Linking Can-do Statements with Language Tests Using Neural Test Theory

Yukie Koyama* and Tetsuo Kimura**

*Nagoya Institute of Technology **Niigata Seiryo University
Can-do Statements

- Common European Framework of Reference for Languages (CEFR) 2001
- to establish a goal of each level of language learning
- to enhance autonomous learners’ self-awareness

- Validity of the CDSs of CEFR: an online language test, DIALANG, as an external evaluation test

Linking CDSs and a test is important for the validation of the CDSs.
What is NTT?

- Not a tele-communication company
Testing Theory

Limitations of NRM

Discrimination Index
Mean score
Standard deviation
Reliability

1930s ~

Classical Test Theory

Lord (1950)

Item Response Theory

Rasch Model
2PLM
3PLM

Neural Test Theory
Shojima (2008)
Latent Variables

Latent Variables Model

- Continuous Variables
  - IRT
  - Factor Analysis
- Ordinal Variables
  - NTT
IRT & NTT

- IRT --- Continuous scale

- NTT --- Ordinal scale, Rank or Level,
Evaluation with an ordinal scale makes the evaluation less precise but improves accountability by making it easier to explain the academic ability corresponding to each grade. Along with this, the specific targets and achievements of the test become clearer.
Objective of this research

- to examine the appropriateness of Neural Test Theory (NTT) in order to analyze Can-do Statements (CDSs) for the purpose of linking them with the corresponding language tests
Case Study 1: EIKEN Can-Do List

- **Participants:** 220 (295) freshmen
  - engineering (64), nursing (54), and social welfare and psychology (102)

- **CDSs on the EIKEN Can-Do List:** 109

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>L</th>
<th>S</th>
<th>W</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Pre 1</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Grade 2</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Grade Pre 2</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Grade 3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>Grade 4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>26</td>
<td>27</td>
<td>30</td>
<td>26</td>
<td>109</td>
</tr>
</tbody>
</table>
Case Study 1: EIKEN Can-Do List

- **Sample CDSs**
  - **Reading**
    - Can find streets, shops, and hospitals, etc. on simple maps written in English. [Grade 3]
    - Can understand the main points of lengthy texts (e.g. required readings and materials for lectures and training courses). [Grade Pre 1]
  - **Listening**
    - Can understand the meaning of simple instructions (e.g. "Open your textbook." / "Close the door, please."). [Grade 4]
    - Can understand a speaker on the telephone, provided the content is simple (e.g. agreeing when to meet, taking short messages). [Grade Pre 2]
Case Study 1 : EIKEN Can-Do List

- **Sample CDSs**
  - **Speaking**
    - Can talk briefly about something that he/she is interested in (e.g. his/her hobbies, club activities). [Grade 3]
    - Can make a complaint about products or services (e.g. about damaged products or unsatisfactory service). [Grade Pre 1]
  - **Writing**
    - Can write sentences using English word order, provided that the sentences are short (e.g. "I went to the park yesterday."). [Grade 4]
    - **G2**: Can describe the details of memorable experiences (e.g. school events, trips). [Grade 2]
Case Study 1 : EIKEN Can-Do List

- **Placement Test**
  - Vocabulary & Grammar (Vgm)
  - Listening comprehension with dialogue (Dlg)
  - Listening comprehension with monologue (Mlg)
  - Reading comprehension (Rdg)

All the items were adopted from the EIKEN Test Grade 3, Grade pre 2, Grade 2 and Grade pre 1 in 2007 and 2008, under the permission of STEP.

Four versions were constructed: equated with common anchor items that were calibrated in the previous study (Kimura, 2009).
Case Study 1: EIKEN Can-Do List

Procedure

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDSs</td>
<td>Dichotomous model NTT-SOM</td>
<td>The number of rank: 5</td>
</tr>
<tr>
<td>PT Vgm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT Listening (Dlg, Mlg)</td>
<td>Graded model NTT-SOM</td>
<td>The target latent rank distribution: not specified</td>
</tr>
<tr>
<td>PT Reading</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NTT-SOM: NTT that uses a self-organizing map mechanism

Software: Exametrika Ver.4.4 (Shojima, 2010)
Case Study 1: EIKEN Can-Do List

Results and Discussion

CDSs’ Difficulty (IRP index \( \beta \)) and EIKEN Grade

\( \beta \) is the location of the latent rank when the IRP value is closest to 0.5, simply expresses item difficulty.

Spearman's rho (\( \rho \))

- Reading ------ .93
- Listening ------ .94
- Speaking ------ .94
- Writing ------ .95
Case Study 1: EIKEN Can-Do List

IRP index $\beta$ and $b$

Kumagai (2007)
Case Study 1 : EIKEN Can-Do List

- Results and Discussion
  - CDSs’ Discrimination power (IRP index $a$) : .02-.34
  - Index $a$ is the maximum difference in the IRP value among all adjoining rank pairs.
Case Study 1: EIKEN Can-Do List

IRP index $\alpha$ and $a$

Kumagai (2007)
Case Study 1: EIKEN Can-Do List

- **Results and Discussion**
  - The estimation of latent ability analyzed by NTT is best described in rank membership profile (RMP).
Case Study 1: EIKEN Can-Do List

Results and Discussion

Self-evaluation based on CDSs and result of placement tests: Spearman's rho (ρ) was low

- CDSs Reading & PT Reading ------- .22
- CDSs Listening & PT Listening ---- .28

Alpha reliability coefficient was high

- CDSs Reading ----- .87
- CDSs Listening ----- .89
- CDSs Speaking ----- .91
- CDSs Writing ------- .92

- PT Vgm -----------------.81
- PT Listening ------ .88
- PT Reading -------- .82
Case Study 1 : EIKEN Can-Do List

- Results and Discussion
  - Self-evaluation based on CDSs and result of placement tests (Listening)

<table>
<thead>
<tr>
<th>Lng</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>R2</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>R3</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>R4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>R5</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
</tbody>
</table>

Over-self-estimation 26%
Under-self-estimation 15%
59%
Case Study 1: EIKEN Can-Do List

- **Results and Discussion**
  - **Self-evaluation based on CDSs and result of placement tests (Reading)**

<table>
<thead>
<tr>
<th>Rdg</th>
<th>CDS</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td></td>
<td></td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>R2</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R4</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>R5</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Over-self-estimation**: 14%
- **Under-self-estimation**: 29%
- **Total**: 57%
Case Study 1: EIKEN Can-Do List

- **Conclusion**
  - **CDSs on the EIKEN Can-Do List are reliable**
    - Alpha reliability coefficient: $\rho = .87 \sim .92$
    - CDSs’ Difficulty (IRP index $\beta$) and EIKEN Grade are highly correlated: $\rho = .93 \sim .95$
  - **Correlations between self-evaluation based on CDSs and result of placement tests are low: $\rho = .22 \sim .28$**
    - Many students over- or under-estimate their skills
    - More tend to over-estimate their listening skills
    - More tend to under-estimate their reading skills
Case Study II

- CDS : English for General Science & Technology
  Developed by English teachers of NIT
  4 skills  5 items per skill
  5 point scale  1. cannot perform at all
  5. can perform easily  (Koyama, 2008)

- Test : Final Examination of NIT EGST course (EXAM)
  100 items  Listening -- 30
  Reading, vocabulary, grammar -- 70
  multiple choice questions
Participants

- First year students of NIT
  - CDS -- 882
  - Exam -- 942
- Mark sheet
NIT CDSs  version 2

e.g.

Reading

1. I can read and understand a menu.
2. I can read a manual of an apparatus and understand how to use it.
3. I can read and understand an abridged edition of a novel.
4. I can read and understand the details of homework.
5. I can read and understand a scientific article of a newspaper.
Procedure

- CDSs and EXAM (an external evaluation test of the CDSs) were analyzed using CTT, Rasch model and NTT.
  (Since the CDS data were polychotomous in this case, the data were analyzed using Andrich’s Rating Scale Model for Rasch model, and Graded Neural Test (GNT) model for NTT.)

- Analysis of NTT -- Exametrika (Shojima, 2008)
- The number of ranks: 5 both for CDS and EXAM
Result 1. CTT (1)

<table>
<thead>
<tr>
<th></th>
<th>Simple Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td><strong>N of Examinees</strong></td>
<td>882</td>
</tr>
<tr>
<td><strong>N of Items</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>Min</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>53</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>52.375</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>130.532</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>11.425</td>
</tr>
<tr>
<td><strong>Alpha Coefficient</strong></td>
<td>0.936</td>
</tr>
</tbody>
</table>
Result 2. CTT (2) Score Distribution

CDS (Total)

Exam 素点分布
Result 3. Rasch model (1)
Result 4. Rasch model (2) Difficulty of CDSs & Ability of Examinees
Result 5. Rasch model (3)
Difficulty of EXAM & Ability of Examinees
Test Reference Profile

the expected score of test-takers in each latent Rank

EXAM is, on the whole, easier for the test-takers’ levels.
### Cross-tabulation (Ranks of CDS x EXAM) 5 × 5

<table>
<thead>
<tr>
<th>EXAM</th>
<th>CDS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RI</td>
<td>R2</td>
<td>R3</td>
<td>R4</td>
<td>R5</td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>49</td>
<td>22</td>
<td>27</td>
<td>26</td>
<td>13</td>
<td>137</td>
</tr>
<tr>
<td>R2</td>
<td>31</td>
<td>21</td>
<td>25</td>
<td>20</td>
<td>27</td>
<td>124</td>
</tr>
<tr>
<td>R3</td>
<td>26</td>
<td>31</td>
<td>42</td>
<td>31</td>
<td>36</td>
<td>166</td>
</tr>
<tr>
<td>R4</td>
<td>28</td>
<td>21</td>
<td>32</td>
<td>36</td>
<td>45</td>
<td>162</td>
</tr>
<tr>
<td>R5</td>
<td>34</td>
<td>45</td>
<td>70</td>
<td>53</td>
<td>84</td>
<td>286</td>
</tr>
<tr>
<td>total</td>
<td>168</td>
<td>140</td>
<td>196</td>
<td>166</td>
<td>205</td>
<td>875</td>
</tr>
</tbody>
</table>
Cross-tabulation
(Ranks of CDS x EXAM)  \(3 \times 3\)

<table>
<thead>
<tr>
<th>EXAM</th>
<th>CDS</th>
<th></th>
<th></th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R1&amp;R2</td>
<td>R3&amp;R4</td>
<td>R5</td>
<td></td>
</tr>
<tr>
<td>R1&amp;R2</td>
<td>123</td>
<td>98</td>
<td>40</td>
<td>261</td>
</tr>
<tr>
<td>R3&amp;R4</td>
<td>106</td>
<td>141</td>
<td>81</td>
<td>328</td>
</tr>
<tr>
<td>R5</td>
<td>79</td>
<td>123</td>
<td>84</td>
<td>286</td>
</tr>
</tbody>
</table>
Relation Between Ranks of CDSs and EXAM

<table>
<thead>
<tr>
<th>EXAM-R1&amp;2</th>
<th>EXAM-R3&amp;4</th>
<th>EXAM-R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS-R5</td>
<td>40</td>
<td>81</td>
</tr>
<tr>
<td>CDS-R3&amp;4</td>
<td>98</td>
<td>141</td>
</tr>
<tr>
<td>CDS-R1&amp;2</td>
<td>123</td>
<td>106</td>
</tr>
</tbody>
</table>
Summary

- Compared with other test models based on the continuous scale, NTT, a test model for graded evaluation based on the ordinal scale, provides us clearer criteria for both CDSs and learners' levels/stages.

- We do not have to think about where to divide the results. Therefore, we can easily connect the learners' levels/stages to CDSs. In other words, we can easily link test results with CDSs when NTT is used for the analysis.
Limitations

- One of the limitations and disadvantages of NTT is that the methodology of item fit analysis or aberrant responses detection has not been developed sufficiently yet. In addition, cumulative of empirical data analysis based on NTT is not enough.

- However, as stated above, NTT has an essential advantage as a testing theory to analyze the test results to be classified into ranks such as those of CDSs.
References


References