



UNDERSTANDING
ACCURACY OF
STANDARDIZED
TESTS:
FOCUS ON
SENSITIVITY,
SPECIFICITY

ADRIANA LAVI, PHD, CCC-SLP

Disclosures

Adriana Lavi, PhD, CCC-SLP

Financial:

Author of *Clinical Assessment of Pragmatics (CAPs) test*

Author of *VideoLearningSquad.com* and *VideoAssessmentTools.com*

Non-financial: No relevant non-financial relationship exists.

Learner objectives



1. Understand accuracy of standardized tests
 2. Learn about sensitivity and specificity
 3. Address most common test-related questions and misconceptions
 4. Discuss impact-related components of a legally defensible and evidence-based assessment
 5. Discuss ways to determine adverse educational and social impact
 6. Get access to report templates, impact rating scales, questionnaires
-

Access to resources

1. Report templates
2. Sample goals
3. Presentation slides
4. Impact related rating scales
5. Online teacher/parent questionnaires

<https://VideoAssessmentTools.com/access>

Most Common Questions

1. Why are there discrepancies in scores between language tests? Why do we NOT get same scores across various language tests? Why do some tests produce scores higher than the others?
 2. Is it better to include children with disabilities in the normative sample or not?
 3. How do we know if the test was validated for the purpose we intend to use the test for?
-

Most Common Misconceptions

1. Standardized tests over or under-identify, therefore we shouldn't use them!
 2. This test is reliable /valid therefore I should use it with ALL my assessments
 3. Standardized tests are inaccurate, they overidentify or under-identify! We should only use dynamic assessment!
 4. We should only use strength-based assessment and observations
-

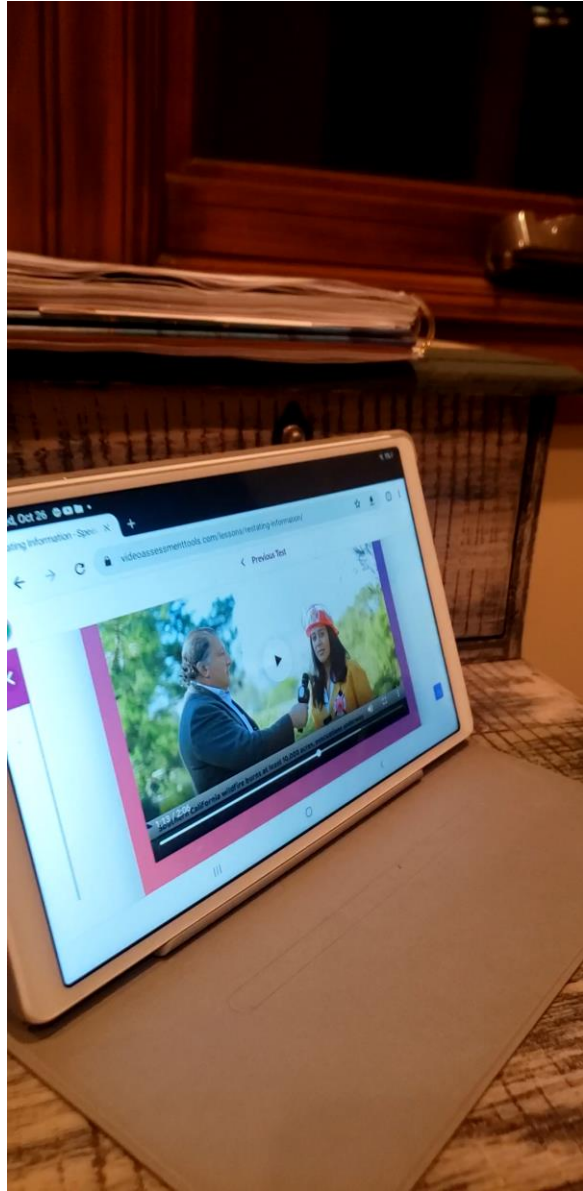
CASE STUDY

- Male student, Age: 8, 2nd grade
- CASL score range: SS=84 to SS=92
- Language Video Assessment Tool:

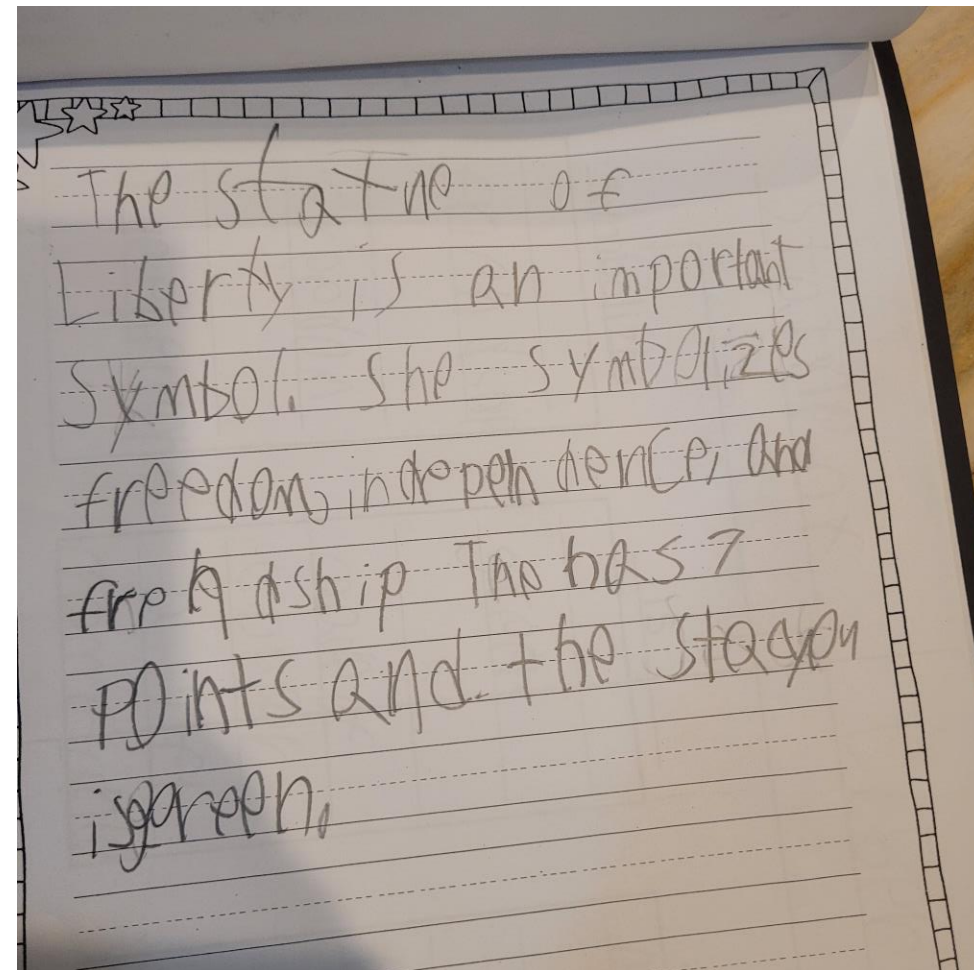
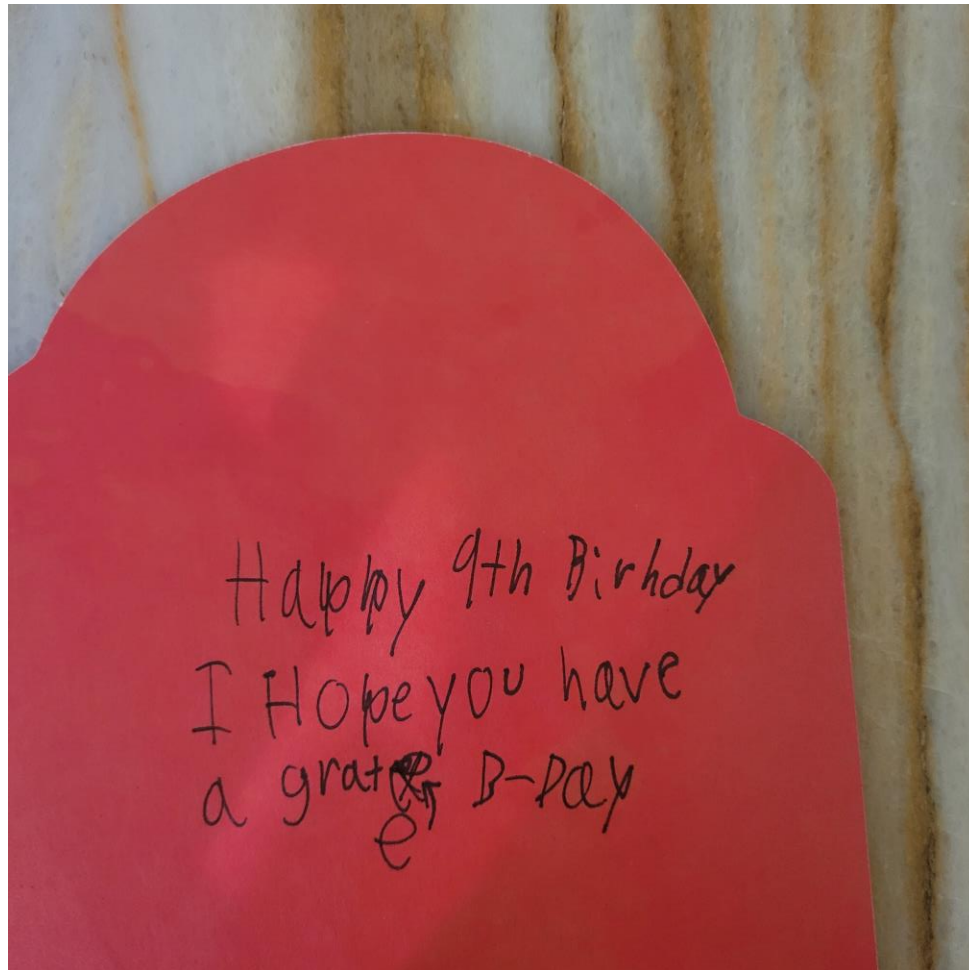
Restating Information: SS=69, Following Directions: SS=92

Listening Comprehension: SS=71, Morphology and Syntax: SS=84

CASE STUDY 2




CASE STUDY 2 – WORK SAMPLES



CASE STUDY I – WORK SAMPLES

My weekend by Aaron
First, I went out my
basketball game. Then,
I ^{took} the football
game. ... →

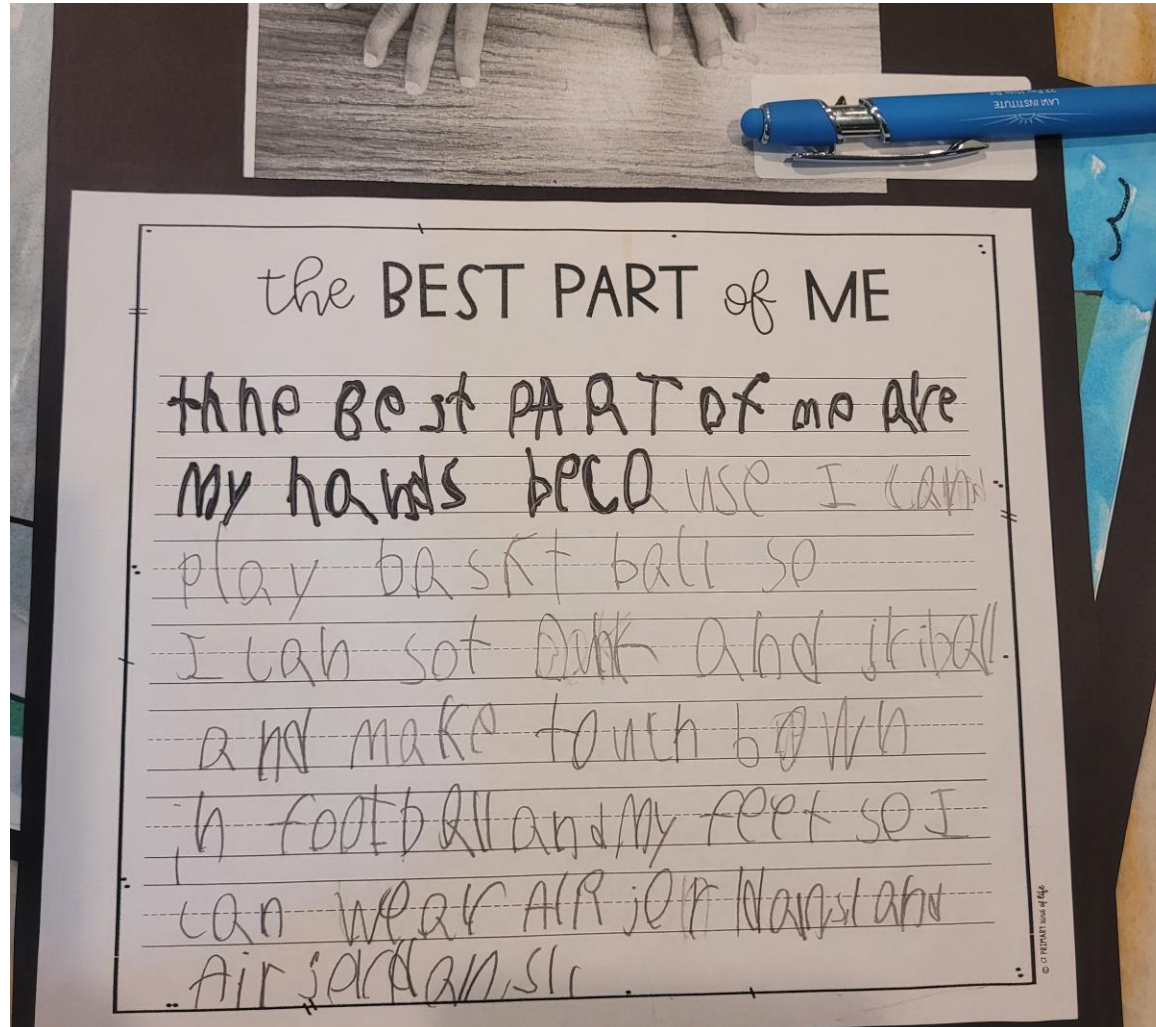


Indent x Title

This weekend was fun. *Nice!*

First, I played games. I had fun
playing games. Next, I had a
^{person} prsin ^{work} come for my mom's work.
The ^{person} prsin is ^{staying} staying for two
and a ^{half} half weeks. Last, I ate
^{dinner} dinner. I had noodles ^{es} and fish.
I had a great weekend.

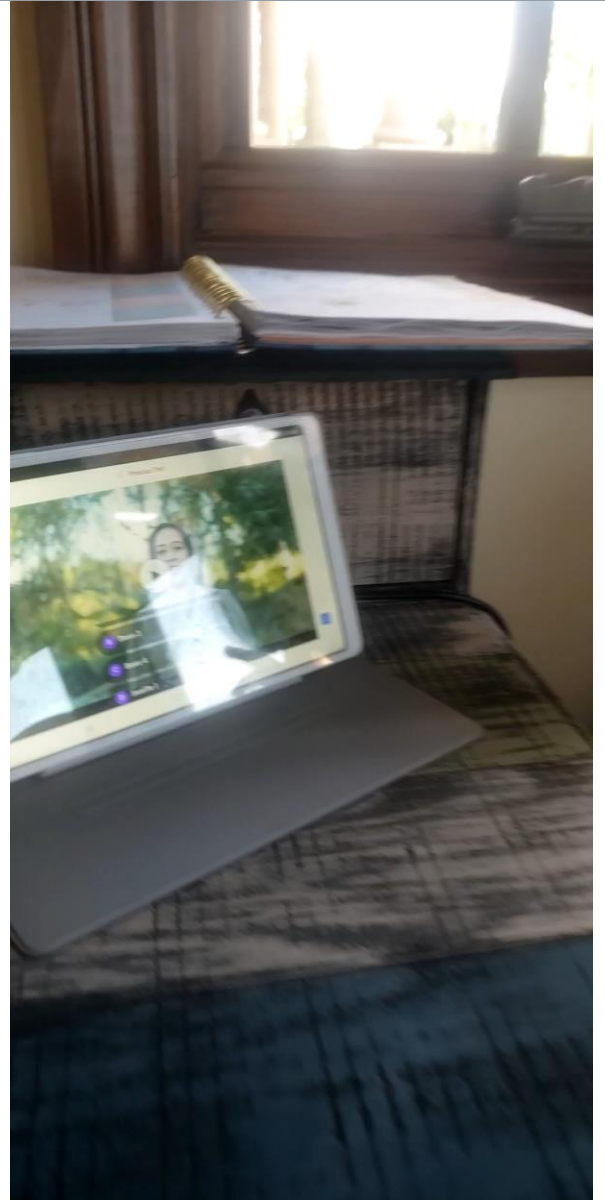
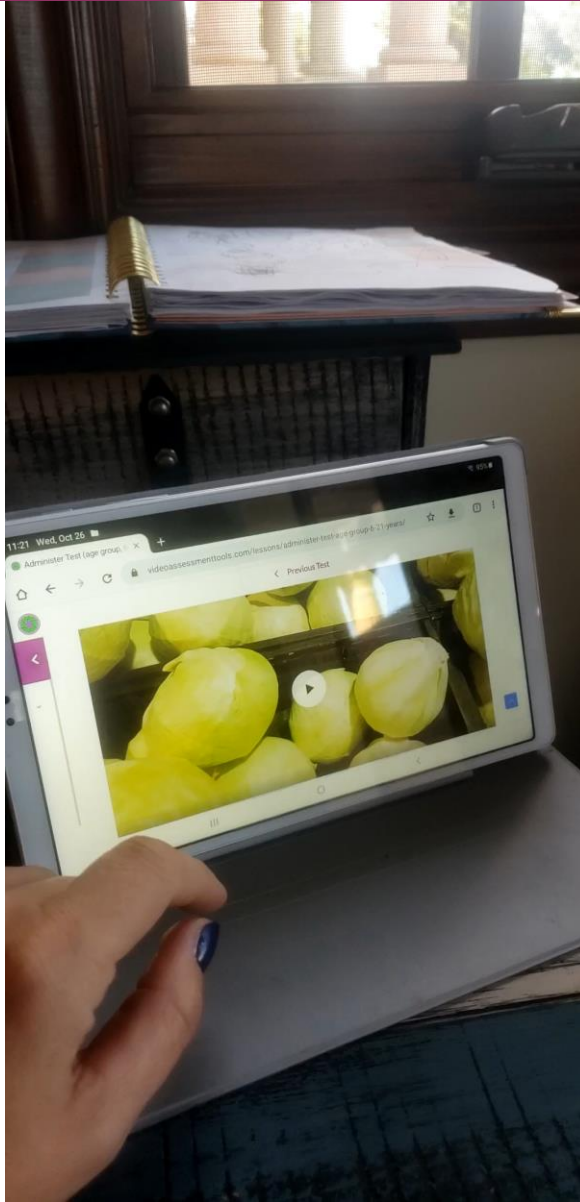
CASE STUDY I – WORK SAMPLES



CASE STUDY 2

- Male student, age: 9, 3rd grade
- GFTA: SS=69
- Articulation and Phonology Video Assessment Tool: SS=76
- Report card: all areas including social-emotional and language arts reported as “3/meeting standard”
- Student self-rating: “ My speech sounds good, I don’t need any help”

CASE STUDY 3



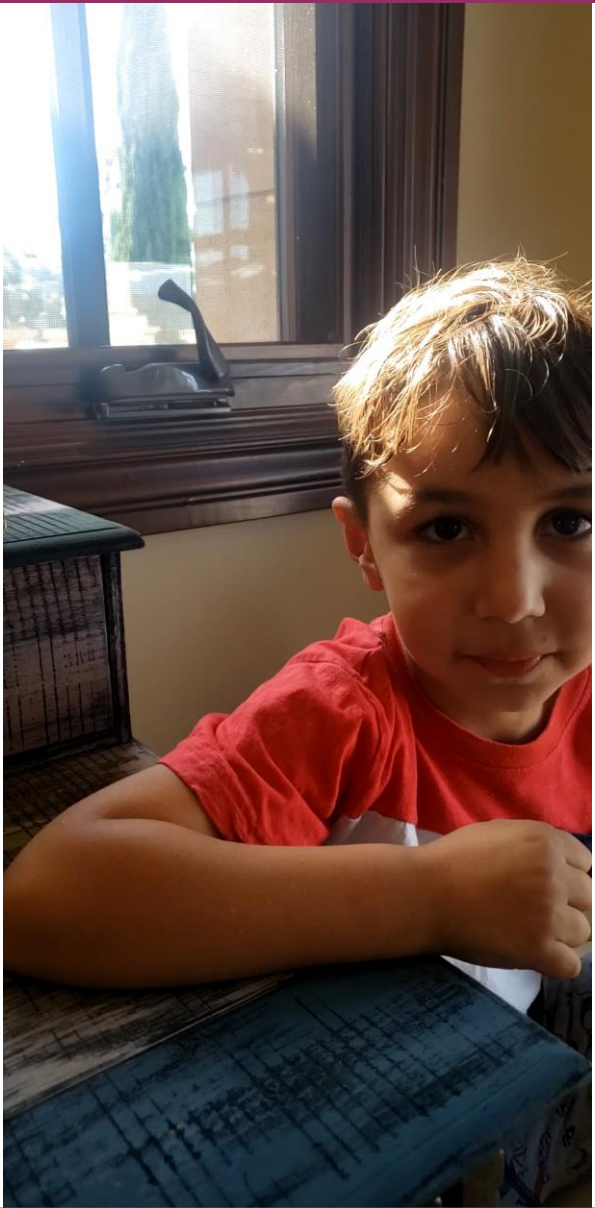
CASE STUDY 3

- Male student, Age: 6, 1st grade
- GFTA: SS=85
- Articulation and Phonology Video Assessment Tool: SS=77
- Student self-rating: “ I tell kids I’m British so they don’t make fun of me”

		Reporting Period	
		1	2
SOCIAL/EMOTIONAL			
Effort		S+	O
Displays impulse control and self-regulation		2+	3
Expresses emotions in an appropriate way		2+	3
Cooperates and shares with others		3	3
Participates positively in group settings		3	3
Engages in meaningful relationships with adults		3	3
LANGUAGE AND LITERACY			
Effort		O	O
Responds to one and two step oral commands		3	3
Speaks clearly to be understood by others		3	3
Uses descriptive language and vocabulary		2+	3
Alphabet knowledge: upper case recognition		3	4
Alphabet knowledge: lower case recognition		3	3
Alphabet knowledge: letter sounds		3	4
Writes letters and draws pictures to communicate		3	3
Writes own name legibly		4	4
Displays comprehension regarding key components of text		NA	3
MATHEMATICS			
Effort		O	O
Counts orally to twenty		3	3
Counts with one to one correspondence to ten		3	4
Writes numbers one through ten		3	4
Understands simple concepts of addition and subtraction		NA	NA
Demonstrates concept of time and days of the week		2	2+
Names primary shapes and describes characteristics		4	4
Makes and describes simple patterns		2	3
Expresses mathematical reasoning and abstract reasoning		NA	3
Recognizes numbers 1-10		4	4
VISUAL AND PERFORMING ARTS			
Effort		O	O
Engages in art activities and completes projects		3	3
Use art tools and resources appropriately		3	3
Sings age appropriate songs from memory		3	3

		Reporting Period	
		1	2
Understands differences between real and imaginary characters		NA	3
Performs simple movements to oral instructions		3	3
PHYSICAL DEVELOPMENT			
Effort		O	O
Shows confidence in locomotor skills (running, jumping, hopping)		3	3
Demonstrates body awareness and personal boundaries during play		3	3
Demonstrates directional awareness (front, behind, over, etc.)		3	3
Shows increased fine motor ability		3	3
Sustains gross motor activity over a period of time		3	3
HEALTH EDUCATION			
Effort		O	O
Practices personal hygiene skills		3	3
Identifies safety rules/symbols at home and at school		NA	3
Identifies healthy food choices		NA	3
HISTORY-SOCIAL SCIENCE			
Effort		O	O
Follows rules and takes turns		3	2+
Accepts responsibility for behavior		3	2+
Understands how to be a good citizen		3	3
Respects rights, diversity, feelings and property of others		3	3
Describes familiar locations in their community		NA	NA
Understands concepts of money		NA	NA
SCIENCE			
Effort		O	O
Demonstrates ability to make predictions		3	3
Demonstrates ability to observe, investigate and describe physical properties (size, weight, shape)		3	3
Describes similarities and differences of living things		NA	3
Describes environmental objects using sensory language		NA	3

CASE STUDY 3



Why should we pay attention to the psychometric properties of the tests we use?

Most Common Question #1

Why are there discrepancies in scores between language tests? Why do we NOT get same scores across various language tests?

TEST SCORE DISCREPANCIES

- **based on normative group characteristics**

The makeup of "normative" groups influences how tests function. If Test A includes people with disabilities in the "normative" group based on the rationale that it better represents the full population, but Test B excludes people with the target disorder from the normative group, Test B will be more sensitive to the disorder, whereas Test A will be more likely to find the child with the disorder to be a member of the "normative group."

- **based on “PURPOSE OF THE TEST”** – the tests were not developed with similar purpose
- **based on “Design of the TEST”** - do not measure similar skills

IDEA

“(A) assessments and other evaluation materials used to assess a child under this section—“

- (i) “are selected and administered so as not to be discriminatory on a racial or cultural basis;
- (ii) “are provided and administered in the language and form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally, unless it is not feasible to so provide or administer;
- (iii) “are used **for purposes** for which the assessments or measures are valid and reliable;
- (iv) “are administered by trained and knowledgeable personnel;
- (v) “are administered in accordance with any instructions provided by the producer of such assessments;

Misconception #1

Tests are reliable and valid

(e.g., this test is reliable /valid therefore I can use it in my assessments)

TESTS ARE RELIABLE AND VALID ONLY RELATIVE TO A PURPOSE

- A test can be valid for one purpose and completely invalid for another purpose. What we are concerned about is the purpose of the assessment. And what is the **interpretation** we are making about a test score?

- Does the student have an impairment?

YES

NO

- What **evidence** is available to support the intended interpretation?
- How **confident** should we be about this interpretation?

(from CEUSmarthub.com, Plante Presentation on Test Psychometrics, 2020)

WHAT IS THE PURPOSE OF YOUR ASSESSMENT?

- Identify an impairment
- Determine strengths and weaknesses
- Identify areas of need/ determine intervention goals
- Document change

So, the tests we select for our assessments, have to be valid and reliable for purpose listed above!

HOW DO WE KNOW IF THE TEST WAS VALIDATED FOR THE PURPOSE WE INTEND TO USE THE TEST FOR?

- Male student, Age: 9, 3rd grade
- CASL score range: SS=84 to SS=92
- Language Video Assessment Tool:

Restating Information: SS=69, Following Directions: SS=92

Listening Comprehension: SS=71, Morphology and Syntax: SS=84

QUESTIONS WE SHOULD BE ASKING...

- What are we trying to get out of this test?
- Why are we giving this test?
- What are we going to learn from it?

Reality:

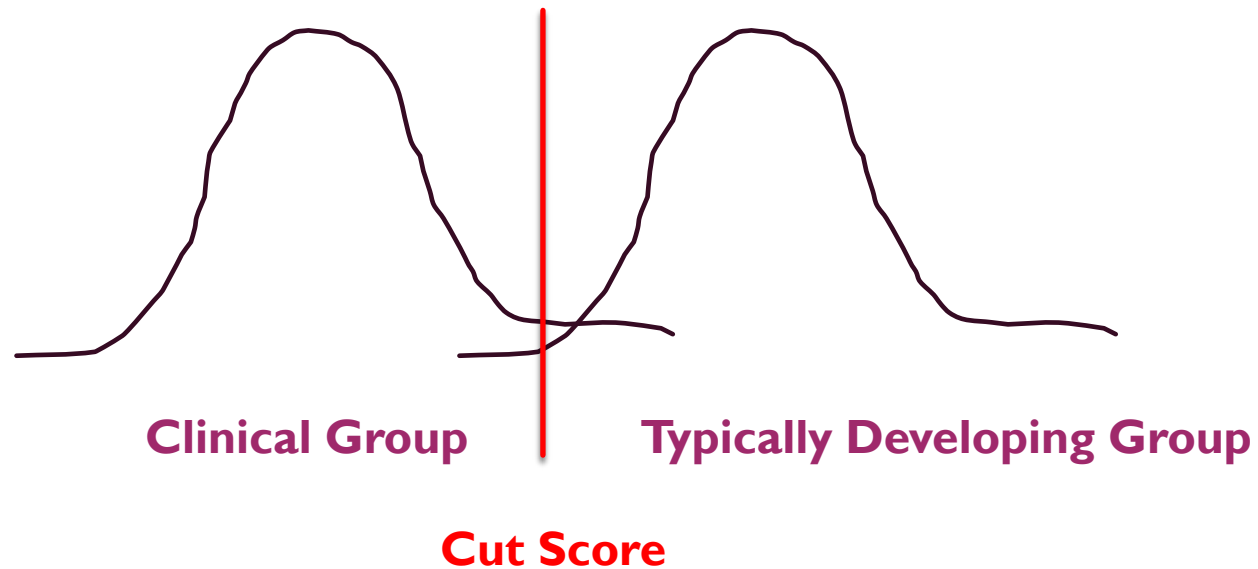
Let's administer a bunch of tests we have here, that are available on the shelf, and we'll sort it out at the end.

QUESTIONS WE SHOULD BE ASKING...

- What we need to do is ONLY look for the evidence that supports our purpose.
- **This can save you and the student so much time!**

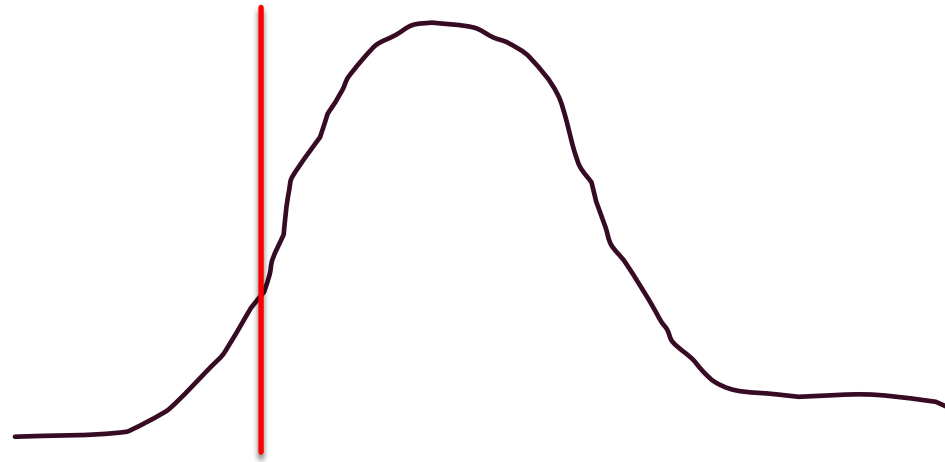
IDENTIFICATION

- How do we identify a disorder?
- **Discriminant Analysis** (statistically discriminates between groups)



WHAT WE LEARNED IN GRAD SCHOOL AND WHAT WE ASSUME...

- We assume that the children with impairment will score at the low end of the curve and this is how we can identify a disorder



- This type of analysis is acceptable for determination of strengths and weaknesses, not for identification of a disorder

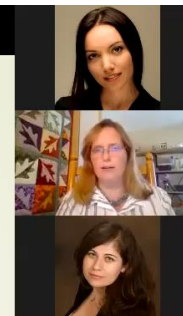
SPAULDING, PLANTE, 2010 STUDY


When is this needed

- ▶ Disorders that overlap with typical behaviors
 - ▶ Anomia
 - ▶ Developmental Language Disorder

When this is NOT needed

- ▶ Disorders that are not already diagnosed
 - ▶ TBI
 - ▶ Down Syndrome
- ▶ Disorders that are not already obvious to the casual listener
 - ▶ Stuttering
 - ▶ Dysarthria





Naturally the question becomes: what do I do if the district eligibility criteria are different from what the test data show?


Discriminant Accuracy

- Sensitivity and Specificity (Dollaghan, 2007)
 - Sensitivity – does the assessment accurately identify those students who truly have a language/reading disorder as having a reading disorder
 - Specificity - does the assessment accurately identify those students who truly do not have any disorders as typical
 - Sensitivity and specificity determine the test's degree of discriminant accuracy, or the ability to distinguish the presence of a disorder
- Vance and Plante (1994) established a criteria for accurate identification of a disorder (discriminant accuracy)
 - 90% should be considered good discriminant accuracy
 - 80% to 89% should be considered fair
 - Below 80%, misidentifications occur at unacceptably high rates” and lead to “serious social consequences” of misidentified children. (p. 21)”
- Most important information about the assessment
 - If the test has low sensitivity and specificity or if that information is missing; NONE of the other psychometric properties matter

Cut scores

- Numerical boundary between what is considered typical and disordered
 - *The formula requires the mean and standard deviation of both a clinical and non-clinical sample, and estimates the score at which a subject has a greater probability of belonging to a clinical sample rather than a non-clinical sample.
- Test Specific –vary from test to test
- Age specific –differ depending on the child's age
- Problem: Often applied arbitrarily without reference to how children actually score on the tests selected for use (Spaulding, Plante, & Farinella, 2006)



- 
- Not all tests have the same cut-off score
 - We must know the test-specific cut-off score
 - But we also need to look at the age range of children who participated in the discriminant analysis. They might say the study included children ages 6-12, but how confident are you using this test with students who are 16 years old.?

- How confident are you to use this test with the following psychometric properties?

Table 5.1 IMPACT Language Functioning Rating Scale sensitivity, specificity and likelihood ratios

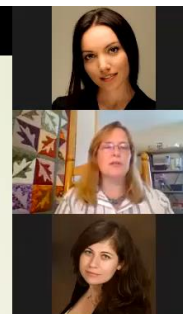
Parent Rating Scale

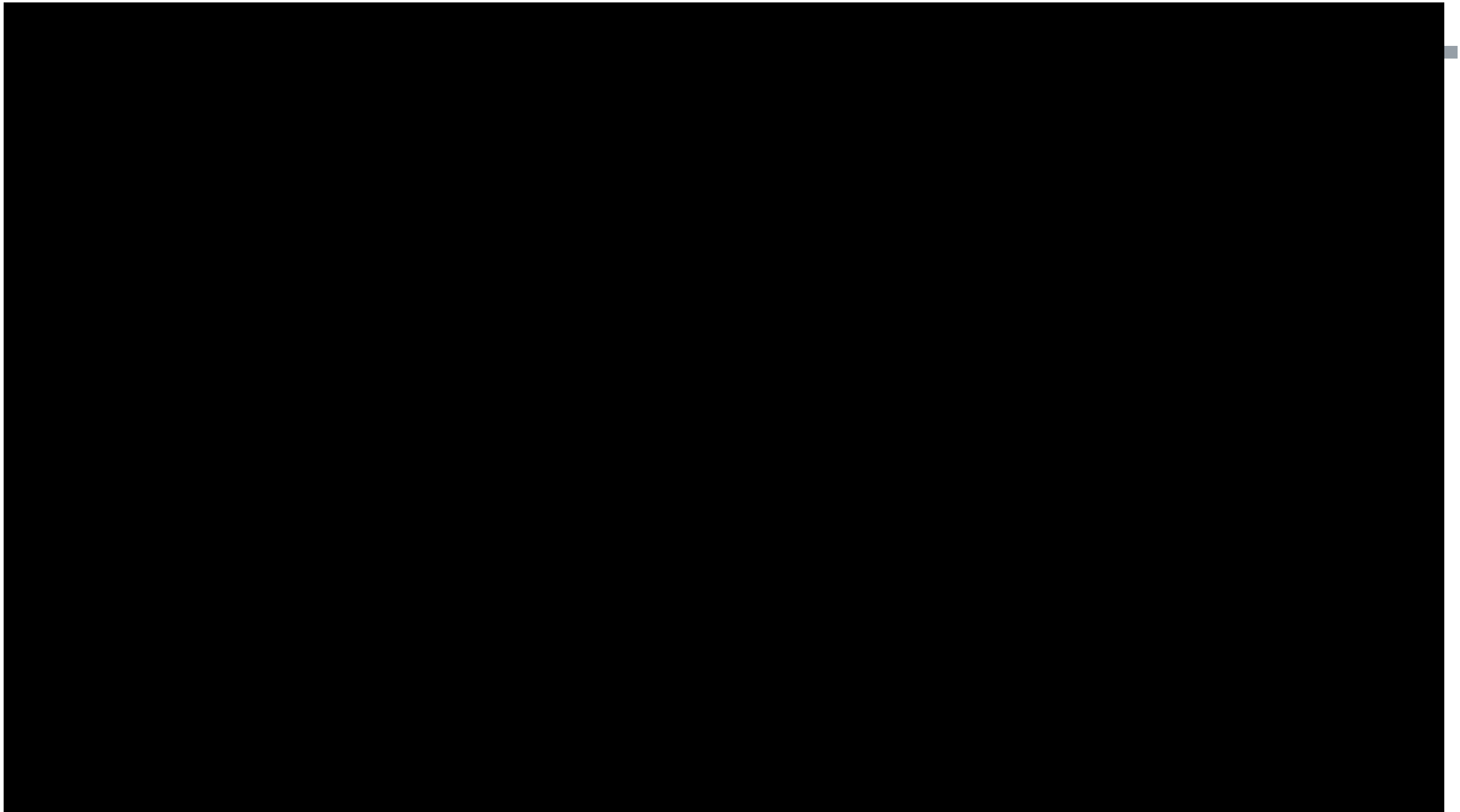
Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
5:0-5:11	77	.88	.83	4.07	.09
6:0-6:11	78	.84	.88	4.11	.11
7:0-7:11	77	.91	.91	5.06	.26
8:0-8:11	77	.95	.84	4.15	.12
9:0-9:11	77	.83	.83	4.09	.09
10:0-10:11	78	.93	.81	4.41	.11
11:0-11:11	77	.88	.84	4.01	.23
12:0-12:11	78	.89	.88	4.06	.12
13:0-13:11	77	.94	.94	5.03	.41
14:0-14:11	77	.88	.93	4.06	.24
15:0-15:11	77	.83	.85	4.11	.27
16:0-21:0	77	.91	.84	4.26	.16

Note: Age groups 16:0-21:0 are reported together as there were no age-related changes detected after the age of 16. Total N=2790; typically developing group n=1403; clinical group=1387

What it looks like: Test form

Age Band	Sum of Identification Core Standard Scores	Cut Score	Sensitivity	Specificity	Decision:
6-7 years		24	84	84	Is the Identification Core composite less than the cut score? <input checked="" type="checkbox"/> Yes This score is consistent with the presence of a language/literacy disorder.
8-11 years	21	34	88	85	<input type="checkbox"/> No This score is not consistent with the presence of a language/literacy disorder.
12-18 years		42	86	90	





PURPOSE: SEVERITY DETERMINATION

- We are not looking at a yes/no question, we are looking at a continuum of skills. We are looking for evidence of relatively even raw score to standard score distributions in the norms.
- Score distribution - we need an even distribution of scores across the normative range for each age of the test

Standard Score	Raw Score
16	21
15	20
14	19
13	18
12	17
11	16
10	13-14
9	12
8	11
7	10
6	9
5	7-8
4	6
3	5
2	3-4

Standard Score	Raw Score
16	12
15	11
14	10
13	
12	9
11	8
10	6-7
9	5
8	
7	
6	4
5	
4	3
3	2
2	

First column- you see a nice and even range of scores all the way down the standard score range.

You can contrast that with the second column where you can see significant gaps.

(from CEUSmarthub.com, Plante Presentation on Test Psychometrics, 2020)

BASEMENT EFFECT

Raw Scores	5-0	5-6	6-0	6-6	7-0	7-6	8-0	8-6	9-0
	5-5	5-11	6-5	6-11	7-5	7-11	8-5	8-11	9-5
0	87	85	84	81	75	70	68	63	60
1	89	86	85	82	78	75	70	64	61
2	97	90	86	85	84	81	75	70	65
3	100	95	93	89	85	84	80	72	70
4	109	100	98	92	91	85	83	74	72
5	111	103	100	97	95	92	85	77	75
6	113	108	106	100	98	95	88	78	77
7	115	112	109	107	100	97	90	85	80
8	118	115	114	110	104	98	92	86	85
9	121	118	115	112	107	100	95	88	86
10	125	123	122	115	110	107	99	92	90
11	130	126	124	119	113	110	100	97	93
12	132	130	128	123	115	112	105	100	97
13	134	131	130	128	121	115	108	106	100
14	136	133	131	130	125	119	115	109	106
15	138	135	132	131	130	123	118	115	112
16	139	137	134	133	131	130	124	121	115
17	140	139	136	135	132	131	130	128	121
18	>140	140	138	137	134	133	132	130	126
19	.	>140	140	139	136	135	134	131	130
20	.	.	>140	140	138	137	136	132	131

- Occurs when test takers in the normative group perform so poorly that those with disorders cannot score lower than normal individuals
- This chart shows how a test works well for severity determination at age: 8 or 9, but not for ages 5 where a child obtains a standard score of 100 if they get a raw score of "0".
- (from CEUSmarthub.com, Plante Presentation on Test Psychometrics, 2020)

%ile Rank	Matching	Selective Analysis	Reordering	Reasoning	Receptive	Expressive	Scaled Score
<1	0-3	0-1			0-3	0-1	1
<1	4	2			4-6	2-3	2
1	5-6	3		0	7	4-5	3
2	7	4		1	8-9	6-8	4
5	8	5	0	2	10-11	9	5
9	9-10	6-7	1	3	12-13	10-11	6
16	11-12	8	2	4	14-15	12-13	7
25	13	9	3	5	16	14-16	8
37	14	10	4-5	6	17-18	17-19	9
50	15	11	6	7	19	20	10
63	16	12	7	8	20-21	21-23	11
75	17	13	8	9	22	24	12
84		14	9	10	23-24	25-26	13
91		15	10	11	25	27-28	14
95		16	11	12	26	29	15
98		17	12	13	27	30	16
99			13	14	28	31-32	17
>99			14	15	29	33-34	18
>99			15	16	30	35	19
>99				17-21	31	36-39	20

In summary,

You need to determine the purpose of your assessment and then look for evidence that supports this purpose.

Purpose: identification → Look for sensitivity, Specificity, Cut scores

Purpose: severity level → Items range in difficulty; Even spread of scores in the norms; No basement effect

On the inclusion of students with disabilities in the normative samples

(Pena & Plante, 2020 [Facebook Group Discussion](#))

- Test developers tend to use the same process as they do for psychological and educational tests. Namely to rank people to represent the full population. For the purpose of ranking, disordered children are used in the sample because it widens the normative range, allowing for more fine-grained divisions and better rank estimates of students who fall -1 SD. However, such tests are not meant for diagnostic purposes, or the determination if a child has a disorder.
- Myth: If a child with a disabling condition is represented in the normative sample than the test is appropriate for usage with that population (e.g., ADHD, ASD, DLD, etc.)
- Reality: For diagnostic purposes there should be no students with disorders included in the normative sample, since our goal is to diagnose impairment for intervention purposes.
- Compromise: During the test development stage it is important to identify items that TD students pass and impaired kids fail for diagnostic accuracy purposes. But disordered students should not be included in the standardization norms because it lowers the mean, increases SD, thereby shifts the cut scores, which results in less likely identification of impaired students (“normalizes the disorder”). The overlap between disordered and typical becomes too great and its much harder to reliably identify those with an impairment.

Assessment Tasks and What They Measure

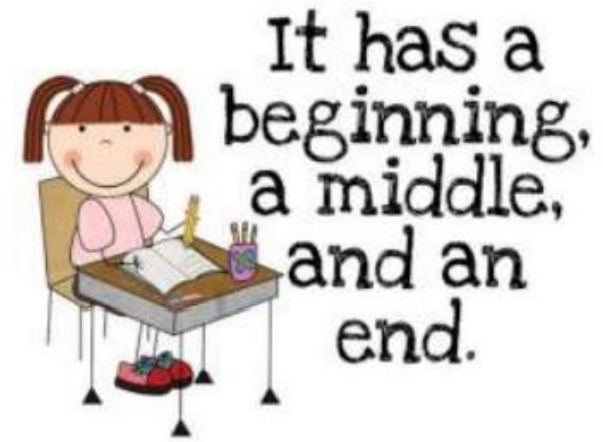
- **Following directions** tasks correlate with working memory functioning and are sensitive to **reading deficits** ([Lahey & Bloom, 1994](#); [Cowan, 1996](#); [Baddeley, 2003](#))
- **Grammatical structure** deficits particularly in the area of tense-marking & agreement incl. past tense '-ed', third person singular '-s', 'be' and 'do' etc., is sensitive to language deficits ([Rice & Wexler, 1996](#); [Loeb and Leonard, 1991](#); [Rice and Wexler, 1996](#); [Oetting and Horohov, 1997](#); [van der Lely and Ullman, 2001](#))
- **Vocabulary breadth, depth, quality as well as manipulation** tasks (e.g., naming definitions, synonyms, relationships among semantically related words, explaining multiple meaning words, etc.) are sensitive language deficits ([McGregor, Oleson, Bahnsen, & Duff, 2013](#); [Marinellie & Johnson, 2002](#); [Norbury, 2005](#); [Sheng & McGregor 2010](#))
 - Children with DLD possess not only “fragile knowledge of the core meaning of individual words, but fragile semantic connections between words” ([Nation, 2014, p.2](#))

Are Vocabulary Tests Useful for School Aged Children

- One-word vocabulary tests are often used in the assessment process to qualify children for speech and language services ([Betz, Eickhoff, & Sullivan, 2013](#))
- Studies have found that single word vocabulary tests have poor psychometric properties and/or are not representative of **linguistic competence embedded in life-activities** ([Gray et al., 1999](#); [Ukrainetz & Blomquist, 2002](#); [Bogue, DeThorne, Schaefer, 2014](#))
- Single word vocabulary tests can overinflate testing scores and not represent the child's true expressive language competence. Even when a student truly has solid or even superior vocabulary knowledge and naming skills, doesn't mean that s/he can **effectively utilize these abilities during the narrative production as well as reading and writing tasks.**

What do Narratives Reveal?

- Sequencing Ability
 - Story order
- Working Memory
 - Recall of relevant details
- Grammar
 - Sentence structure errors, run-on sentences, etc.
 - Use of temporal markers and cohesive ties to connect the story
- Vocabulary
 - Immature vs. age-level
 - Word retrieval issues vs. lexical fluency
- Pragmatics and perspective taking
 - Topic cohesion /coherence
 - Use of anaphoric references
 - Insight into character's feelings, beliefs, thoughts



Clinical Assessment of Narrative Abilities

- Ask students to summarize a read book or a viewed movie
 - Quick and efficient way to assess multiple areas of language
 - Provide more detailed information regarding macrostructural (story grammar elements, perspective taking, etc.) and microstructural elements (vocabulary, syntax, and grammar) as well as child's thought processes and socio-emotional functioning
- Preschool (3-6 years old)
 - Wordless picture books
- Early Elementary (7-12 years old)
 - Picture books
- Middle School/High School (13-18 years old)
 - Delayed retelling favorite book or movie

COMMON QUESTIONS AND MISCONCEPTIONS

- Why do some tests produce scores higher than the others?
- Is it better to include children with disabilities in the normative sample or not?
- Standardized tests are inaccurate, they overidentify or under-identify! We should only use dynamic assessment!
- We should only use strength-based assessment and observations

FEDERAL LAW (20 USC §1414(B)) REQUIRES SCHOOL DISTRICTS TO DO THE FOLLOWING:

1. Use a variety of assessment tools and strategies to obtain relevant, functional and developmental information and academic instruction;
2. Include information provided by the parent that may assist in determining whether the child is a child with a disability and the content of the child's IEP;
3. Include information related to enabling the child to be involved in and progress in the general curriculum, or, for preschool children, to participate in appropriate activities;
4. **Not use any single procedure as the sole criterion for determining whether a child is a child with a disability or determining an appropriate educational program for the child, and to use technically sound instruments that may assess the relative contribution of cognitive and behavioral factors, in addition to physical or developmental factors.**

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

Three prongs for eligibility (an assessment under the IDEA needs to accomplish):

- (1) Have an impairment, that
- (2) results in an education impact, that
- (3) requires specially designed instruction (34 CFR S300.8)

- IDEA Definition for SLI:

34 C.F.R. §300.7 Child with a disability. (c) Definitions of disability terms. (11) Speech or language impairment means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that *adversely affects a child's educational performance*.

2 TYPES OF DECISIONS

- In terms of evidence based assessment, this means that 2 types of decisions need to be accomplished:
 1. A child has a disability (this is a straight forward yes/no question)
 2. The nature and extent of the special education and related services that the child needs

The second question is complex, because we are answering a few questions compacted into one question. For example:

Are special education services needed even if the student has a disability?

What is the nature of the services?

How does the disability affect the student's ability to access the curriculum?

How does the disorder interact with the curricular demands in ways that impedes access to regular education?

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

- IDEA does not allow the use of any one measure or assessment as the sole criterion in determining if a child has a disability or in determining an appropriate education program (U.S. Department of Education, 2006. CFR 300.304 b. 2).
- Thus, it is required that IEP teams use a variety of both formal and informal assessment tools (U. S. Department of Education, 2006; 34 CFR §300.304 b).
- For example, school-based SLPs can conduct classroom observations, checklists, play-based assessments, language samples, standardized and norm reference tests, narrative assessments, and speech intelligibility measures.
- IDEA (2004) states that when assessing a student for a speech or language impairment, we need to determine **whether or not the impairment will negatively impact the child's educational performance.**

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

Neither federal nor state law defines the term
“adversely affect educational performance.”

So, a review of the court cases interpreting this
phrase is necessary to understand how it has been
applied

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

- Courts have interpreted the phrase to mean that education is adversely affected if, without certain services, the child's condition would prevent her from performing academic and nonacademic tasks and/or from being educated with non-disabled peers. [*Yankton School District v. Schramm*, 93 F.3d 1369 (8th Cir. 1996).]
- In California, the administrative hearing office has found poor grades to be a primary indicator of an adverse effect on educational performance. [*Lodi Unified Sch. Dist.*, SN 371-00; *Capistrano Unified Sch. Dist.*, SN 686-99, 33 IDELR 51; *Ventura Unified Sch. Dist.*, SN 1943-99A; *Murrieta Valley Unified Sch. Dist.*, SN 180-95, 23 IDELR 997.]
- Poor grades and falling behind academically are also examples of adverse effect on educational performance. [*Enterprise Elem. Sch. Dist.*, SN 1055-89.] In addition, a student's condition, which caused declining grades and conduct at school, resulted in an adverse effect on educational performance. [*Sierra Sands Unified Sch. Dist.*, SN 1367-97, 30 IDELR 306.]

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

- Although grades and, standardized test scores may be one measure of educational performance, the law and the courts take a broader view.
- Although some students test well when taking standardized tests, the law does not require poor standardized test scores in order to find an adverse effect on educational performance. The courts have established that a child's educational needs include academic, social, health, emotional, communicative, physical, and vocational needs. [*Seattle School Dist. No. 1 v. B.S.*, 82 F.3d 1493, 1500 (9th Cir. 1996).]

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

- Federal special education law also distinguishes between “educational” performance and “academic” performance and establishes that “educational” performance is a broad concept.
- Congress and the California Legislature used the broader term “educational performance” in eligibility definitions. In addition to grades and standardized tests scores, schools must consider how a child’s emotional, health or other conditions adversely affect her non-academic performance in social, behavioral and other domains as well.

INDIVIDUALS WITH DISABILITIES EDUCATION ACT (IDEA, 2004)

- For example, a response to ASHA's request "The extent of a child's mastery of the basic skill of effective oral communication is clearly includable within the standard of 'educational performance' set by the regulations – that is, academic failure is not a prerequisite for services. It remains the Department's position that the term 'educational performance' is not limited to academic performance. Services cannot be denied as a matter of policy because the adverse effect on educational performance is not reflected in grades or academic achievement."
- 1. Articulation errors drawing negative and undue attention to the child.
- 2. The negative social stigma as the bright child sounds less mature and may appear less intelligent than peers.
- 3. Embarrassment and potential fear of class participation due to articulation errors, and possibly lack of verbal interaction/participation in class, even when intelligible.
- 4. Reduced confidence in reading aloud in class or in small group settings due to articulation errors.
- 5. The potential of being bullied or shunned by peers because of sounding "different" than peers.

COMMON MISCONCEPTIONS/MISUNDERSTANDINGS

- **use of 2 standardized tests for eligibility purposes** because IDEA says to not use single measures! - THE IDEA SAYS NOTHING ABOUT USING MULTIPLE TESTS or using 2 of the same types of tools/strategies, THE IDEA SAYS TO USE A VARIETY OF TOOLS/STRATEGIES!
- use of “severity rating” as a criterion for eligibility - THE IDEA SAYS NOTHING ABOUT SEVERITY as it relates
to eligibility

THIS IS WHAT THE LAW ACTUALLY SAYS:

- (b)(2)(A) use a variety of assessment tools and strategies....
- (b)(2)(B) not use a single measure or assessment as a single criterion...
- (b)(2)(C) use technically sound instruments that may assess...
- (b)(3)(A)(i) ...not to be discriminatory...
- (b)(3)(A)(ii) ..in the language and form most likely to yield accurate information...
- (b)(3)(A)(iii) ... are valid and reliable;
- (b)(3)(A)(v) are administered in accordance with any instruction by producer...
- (b)(3)(D)assessment tools and strategies that provide relevant information that directly assists persons in determining the educational needs...

(20 U.S.C. §1414(b))

ADDITIONAL PROCEDURE: REQUIREMENTS

- A) review existing evaluation data on the child, including—
 - (ii) current classroom-based, local, or State assessments, and classroom-based observations; and
 - (iii) observations by teachers and related services providers; and..
- (20 U.S.C. §1414(c)(1))

How can we analyze the impact of a speech and language disorder in an objective and fair way?

HOW CAN WE JUDGE THE IMPACT OF A SPEECH AND LANGUAGE DISORDER IN AN OBJECTIVE AND FAIR WAY?

Language/Speech
Samples,
Narrative
Analysis

Report Cards,
Work Samples,
State Testing

Parent & Teacher
Input

Curriculum
based measures

Clinical
Impressions/
Observations

WORK SAMPLES: ACADEMIC MATERIALS, ASSIGNMENTS

- Extremely helpful and important in determining impact

Analysis of school performance includes reviewing educational records, collecting evidence of academic performance (including documents from class assignments, independent and group work, homework, class tests, and portfolios of class performance), and completing observations across a variety of educational contexts (classes, playground, extra-curricular activities, lunch, etc.). These observations provide insight into the student's speech language performance during real communication tasks.

(Virginia Department of Education, 2011)

WORK SAMPLES: ACADEMIC MATERIALS, ASSIGNMENTS

- Classwork that demonstrates limited ability when compared to the performance of grade level peers on the same measure

STANDARDIZED TESTS

- Researchers are suggesting that norm-referenced measures should have at least 80% accuracy in discriminating language abilities (**Spaulding, Plante, & Farinella, 2006**).
- Speech-language pathologists should review assessment instruments and consider the diagnostic accuracy, sensitivity, and specificity prior to use in educational evaluations (**Spaulding et al., 2006**).
- “Standardized speech-language tests measure decontextualized communication skills using formalized procedures. Administered outside the normal contexts in which the child communicates, they capture neither the complexities nor the subtle nuances of the communication process,” (**Connecticut State Department of Education, 2008**, p. 23).

IMPORTANCE OF OBSERVATIONS AND RATIONALE FOR A RATING SCALE

- A speech and language evaluation should include systematic observations and a contextualized analysis that involves multiple observations across various environments and situations (Westby et al., 2003).
- According to IDEA (2004), such types of informal assessment must be used in conjunction with standardized assessments.
- Section. 300.532(b), 300.533 (a) (I) (I, ii, iii); 300.535(a)(I) of IDEA states that, “assessors must use a **variety of different tools and strategies** to gather relevant functional and developmental information about a child, including **information provided by the parent, teacher,** and information obtained from classroom-based **assessments and observation.**”
- By using both formal and informal assessments, clinicians are able to capture a larger picture of a student’s speech and language abilities.



STANDARDIZED EXAMPLES OF IMPACT DETERMINATION



THE IMPACT MODEL (Lavi, 2020)

It is designed to analyze the real-life authentic observations of clinicians, parents, and teachers.

Developed based on current literature and examination of real-world challenges faced by individuals with speech and language impairments.

Uses a contextualized, whole language approach to see the impact and the outcome of a speech and/or language impairment on education and social interactions.

RESEARCH AND DEVELOPMENT OF THE IMPACT RATING SCALES

- We began by conducting a thorough research review for each scale's focus (i.e., Social Communication, Articulation and Phonology, Language Functioning).
- Next, we analyzed the most predictive areas in education and social interactions that are affected by poor articulation and phonology, oral expression and spoken language comprehension, and social communication, respectively.
- Additionally, we asked teachers and parents to complete surveys to provide their input on the potential impact of deficits in these areas.

RESEARCH AND DEVELOPMENT OF THE IMPACT RATING SCALES

- Based on our research review, analysis, and input from teachers and parents, we developed and compiled a list of questions.
- A pilot study was then conducted with over 100 students for each of our rating scales.
 - Items were reviewed for content quality, clarity and lack of ambiguity, and sensitivity to cultural issues.
 - Once the pilot studies were validated, some questions were eliminated and supplemental questions were added.
 - Then, a final list of questions was prepared and finalized for each rating scale.
 - The scales were then normed in the second phase of the standardization project.

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- By observing a child's language via informal observation, examinees (i.e., clinician, teacher, and parent) can observe how the child understands language and uses language (e.g., express needs and wants, make requests, converse with peers/friends, etc.), as well as the potential impact a language disorder may have on a child's academic and social life.
- This information can help determine what areas the child has deficits in and how deficits in these areas may impact the child in both the classroom and in the home environment.

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Spoken Language Comprehension

Oral Expression

Language Processing and Integration

Language & Literacy

Social Language Skills

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Spoken Language Comprehension

- The spoken language comprehension rating scale items look at how well an individual understands spoken language. For example, rating scale items look at a child's ability to understand grade level stories, vocabulary, narratives, and his/her ability to answer questions regarding a given story. Additional test items in this area look at an individual's ability to follow along with a conversation, lecture, or discussion, and the ability to recognize when something he/she hears does not make sense.
- *Sample Spoken Language Comprehension Item:* After listening to a lesson, discussion, or story, is the student able to answer who, what, where, and when questions? For example, is the student able to recall the characters, setting, time, place, and what was happening in the story?

Spoken Language Comprehension

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Oral Expression

- The *oral expression* rating scale items look at how well an individual is able to use spoken language. For example, test items investigate if the individual is able to appropriately ask and answer questions, initiate conversations, use narrative storytelling, grade level vocabulary, correct word order, and grammar. Additional test items in this area look at an individual's ability to add comments and questions to a conversation, maintain the topic, form thoughts and ideas, problem solve, negotiate, and use critical thinking skills.
- *Sample Oral Expression Item:* Does the student experience difficulty asking or answering questions in class? For example, does he/she have trouble responding to teacher or peer comments during classroom activities?

Oral Expression

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Language Processing and Integration

- The *language processing and integration* rating scale items look at how an individual follows multi-step instructions, understands figurative language, analogies, and inferences, and sequences details or events. Additionally, rating scale items look at whether an individual's ability to comprehend and use spoken language impacts his/her reading abilities.
- *Sample Language Processing and Integration Item:* Does the student have a difficult time making inferences/implied meaning from given information? For example, does the student have a difficult time “reading between the lines,” making connections, or drawing conclusions?

Language Processing and Integration

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Language and Literacy

- The *language and literacy* rating scale items look at an individual's ability to comprehend and understand what he/she is reading, to distinguish between the main idea and supporting details, and to use his/her own experiences to predict what might happen in grade-level stories. Additionally, literacy rating scale items look at an individual's writing abilities.
- *Sample Language Processing and Integration Item:* Does the student demonstrate an understanding of grade level stories and literature? For example, is the student able to follow along with stories that are read in class and is he/she able to comprehend what is going on in the story?

Literacy

IMPACT LANGUAGE FUNCTIONING RATING SCALE

Social Interactions

- The *social interactions* rating scale items look at how spoken language comprehension and use may impact an individual's social interactions. For example, rating scale items may look at whether an individual is aware of his/her language deficits and how he/she expresses their feelings towards their language disorder. Additionally, rating scale items investigate an individual's confidence regarding his/her communication and how this impacts their participation in conversations and activities with peers, friends, and family.
- *Sample Language Processing and Integration Item:* Does the student's ability to understand and use language make it difficult for him/her to participate fully in school related clubs or activities? For example, does the student's language skills hold them back from joining drama club or yearbook club?

Social Language Skills

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- *Language impairment* involves difficulty in the understanding and/or use of spoken, written, and/or other symbol systems. The disorder may involve:“(1) the form of language (phonology, morphology, syntax); (2) the content of language (semantics); and/or (3) the function of language in communication (pragmatics) in any combination” (ASHA, 2016).
- Spoken language comprehension and oral expression, refers to the understanding and the use of spoken language across various contexts and social situations.
- Approximately 7% of children have deficits in language comprehension or language use and these difficulties can persist into the school-age years and interfere with communication, academics, and social interactions (Tomblin, Records, Buckwalter, Zhang, Smith, & O’Brien, 1997).

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Previous research has suggested that language disorders can be detrimental to a child's development and children whose language falls behind their peers are at an increased risk of academic failure (Durkin, Conti-Ramsden, & Simkin, 2012; Johnson, Beitchman, & Brownlie, 2010), behavioral and psychiatric problems (Conti-Ramsden, Mok, Pickles, & Durkin, 2013, Snowling & Hulme, 2006), unemployment, economic disadvantage, (Parsons, Schoon, Rush, & Law, 2011), and social impairment (Clegg, Hollis, Mawhood, & Rutter, 2005).
- Additionally, longitudinal studies have revealed that language impairments that persist into school age remain in adolescence (Conti-Ramsden & Durkin 2007) and adulthood (Johnson, Beitchman, & Brownlie, 1999; Clegg, Hollis, Mawhood, & Rutter, 2005), often with accompanying literacy deficits (Clegg, Hollis, Mawhood, & Rutter, 2005, Snowling & Hulme, 2000).

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Listening comprehension is a high-order skill that involves both language and cognitive abilities (Florit, Roch, & Levorato, 2013; Kim & Phillips, 2014; Lepola, Lynch, Laakkonen, Silven, & Niemi, 2012). Specifically, listening comprehension refers to one's ability to comprehend spoken language (e.g., conversations, stories/narratives) by extracting and constructing meaning.
- Research has showed that listening comprehension is critical to reading comprehension (Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015; Kim, 2015; Kim, Wagner, & Lopez, 2012; Kim & Wagner, 2015). When children present with reading comprehension deficiencies, there is a heavy focus on word recognition difficulties, including dyslexia and learning disabilities. Difficulties with word recognition are linked to weakness in the phonological domain of language and are often identified early on in the pre-school years (Catts, Fey, Zhang, & Tomblin, 2001).

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- On the other hand, some children demonstrate reading comprehension difficulties despite adequate word reading abilities (Catts, Adlof, & Ellis Weismer, 2006; Nation, Clarke Marshall, & Durand, 2004).
- This group of individuals is known as poor comprehenders.
 - Poor comprehenders are able to read text accurately and fluently at age-appropriate levels, however, they have difficulty understanding what they are reading (Cain & Oakhill, 2007; Nation, 2005).
 - For example, when reading, poor comprehenders have weaknesses in the areas of semantics, syntax (Catts, Adlof, & Ellis Weismer, 2006; Nation & Snowling, 1998; Nation, Snowling, & Clarke, 2007) and more complex parts of language such as idioms, inferencing, comprehension monitoring, and knowledge of text structure (Oakhill, 1984; Cain & Towse, 2008; Cain, Oakhill, & Bryant, 2004; Oakhill & Yuill, 1996).

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Additionally, when we consider narrative comprehension, children with language disorders are less likely to provide correct answers to literal or inferential questions about stories that have been read to them (Gillam, Fargo, & Robertson, 2009; Laing & Kamhi, 2002).
- Since reading comprehension takes time to develop, it is difficult to demonstrate reading comprehension deficits in children before they are able to read accurately and fluently. Thus, these students' reading comprehension deficits may go unnoticed until later grades.
- As such, it is critical that language deficits are identified as early on in development as possible.

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- There is also a strong relationship between oral language abilities and reading ability (Hulme & Snowling, 2013).
- Nation, Clarke, Marshall, and Durand (2004) investigated poor comprehenders' spoken language skills.
 - The results of this study found that these students were less skilled than those in the typically developing group on semantic tasks (e.g., vocabulary and word knowledge), morphosyntax (e.g., past tense inflection, sentence comprehension) and aspects of language use (e.g., understanding figurative language).
- Research also suggests that students with expressive language difficulties are four to five times more likely than their peers to present with reading difficulties (Catts, Fey, Zhang, & Tomblin, 2001).
 - For example, Zielinski, Bench, and Madsen (1997) explored expressive language delays in preschoolers and found that these children were more likely to have difficulties with reading performance.

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Poll and Miller (2013) also reported that when children are 8 years old, expressive language delays could be a significant risk factor for poor oral language and reading comprehension.
- Lee (2011) discovered that expressive language development predicts comprehension of reading passages in both third and fifth grade students.
- Vocabulary can also play an important role early on in development as was demonstrated in Duff, Reen, Plunkett, and Nation's (2015) study that found infant vocabulary between 16 and 24 months is predictive of reading comprehension early on in school instruction years.
- Additionally, Pysridou, Eklund, Poikkeus, and Torppa's study (2018) found that expressive language ability at age 2–2.5 years old is associated with reading comprehension in ages 8–16 years old.

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Listening comprehension and oral language abilities can also be important when we consider writing development (Kim, Al Otaiba, Wanzek, & Gatlin, 2015; Hulme & Snowling, 2013).
- Children with language impairments have been found to show grammatical errors (Gillam & Johnston, 1992; Scott & Windsor, 2000; Windsor, Scott, & Street, 2000) and spelling errors in their written texts.
- The spelling errors are similar to those found in children with dyslexia (Puranik, Lombardino, & Altmann, 2007), however, an individual's ability to create and think of new ideas appears to be specific to difficulties within the language system (Bishop & Clarkson, 2003; Puranik, Lombardino, & Altmann, 2007).

THEORETICAL & CONTEXTUAL BACKGROUND

IMPACT LANGUAGE FUNCTIONING RATING SCALE

- Numerous studies have explored the difficulties that school-age children with language impairment have with telling stories. For example, when compared to typically developing children, children with language deficits tend to compose stories that contain fewer words and utterances (Moyano & McGillivray, 1988 [as cited in Hughes, McGillivray, & Schmidek, 1997]), fewer story grammar components (Paul, 1996), reduced sentence complexity (Gillam & Johnston, 1992), fewer complete cohesive ties (Liles, 1985), increased grammatical errors (Liles, Duffy, Merritt, & Purcell, 1995; Norbury & Bishop, 2003), and poorer overall story quality (Gillam, McFadden, & van Kleeck, 1995; McFadden & Gillam, 1996).

PSYCHOMETRIC PROPERTIES

- When selecting an assessment for an evaluation, it is important to consider whether it is truly a *good* assessment tool.
- A good assessment is one that produces results that will benefit the individual being tested or society as a whole (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014).
- There are a few ways we can examine whether a test is considered a good and strong assessment. We can take a look at the standardization, normative information, and the psychometric properties of each test.

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Normative Sample

- Previous research has suggested that the inclusion of children with disabilities in a normative sample can have a negative impact on a test's ability to differentiate between children with disorders and children who are typically developing (Peña, Spaulding, & Plante, 2006). Thus, normative data for the *IMPACT Rating Scales* was based solely on typically developing children to allow for high sensitivity and specificity.
- Since the purpose of the *IMPACT Rating Scales* is to help to identify speech and language disorders and the impact of these disorders, it was critical to exclude students from the normative sample who had diagnoses that are known to influence each area of speech and language (Peña, Spaulding, & Plante, 2006).
- For example, students who had previously been diagnosed with a specific language impairment or learning disability were not included in the normative sample for the *IMPACT Rating Scales*. Further, students were excluded from the normative sample if they were diagnosed with autism spectrum disorder, intellectual disability, hearing loss, neurological disorders, or genetic syndromes.

Normative Sample

Table 4.1
Representation of the Sample, by Age Group

Age Group	Age	<i>N</i>	%
1	5-0 to 5-11	84	8
2	6-0 to 6-11	97	9
3	7-0 to 7-11	112	10.5
4	8-0 to 8-11	116	11
5	9-0 to 9-11	102	9.5
6	10-0 to 10-11	85	8
7	11-0 to 11-11	95	9
8	12-0 to 12-11	96	9
9	13-0 to 13-11	86	8
10	14-0 to 14-11	85	8
11	15-0 to 21-0	106	10
Total Sample		1064	100%

1064 typically developing examinees across 11 age groups (in 17 states (Arizona, California, Colorado, Nevada, Idaho, Illinois, Iowa, Kansas, Ohio, Minnesota, Florida, New York, Pennsylvania, Florida, South Carolina, Texas, Washington)).

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Sensitivity and Specificity

- Strong sensitivity and specificity (i.e., 80% or stronger) is needed to support the use of a test in its identification of the presence of a disorder or impairment.
- Sensitivity measures how well the assessment will accurately identify those who truly have a speech sound disorder (Dollaghan, 2007).
- Specificity measures the degree to which the assessment will accurately identify those who do not have a speech sound disorder, or how well the test will identify those who are “typically developing” (Dollaghan, 2007).

Normative Sample

Table 5.1 IMPACT Language Functioning Rating Scale sensitivity, specificity and likelihood ratios

Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
5:0-5:11	76	83	79	3.48	.14
6:0-6:11	77	81	80	3.92	.09
7:0-7:11	78	84	79	4.34	.11
8:0-8:11	78	80	83	5.13	.21
9:0-9:11	77	82	80	3.32	.12
10:0-10:11	77	83	81	5.17	.17
11:0-11:11	76	82	84	4.16	.09
12:0-12:11	77	88	91	5.11	.08
13:0-13:11	78	91	88	5.13	.18
14:0-14:11	78	86	91	4.33	.19
15:0-15:11	78	82	79	5.87	.21
16:0-21:0	77	86	88	6.21	.13

Age groups 16:0-21:0 are reported together as there were no age-related changes detected after the age of 16.

Scaled Score Means (and Standard Deviations) of Subtests for Two Clinical Groups and a Demographically Matched Typically Developing Group, (N= 212)

	SLC Impairment group (n=67)	EL Impairment group (n=41)	TD group (n=71)	p-value*
Clinician ^{a,b,c} <small>~~~~~</small>	.82	.79	.81	.87
Teacher ^{a,b,c} <small>~~~~~</small>	.91	.84	.83	.68
Parent ^{a,b,c} <small>~~~~~</small>	.84	.81	.78	.73

| Abbreviation: SLC, Spoken language comprehension; EL, Expressive Language; and TD, Typically Developing

*Kruskal-Wallis Analysis of Variance test

^a significant difference between ASD and TD groups

^b significant difference between SCD and TD groups

^c significant difference between SCD and ASD groups

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Content Validity

- The validity of a test determines how well the test measures what it purports to measure. Validity can take various forms, both theoretical and empirical. This can often compare the instrument with other measures or criteria, which are known to be valid (Zumbo, 2014). Expert opinion was elicited for all of the IMPACT Rating Scales.
- For example, 29 speech language pathologists (SLPs) reviewed the *IMPACT Articulation and Phonology Rating Scale*.
 - All SLPs were licensed in the state of California, held the Clinical Certificate of Competence from the American Speech-Language-Hearing Association, and had at least 5 years of experience in assessment of children with speech sound disorders.
 - Each of these experts was presented with a comprehensive overview of the rating scale descriptions, as well as rules for standardized administration and scoring.
 - They all reviewed 6 full-length administrations.

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Content Validity cont'd

- Following this, they were asked 30 questions related to the content of the rating scale and whether they believed the assessment tool to be an adequate measure of speech sound disorders. For instance, their opinion was solicited regarding whether the questions and the raters' responses properly evaluated the impact of speech sound disorders on educational performance and social interaction. The reviewers rated each rating scale on a decimal scale.
- All reviewers agreed that the *IMPACT Articulation and Phonology Rating Scale* is a valid informal observational measure to evaluate speech skills and to determine the impact on educational performance and social interaction, in students who are between the ages of 5 and 21 years old.

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Criterion Validity

- Criterion validity measures how well one measure predicts an outcome for another measure.
- In assessing criterion validity, the *IMPACT Articulation and Phonology Rating Scale* was correlated to other measures of articulation and phonology: *Arizona Articulation and Phonology Scale - Fourth Edition* (Arizona-4; Fudala & Stegall, 2017) and the *Diagnostic Evaluation of Articulation and Phonology* (DEAP; Dodd, Holm, Crosbie, & Ozanne, 2003).
- Time between test administrations ranged from the same day to 5 days.
- The concurrent validity was assessed using Pearson's correlation among all measures. Correlation coefficients of ≥ 0.7 are recommended for same-construct instruments while moderate correlations of ≥ 0.4 to ≤ 0.70 are acceptable. The level of significance was set at $p \leq 0.05$. When assessing validity, the *IMPACT Articulation and Phonology Rating Scale* was substantially correlated with the *DEAP* and the *Arizona-4*: 0.87, and 0.83 respectively, $p < 0.001$.

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

Response Bias

- Research has also suggested that we consider the potential impact of biases when evaluating an assessment tool.
- Responses to questionnaires, tests, and scales, may be biased for a variety of reasons. For example, response bias may occur consciously or unconsciously and when it does occur, the reliability and validity of our measure will be compromised.
- The *IMPACT Rating Scales* use balanced set of questions in order to protect against response biases.
- A balanced scale is a test or questionnaire that includes some items that are positively keyed and some items that are negatively keys.

PSYCHOMETRIC PROPERTIES OF THE IMPACT RATING SCALES

- Here is an example taken from the *IMPACT Social Communication Rating Scale*. Items on this scale are rated on a 4-point scale ("never," "sometimes," "often," and "typically"). Now, imagine if we asked a teacher to answer the following two items regarding one of their students:
 1. Appears confident and comfortable when socializing with peers.
 2. Does not appear overly anxious and fidgety around group of peers.
- Both of these items are positively keyed because a positive response indicates a stronger level of social language skills. To minimize the potential effects of acquiescence bias ("yea-saying and nay-saying" when an individual consistently agrees or disagrees [Danner & Rammstedt, 2016]), the test creator may revise one of these items to be negatively keyed. For example:
 1. Appears confident and comfortable when socializing with peers.
 2. Appears overly anxious and fidgety around group of peers.
- Now, the first item is keyed positively and the second item is keyed negatively. The revised scale, which represents a balanced scale, helps control acquiescence bias by including one item that is positively keyed and one that is negatively keyed.
- To read more about the psychometric properties of each *IMPACT Rating Scale*, please review the technical manual for each scale.

CONTACT

- Email: adriana@laviinstitute.com