Using CEFR-linked exams in the university entrance examination system in Japan

Rie KOIZUMI (Juntendo University)

rkoizumi@juntendo.ac.jp


September 22
I appreciate the conference organizing committee members, especially Dr. Nick Zhiwei Bi for inviting me and organizing this conference.

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant-in-Aid for Scientific Research (C), Grant Number 26370737.
Conference theme

- Aligning standards with assessments
- In the Japanese context, aligning CEFR with
  - university entrance examinations
  - standardized tests made by private test institutions
  - scholastic assessment
  - classroom assessment
- Focus on L2 English exams
Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001)

- Transparent, coherent, and comprehensive basis for guiding the teaching, learning, and assessment
- Proficient language user C2
- C1
- Independent language user B2
- B1
- Basic language user A2
- A1
C2: Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

B1: Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

A1: Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.
## Adjusting CEFR to Japan: CEFR-J
(Negishiki & Tono, 2014)

<table>
<thead>
<tr>
<th>PreA1</th>
<th>A1.1</th>
<th>A1.2</th>
<th>A1.3</th>
<th>A2.1</th>
<th>A2.2</th>
<th>B1.1</th>
<th>B1.2</th>
<th>B2.1</th>
<th>B2.2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
<td>A2</td>
<td></td>
<td></td>
<td></td>
<td>B1</td>
<td></td>
<td>B2</td>
<td></td>
<td>C1</td>
<td>C2</td>
</tr>
</tbody>
</table>

- Proficient language user  C2
- Independent language user  B2 → 2 levels
- Basic language user  A2 → 2 levels
- Pre-basic language user  Pre-A1
- Basic language user  A1 → 3 levels

B1 → 2 levels
Use of CEFR in the educational policy (based on e.g., Central Council for Education, 2016)

- From elementary school to senior high school
- Teaching:
  - Current: A1 to A2- → Future: A1+ to B1
  - Relate national curriculum goals to CEFR levels and Can-Do descriptors
  - Use modified CEFR Can-Do descriptors for teaching goals in classroom
- Testing: Relate university entrance exams to CEFR levels
Current types of university entrance examinations (Kuramoto & Koizumi, 2018)

- (1) General examinations
  - (a) National Center Test for University Admissions (Center Test)
    - Required for national and public university admissions
    - Often required for private university admissions
    - Largest No. of test takers: 550,000
  - (b) University-developed tests
- (2) Recommendation-based examinations
- (3) Admissions Office (AO) examinations
- (4) Special selection examinations
Center Test (until 2019; Sasaki, 2008; Watanabe, 2013)

- Developed by the National Center for University Entrance Examinations (National Center)
- Construct: reading (R), listening (L), knowledge
- Format: multiple-choice (MC) with single answers
- Problems:
  - Difficult to assess the ability to think logically and express effectively
  - Having negative washback on teaching/learning
Reform of the Center Test
(see Hakui & Osugi, 2017; McCrostie, 2017)

- University Entrance Common Test (Common Test) from 2020 to 2023
  - Construct: RL + writing (W), speaking (S) + ability to think logically and express effectively
- (1) RL test (at the trial stage)
  - Developed by the National Center
  - Format: MC with no, single, or multiple answers
- (2) RLWS tests developed by private test institutions (private RLWS tests)
RL test in the Common Test
(National Center, 2018)

- Constructed based on clear test specifications with CEFR levels and Can-Do descriptors
- Covers A1 to B1 levels
- More difficult with more diverse text types
- Requires more deep comprehension
- Listen *once* or *twice* (vs. *twice* now)
- Relationship between the RL test and CEFR
  - Analysis based on content and expert judgment
  - Not yet empirically examined
RL test in the Common Test

- Feedback from experts and the public
  - ○ Appropriate format similar to real-life situations
  - ○ Clear test specifications leads to good practices.
  - ○ Test takers listen to passages once or twice. This leads to positive washback.
  - ○ & × Abolishing pronunciation and grammatical knowledge items
  - × MC with multiple or no answers has low discrimination (Japan Association for Research on Testing, 2018).
Private RLWS test in the Common Test
(Ministry of Education, 2018)

- **8 tests used**: Cambridge English, Eiken, GTEC, IELTS, TEAP, TEAP CBT, TOEFL iBT, and TOEIC LR & SW,
  - Construct: broadly the same: RLWS, but a little different
  - Purposes, formats, prices, access to tests vary across tests.
- Test takers can take them twice in the year of entrance exams.
- Their scores are sent to the National Center and then sent to universities students apply to.
Example 1: TEAP (Test of English for Academic Purposes)

- Developed by the Eiken Foundation of Japan and Sophia University [http://www.eiken.or.jp/teap/]
- Construct: academic English proficiency required for learning and researching at universities
- CEFR: about A2 to C1
- S: 10-min face-to-face, one-on-one interviews
- S: 0 to 30 points
TEAP speaking tasks and criteria

- Part 1: Interview
- Part 2: Role play (interviewing an examiner)
- Part 3: Monologue
- Part 4: Extended interview
- 5 criteria
  - (a) Pronunciation, (b) Grammatical Range and Accuracy, (c) Lexical Range and Accuracy, (d) Fluency, (e) Interactional Effectiveness
Example 2: GTEC CBT (Global Test of English Communication Computer Based Testing)

- Developed by the Benesse Corporation and Center for Entrance Examination Standardization (http://www.benesse-gtec.com/cbt/en)
- Construct: English proficiency in four skills for academic purposes
- CEFR: about A1 to C1
- S: 20-min computer-based test
GTEC CBT speaking tasks and criteria

- **Part 1: Listening and responding (6 items)**
  - E.g., Where do people like to travel to in your country?
  - When is the best time of year to visit your country?
- **Part 2: Delivering and asking for information (3 items)**
- **Part 3: Expressing your opinion (3 items)**
  - Do you think technology has changed the way we live?
- 23 criteria in total. Each item has 1 to 5.
Private RLWS test in the Common Test

- **8 tests used**: Cambridge English, Eiken, GTEC, IELTS, TEAP, TEAP CBT, TOEFL iBT, and TOEIC LR & SW,
  - **Construct**: broadly the same: RLWS, but a little different
  - **Purposes, formats, prices, access to tests vary across tests.**
- Test takers can take them twice in the year of entrance exams.
- Their scores are sent to the National Center and then sent to universities students apply to.
Score concordance table: CEFR and eight exams (Ministry of Education, 2018)

<table>
<thead>
<tr>
<th>CEFR</th>
<th>Cambridge</th>
<th>Eiken</th>
<th>GTEC</th>
<th>IELTS</th>
<th>CBT</th>
<th>TEAP</th>
<th>TOEFL</th>
<th>TOEIC</th>
<th>LR &amp; SW</th>
</tr>
</thead>
</table>

- [http://www.mext.go.jp/b_menu/houdou/30/03/__icsFiles/afieldfile/2018/03/26/1402610.pdf](http://www.mext.go.jp/b_menu/houdou/30/03/__icsFiles/afieldfile/2018/03/26/1402610.pdf)
How to use private RLWS tests: Examples
(Japan Association of National Universities, 2018)

1. Requirement for applying to a university
   - E.g., University A requires test candidates to have CEFR A2 level.

2. Addition of private RLWS test scores to RL test scores. Use the combined score for rank ordering test takers
   - Option 1: Use converted CEFR scores
     - E.g., RL test, 200 points + private test, 20 points
   - (Option 2: Use finely divided scores)

3. Both
   - Each university has to decide how to use them.
Positive feedback on the private RLWS test system (e.g., Negishii, 2017; Yasukochi, 2018)

- Positive washback is expected.
  - Teachers can focus more on WS in class, and overall RLWS abilities will increase.
- Clear goals can motivate students.
  - We can easily compare current levels with international standards and set the next goals.
- Obtained test scores can also be used for studying abroad.
Criticism toward the private RLWS test system (e.g., Abe, 2018; Haebara, 2018; Terasawa, 2018; Yanase, 2018) 1/4

- Doubts about exams (e.g., test quality, scoring quality)
- Limits on comparability of test scores
- Limits on discrimination
  - Most test takers are at A1 or A2 levels.
- Little positive washback
  - What is assessed on tests is limited (e.g., S monologue; formulaic type of S).
  - Causality (RLWS tests → better teaching of WS → increase in WS abilities) has not been tested.
Comparability of private test scores

- Each exam has a little different constructs.
- Uses a different format, and a different standard-setting method to link with CEFR.
- May have a different standard for CEFR levels.
- They may not be strictly comparable but will be used as comparable.
  - They are linked to CEFR but not equated.
  - IELTS’s A2 may not be the same as TOEFL iBT’s B1. (see Green, 2018; Harsh, 2018)
Criticism 2/4

- **Negative effects on English education**
  - What is on tests is not always consistent with the national curriculum and what is actually taught.
  - Most English classes may focus on private test preparation.
  - Less class time on reading, which is the basis of the English proficiency

- **Negative effects on high school education in general**
  - Schools may need to decrease the number of school activities, such as sports festival.
Criticism 3/4

Financial burden on test takers

- Students from high socioeconomic families are likely to obtain higher scores because of more opportunities to practice tests.

Different opportunities for test takers in urban/non-urban areas

- Students living in non-urban areas are likely to find fewer opportunities to take a test. They need to pay extra money to go to the test site.
Criticism 4/4

- Difficult to get sufficient test centers for 550,000 candidates taking tests twice
- Difficult to provide adequate accommodation to those with disability
- The decision-making process was not clear.
  - Company-related members were involved from the start.
  - Thorough discussions considering feasibility and negative washback were not made before the decision.
Serious, continuing discussions

- Conference at the U. of Tokyo (February 2018)
- Japan Association of National Universities proposed ways to use both RL and RLWS test scores. (June 2018; Japan Association of National Universities, 2018)
- Tokyo U. announced no use of private tests. They retracted their plan (March & April 2018)
- Publication on this issue (Haebara, June 2018)
- Symposia at Conferences (June to August 2018)
- Tokyo U. announced there have been internal discussions on this issue. They may not use private tests. (July 2018)
- Responses from the National Center (August 2018)
- Daily tweets on this matter (up to now)
Exploring future directions 1/3

- Negative impact on learning
  - Narrowing of the curriculum; Less time on R
- → Conduct teacher training on teaching and assessment
- → Enhance assessment literacy of test users (university admission officers, policymakers; e.g., how to interpret and use results)
- → Encourage test companies to conduct validation studies and publicize information on tests themselves, test development processes, measurement error, etc.
Exploring future directions 2/3

- Limits on comparability & discrimination
- → Give lighter weighting of the scores?
  - Use it only as the requirement for applying to universities
  - Give only small points to the total scores
  - Make a finer conversion table, by obtaining more data and examining relationships across tests?
Limits on fairness

The National Center will develop an SW test for the Common Test (e.g., Fujiwara, 2018).

- In collaboration with test companies
- To avoid a situation in which collusion with one company is a suspect, ask one private test company to create an S test and another to create a W test (Koizumi, 2018)

The English Common Test will be abandoned. Each university decides what to do in their English exam (e.g., Fujiwara, 2018).

Use SW school-based assessment (SBA) for the Common Test (A professor at a national university, personal communication, August 18, 2018)

- Hong Kong: 15% for S SBA (Hong Kong Diploma of Secondary Education Examination, 2018)
Issues to consider when the National Center creates an SW test

- Use the national curriculum as the basis
- Select appropriate tasks and procedures
- Decide on how to elicit and record performance
  - Use of tablets/computers
- Scoring
  - Automated scoring: not feasible at the moment
  - Double human scoring; quality control and monitoring of raters
Issues to consider when the National Center creates an SW test

- Prices
  - Prices should not be increased drastically.
- Learn from China and other countries with a large number of test takers
Using CEFR-linked exams in the university entrance examination system in Japan

- 2020 to 2023: The system in the today’s talk
- 2024 and onward?: Continue with, modify, or abandon this system?
- Ministry of Education needs to increase financial and human efforts to improve Japan’s assessment system & teacher training and to enhance assessment literacy of teachers and the public, to make future reforms meaningful.
References


Negishi, M (根岸雅史). (2017).『テストが導く英語教育改革「無責任なテスト」への処方箋』三省堂


