A modern view of validity for the Early Years Evaluation in a francophone population

Robert Laurie
Director
Assessment and Evaluation
New Brunswick Department of Education
Assumptions

• Basic knowledge of the EYE-TA and the five child developmental domains it measures.

• When it comes to screening for vulnerable children « the earlier the better ».

• Receiving support services is a right not a favour.

• We can, we do and we will make a difference.
Presentation overview

• Traditional View of Validation Studies
• Reliability of the EYE-TA
• Validity of the EYE-TA
• DIF analysis (English-French)
• Modern View of Validity
• What this means for NBf
• Questions/Comments
Traditional View of Validation Studies

Recipe

- Mix one part Cronbach’s coefficient alpha with one part construct validity
- Add a dash of content validity
- Interpret for a few minutes
- Smile and go on with your life 😊
Types of reliability

- Test-retest
- Parallel (alternate form)
- Inter-rater
- Internal consistency
  - Split-half
  - Kuder-Richardson 20
- Cronbach's alpha
Reliability results

Awareness of self and environment \( \alpha = 0.86 \)

Social skills, behaviour and approaches to learning \( \alpha = 0.85 \)

Cognitive skills \( \alpha = 0.84 \)

Language and communication \( \alpha = 0.92 \)

Physical development \( \alpha = 0.81 \)
Types of validity

- Discriminant
- Convergent
- Construct
- Content
- Criterion-related
- Concurrent
- Predictive
Validity results

- Content
- Construct
- Concurrent and discriminant
Validity results

• **Content validity** *(the extent to which a test’s items address the full range of what is being tested)*

• Obtained from items being built in accordance with known research.

• Discussions with kindergarten teachers during training sessions.

• Result is that there is a strong buy-in by teachers with regard to the content.
  – « It really does represent child development ».
Validity results

• **Construct validity** *(Groups of items which correlate highly with each other; obtained through factor analysis)*

• Identifies factors (domains and sub domains) which the assessment measures.
Validity results

Awareness of self and environment (0 sub domains)

Social skills, behaviour and approaches to learning (4 sub domains)
- aggression & attentiveness
- depression & anxiety
- mean & cruel (1 item)
- persistence (1 item)
Validity results

Cognitive skills (2 sub domains)
  - reading
  - mathematics

Language and communication (0 sub domains)

Physical development (2 sub domains)
  - gross & fine motor skills
  - gross & fine motor skills
Validity results

• Concurrent and discriminant
  – Students were assessed using the PPVT-R and the WIAT-II reading sub-test
  
  – High correlations and coefficients of determination or explained variance \((r^2)\) is evidence of concurrent validity
  
  – Low correlation and coefficients of determination or explained variance \((r^2)\) is evidence of discriminant validity
## Validity results

<table>
<thead>
<tr>
<th>Domains</th>
<th>Test</th>
<th>Results</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness of self and environment</td>
<td>PPVT-R</td>
<td>$r = .58$, $r^2 = .34$</td>
<td>Concurrent</td>
</tr>
<tr>
<td>Social skills and behaviour</td>
<td>PPVT-R</td>
<td>$r = .32$, $r^2 = .10$</td>
<td>Discriminant</td>
</tr>
<tr>
<td>Cognitive skills</td>
<td>PPVT-R</td>
<td>$r = .55$, $r^2 = .30$</td>
<td>Concurrent</td>
</tr>
<tr>
<td></td>
<td>WIAT-II</td>
<td>$r = .65$, $r^2 = .42$</td>
<td>Concurrent</td>
</tr>
<tr>
<td>Language and communication</td>
<td>PPVT-R</td>
<td>$r = .67$, $r^2 = .44$</td>
<td>Concurrent</td>
</tr>
<tr>
<td></td>
<td>WIAT-II</td>
<td>$r = .45$, $r^2 = .21$</td>
<td>Concurrent</td>
</tr>
<tr>
<td>Physical development</td>
<td>PPVT-R</td>
<td>$r = .31$, $r^2 = .09$</td>
<td>Discriminant</td>
</tr>
<tr>
<td></td>
<td>WIAT-II</td>
<td>$r = .30$, $r^2 = .09$</td>
<td>Discriminant</td>
</tr>
</tbody>
</table>
Differential Item Functioning (DIF)

- Number of DIF items = 7/47 (4 in favour of Ang, 3 in favour of Fr)

<table>
<thead>
<tr>
<th>DIF Items</th>
<th>Favoured Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ASE) A3</td>
<td>Anglophones</td>
</tr>
<tr>
<td>(SSB) B14</td>
<td>Francophones</td>
</tr>
<tr>
<td>(CS) C4</td>
<td>Anglophones</td>
</tr>
<tr>
<td>(CS) C5</td>
<td>Francophones</td>
</tr>
<tr>
<td>(CS) C8</td>
<td>Francophones</td>
</tr>
<tr>
<td>(LC) D4</td>
<td>Anglophones</td>
</tr>
<tr>
<td>(PD) E8</td>
<td>Anglophones</td>
</tr>
</tbody>
</table>

- Overall test = no problem
Overall validity of the ÉPE-AE

• Very reliable in all domains and the instrument as a whole ($\alpha = 0.91$).
• Strong validity results
  – Content
  – Construct
  – Concurrent
  – Discriminant
• Few DIF items which are not cause for concern in either a specific domain or the instrument as a whole.
Therefore, the ÉPE-AE is a very valid assessment instrument.

But what about the modern view of validity?
Modern view of validity

• Validity is not a property of the instrument

• Validity is a property of the inferences made from the results

• Validity is dependant of context (Messick’s progressive validity framework)
<table>
<thead>
<tr>
<th>FUNCTION OR OUTCOME</th>
<th>JUSTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Evidential basis</strong></td>
</tr>
<tr>
<td>Interpretation</td>
<td><strong>Construct validity</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Value implication</strong></td>
</tr>
<tr>
<td>Applied Use</td>
<td><strong>Construct validity</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Relevance &amp; utility</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Value implications</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Social consequences</strong></td>
</tr>
</tbody>
</table>
What this means for NBf

Add to the EYE-TA’s construct validity

1. Calculate the EYE-TA’s sensitivity and specificity of based on our grade 2 literacy assessment.

2. **Sensitivity**: the extent to which the EYE-TA correctly identifies vulnerable children.

3. **Specificity**: the extent to which the EYE-TA correctly identifies children who are not considered to be vulnerable.
What this means for NBf

• Follow this cohort to determine each domain’s predictive validity based on a number of future outcomes (academic, physical, behaviour, etc.)

• Target future interventions based on predictive validity scores instead of waiting for signs of negative outcomes. Make use of a leading indicator instead of a trailing indicator.
Major focus for next 6-12 months

• Address the « social consequences » issues in the Messick validity framework.
  – Ensure educators, parents, stakeholders, and all interested parties understand what the EYE measures and what it doesn’t, it’s purpose, what it’s not used for, etc.

• Work with school districts and teachers to ensure communication with parents is clear, respectful, and helpful.
Major focus for next 6-12 months

• Provide training for the school district consultants and kindergarten teachers.

• Work with kindergarten teachers to ensure sharing of data with methods and resource teachers.

• Consider developing a more detailed student level report for educators.
Thank you

Robert.Laurie@gnb.ca