Curriculum-Based Language Evaluations for School-Aged Students

• Jayne Brandel has a relevant financial relationship to disclose. I am receiving an honorarium.

Your learning objectives
• Critically evaluate a test’s content in regards to student concerns.
• Interpret test results taking into account the test’s diagnostic accuracy and content.
• Select research tested, non-standardized, curricular-based activities to supplement the standardized testing.
The Evaluation

What should be included?

Language Disorders
Later Characteristics—especially Reading Comprehension and Writing
## Oral and Written Language

- **Semantics**
  - Multiple meanings
  - Similes and metaphors
  - Limited Tier 2 word use

- **Morphosyntax**
  - Grammaticality in writing
  - Simple sentences
  - Disorganized sentences

- **Pragmatics**
  - Organization of different genres
  - Expository
  - Perspective taking
  - Nonverbal cues
  - Pitch
  - Stress
  - Pausing

## The interview

- **Teacher**
  - Other relevant school personnel

- **Student**

- **Family**
  - Phone and email if needed

## Questions to answer

- **Why was the student referred?**
  - What specific tasks in school is the student not performing like peers?
  - What does the student do instead?

- **What language skills underlie that task(s)?**
  - What clues about the underlying language deficit can you see in what the student does?
• What are the language demands of the curriculum?
• What are the underlying skills that are causing concerns across the curriculum?

Language Underpinnings

Sarah can’t follow directions

Follow Directions
Johnny (2\textsuperscript{nd} grade) sounds immature.

Tabitha’s writing is disorganized and simple (middle school).

Beginning the evaluation
Observation

• Live observations (multiple classrooms)
  • Use that list of concerns from the interview
  • Take data on the observable behaviors
    • Take notes on what the student does

• Review Class Artifacts
  • Quantify your analysis regarding the areas/skills of concern

Testing

Include

Standardized Test   Additional Activities
Standardized Test

What we have emphasized

Psychometric Properties

Reliability

Test-Retest Interjudge Split-Half Construct

Validity

Content Criterion-related

But what do these mean?
Reliability
Test-retest Reliability and Interjudge Reliability

How consistent is student performance?

CAN I TRUST THE SCORE?
Impact on the evaluation...

- Test-retest is important because of the Standard Error of Measure (SEM).
- Test performance is an estimate.
  - I should have reported the confidence interval in my report.
    - 90% is good when available.
    - 95% can be calculated based on +/- 2 SEMs around the earned score. Look in the manual in the reliability section.

Questions? Concerns?
Do we get the same score?

- Interjudge reliability

Reliability recap...

Validity

I DON'T TRUST JOGGERS. THEY'RE ALWAYS THE ONES THAT FIND THE DEAD BODIES. I'M NO DETECTIVE, JUST SAYIN'.

I DON'T TRUST JOGGERS. THEY'RE ALWAYS THE ONES THAT FIND THE DEAD BODIES. I'M NO DETECTIVE, JUST SAYIN'.
Content/Construct Validity

What does that really mean?

How do you complete a content analysis?
Pair with research tested tasks

Oral Language

Grammatical Morphology

• TEGI - available online for free
  https://cldp.ku.edu/rice-wexler-tegi
Narrative Language

- TNL Alien Story (transcribed and analyzed)
- Test of Narrative Retell (PreK through 3rd grade)
- Frog Stories by Mercer Mayer (Einstein & Phillips, 2009; McCabe et al., 2008)
- Fable Elicitation Task (adolescents) (Nippold, Frantz-Kasper, Cramond, Kirk, Hayward-Hayhew, & McKinnon, 2014)

Expository Language

- Favorite Game or Sport Task (Nippold, Hesketh, Duthie, & Mansfield, 2000)
- Peer Conflict Resolution Task (Nippold, Mansfield, & Bhow, 2007)

Written Language
Narrative

- “Your Backpack Turned Into Wings” (Koutsoftas & Gray, 2012)
- “What Happened One Day” (Sun & Nippold, 2012)

Expository Language

- “Money for a Field Trip” (Koutsoftas & Gray, 2012)
- “The Nature of Friendship” (Nippold & Sun, 2010)

Persuasive

- “The Circus Controversy” (Nippold, Ward-规章制度, & Fanning, 2006)
Sounded good to me...right?!

Calculated Score – Predicted Score
Predicted Standard Deviation

| 4.6 – 4.88 | 0.72 |

-0.39 SD (0.39 standard deviations below student’s peers)

College of Education and Human Services

Interpreting the test results...
Now, I have the best available tasks, but what is normal?

Criterion related validity
Why the cutoff score matters?
Changing your cutoff score...

- Changes the diagnostic accuracy...
  - Sensitivity/Specificity
  - PPP/NPP

What is diagnostic accuracy?

<table>
<thead>
<tr>
<th>TRUE ABILITY (LANGUAGE)</th>
<th>Disorder +</th>
<th>Typical -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered</td>
<td>TRUE POSITIVE</td>
<td>-</td>
</tr>
<tr>
<td>Typical</td>
<td>-</td>
<td>TRUE NEGATIVE</td>
</tr>
</tbody>
</table>

In reality...

<table>
<thead>
<tr>
<th>TRUE ABILITY (LANGUAGE)</th>
<th>Disorder +</th>
<th>Typical -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered</td>
<td>-</td>
<td>FALSE POSITIVE</td>
</tr>
<tr>
<td>Typical</td>
<td>FALSE NEGATIVE</td>
<td>-</td>
</tr>
</tbody>
</table>
Let’s look at an example...

- **Test of Narrative Language-2**
  - Evaluates
    - Narrative comprehension
    - Narrative expression

### Sensitivity

<table>
<thead>
<tr>
<th>TRUE ABILITY (LANGUAGE)</th>
<th>Disorder</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Typical -</td>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

### Specificity

<table>
<thead>
<tr>
<th>TRUE ABILITY (LANGUAGE)</th>
<th>Disorder</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Typical -</td>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>
TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder +</th>
<th>Typical -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Typical -</td>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Positive Predictive Power (PPP)

True Positive / True Positive + False Positive

TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder +</th>
<th>Typical -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Typical -</td>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Negative Predictive Power (NPP)

True Negative / True Negative + False Negative

If the TNL-2 were THE PERFECT TEST

TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder +</th>
<th>Typical -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>Typical -</td>
<td>0</td>
<td>257</td>
</tr>
</tbody>
</table>
CUTOFF = SS of 70

TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Typical -</td>
<td>44</td>
<td>257</td>
</tr>
</tbody>
</table>

Sensitivity = 13%
Specificity = 100%
PPP = 100%
NPP = 85%

CUTOFF = SS of 82

TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Typical -</td>
<td>29</td>
<td>257</td>
</tr>
</tbody>
</table>

Sensitivity = 43%
Specificity = 100%
PPP = 100%
NPP = 89%

CUTOFF = SS of 85

TRUE ABILITY (LANGUAGE)

<table>
<thead>
<tr>
<th></th>
<th>Disorder</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered +</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>Typical -</td>
<td>23</td>
<td>253</td>
</tr>
</tbody>
</table>

Sensitivity = 55%
Specificity = 98%
PPP = 98%
NPP = 91%
In Summary for the TNL-2...

<table>
<thead>
<tr>
<th>CUTOFF SCORE</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>13%</td>
<td>100%</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>82</td>
<td>43%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
</tr>
<tr>
<td>90</td>
<td>55%</td>
<td>98%</td>
<td>98%</td>
<td>91%</td>
</tr>
<tr>
<td>92</td>
<td>55%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
</tr>
</tbody>
</table>
So what about other tests?

### CELF-5

<table>
<thead>
<tr>
<th>CUTOFF SCORE</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>100%</td>
<td>91%</td>
<td>56%</td>
<td>100%</td>
</tr>
<tr>
<td>80</td>
<td>97%</td>
<td>97%</td>
<td>96%</td>
<td>95%</td>
</tr>
<tr>
<td>77</td>
<td>95%</td>
<td>96%</td>
<td>96%</td>
<td>74%</td>
</tr>
<tr>
<td>70</td>
<td>57%</td>
<td>100%</td>
<td>100%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Base Rate: 70%

### PLS-5

<table>
<thead>
<tr>
<th>Cutoff Score</th>
<th>Auditory Comprehension</th>
<th>Expressive Communication</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>PPP</td>
<td>83%</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>NPP</td>
<td>77%</td>
<td>83%</td>
</tr>
<tr>
<td>78</td>
<td>PPP</td>
<td>93%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>NPP</td>
<td>66%</td>
<td>74%</td>
</tr>
<tr>
<td>70</td>
<td>PPP</td>
<td>69%</td>
<td>94%</td>
</tr>
<tr>
<td></td>
<td>NPP</td>
<td>69%</td>
<td>63%</td>
</tr>
</tbody>
</table>
### TOLD-P:4

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Spoken Language</td>
<td>75%</td>
<td>87%</td>
<td>71%</td>
</tr>
</tbody>
</table>

Cutoff score: Standard Score 90

---

### TOLD-I:4

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Spoken Language</td>
<td>77%</td>
<td>89%</td>
<td>71%</td>
</tr>
</tbody>
</table>

Cutoff score: Standard Score 90

---

### Curriculum-Based Evaluations

---

---
In summary, you should be able to

• Critically evaluate a test’s content in regards to student concerns.
• Interpret test results taking into account the test’s diagnostic accuracy and content.
• Select research tested, non-standardized, curricular-based activities to supplement the standardized testing.

References


