A vocabulary size test for non-literate learners of Dutch as a second language: Design and validation

Nel de Jong en Annebel Prins

Vocab@Leuven, 3 juli 2019
Vocabulary size test for non-literate L2 learners

- Vocabulary size predicts levels of language skills: reading, writing, listening, speaking (e.g. Milton, 2013; Staehr, 2008)
- Vocabulary size tests often use a written yes/no or multiple-choice format
- But there are no vocabulary size tests specifically for non-literate L2 learners

- Test goal: formative (progress) or summative (end-of-course)
Non-literate L2 learners

- Little or no understanding of script in L2 (or in L1)
- Lack of school experience
- Lack of metalinguistic awareness
- Reliance on personal experience
- Lack of computer skills

Van de Craats, Kurvers & Young-Scholten (2006)
Testing non-literate learners of DSL

- Motivational
- Credible and recognizable tasks
- Independent of learning strategies, computer skills and experience in test taking

Dalderop, Janssen-van Dieten & Stockmann (2009)
Vocabulary size

- Level A0-A2 of the CEFR (2,000 words)
- Meaning recognition of spoken words (cf. Laufer & Goldstein, 2004; Schmitt, 2010)
## Selection of target words

<table>
<thead>
<tr>
<th>Frequency band</th>
<th>Word category</th>
<th>Percentages</th>
<th>Number of target words</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (0-500)</td>
<td>Verb</td>
<td>33%</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Noun</td>
<td>39%</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Adjective</td>
<td>28%</td>
<td>4</td>
</tr>
<tr>
<td>B (501-1,000)</td>
<td>Verb</td>
<td>24%</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Noun</td>
<td>56%</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Adjective</td>
<td>20%</td>
<td>3</td>
</tr>
<tr>
<td>C (1,001-1,500)</td>
<td>Verb</td>
<td>28%</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Noun</td>
<td>53%</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Adjective</td>
<td>19%</td>
<td>3</td>
</tr>
<tr>
<td>D (1,501-2,000)</td>
<td>Verb</td>
<td>30%</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Noun</td>
<td>56%</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Adjective</td>
<td>14%</td>
<td>2</td>
</tr>
</tbody>
</table>
Selection of distractors

- 3 distractors from the same frequency band (cf. Linger, 2018)
- Thematic similarity (cf. Beglar, 2010)
- Not too many shared visual characteristics (cf. Ishii, 2015)
Validity

- Consistency of tasks
- Minimal computer skills
- Target selection based on frequency
- Motivational
- Connection to the real world / own experience
- Mirroring the frequency of word categories
- No bias
Pilot

Test quality

- Reliability (Cronbach’s alpha): .91
- Identifying items to improve
  - 2 low $R_{ir}$, 17 somewhat low $R_{ir}$
  - 8 items: correct answer chosen less often than one of the distractors

- Changes made to test:
  - Different photographs for correct answer and/or distractors
  - Additional practice item
  - Computerized online version
Research questions

- Is the WTN-Alfa reliable and valid as a test of vocabulary size in non-literate learners of Dutch as a second language?
- Do individual differences explain variance in the scores on the WTN-Alfa, and if so, which individual differences?
Method
Participants

- 105 participants
- Age: 37.6 (SD = 12.4)
- Years of residence: 2.3 years (SD = 2.5)
- Months in Dutch literacy course: 11.3 months (SD = 10.5)
- 62 monolingual; 42 multilingual
- 36 non-literate; 68 non-Roman literate
Procedures

One session in computer room with regular teacher
- Background questionnaire
- Test
- Test results
Scores

- Mean: 40.6 (max. 60)
- Range: 13 - 57
- SD: 10.2
RQ1: Reliability and validity
Reliability

- Rasch person reliability: .90
- Rasch item reliability: .96
### Representativeness

- Sufficient number of items
- Sufficient spread in item difficulty
- Item strata: 6.87
- Gap in more difficult items
Technical quality of individual items

- Standardized infit measure for items:
  - > 2.0 (underfitting): raken, jong, schrikken, vechten, heuvel (8.3% of items)
  - < -2.0 (overfitting): bellen

- Standardized infit measure for participants:
  - > 2.0 (underfitting): seven participants
  - < -2.0 (overfitting): one participant

- A few too many underfitting items and participants
- Few overfitting items and participants
Substantive aspect of construct validity

- No effect of frequency
RQ2: Individual differences
## Individual differences

<table>
<thead>
<tr>
<th>Measure</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>38.98</td>
<td>3.36</td>
<td></td>
<td>&lt;.00</td>
</tr>
<tr>
<td>Age at test</td>
<td>- .26</td>
<td>.07</td>
<td>- .32</td>
<td>&lt;.00</td>
</tr>
<tr>
<td>Mono/multilingual</td>
<td>.35</td>
<td>1.77</td>
<td>.02</td>
<td>.85</td>
</tr>
<tr>
<td>Literacy type</td>
<td>2.86</td>
<td>2.25</td>
<td>.13</td>
<td>.21</td>
</tr>
<tr>
<td>Length of residence</td>
<td>.65</td>
<td>.35</td>
<td>.17</td>
<td>.07</td>
</tr>
<tr>
<td>Time in literacy course</td>
<td>.47</td>
<td>.09</td>
<td>.45</td>
<td>&lt;.00</td>
</tr>
<tr>
<td>Level of education</td>
<td>2.84</td>
<td>1.31</td>
<td>.22</td>
<td>.03</td>
</tr>
</tbody>
</table>

Adjusted $R^2 = .35; F(6,83) = 8.852, p < .001
Discussion and conclusion
Discussion and conclusion

- WTN-Alfa is a reliable test
- WTN-Alfa is valid as a test of vocabulary size (meaning recognition) in non-literate/non-Roman-literate learners of Dutch as a second language
  - Enough items and good spread but small gap
  - But no effect of frequency
WTN-Alfa scores are predicted by age at test, length of time in literacy course and level of education but not by multilingual status, literacy status and length of stay in the country.

- Length of time in literacy course and length of stay in the country both reflect amount of exposure, but exposure in literacy course seems to have a stronger effect.
- Level of education may reflect test taking skills or study skills.
- Age at test: younger learners score higher; due to study skills or motivation?

- Literacy type did not predict test scores ➔ most words learned by listening, not reading?
- Multilingual status ➔ the additional language is often Arabic or Tigrinya (no/few cognates with Dutch); language learning experience doesn’t seem to contribute.
Thank you!

- nel.de.jong@vu.nl
- annebel_prins@hotmail.com
References