

## Limitations of Norm-Referenced Tests

August 25, 2021

Andrea Bertone, MS, CCC-SLP  
Education Consultant

John Hellmuth, PhD, CCC-SLP  
University of Wisconsin-Milwaukee



## New Rule Effective August 1, 2021

Individualized Education Program (IEP) teams must use the new criteria to identify a speech or language impairment for referrals for special education dated on or after August 01, 2021.

The updated rule may be found here:  
[https://docs.legis.wisconsin.gov/code/register/2021/786a3/register/cr/cr\\_20\\_074\\_rule\\_text/cr\\_20\\_074\\_rule\\_text](https://docs.legis.wisconsin.gov/code/register/2021/786a3/register/cr/cr_20_074_rule_text/cr_20_074_rule_text)



## Housekeeping – Meeting Resources

- In the [Participant Folder](#) you will find the google slide deck as well as other relevant resources for the meeting.
- Previously shared resources:
  - [John's Youtube video intro](#)
  - [Significant Discrepancy Document](#)
- Once finalized, the recorded meeting and resources will be all be available to view on the [WI DPI Speech-Language Website](#)
- If you have any questions, feel free to put them in the chat or on [Jamboard](#)
- Future webinars in this series: [link](#) to flyer
- Must read: [Evaluation and Eligibility in Schools](#) by Marie Ireland and Barbara Conrad



---

---

---

---

---

---

---

---

---

---

---

---

## Today's Learning Objectives

1. Participants will review the comprehensive assessment model.
2. Participants will delve into the limitations of norm-referenced tests.
3. Participants will engage in case study work to design more comprehensive assessments.
4. Participants will learn of resources and next steps for further learning and support.



## Learning Objective #1

Participants will review the comprehensive assessment model.



## What is an Evaluation for Special Education?

“Evaluation means procedures used . . . to determine whether a child has a disability and the nature and extent of the special education and related services that the child needs.”

34 CFR 300.15



---

---

---

---

---

---

---

---

---

---

---

---

## Planning Assessment

“No single measure or assessment is used as the sole criterion”

IDEA 34 CFR §300.304 (b)(2)



## Comprehensive Assessment Model

### 4 Part Model for Comprehensive Assessment

#### Academic Activities:

- Artifact analysis
- Curriculum-based assessment
- Observations in school (natural settings)
- Educational records

#### Speech-Language Probes:

- Case history
- Interviews
- Language/Narrative samples
- Stimulability
- Dynamic assessment
- Play-based assessment

#### Contextualized Tests:

- Norm-referenced measures of academic achievement
- Curriculum benchmarks

#### Decontextualized Tests:

- Norm-referenced speech-language tests (parsed skills: articulation, semantics, syntax, morphology, fluency, etc.)

Ireland, Marie. "The Real Requirements Behind Eligibility Decision Making in Schools". Lecture. ASHA Connect, Chicago, IL, July 19, 2019.

## Limitations of Norm-Referenced Tests

**\*Decontextualized tests are just one-fourth of the model**

Norm-referenced tests:

- Vary in their accuracy
- Assumes all children have had the same experiences and opportunities
- Skills evaluated are associated with culture or socioeconomic status, such as vocabulary, rather than a true language impairment.

Kate Crowley, Leader's Project, Columbia University  
<https://www.leadersproject.org/2012/11/26/aspbt-pg-dynamic-assessment/>



---

---

---

---

---

---

---

---

---

---

---

---

## Current Research

- Over-reliance on formal assessments in identification of speech-language impairments (Fulcher-Rood, Castilla-Earls, & Higginbotham 2018; Ireland & Conrad 2016)
- Test selection is often not dependent on student need or psychometric properties of the test (Merrell & Plante 1997; Betz, Eickhoff, & Sullivan 2013; Fulcher-Rood, Castilla-Earls, & Higginbotham 2020).



## Issues with Norm-Referenced Test Data

- Does not address disorder within dialect
- Normative sample is limited
- Documentation of educational impact is missing
- Disability determination is not in compliance if solely based on standardized tests



## Converging Evidence Framework



Castilla-Earls, Bedore, Fabiano-Smith, Pruitt-Lord, Rojas, Peña and Restrepo 2020

---

---

---

---

---

---

---

---

---

---

---

---

## Considerations for Students Who Are English Learners

- Evaluate the child in both languages *unless it is not feasible to do so*.
- Recommendation to use a translator to obtain case history to determine extent of knowledge and use of home language.
- Interview school staff (gen ed teacher; EL teacher) regarding language use across settings.
- Use dynamic assessment to determine student's modifiability.

Orellana et al. 2019;  
Roseberry-McKibbin 2021



## Top 4 Ways to Make Your Assessment Culturally Relevant

1. Review of existing data: ask questions about the whole child.
2. Do more than standardized norm-referenced assessments: *Start with functional performance and add standardized assessment if needed.*
3. Include parent information: consider cultural expectations vs. school expectations for behavior and participation.
4. When considering norm-referenced assessment, look at normative sample and test items for potential bias.



## Learning Objective #2

Participants will delve into the limitations of norm-referenced tests.



---

---

---

---

---

---

---

---

---

---

---

---

## John's Disclosures

- Primary research on descriptive assessments
- Think a lot about limitations of norm-referenced tests
- I do believe norm-referenced tests have a time and a place



## Plan for today

- Quick overview of Norm Referenced Tests
- 4 major limitations
  - Biases
  - Psychometrics
  - Decision making
  - Functional Communication
- Put questions in chat - answered between each section



## Standardized Tests

- Uniform and consistent
- Same directions, items, scoring, prompts, etc.



---

---

---

---

---

---

---

---

---

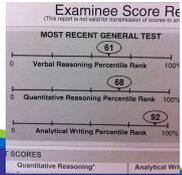
---

---

---

## Norm-referenced Tests

- Compare performance to peers
- Get measure of "relative standing" (percentile, standard score)



## Standardized vs. Norm-referenced Tests

- All norm-referenced tests are standardized
- All standardized tests are not necessarily norm referenced

## Why we love norm-referenced tests

- Standardized administration promotes consistency (aka, *reliability*)
- Can "objectively" quantify a student's performance
  - 75th percentile
  - Standard score of 71
  - > -1.75 SD below the mean

---

---

---

---

---

---

---

---

---

---

---

---

## Moving Forward

- Will examine several limitations of NR tests
- Will not lose sight of how they can effectively be used in comprehensive assessment



## Limitation 1: Many potential biases



## More Obvious Biases



---

---

---

---

---

---

---

---

---

---

---

---

## English Proficiency

- English tests not appropriate for students with limited English proficiency
- Translations quite problematic
- Bilingual student with English proficiency
  - English performance < Monolingual English-speaking peers



## Dialect Spoken

- Many norm-referenced tests examine inflectional morphology (e.g., walk/ing; talk/s)
- Dropping morphemes a systematic rule within African American English



## Obvious Biases

- Norm-referenced tests add little value
- Standard scores should not be reported
- Be sure to check out Courtney Seidel's upcoming presentations!



---

---

---

---

---

---

---

---

---

---

---

---

## Less Obvious Biases



## Format Bias

- Student not as familiar with the format of the test
- Less experience means poorer performance
- Scores artificially low



## True for us!



---

---

---

---

---

---

---

---

---

---

---

---

## Western Values (Helms, 1992)

- Individual achievement
- Competitiveness
- Value of speed
- Adhering to time schedules
- Emphasis on logical thought & objectivity
- Belief that performance can be objectified



## Western Values

*Most tests assume that children share common beliefs/values*

*If a student does not share beliefs/values, tasks will be unfamiliar and test is biased*



## Value of Speed

**Expressive Vocabulary Test: Must answer in 5 seconds**

	Student shares value of speed
Strong vocabulary	Answer quickly
Weak vocabulary	Answer slowly because of poor lexical representations



---

---

---

---

---

---

---

---

---

---

---

---

## Value of Speed

Expressive Vocabulary Test: Must answer in 5 seconds

	Student shares value of speed	Does not share value of speed
Strong vocabulary	Answer quickly	Answer slowly because this is really weird
Weak vocabulary	Answer slowly because of poor lexical representations	Answer slowly because this is weird (and hard)



## Individualistic vs Collective Cultures

- **Individualistic culture**
  - Students expected to perform their best
  - Primary interest on personal achievement
- **Collective culture**
  - Students more concerned about group performance
  - Emphasis on helping struggling students



Individualistic  
"I need to do my best  
so they know I'm smart"



Collective  
"This is kind of weird, I wonder  
how the others would like it"



---

---

---

---

---

---

---

---

---

---

---

---

## Frustration Tolerance

- Assumption is that children tolerate a certain level of frustration to achieve ceiling
- Most students from the dominant culture have experience and know this expectation
- What if this is something new for your student?



## Format of Test Items

Is your student familiar with the testing format?

- Interacting with picture probes
- Asking "known" information questions
- Playing with an unfamiliar adult



## C'mon, really?

"All my students can finish my tests"

Not an either/or issue - relative impact on performance



---

---

---

---

---

---

---

---

---

---

---

---

## Fagundes, Haynes, Haak, & Moran (1998)

- N=24 students (12 African American & 12 white)
- Preschool Language Assessment Instrument
  - Traditional administration (less familiar format for AA children)
  - Thematic administration (more familiar format for AA children)



## Biases: What do I do?



## New Rule: Choose at Least Two

5. Following consideration of the child's age, culture, language background, or dialect, the child demonstrates a language impairment in the area of language form, content or use, as evidenced through an observation in a natural environment and by measurement of at least two of the following:
- a. Language sample analysis.
  - b. Dynamic assessment.
  - c. Developmental scales or another criterion-referenced assessment.
  - d. Significant discrepancy from typical language skills on a norm-referenced assessment of comprehensive language.



---

---

---

---

---

---

---

---

---

---

---

---

### If you have a student who is an English Learner ....

- Tests for monolingual English speakers are not appropriate (student would not be represented in comparison sample)
- How well can the student learn new language features?  
*Dynamic Assessment*
- How is the student doing in school? *Observations, interviews, academic tests, writing samples*



### "This student only talks with their teacher...."

- Hey, you're an SLP, she will talk to you!
- If no SLP magic, then what?
- Can you analyze language used with teacher or parent?  
*Language sample analysis*
- Can you observe across different settings?
- Does the student's communication change with different levels of scaffolding? *Dynamic assessment*



### Student took a test, but didn't seem to "get it"

- Can we do digging to see if any format biases?
- Can we look at some specific language features in more detail? *Probes; dynamic assessment*
- Can the student describe technical processes and participate in science class? *Language Sample Analysis; academic record; teacher interview*



---

---

---

---

---

---

---

---

---

---

---

---

## Any questions on bias?



## Limitation 2: Not as accurate as we think



## Psychometrics Crash Course

- **Reliability - Data are stable & repeatable**
  - Similar when take a test on Monday & Friday
  - Similar across odd and even questions
- **Validity - Measure what you think you are measuring**
  - Score on CELF similar to performance on narrative
  - Performance on TACL related to listening in the classroom



---

---

---

---

---

---

---

---

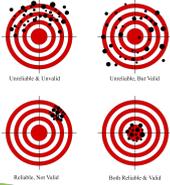
---

---

---

---

## Reliability & Validity



## Diagnostic Accuracy

Do test data result in an accurate diagnosis?

- Sensitivity (*true positive*) = accurate diagnosis of impairment
- Specificity (*true negative*) = accurate diagnosis of no impairment



## Diagnostic Accuracy

	Child actually has Dx	Child does not have Dx
Fail test	True Positive	
Pass test		True Negative



---

---

---

---

---

---

---

---

---

---

---

---

## Diagnostic Accuracy

	Child actually has Dx	Child does not have Dx
Fail test	True Positive	False Positive
Pass test	False Negative	True Negative



## Cut-off Score

- Perform above the score = typically developing
- Perform below the score = impairment/disorder
- Historically,  $-1.75$  SD in Wisconsin



## Plante & Vance (1994)

- 20 children with DLD; 20 with normal language
- Administered 4 tests with decent published psychometrics
- Found best cut-point to differentiate groups



---

---

---

---

---

---

---

---

---

---

---

---

## Things to Note

- Pretty old study
- Used non-mainstream tests
- Tests have gotten better



## More Recent Data

Spaulding et al. (2006) reviewed 43 language tests for sensitivity & specificity



## What is the gold standard?

	Experimental Measure	Common Gold Standard
Validity	Norm Referenced Test	Other Norm Referenced Test
Validity (2nd edition)	New version of test	Old version of test
Diagnostic Accuracy	Norm Referenced Test	High vs. Low performance on other NR test



---

---

---

---

---

---

---

---

---

---

---

---

## What is the gold standard?

	Experimental Measure	Common Gold Standard	Less Common Gold Standard
Validity	Norm Referenced Test	Other Norm Referenced Test	Functional language skills
Validity (2nd edition)	New version of test	Old version of test	Ability to communicate
Diagnostic Accuracy	Norm Referenced Test	High vs. Low performance on other ILL test	Results of comprehensive clinical assessment



## Test of Language Development: P4

- Sensitivity = .74; Specificity = .87
- Based on +/- Standard Score of 90 on TOLD
- Gold standard +/- 90 on Pragmatic Language Outcome Scale



## Why do we see differences across tests?

### Differences in the norming sample

- Race, ethnicity, SES, language spoken, dialect spoken, inclusion of children with disabilities, luck of the draw...
- Different samples have different performance
- Especially noticed with culturally & linguistically diverse students!



---

---

---

---

---

---

---

---

---

---

---

---

## Why is performance so different for some students?

### Biases in tests

- If not familiar with a test (or task within test), will perform worse
- If get punished for dialect or experience, will score worse



## What am I supposed to do?

### Learn about your tests

- reliability & validity
- sensitivity & specificity
- representation of your students
- account for culture, language, dialect



## Use Clinical Expertise

If you "know" about a certain test, you are probably accurate

- Test A scores my students too high
- Test B scores my students too low
- *Look for data to support this (test manual; articles)*



---

---

---

---

---

---

---

---

---

---

---

---

## Understand the Limitations of Norm-Referenced Tests

- In general, modern tests have good psychometrics (*though not for culturally & linguistically diverse students*)
- Can't rely on this one piece of data
- *Need to corroborate with other data!*



**Considerations for Language Impairment**

This worksheet can be used by IEP teams in considering whether a student meets criteria for a speech or language impairment. For each item, circle the item that best represents the student's performance. Students must demonstrate the presence of a delay in communication skills that also has an educational impact in order to meet the criteria for a speech or language impairment in Wisconsin. When a valid comparison to a normative sample cannot be made or a student has significant impairments, consider completion of the [Functional Communication Summary](#) worksheet.

	Academic Activities <small>(e.g., writing samples, reading fluency records, observations in natural settings, records)</small>	Academic Tests and Measurements <small>(e.g., districtwide and statewide assessments, curriculum benchmarks)</small>	Speech Language Pathology Probes <small>(e.g., language sample, intelligibility, fluency, dynamic assessment, play-based assessment, observations, case history)</small>	Speech Language Pathology Norm-Referenced Tests and Measurements <small>(e.g., norm-referenced assessments with appropriate variability and specificity)</small>
<b>No Apparent Impact</b>	Performs similarly to peers in most areas	Performs similarly to peers in most areas	May indicate differences from Standard American English Demonstrates improvements during dynamic assessment	3 or 2 composite scores* at or above: • mean to -1 SD • +0.50 • +17th percentile
<b>Minimal Impact</b>	Evidence of struggle with one or more areas when compared to peers Evidence of occasional difficulty with test/skills	Evidence of struggle with one or more areas when compared to peers	May indicate differences from Standard American English Demonstrates improvements during dynamic assessment Occasional difficulty with pragmatic, semantic or syntax, morphological skills	3 or 2 composite scores* documenting: • -1 to -1.5 SD • +0.50-77.5% • 16th-7th percentile
<b>Moderate Impact</b>	Evidence of struggle in most areas when compared to peers Evidence of difficulty with test/skills	Evidence of struggle in most areas when compared to peers	Demonstrates limited improvement during dynamic assessment Frequent difficulty with pragmatic, semantic or syntax, morphological skills	3 or 2 composite scores* documenting: • -1.5 to -2 SD • -7.5 to -50% • 4th-30th percentile
<b>Substantial Impact</b>	Evidence of very limited ability in most areas Evidence of absence of test/skills	Evidence of very limited ability in most areas	Demonstrates very limited improvement during dynamic assessment Extreme difficulty with pragmatic, semantic or syntax, morphological skills	3 or 2 composite scores* documenting: • -2 or greater SD • 4th or below 5% • below 2nd percentile



## Scenario 1

- Test seems appropriate for student
- Administer a favorite test
- Student scores 78 (~1.5 SD below mean)
- **I'm confident in the data, BUT**
  - *Need to corroborate with other data*
  - *Need other data to describe the nature of the disorder & develop IEP goals*




---



---



---



---



---



---



---



---



---



---



---



---

Considerations for Language Impairment			
<p>This worksheet can be used by IEP teams in considering whether a student meets criteria for a speech or language impairment. For each item, circle the item that best represents the student's performance. Students must demonstrate the presence of a delay in communication skills that also has an <b>educational impact</b> in order to meet the criteria for a speech or language impairment in Wisconsin. When a valid comparison to a normative sample cannot be made or a student has significant impairments, consider completion of the <a href="#">Functional Communication Summary</a> worksheet.</p>			
	<p><b>Academic Activities</b>            (e.g., writing samples, reading routine, report observations in natural settings, records)</p>	<p><b>Academic Tests and Measurements</b>            (e.g., districtwide and statewide assessments, curriculum benchmarks)</p>	<p><b>Speech-Language Pathology Probes</b>            (e.g., language sample, intelligibility, comprehensibility, dynamic assessment, play-based assessments, case history, case history)</p>
	<p><b>No Apparent Impact</b>            Performs similarly to peers in most areas.</p>	<p>Performs similarly to peers in most areas.</p>	<p>May indicate differences from Standard American English.            Demonstrates improvements during dynamic assessment.</p>
	<p><b>Minimal Impact</b>            Evidence of struggle with one or more areas when compared to peers.            Evidence of occasional difficulty with 'vocal' skills.</p>	<p>Evidence of struggle with one or more areas when compared to peers.            Evidence of occasional difficulty with 'vocal' skills.</p>	<p>May indicate differences from Standard American English.            Demonstrates limited improvements during dynamic assessment.            Occasional difficulty with pragmatic, semantic or syntax, morphological skills.</p>
	<p><b>Moderate Impact</b>            Evidence of struggle in most areas when compared to peers.            Evidence of difficulty with 'vocal' skills.</p>	<p>Evidence of struggle in most areas when compared to peers.            Evidence of difficulty with 'vocal' skills.</p>	<p>Demonstrates limited improvement during dynamic assessment.            Frequent difficulty with pragmatic, semantic or syntax, morphological skills.</p>
	<p><b>Substantial Impact</b>            Evidence of very limited ability in most areas.            Evidence limited or absence of 'vocal' skills.</p>	<p>Evidence of very limited ability in most areas.</p>	<p>Demonstrates very limited improvement during dynamic assessment.            Extreme difficulty with pragmatic, semantic or syntax, morphological skills.</p>
			<p><b>Speech-Language Pathology Norm-Referenced Tests and Measurements</b>            (e.g., norm-referenced assessments with appropriate sensitivity and specificity)</p>
			<p>1 or 2 composite scores* at or above:            • mean to 1 SD            • 85-90 #            • 17th percentile</p>
			<p>1 or 2 composite scores* documenting:            • 1 to 1.5 SD            • 68 to 77.5 #            • 16th-7th percentile</p>
			<p>1 or 2 composite scores* documenting:            • 1.5 to 2 SD            • 50-75 #            • 16th-3rd percentile</p>
			<p>1 or 2 composite scores* documenting:            • 2 or greater SD            • 69 or below 5 #            • below 3rd percentile</p>

## Scenario 2

- Student didn't seem to "get" the norm-referenced test
- Standard score = 78
- Language sample, teacher report, academic review seem pretty good
- *Norm-referenced data may not be accurate, and that's OK*

Considerations for Language Impairment			
<p>This worksheet can be used by IEP teams in considering whether a student meets criteria for a speech or language impairment. For each item, circle the item that best represents the student's performance. Students must demonstrate the presence of a delay in communication skills that also has an <b>educational impact</b> in order to meet the criteria for a speech or language impairment in Wisconsin. When a valid comparison to a normative sample cannot be made or a student has significant impairments, consider completion of the <a href="#">Functional Communication Summary</a> worksheet.</p>			
	<p><b>Academic Activities</b>            (e.g., writing samples, reading routine, report observations in natural settings, records)</p>	<p><b>Academic Tests and Measurements</b>            (e.g., districtwide and statewide assessments, curriculum benchmarks)</p>	<p><b>Speech-Language Pathology Probes</b>            (e.g., language sample, intelligibility, comprehensibility, dynamic assessment, play-based assessments, interviews, case history)</p>
	<p><b>No Apparent Impact</b>            Performs similarly to peers in most areas.</p>	<p>Performs similarly to peers in most areas.</p>	<p>May indicate differences from Standard American English.            Demonstrates improvements during dynamic assessment.</p>
	<p><b>Minimal Impact</b>            Evidence of struggle with one or more areas when compared to peers.            Evidence of occasional difficulty with 'vocal' skills.</p>	<p>Evidence of struggle with one or more areas when compared to peers.            Evidence of occasional difficulty with 'vocal' skills.</p>	<p>May indicate differences from Standard American English.            Demonstrates limited improvements during dynamic assessment.            Occasional difficulty with pragmatic, semantic or syntax, morphological skills.</p>
	<p><b>Moderate Impact</b>            Evidence of struggle in most areas when compared to peers.            Evidence of difficulty with 'vocal' skills.</p>	<p>Evidence of struggle in most areas when compared to peers.            Evidence of difficulty with 'vocal' skills.</p>	<p>Demonstrates limited improvement during dynamic assessment.            Frequent difficulty with pragmatic, semantic or syntax, morphological skills.</p>
	<p><b>Substantial Impact</b>            Evidence of very limited ability in most areas.            Evidence limited or absence of 'vocal' skills.</p>	<p>Evidence of very limited ability in most areas.</p>	<p>Demonstrates very limited improvement during dynamic assessment.            Extreme difficulty with pragmatic, semantic or syntax, morphological skills.</p>
			<p><b>Speech-Language Pathology Norm-Referenced Tests and Measurements</b>            (e.g., norm-referenced assessments with appropriate sensitivity and specificity)</p>
			<p>1 or 2 composite scores* at or above:            • mean to 1 SD            • 85-90 #            • 17th percentile</p>
			<p>1 or 2 composite scores* documenting:            • 1 to 1.5 SD            • 68 to 77.5 #            • 16th-7th percentile</p>
			<p>1 or 2 composite scores* documenting:            • 1.5 to 2 SD            • 50-75 #            • 16th-3rd percentile</p>
			<p>1 or 2 composite scores* documenting:            • 2 or greater SD            • 69 or below 5 #            • below 3rd percentile</p>

## Questions on Psychometrics & Diagnostic Accuracy?



## Limitation 3: Decisions Aren't Always Clear Cut



## What are norm-referenced tests designed for?

- Show an individual's *relative standing* compared to a reference group
- aka, rank-order a group of individuals



---

---

---

---

---

---

---

---

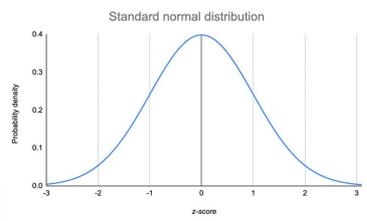
---

---

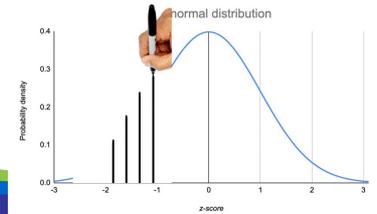
---

---

## Relative Standing



## Determine a Cut-Point



## Significant Discrepancy

At what point is performance significantly different than peers?

*Not a clear question*



---

---

---

---

---

---

---

---

---

---

---

---

## What is significant?

- **General research criterion is -1.25 SD**
  - ~7.5% of population has DLD/SLI
  - Many of these kids do not receive services
- **Historical WI criterion is -1.75 SD**
- **Common criterion is -1.5 SD**
- **All of these are arbitrary**



## Are the numbers accurate?

- **Many potential biases**
- **Acknowledged imprecision**
  - Confidence interval
  - Standard error of measurement



## How I approach interpreting standard scores

- Is the test appropriate?
- Is student clearly high?
- Is student clearly low?
- In-between (around 1.5)



---

---

---

---

---

---

---

---

---

---

---

---

## When Performance is Clearly “Normal”

- Student scores on low end of normal (e.g.,  $-0.75$  Std Dev,  $\pm 0.2$ )
- I have confidence that the test data were valid
- Converging data from other sources
  - Teacher report: *He's not the strongest student, but other students certainly struggle more*
  - Academic record: *Low end of normal, consistent with my testing*



## When Performance is Clearly Low

- Student scores low (e.g.,  $-2.0$  Std Dev,  $\pm 0.2$ )
- Confident that data are accurate
- Converging evidence
  - Teacher notes substantial struggles
  - Receiving other special education services
  - Teacher report and other assessment data show aspects of educational impact (*this is what I'm REALLY interested in with this student!*)



## When Performance Is In the Middle

- Student scores beyond low end of normal, but above the lowest percentiles (e.g.,  $-1.25$  -  $-1.75$  Std Dev)
- Detective hat
  - How well do I trust the norm referenced data?
  - What are the converging evidence? *Dynamic assessment, language sample, academic performance*
  - *What is the educational impact?*



---

---

---

---

---

---

---

---

---

---

---

---

**Considerations for Language Impairment**

This worksheet can be used by IEP teams in considering whether a student meets criteria for a speech or language impairment. For each item, circle the item that best represents the student's performance. Students must demonstrate the presence of a delay in communication skills that also has an **educational impact** in order to meet the criteria for a speech or language impairment in Wisconsin. When a valid comparison to a normative sample cannot be made or a student has significant impairments, consider completion of the [Functional Communication Summary](#) worksheet.

	Academic Activities	Academic Tests and Measurements	Speech Language Pathology Probes	Speech Language Pathology Norm Reference Tests and Measurements
	e.g., writing samples, reading running record, observations in natural settings, records	e.g., district-wide and statewide assessments, curriculum benchmarks	e.g., language sample intelligibility, fluency/ability dynamic assessment, play-based assessments, case review, case history	e.g., norm-referenced assessments with appropriate validity and specificity
<b>No Apparent Impact</b>	Performs similarly to peers in most areas.	Performs similarly to peers in most areas.	May indicate differences from Standard American English. Demonstrates improvements during dynamic assessment.	1 or 2 composite scores* at or above: <ul style="list-style-type: none"> <li>mean to 1 SD</li> <li>• 85/50.4</li> <li>• 17th percentile</li> </ul>
<b>Minimal Impact</b>	Evidence of struggle with one or more areas when compared to peers. Evidence of occasional difficulty with 'meta' skills.	Evidence of struggle with one or more areas when compared to peers.	May indicate differences from Standard American English. Demonstrates improvements during dynamic assessment. Occasional difficulty with pragmatic, semantic or syntax- morphological skills.	1 or 2 composite scores* documenting: <ul style="list-style-type: none"> <li>• 1 to 1.5SD</li> <li>• 68 to 77.5% #</li> <li>• 10th-7th percentile</li> </ul>
<b>Moderate Impact</b>	Evidence of struggle in most areas when compared to peers. Evidence of difficulty with 'meta' skills.	Evidence of struggle in most areas when compared to peers.	Demonstrates limited improvement during dynamic assessment. Frequent difficulty with pragmatic, semantic or syntax- morphological skills.	1 or 2 composite scores* documenting: <ul style="list-style-type: none"> <li>• 2.5 to 2SD</li> <li>• 15 to 24%</li> <li>• 2 to 25.8 percentile</li> </ul>
<b>Substantial Impact</b>	Evidence of very limited ability in most areas. Evidence of limited or absence of 'meta' skills.	Evidence of very limited ability in most areas.	Demonstrates very limited improvement during dynamic assessment. Extreme difficulty with pragmatic, semantic or syntax- morphological skills.	1 or 2 composite scores* documenting: <ul style="list-style-type: none"> <li>• 3 or greater SD</li> <li>• 0 or below 50.4</li> <li>• below 2nd percentile</li> </ul>

## Questions about Significant Discrepancy?

## Limitation 4: Functional Communication is Not Assessed

---

---

---

---

---

---

---

---



---

---

---

---

---

---

---

---



---

---

---

---

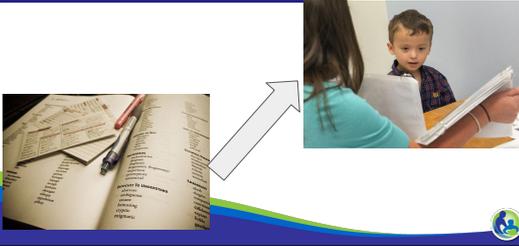
---

---

---

---

## Do GRE scores describe clinical expertise?



## My favorite task: Sentence Repetition

- Generally, good sensitivity & specificity
- Not just index of working memory - language definitely involved (Klem et al., 2014)

*"Sentence repetition appears to be a valuable tool for language assessment because it draws upon a wide range of language processing skills."*

## Skills for Sentence Repetition

- Strong verbal working memory
- Comprehend & produce complex syntax
- Include inflectional morphemes
- Recall complex words

---

---

---

---

---

---

---

---

---

---

---

---

## What Sentence Repetition is Missing

- How well do you use complex syntax when telling stories?
- How well can you express your needs to your peers?
- How effective are you when telling an amusing story?



## How are NR test items chosen?

What I used to think...

- Super important skills
- Something I should work on in therapy
- Essential component of communication



## How NR test items are chosen

*Have the ability to distinguish between high & low performers*



---

---

---

---

---

---

---

---

---

---

---

---

## Inflectional Morphology

One of the strongest indicators of a language disorder\*



## What if the Student Uses AAE?



## Think of a task from your favorite NR test

- What do you think it is measuring?
- What *else* could it be measuring?
- How functional is the task?



---

---

---

---

---

---

---

---

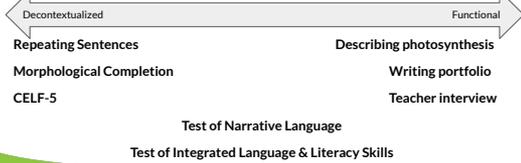
---

---

---

---

## Continuum of Functionality



## Norm Referenced Test Temptations



## Determining Disability Based on Norm-Referenced Tests Alone

- You get “objective” and “accurate” data
- Tests are efficient and fun
- It’s what we have trained you to do
- Old criteria put a lot of emphasis on NR tests

*We need to understand limitations & look for educational impact*

---

---

---

---

---

---

---

---

---

---

---

---

## Build a Profile Based on Test Scores

Different tests *appear* to look at different skills

- Expressive one word picture vocabulary test
- Receptive one word picture vocabulary test
- Test of Auditory Comprehension of Language
- Structured Photographic Expressive Language Test
- Test of Problem Solving



## Problems with Profiling

Usually no real differences across (sub)tests  
*Often overlooks functional communication & educational impact*



## Develop Goals with Norm-Referenced Items

- We know which items were incorrect
- The developers picked really important items, right?
- *Items chosen because they differentiate kids, not because of educational impact*



---

---

---

---

---

---

---

---

---

---

---

---

**Questions about Functional Communication & other Limitations?**



**Learning Objective #3**

Participants will engage in case study work to design more comprehensive assessments.



**Moving Forward with Comprehensive Assessment**



---

---

---

---

---

---

---

---

---

---

---

---

### How important are the norm-referenced data?

- Starting with academic reports & referral information, how likely will student receive services?
- If completing re-evaluation (or student already has a diagnosis), how likely is it that things have changed?



### Do I anticipate getting valid data?

- **Are there any obvious biases (English proficiency; dialect spoken)?**
  - If so, are there test options (e.g., Bilingual English-Spanish Assessment)?
- **Reasons to suspect less obvious biases?**



### Is the test appropriate?

- Was it designed to assess the predicted impairment area?
- Does the test provide reliability, validity, and diagnostic accuracy data?
- Any other sources of data for the test?
- Is your student represented well in the norming sample?
- Do you have a good track record with the test?



---

---

---

---

---

---

---

---

---

---

---

---

## How are you going to document educational impact?

Remember, norm referenced tests....

- Developed to rank-order students
- Tend to focus on decontextualized language
- Provide minimal information on functional communication



## John (9 years old)

- Minimally verbal student with autism
- Identifies as Hebrew, moved from India 6 years ago
- Mom speaks only English
- Attends public school, in "life skills" classroom
- Likes physical activity and art class
- Many sensory & motor difficulties
- Completed battery of testing, observation, & interviews

From Vidal, McAllister, & DeThorne, 2020



---

---

---

---

---

---

---

---

---

---

---

---

## Testing results

Vineland adaptive behavior: < 1st %tile (Age Equiv < 3 yrs)

Macarthur-Bates CDI: 80 word; < 1st %tile (Age Equiv = 18 mths)

Speech sample

- Limited phonetic repertoire
- 27 speech-like utterances
- Responded to simple questions; imitated some words
- Consistently names classmates when asked about friends



## Testing results (cont.)

Observation: Basic requests using multiple modalities

Interviews

- Peers & teachers report communicating with vocalizations, words, & gestures
- Can ask for bathroom & other basic requests
- Will answer "yes/no" with multiple prompts
- Peers observe him screaming; do not report seeing him use AAC
- Peers did not realize that iPad is an AAC device
- Teachers: has trouble sitting & listening



## 7-minute art class observation

- Vocalizations, gestures, & gaze - intent unknown
- Try several options (break, play ball, paint) - no response
- Paraprofessional gets AAC device and asks if he wants to play with peer named Gold
- John not interested, selects Elise on AAC device (peer)
- Appeared to want to play with Elise the entire time



---

---

---

---

---

---

---

---

---

---

---

---

## Conclusions

- Expressive single words with some 2-word combinations & gestures
- Restrictions in language comprehension
- Poor intelligibility & motor speech issues
- Gross motor a relative strength



## Conclusions

- **People likely underestimate John's ability**
  - Sensory-motor difficulties
  - See potential when provided support
- **Peers and teachers do not recognize full potential**
- **AAC powerful, but underused**



## Considerations for John

- What did we learn from norm-referenced tests?
- What would we have learned from other norm-referenced tests?
- What assessment data was most powerful?
- What else would you want to know?



---

---

---

---

---

---

---

---

---

---

---

---

Chris (5 years old)

CELF-P2: 77  
PLS-5: 82  
SPELT-3: 79

From Fulcher-Rood et al., 2019



### Minimal parent and teacher concerns

In comparison with other children of the same age, do you think that your child has problems expressing him/herself or being understood?	No
Does your child understand most of what he/she is told?	Yes
In comparison to children of the same age, is it difficult for your child to learn new ideas?	No
In comparison to children of the same age, does your child have a very low or limited vocabulary?	No



### Narrative measures within normal limits

Supplemental Material S18. Language sample, story retell, Cwe 4, incongruent dan.

C The dog and frog and turtle are looking at him getting dressed.  
C He said bye.  
C And he pet the doggy.  
C And he looked at the window.  
E Mhm.  
E How about here?  
C They're waiting they're waiting for the dinner.  
C The frog is jumping in there.  
E Is there?  
E Alright.  
E How about here?  
C He jumped in there.  
C Would he jump?  
E Mhm.  
C And he looked in there.  
C He jumped on the man's head.  
C And he he jumped on his face.  
E Mhm.  
E And then what happened here?  
C He broke the -  
C He said broke he broke my music.  
E That's fine.  
E And what did the frog do?  
C He jumped in the dinner.  
E So what happened here?  
C The girl saw the frog.  
C And the frog was on underneath the salad.  
E Mhm.



---

---

---

---

---

---

---

---

---

---

---

---

## Chris - Considerations

- Would he likely qualify for services?
- Why may we have seen differences between norm-referenced tests and other tests?
- What else could tell us about the educational impact?



## Final Questions?



## Comprehensive Assessment Model

### 4 Part Model for Comprehensive Assessment

#### Academic Activities:

- Artifact analysis
- Curriculum-based assessment
- Observations in school (natural) settings
- Educational records

#### Speech-Language Probes:

- Case history
- Interviews
- Language/Narrative samples
- Stimulability
- Dynamic assessment
- Play-based assessment

#### Contextualized Tests:

- Norm-referenced measures of academic achievement
- Curriculum benchmarks

#### Decontextualized Tests:

- Norm-referenced speech-language tests (parsed skills: articulation, semantics, syntax, morphology, fluency, etc.)

Ireland, Marie. "The Real Requirements Behind Eligibility Decision Making in Schools". Lecture. ASHA Connect, Chicago, IL, July 19, 2019.

---

---

---

---

---

---

---

---

---

---

---

---



## Resources to Come

- Wednesday, September 8, 2021 – 3:30-5:00pm  
**Fluency Assessment with Laura Johnson**
- Wednesday, October 6, 2021 – 3:30-5:00pm  
**Best Practices when Assessing English Learners I with Courtney Seidel, UW-Madison**
- Wednesday, November 3, 2021 – 3:30-5:00pm  
**Best Practices When Assessing English Learners II with Courtney Seidel**
- Wednesday, December 1, 2021- 3:30-5:00 pm  
**Dynamic Assessment Part 2**
- Wednesday, February 2, 2022 – 3:30-5:00pm  
**Language Assessment**
- Wednesday, March 2, 2022 – 3:30-5:00pm  
**Assessment of Voice**



## References

- Fagundes, Deana D., William O. Haynes, Nancy J. Haak, and Michael J. Moran. 1998. "Task Variability Effects on the Language Test Performance of Southern Lower Socioeconomic Class African American and Caucasian Five-year-olds." *Language, Speech, and Hearing Services in Schools*, 29(3): 148-157. <https://doi.org/10.1044/0161-1661.2903.148>
- Fukcher-Rood, Katrina, Amy Castilla-Earls, and Jeff Higginbotham. 2019. "Diagnostic Decisions in Child Language Assessment: Findings From a Case Review Assessment Task." *Language, Speech, and Hearing Services in Schools*, 50(3): 385-398. <https://doi.org/10.1044/2019.JSHSS-18-0044>
- Helm, Janet E. 1992. "Why Is There No Study of Cultural Equivalence in Standardized Cognitive Ability Testing?" *American Psychologist*, 47(9): 1083-1101. <https://doi.org/10.1037/0003-066X.47.9.1083>
- Klem, Marianne, Monica Melby-Lervåg, Bente Hagtvet, Solveig-Alma Halaas Lyster, Jan-Eric Gustafsson, and Charles Hulme. 2015. "Sentence Repetition Is a Measure of Children's Language Skills Rather Than Working Memory Limitations." *Developmental Science*, 18(1): 146-154. <https://doi.org/10.1111/desc.12202>
- Orellana, Carla I., Rebekah Wada, and Ronald B. Gillam. 2019. "The Use of Dynamic Assessment for the Diagnosis of Language Disorders in Bilingual Children: A Meta-Analysis." *American Journal of Speech Language Pathology* 28(3): 1298-1317. [https://doi.org/10.1044/2019\\_AJSLP-18-0202](https://doi.org/10.1044/2019_AJSLP-18-0202)



## References

- Plante, Elena and Rebecca Vance. 1994. "Selection of Preschool Language Tests: A Data-Based Approach." *Language, Speech, and Hearing Services in Schools* 25(1): 15-24. <https://doi.org/10.1044/0161-1661.2501.15>
- Roseberry-McKibbin, Celeste. 2021. "Utilizing Comprehensive Pre-Assessment Procedures for Differentiating Language Difference from Language Impairment in English Learners." *Communication Disorders Quarterly* 42(2): 93-99.
- Spaulding, Tammie J., Elena Plante, and Kimberly A. Farinella. 2006. "Eligibility Criteria for Language Impairment: Is the Low End of Normal Always Appropriate?" *Language, Speech, and Hearing Services in Schools* 37(1): 61-72. <https://doi.org/10.1044/0161-1661.0206.007>
- Vidal, Verónica, Anita McAllister, and Laura DeThorne. 2020. "Communication Profile of a Minimally Verbal School-age Autistic Child: A Case Study." *Language, Speech, and Hearing Services in Schools*, 51(3): 671-686. <https://doi.org/10.1044/2020.JSHSS-19-00021>



---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

