME GUIDE will help you to:

1) Determine the diagnostic accuracy (sensitivity and specificity) of standardized tests

2)To determine if the sensitivity and specificity rates are acceptable.

3)To determine the standard deviation cut-off score that is appropriate to use for the test in order to increase the chances of making an accurate diagnosis.

Name of Test _____

1) Is there a **Validity** section in the Table of Contents?

- If NO, you will be hand searching through this entire manual. Start your search and skip to question 4.
- If YES, go to question 2.

2) Is there a section under **Validity** that either says *Clinical Validity*, *Discriminant Analysis*, or indicates *Special Populations*?

- If YES, turn to this section in the manual, and skip to question 4.
- If NO, go to question 3.

3) Is there a section under **Validity** that says *Construct Validity*?

- If YES, turn to this section and go to question 4.
- If No, turn to the Validity section of the manual and go to question 4.

TAKE HOME GUIDE (continued)

4) Is sensitivity and specificity information available in the manual and calculated for you?

- If YES, indicate sensitivity and specificity rates and skip to question 6.

Sensitivity: _____ **Specificity:** _____

- If NO, go to question 5.

5) Is sensitivity and specificity information able to be calculated from information provided in the test manual?

- If YES, calculate the sensitivity and specificity rates using the formula and go to question 6.

Sensitivity: _____ **Specificity:** _____

- If NO, you are done. You don't know the diagnostic accuracy of this test.

TAKE HOME GUIDE (continued)

6) What is the standard deviation (or standard score) cut-off score to obtain that level of sensitivity and specificity? _____

- If there IS one, go to question 8.
- If there IS NOT one, this test is not useful for diagnosing language impairment because you don't know what cut-off to use.

7) Is the level of sensitivity and specificity .80 (80%) or better?

- If YES, this is an acceptable level of diagnostic accuracy. Use the cut-off they provide.
- If NO, this is not an acceptable level of diagnostic accuracy to use the test for this purpose.

Contact information

Tammie Spaulding: tammie.spaulding@uconn.edu

Luanne Andersson: landersson@ithaca.edu