State of the Science of Health Literacy Measures: Validity Implications for Minority Populations

6th Health Literacy Annual Research Conference
November 3-4, 2014

Tam H Nguyen, PhD, MSN/MPH, RN
Hae-Ra Han, PhD, RN
Hyun Jeong Park, PhD, FNP, RN
Michael Paasche-Orlow, MD, MA, MPH
Kim, Miyong, PhD, RN
- Limited HL is a problem for many minority populations

- Existing reviews have not systematically assessed the validity of current measures for minority populations (Mancusco, 2009; Luk & Aslani, 2011)

- Using tools that are not well-validated for a given population can lead to measurement error (Nunnelly & Bernstein, 1994)
Systematically review the literature and examine the validity of existing HL measures for a wide range of minority populations
3 databases searched

January 1966 and September 2013

Terms/synonyms related to “health literacy” AND “measures”

Included: Articles that reported original validation data on a HL measure, shorted measure, or translated measure

Excluded: If psychometrics were not reported
Methods:

Example of PubMed search

(health literacy [mh] OR "health literacy" OR (health [tiab] AND (literate [tiab] OR literacy [tiab])) OR numeracy [tiab])

AND

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Instrument Name</th>
<th>Description (Items)</th>
<th>Scoring</th>
<th>Time, min.</th>
<th>Reliability</th>
<th>Validity</th>
<th>Sample</th>
<th>Domain Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davis et al, 1991</td>
<td>Rapid Estimate of Adult Literacy in Medicine (REALM)</td>
<td>General list of medical words in descending levels of difficulty (125-items)</td>
<td>Possible range: 0-125 Low HL: 0-78; ≤3rd grade; 79-103; 4th-6th grade; 104-114; 7th-8th grade; 115-125; ≥9th grade</td>
<td>~3-5</td>
<td>Test-retest: r = .98; Inter-rater: r = .99</td>
<td>Convergent: SORT-87, r = 95; PIAT-R, r = 94*.81*</td>
<td>(n=207)</td>
<td>Print</td>
</tr>
<tr>
<td>Davis et al, 1993</td>
<td>REALM (Shortened)</td>
<td>General list of medical words in descending levels of difficulty, shortened (66-items)</td>
<td>Possible range: 0-66 0-6: ≤3rd grade; 6-18: ≤5th grade; 19-44: 6th-8th grade; 45-60: 7th-8th grade; 61-66: ≥9th grade</td>
<td>~1-3</td>
<td>Test-retest: r = .99</td>
<td>Convergent: SORT-87, r = 96; WRAT-87, r = 88; PIAT-R, r = 97*</td>
<td>(n=203)</td>
<td>Print</td>
</tr>
<tr>
<td>Bass et al, 2003</td>
<td>REALM-R (Revised)</td>
<td>General list of medical words in descending levels of difficulty, shortened (8-items)</td>
<td>Possible range: 0-8 Total score = THL</td>
<td>&lt;2</td>
<td>Internal Consistency: α = .91</td>
<td>Convergent: WRAT-87, r = .64</td>
<td>(n=157)</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Arouallah et al, 2007</td>
<td>REALM-SF (Short Form)</td>
<td>General list of medical words in descending levels of difficulty, shortened (7-items)</td>
<td>Possible range: 0-7 0-≤3rd grade; 1-3: 4th-6th grade; 4-6: 7th-8th grade; 7: ≥9th grade</td>
<td>&lt;1</td>
<td>N/A</td>
<td>Convergent: REALM-87, r = .96; WRAT-87, r = .83; CFA: R² = .92; 16% ² 14.5% (p = .41)</td>
<td>(n=1500)</td>
<td>Print</td>
</tr>
<tr>
<td>Hanson-Diven, 1997</td>
<td>Medical Achievement Reading Test (MART)</td>
<td>General medical word recognition test modeled after the WRAI, with small font size and glossy covering to resemble prescription bottles</td>
<td>Possible range: 0-42 “Converted into grade level equivalences similar to the WRAI”</td>
<td>~5</td>
<td>Internal Consistency: α = .98</td>
<td>Convergent: WRAT-87, r = .98</td>
<td>(n=405)</td>
<td>Print</td>
</tr>
<tr>
<td>Lee et al, 2006</td>
<td>Rapid Estimate of Health Literacy for Spanish-speaking Adults (SAHLSA-50)</td>
<td>Spanish general medical word recognition and matching test (50-items)</td>
<td>Possible range: 0-50 ≥37 = inadequate HL</td>
<td>~3-6</td>
<td>Test-retest: r = .86; Internal Consistency: α = .92</td>
<td>Convergent: TOFHLA, r = 65; REALM, r = .76; Concurrent: Physical health status, r = .17</td>
<td>(n=403)</td>
<td>Print</td>
</tr>
</tbody>
</table>

**Methods:**

Example of extraction tables
93 HL measures identified
- 47 general, 46 content/context specific
- 20 REALM or REALM-like
- 14 TOFHLA or TOFHLA-like

65 were English language measures
28 were non-English measures

Reliability was generally strong across all measures
Weaker evidence of validity
- 82 used Classical Test Theory
- 11 used Item Response Theory/Rasch alone or with CTT

Results
Of the 65 English language measures:

- 15 measures did not specify the racial/ethnic characteristic of sample

- Of the remaining 50 measures,
  - 7 did not include any blacks (14%)
  - 29 did not include any Hispanics (58%)
  - 33 did not include any Asians (66%)

- When Hispanic and Asian Americans were included, they usually accounted for a small % of overall sample
Results

- Of the 28 “other” language measures
  - Translations for REALM (n=3) and TOFHLA (n=9)
    - Issues with phonetic structure of language
    - Issues with cultural equivalence
  - New HL measures in languages other than English (n=16)
    - Poor description of subgroups sampled
    - For some cultural groups this is more/less important
Conclusions

- Many HL measures exist

- Most have not been properly validated for minority populations.
  - Challenges exist with translations & cultural equivalence
  - Problematic b/c most measures are validated using CTT vs IRT

- To address this issue researchers/clinicians have started to translate and developing new ones.
Practice Implications

- 93 HL measures -> public repository

- Pilot testing should be done if not yet validated

- Use Item Response Theory/Rasch Modeling with Classical Test Theory to guide instrument development and refinement

- Think about how to collaboratively “re-engineer” HL measurement
Questions?

tam.nguyen@bc.edu

Thank you
- Item Difficulty, Discrimination, and Spread

**Example of Item Spread & Density**

[Graph showing item spread and density with TS-REALD and HLSI curves]

Extra slide (IRT/RASCH)
Examining “Differential Item Functioning” (DIF)