A proof-of-concept study for measuring receptive written vocabulary size in second language Chinese

Eric Pelzl
Second Language Acquisition Program, University of Maryland, College Park

**Goal**
Develop measures of second language (L2) Chinese receptive written vocabulary size

1) Compare two approaches:
   - **Vocabulary self-assessment (VSA)** (aka ‘Yes/No vocabulary test’ [4])
     - efficient, easy to design, easy to score
     - subjective—test takers over/underestimate vocabulary knowledge
     - pseudowords can introduce other difficulties ([4,5])
   - **Vocabulary translation judgment (VTJ)**
     - could either supplement or replace the VSA
     - more objective measure
     - not as easy to design
     - little available research on this format

2) Examine relationship between vocabulary frequency and item difficulty ([6])—can we predict vocabulary size in L2 Chinese?

**Methods**
VSA and VTJ administered online to L2 learners of Chinese (n=48)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min-Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at test time</td>
<td>26</td>
<td>9.22</td>
<td>18-61</td>
</tr>
<tr>
<td>Age at which Mandarin learning began</td>
<td>18.35</td>
<td>5.72</td>
<td>6-35</td>
</tr>
<tr>
<td>Semesters of formal study (monthly)</td>
<td>7.79</td>
<td>4.71</td>
<td>1-16</td>
</tr>
<tr>
<td>Time in immersion (months)</td>
<td>10.19</td>
<td>7.11</td>
<td>0-240</td>
</tr>
</tbody>
</table>

Each test contained 105 vocabulary items sampled from a range of frequencies in the SUBTLEX-CH corpus ([5])

Each test-taker took both the VSA and the VTJ

Results were analyzed using a Rasch model with estimates of person ability and item difficulty

**Results**

1) VSA vs. VTJ
   - Considering the relatively small sample, both tests obtained reasonable model fit and item reliabilities
   - VSA: RELIABILITY=.92, SEPARATION=.33
   - VTJ: RELIABILITY=.88, SEPARATION=.70
   - Correlation of Rasch logits for the two measures was large and significant (r = .80, p < .001) suggesting both instruments measure the same underlying construct
   - Appears to be a strong ‘no’ bias for answers on VTJ

2) Vocabulary frequency and difficulty
   - Correlations of VSA and VTJ item difficulty ranks with word frequency ranks showed only a modest relationship
   - VSA: r=.36, p<.001
   - VTJ: r=.33, p<.001

**Conclusions**

1) VSA vs. VTJ
   - Both seem like promising formats for creating tests that separate persons via vocabulary knowledge, particularly for low stakes testing situations
   - VTJ needs further work to be sure ‘no’ bias does not skew results

2) Vocabulary frequency and difficulty
   - Extrapolating L2 vocabulary size based on the current corpus seems problematic
   - In contrast to English ([3]), Chinese may lack quality resources for L2 learning—comparisons with other corpora are ongoing, but do not seem very promising
   - The unit of measurement (‘word’) may need to be worked out more fully in Chinese to get accurate relationships between frequency and difficulty
   - Word vs. character frequency?

**References**


**Contact**
epelzl@umd.edu
http://hercproceed.umd.edu/~epelzl/index.html

**Acknowledgments**
Thanks to Steve Ross and Payman Vafaee for helpful discussion and advice