The FBI develops a Translation Aptitude Test (TAT)

Maria M. Brau
Chief, Language Testing and Assessment Unit
TOPICS

- Background
- Translation Ability Defined
- Construct
- Test Format
- Sample Items
- Piloting
- Results
- Proof of Concept
Background

- FBI Applicant Test Battery:
  - DLPT Listening/Reading
  - Verbatim Translation Exam (VTE) FL to Eng
  - 2 SPTs: FL and Eng

- Recruiter’s Questions:
  - “What if there is no VTE? What can we give them to find out if they have some ability to translate? How about a language–neutral test?”
Ability to Translate

- Prerequisite for any translation task: the translator must be able to
  1) read and comprehend the source language and
  2) write in the target language.

- Given these prerequisites, the translator must also be able to
  3) choose the equivalent expression in the target language that both fully conveys and best matches the meaning intended in the source language (normally referred to as congruity judgment).*

*ILR SLDs for Translation Performance
Construct

Language–neutral?
- Meaning no foreign language?
- ∴ English to English

Ability to Translate?
- English to English equivalents?
- ∴ PARAPHRASE

Hypothesis:
- Paraphrase process similar to translation process
Paraphrase Part I:
  ◦ Simplification: Revise a text to make it easier to understand
    • To serve as warm-up

Paraphrase Part II:
  ◦ Equivalents: Reword a text without changing the original meaning
    • Process similar to translation
Paraphrase Instructions

- Do not use any words from the original, except for “the.”
- Do not add or omit any content.
- Make sure your paraphrase is not awkward or difficult to understand.
Paraphrase Test Items

- Ten items
  - Eight short (one line) sentences
  - One two-line sentence
  - One three-line sentence
- All items authentic
Original text:
- The sooner the better.
- Paraphrase:
  - Earlier is preferable.

Original Text:
- I don’t know too much about this.
- Paraphrase:
  - It is not my area of expertise.
In an evening interview, the president explained why he was resigning.

During a late afternoon question and answer session, the Chief Executive clarified the reason for his stepping down.

Around dinner time, the Chief Executive met with a reporter and made clear his reason for leaving office before finishing his term.
Lack of an overall picture makes it difficult to detect long-term trends.

Without a broad perspective, finding out developmental patterns through several time periods is not easy.
Before choosing a course of action one should be aware of two things: the implications of the action and the range of possible options.

Prior to making a decision better check: (1) likely consequences; (2) feasible alternatives.
Piloting Procedure

- **Administer:**
  - FBI ILR–based Verbatim Translation Exam (VTE)
    - Foreign Language to English
  - Translation Aptitude Test (TAT)
    - English to English Paraphrase
  - **Requisite:**
    - DLPT (Reading) 2+ or higher

- **Rate:**
  - VTE and TAT with similar rating procedures

- **Compare:**
  - VTE scores to TAT scores
Rating Procedures for VTE

- ILR translation text level assigned to each passage
  - Passage 1: Level 2
  - Passage 2: Level 2+
  - Passage 3: Level 3
  - Passage 4: Level 4

- Translation for each of 4 passages is rated as:
  - SM: Successfully matches the source
  - MM: Mostly matches
  - IM: Intermittently matches
  - RM: Rarely matches

- Final ILR score
Each sentence paraphrase rated as:
  ◦ SM, MM, IM, or RM

No ILR Descriptions for Paraphrase

∴ Number values were used:
  ◦ 4 = SM
  ◦ 3 = MM
  ◦ 2 = IM
  ◦ 1 = RM
Piloting Participants

- Subjects: 164  Usable Tests: 144
  - NSA: 128  NSA: 111
  - DIA: 25  DIA: 19
  - FBI: 17  FBI: 14

- Languages
  - Albanian, Arabic, Bulgarian, Chinese, Croatian, Farsi, French, German, Hebrew, Hindi, Italian, Japanese, Korean, Pashto, Portuguese, Russian, Spanish, Turkish, Urdu
# Participant VTE Ratings

<table>
<thead>
<tr>
<th>VTE Rating</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1+</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>2+</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>3+</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>144</td>
</tr>
</tbody>
</table>
## ILR Level Ranges for TAT Scores

<table>
<thead>
<tr>
<th>VTE Rating</th>
<th>Number of Participants</th>
<th>TAT Range</th>
<th>Below Range</th>
<th>In Range</th>
<th>Above Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1+</td>
<td>34</td>
<td>0 – 21</td>
<td>0</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>22 – 26</td>
<td>8</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>2+</td>
<td>43</td>
<td>22 – 26</td>
<td>8</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>3/3+</td>
<td>17</td>
<td>27 – 40</td>
<td>1</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the ILR SLDs for Translation Performance, ‘translation’ does not occur until the Level 3.
TAT Inter-rater Reliability

- 2 raters
- 144 cases
- 288 decisions
- Krippendorff’s alpha (interval): 0.90
  - High inter-rater reliability (> 0.80) indicates that there was a measurable construct to be evaluated.
Results: Correlation

- Is there a relationship between paraphrasing (TAT) and translating (VTE)?
- Pearson Correlation
  - $R = 0.52$
    - A moderate positive correlation: there is a tendency for high TAT scores to go with high VTE scores and vice versa.
  - $P < .001$
    - Indicates that the correlation is significant.
Scatter Plot of VTE and TAT Scores

TAT/VTE Score Scatter Plot

VTE Score vs. TAT Score graph with data points plotted and three distinct groups highlighted by circles.
Scatter Plot of VTE and TAT Scores

TAT/VTE Score Scatter Plot

- VTE Score
- TAT Score
- TAT/VTE Score Scatter Plot
Proof of Concept

- There is a relationship between paraphrasing and congruity judgment (translation).
- The TAT construct shows promise for predicting translation ability if...
  - It had more items.
  - The items were aligned with the ILR.