

Linking a test of English for academic purposes to CEFR and its implication to the Japanese college entrance exam reform

Keita Nakamura

Eiken Foundation of Japan

Overview

- 1. Research context**
- 2. The goals of this project**
- 3. Workshop design**
- 4. Workshop result**
- 5. New CEFR reference table**
- 6. Critical issues related to the implementation of the new system**

Research context: Educational Reform

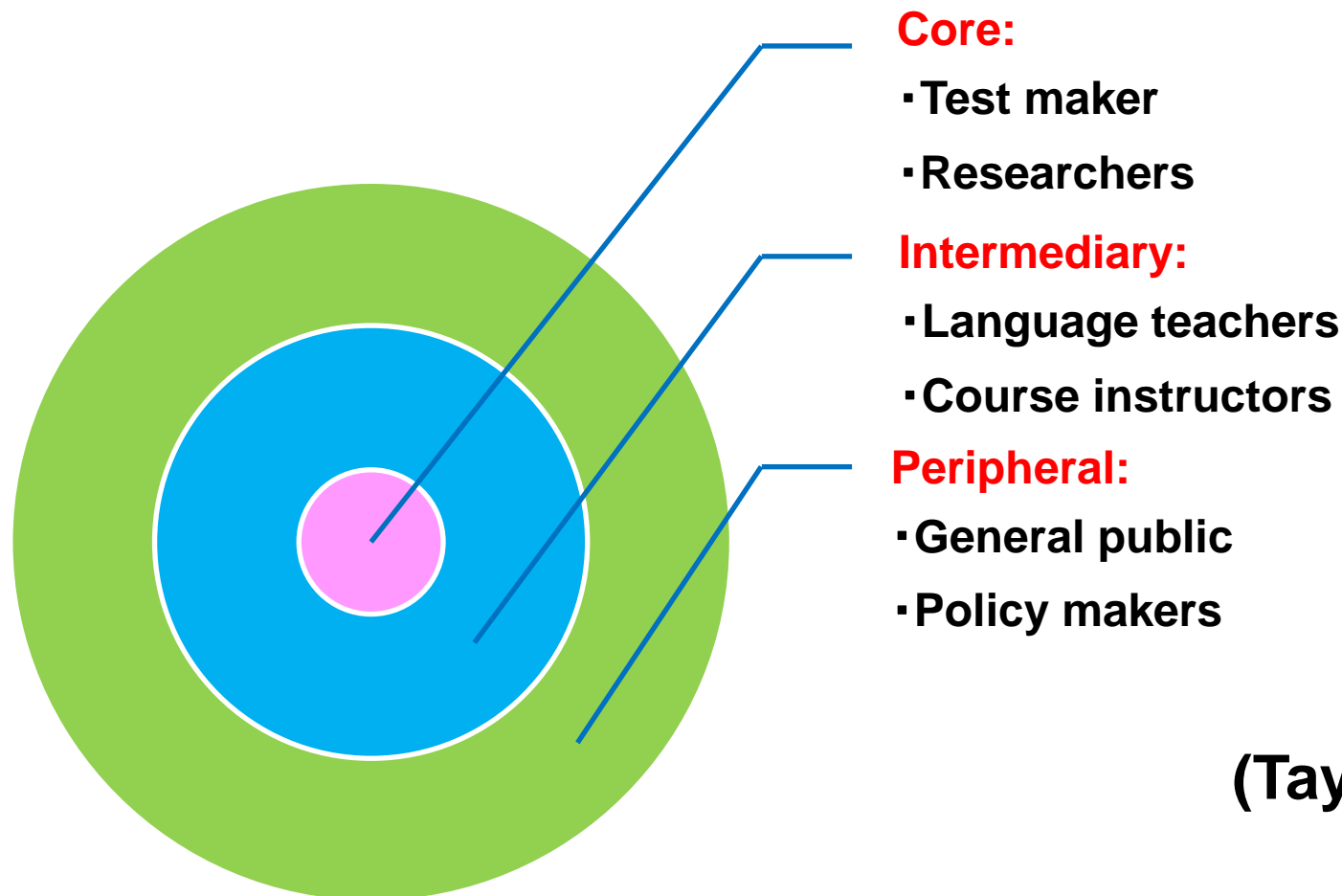
Global Educational Reform Movement (GERM)

“Reform policies and related accountability assessments require a coherent system that makes explicit the **interconnections among policy goals, testing functions, validity research, and the groups/systems impacted.**”

(Chalhoub-Deville, 2016)

Language Assessment Literacy (LAL)

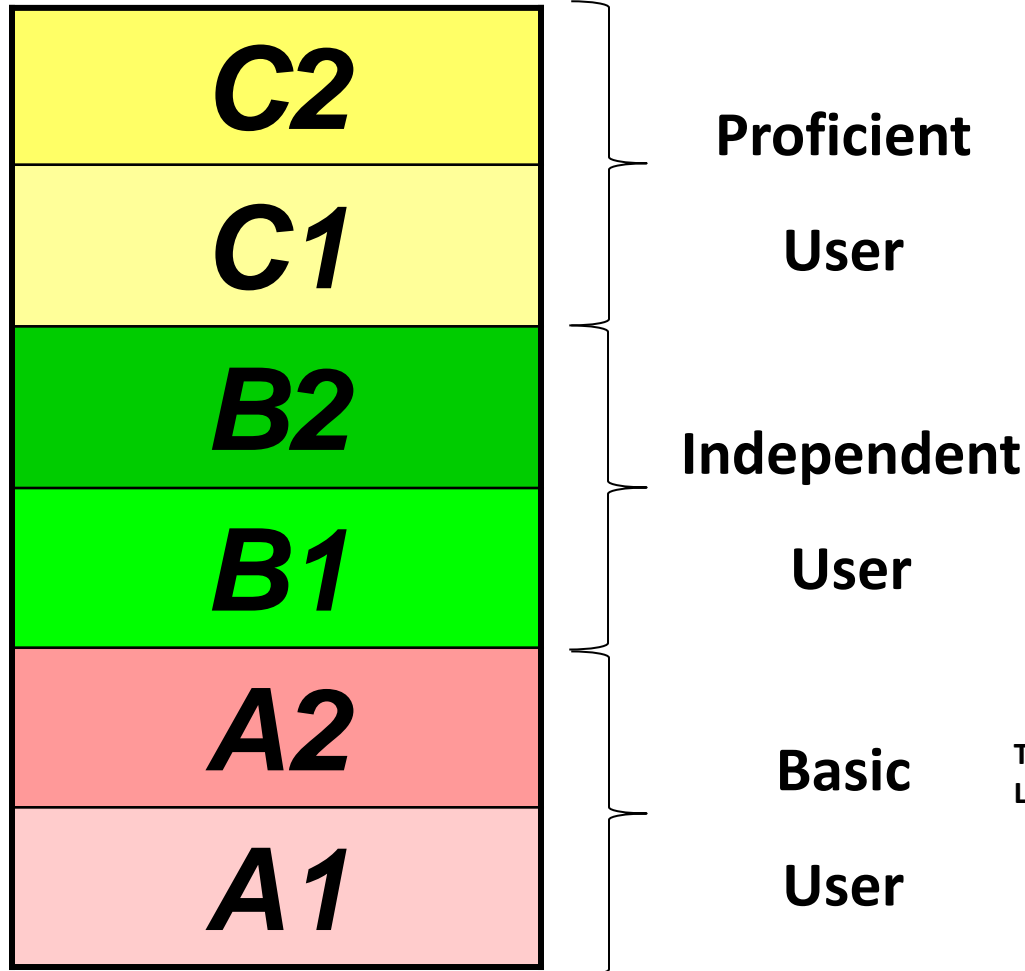
Stakeholders' familiarity with measurement practices and the application of this knowledge to issues of assessing language.
(Modified from Malone, 2013)



(Taylor, 2013)

Research context: CEFR

6 main levels



The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (Council of Europe, 2001)

Research context: Educational Reform in Japan

Ministry of education's new plan to revise English education (MEXT, 2014)

- 1) Revision concerns a focus on the balanced teaching and learning of the four English skills of reading, writing, listening, and speaking.

High school graduates are expected to achieve a level of English competency equivalent to CEFR level B1-B2, while junior high school students are expected to achieve CEFR level A1-A2.

	Before	After
Elementary School	35 hours a year at G5/6	35 hours a year at G3/4 70 hours a year at G5/6 600-700 words
Junior High School	1200 words	1600-1800 words + ES
Senior High School	1800 words	1800-2500 words +ES+SH

Research context: Educational Reform in Japan

Ministry of education's new plan to revise English education (MEXT, 2014)

2) Assessment of the four English skills in university entrance examinations.

Universities are allowed to utilize externally developed standardized English tests of four skills in lieu of an in-house test. Each university is also expected to utilize multiple sources of information, such as high school GPA, interviews, and essays, when making admission decisions.

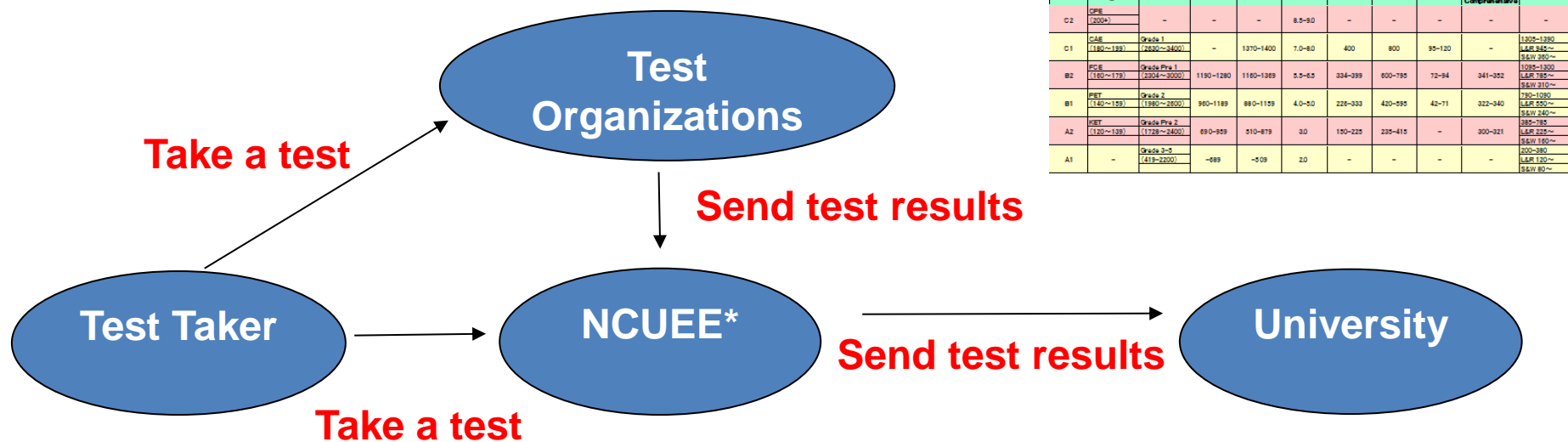
CEFR	Cambridge English	EIKEN	GTEC	GTEC CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEFL Junior Comprehensive	TOEIC L&R TOEIC S&W
C2	CPE (200+)	–	–	–	8.5–9.0	–	–	–	–	–
C1	CAE (180~199)	Grade 1 (2630~3400)	–	1370–1400	7.0–8.0	400	800	95–120	–	1305–1390 L&R 945~ S&W 360~
B2	FCE (160~179)	Grade Pre 1 (2304~3000)	1190–1280	1160–1369	5.5–6.5	334–399	600–795	72–94	341–352	1095–1300 L&R 785~ S&W 310~
B1	PET (140~159)	Grade 2 (1980~2600)	960–1189	880–1159	4.0–5.0	226–333	420–595	42–71	322–340	790–1090 L&R 550~ S&W 240~
A2	KET (120~139)	Grade Pre 2 (1728~2400)	690–959	510–879	3.0	150–225	235–415	–	300–321	385–785 L&R 225~ S&W 160~
A1	–	Grade 3–5 (419–2200)	–689	–509	2.0	–	–	–	–	200–380 L&R 120~ S&W 80~

Research context: Educational Reform in Japan

Ministry of education's new plan to revise English education (MEXT, 2014)

3) Revision of the National Center Test(NCT)

Current 2-skill (Reading/Listening) NCT will be continued until 2023 when it will be abolished. Until 2023, applicants are given three choices a) take the NCT, b) take the authorized standardized tests, or c) take both. Students can take the standardized tests up to two times for admission purposes.



*NCUEE

National Center for University Entrance Examinations

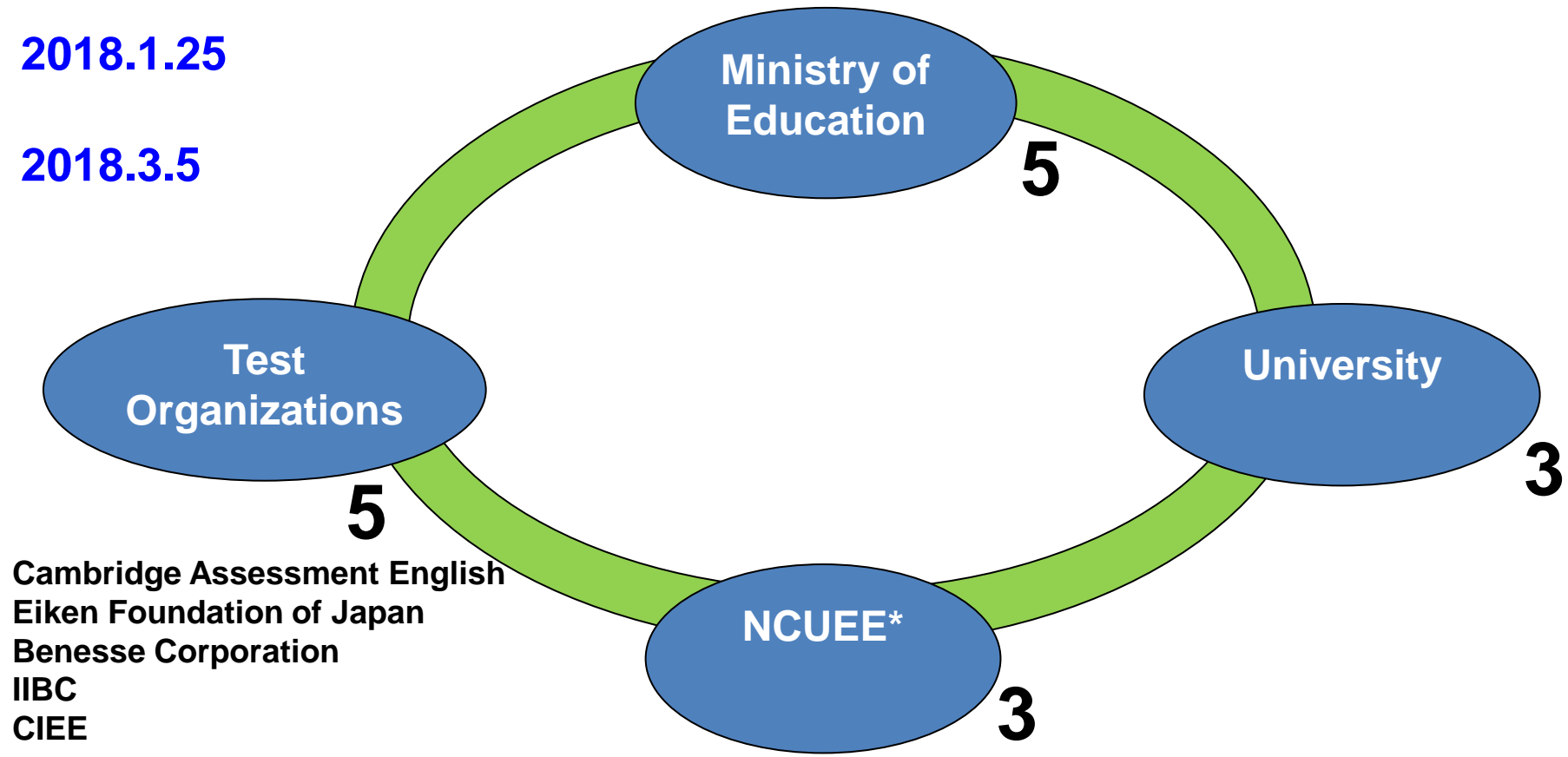
Research context: Groups/Systems impacted

Stakeholders of working group held by the ministry 2017-2018

2017.9.25

2018.1.25

2018.3.5



*NCUEE
National Center for University Entrance Examinations

Research context: Groups/Systems impacted

Meetings of the working group held by the ministry 2017-2018

2017.9.25

- Testing organizations explained the linking process and result
- Members discussed the method and process to update the CEFR reference table

Issues:

⇒Each test has its own original construct and scoring scale (pass/fail)

⇒Each test is linked to the CEFR by using different standard setting method

CEFR	Cambridge English	EIKEN	GTEC	GTEC CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEFL Junior Comprehensive	TOEIC L&R TOEIC S&W
C2	CPE (200+)	—	—	—	8.5–9.0	—	—	—	—	—
C1	CAE (180~199)	Grade 1 (2630~3400)	—	1370–1400	7.0–8.0	400	800	95–120	—	1305–1390 L&R 945~ S&W 360~
B2	FCE (160~179)	Grade Pre 1 (2304~3000)	1190–1280	1160–1369	5.5–6.5	334–399	600–795	72–94	341–352	1095–1300 L&R 785~ S&W 310~
B1	PET (140~159)	Grade 2 (1980~2600)	960–1189	880–1159	4.0–5.0	226–333	420–595	42–71	322–340	790–1090 L&R 550~ S&W 240~
A2	KET (120~139)	Grade Pre 2 (1728~2400)	690–959	510–879	3.0	150–225	235–415	—	300–321	385–785 L&R 225~ S&W 160~
A1	—	Grade 3–5 (419–2200)	—689	—509	2.0	—	—	—	—	200–380 L&R 120~ S&W 80~

Research context: Groups/Systems impacted

Meetings of the working group held by the ministry 2017-2018

2018.1.25

- Testing organizations updated the linking/scoring process and result
- Members discussed the necessary information for the CEFR reference table/score reporting process to be appropriately used

Issues:

⇒Some tests output pass/fail, other tests output scores

⇒Some tests have balanced skill scores, others have weighted skill scores

CEFR	Cambridge English	EIKEN	GTEC	GTEC CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEFL Junior Comprehensive	TOEIC L&R TOEIC S&W
C2	CPE (200+)	—	—	—	8.5–9.0	—	—	—	—	—
C1	CAE (180~199)	Grade 1 (2630~3400)	—	1370–1400	7.0–8.0	400	800	95–120	—	1305–1390 L&R 945~ S&W 360~
B2	FCE (160~179)	Grade Pre 1 (2304~3000)	1190–1280	1160–1369	5.5–6.5	334–399	600–795	72–94	341–352	1095–1300 L&R 785~ S&W 310~
B1	PET (140~159)	Grade 2 (1980~2600)	960–1189	880–1159	4.0–5.0	226–333	420–595	42–71	322–340	790–1090 L&R 550~ S&W 240~
A2	KET (120~139)	Grade Pre 2 (1728~2400)	690–959	510–879	3.0	150–225	235–415	—	300–321	385–785 L&R 225~ S&W 160~
A1	—	Grade 3–5 (419–2200)	—689	—509	2.0	—	—	—	—	200–380 L&R 120~ S&W 80~

Research context: Groups/Systems impacted

Meetings of the working group held by the ministry 2017-2018

2018.3.5

- Testing organizations reported the linking/scoring result
- Members discussed the necessary information for the CEFR reference table/score reporting process to be appropriately used

Issues:

⇒Each testing organization's continuous commitment to revise and re-do the standard setting

CEFR	Cambridge English	EIKEN	GTEC	GTEC CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEFL Junior Comprehensive	TOEIC L&R TOEIC S&W
C2	CPE (200+)	—	—	—	8.5–9.0	—	—	—	—	—
C1	CAE (180~199)	Grade 1 (2630~3400)	—	1370–1400	7.0–8.0	400	800	95–120	—	1305–1390 L&R 945~ S&W 360~
B2	FCE (160~179)	Grade Pre 1 (2304~3000)	1190–1280	1160–1369	5.5–6.5	334–399	600–795	72–94	341–352	1095–1300 L&R 785~ S&W 310~
B1	PET (140~159)	Grade 2 (1980~2600)	960–1189	880–1159	4.0–5.0	226–333	420–595	42–71	322–340	790–1090 L&R 550~ S&W 240~
A2	KET (120~139)	Grade Pre 2 (1728~2400)	690–959	510–879	3.0	150–225	235–415	—	300–321	385–785 L&R 225~ S&W 160~
A1	—	Grade 3–5 (419–2200)	—689	—509	2.0	—	—	—	—	200–380 L&R 120~ S&W 80~

Research context : TEAP

Background of TEAP

TEAP (Test of English for Academic Purposes)

The characteristics of TEAP

Designed basically for third year high school students.

Used for university admissions purposes.

Designed to measure proficiency levels CEFR A2-B2.

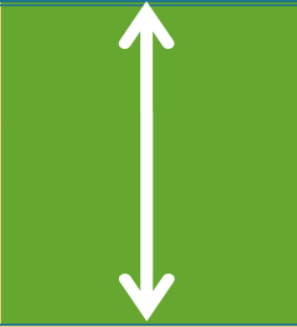
Administered three times a year.

Four-skill test



Research context : TEAP

Levels measured by TEAP

CEFR	EIKEN	MEXT ¹ benchmarks	TOEFL® iBT ²	TEAP
C2				
C1	1		100	
B2	Pre-1	English teachers	80	
B1	2	High school graduates	45	
A2	Pre-2		32	
A1	3, 4, 5			

1 Ministry of Education, Culture, Sports, Science and Technology

2 Based on comparability studies undertaken by Eiken at the University of Hawaii in conjunction with Professor J.D. Brown & Dr. Yao Hill.

TOEFL® is a registered trademark of Educational Testing Service (ETS), Inc. This table has not been endorsed or approved by ETS.

Research context : TEAP



Test Structure

Skill	Test format	Number of items	Time
Reading	Multiple Choice	60	70 min
Listening	Multiple Choice	50	50 min
Writing	Direct Response	2	70 min
Speaking	One to One Interview	4	10 min

Time	70 min	Items	60
Format	Multiple choice using a mark sheet		

Task	Format	Skill focus
Part 1 (20 items)	Vocabulary and word usage	Language knowledge necessary to comprehend texts of an academic nature which students are likely to encounter in the context of their university studies.
Part 2A (5 items)	Reading graphs and charts	Interpreting and drawing inferences from visual information such as graphs and charts which students are likely to encounter in the classroom.
Part 2B (5 items)	Reading advertisements and notices	Comprehending important information from notices, announcements, e-mails, etc. which students are likely to encounter on campus and which relate to the context of teaching and learning.
Part 2C (10 items)	Reading short texts	Comprehending important information at the paragraph level in texts of an academic nature which students are likely to encounter in the classroom.
Part 3A (8 items)	Reading extended texts	Comprehending text-level information such as logical sequence in longer texts of an academic nature which students are likely to encounter in the context of their university studies.
Part 3B (12 items)	Reading extended texts (including graphs and charts)	Comprehending information and ideas in, and drawing inferences from, extended texts of an academic nature which students are likely to encounter in the context of their university studies, including the integration of information from both the text and visual information such as graphs and charts.



TEAP Listening

Full Score: 100

Research context : TEAP



Time	Approx. 50 min	Items	50
Format	Audio description by CD ・ Multiple choice using a mark sheet		

Task	Format	Skill focus
Part 1A (10 items)	Listening to short dialogues	Understanding dialogues between students and persons with whom students are likely to converse in the context of their university studies (e.g., professors, academic advisors, exchange students).
Part 1B (10 items)	Listening to short monologues	Grasping important information from brief lectures and announcements relevant to academic subjects or the university context.
Part 1C (5 items)	Listening to short monologues	Grasping important information from brief lectures and announcements relevant to academic subjects or the university context and interpreting visual information such as graphs and charts which students are likely to encounter in the classroom.
Part 2A (9 items)	Listening to long dialogues	Grasping important information in long dialogues between students and persons with whom students are likely to converse in the context of their university studies (e.g., professors, academic advisors, exchange students). Includes both two- and three-person dialogues.
Part 2B (16 items)	Listening to long monologues (including graphs and charts)	Understanding monologues relevant to academic subjects or the university context and interpreting visual information such as graphs and charts which students are likely to encounter in the classroom.

Time	70 min	Items	2
Format	Writing on an answer sheet		
Rating method	Rating by certified raters		



Task	Format	Skill focus
Task A (1 item)	Summary	Summarizing a text which has been written on a topic with academic relevance such as might be encountered in the classroom.
Task B (1 item)	Essay	Writing an essay which synthesizes information and ideas from multiple input texts, including graphs and charts, written on topics of academic relevance such as might be encountered in the classroom.

Task A : MAIN IDEAS, COHERENCE & COHESION, LEXICAL RANGE & ACCURACY, GRAMMATICAL RANGE & ACCURACY (4 perspectives)

Task B: MAIN IDEAS, COHERENCE , COHESION, LEXICAL RANGE & ACCURACY, GRAMMATICAL RANGE & ACCURACY (5 perspectives)

**For each perspective, a graded point scale that corresponds to CEFR is used.
0(BelowA2),1(A2),2(B1),3(B2)**

Total raw score for Task A : 12

Total raw score for Task B : 15

Total raw score for Writing :27

Task A Test Characteristics

Writing a summary of approximately 70 words after reading a text such as an editorial article.

Task B Test Characteristics

Writing an essay of approximately 200 words after identifying the main points of argument from multiple input texts, including graphs and charts, and synthesizing them.

Time	Approx. 10 min	Items	4
Format	A one to one interview with an examiner		
Rating method	Rating by certified raters		
N.B.	The test is recorded and used for rating		



Task	Format	Skill focus
Part A (1 item)	Short interview	Providing specific personal information at different temporal frames (past, present, and future) such would be included in a self-introduction.
Part 2 (1 item)	Role-play (test taker interview interlocutor)	Initiating interaction and managing the c
Part 3 (1 item)	Monologue	Giving a speech on a topic relevant to academic subjects or the university context.
Part 4 (2-4 item)	Extended interview	Expressing and justifying opinions on topics with academic relevance or the university context such as might be addressed in a discussion or debate.

PRONUNCIATION, GRAMMATICAL RANGE & ACCURACY, LEXICAL RANGE & ACCURACY, FLUENCY, INTERACTIONAL EFFECTIVENESS
(5 perspectives)

For each perspective, a graded point scale that corresponds to CEFR is used.
0(BelowA2),1(A2),2(B1),3(B2)

Total raw score for Speaking : 15

The Goals of this project

Following the current discussion on the reform of the college entrance exam system:

1. To investigate the relationship between TEAP and the CEFR
2. To report the process and result of the linking for further validation

In order to provide more evidenced based information to the relevant stakeholders on the linking process and result of TEAP to the CEFR

Workshop design

① Pre-workshop training

- Learn about CEFR and TEAP

Familiarisation/Specification

② Workshop

- Review of ①
- Standardization (CEFR items)
- Benchmarking (TEAP items)
 - 1st round
 - Feedback and discussion
 - 2nd round
- Standard Setting

Training/Standardization

Standard Setting

Workshop design

Standard Setting Method for Reading and Listening

The Bookmark Method

Panelists will be given an *Ordered Item Booklet* (OIB). This contains all the items from a TEAP Reading/Listening test arranged in increasing order of difficulty (IRT item measure).

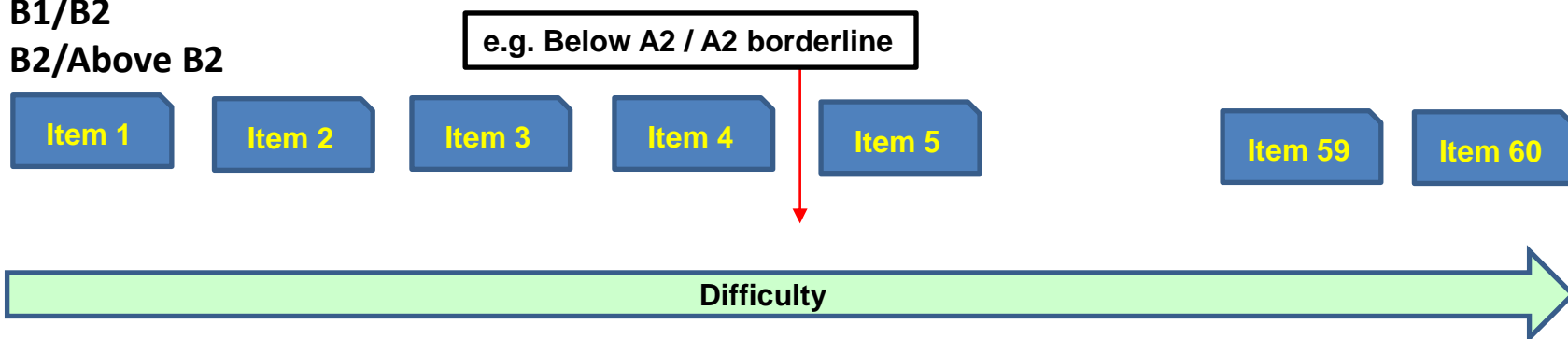
The task of panelists is to decide where the borderline between each of the CEFR bands being considered is located in the OIB. For the TEAP standard setting workshops, panelists will be asked to make judgments for the following standards:

Below A2/A2

A2/B1

B1/B2

B2/Above B2



Easiest item

Most difficult item

Workshop design

Standard Setting Method for Reading and Listening

Panelists' tasks:

1. Start with the lowest standard (Below A2/A2)
2. Go through the OIB from easy to difficult.
3. For each item, judge whether a minimally acceptable person (a borderline A2 person in the case of the lowest standard) could master the item.
4. Select the item number that corresponds to the first item in the OIB that you judge a minimally acceptable person could master.
5. Repeat for all standards (A2/B1, B1/B2, B2/Above B2).

Some important points:

- Carry out the bookmarking on paper before inputting your final decisions into the online forms.
- Read all items in the OIB before finalizing your bookmarks. (You may find after assigning later bookmarks that you want to go back and adjust earlier ones.)

Workshop design

Standard Setting Method for Writing and Speaking

The Contrasting Groups Method

For each standard to be established, panelists will be given a booklet containing 10 student performances. The performances in each booklet will be in random order (in terms of ability).

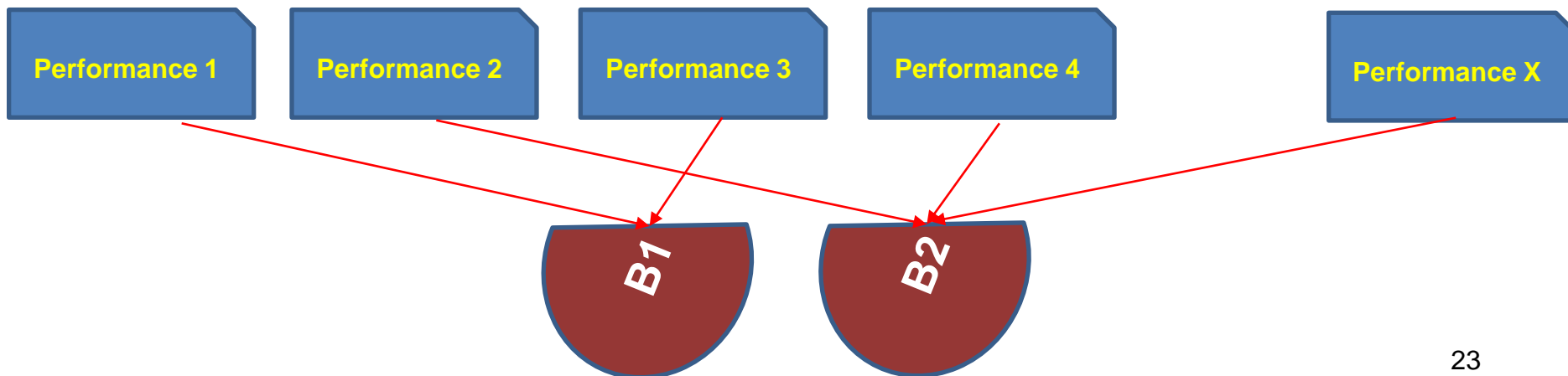
The task of panelists is to assign each student performance to one of two categories. For example, to establish the standard for B1/B2, every student performance will be categorized by panelists as either B1 [or lower] or B2 [or higher]. For the TEAP standard setting workshops, panelists will be asked to categorize performances for the following standards:

Below A2/A2 [or higher]

A2 [or lower]/B1 [or higher]

B1 [or lower]/B2 [or higher]

B2 [or lower]/Above B2



Workshop design

Standard Setting Method for Writing and Speaking

Panelists' tasks:

1. Start with the booklet for the lowest standard (Below A2/A2 [or higher])
2. Review each performance in the booklet and categorize it as one of the two given groups (Below A2 and A2 [or higher] in the case of the lowest standard).
3. Record the group you choose for each performance using the online form provided.
4. Repeat for all standards.

Workshop design

Timeline of the four-day workshops

Timeline	Day 1 Reading	Day 2 Listening	Day 3 Writing	Day 4 Speaking
9:00-9:10	Registration & Overview of the 4-day session& Day 1	Registration & Overview of Day 2	Registration & Overview of Day 3	Registration & Overview of Day 4
9:10-10:40	Familiarization of CEFR scales	Familiarization of CEFR scales	Familiarization of CEFR scales	Familiarization of CEFR scales
10:40-10:50	Coffee break	Coffee break	Coffee break	Coffee break
10:50-12:20	CEFR Task SS Training	CEFR Task SS Training	CEFR Task SS Training	CEFR Task SS Training
12:20-13:10	Lunch break	Lunch break	Lunch break	Lunch break
13:10-14:40	TEAP SS Training	TEAP SS Training	TEAP SS Training	TEAP SS Training
14:40-14:50	Coffee Break	Coffee Break	Coffee Break	Coffee Break
14:50-16:20	Standard setting (R)	Standard setting (L)	Standard setting (W)	Standard setting (S)
16:20-16:30	Closing of Day 1 and overview of Day 2	Closing of Day 2 and overview of Day 3 & 4	Closing of Day 3 and overview of Day 3	Closing of Day 4 and all sessions

Training (Benchmarking)

Developing an ability to relate **CEFR/ Target** tasks and performances to those levels

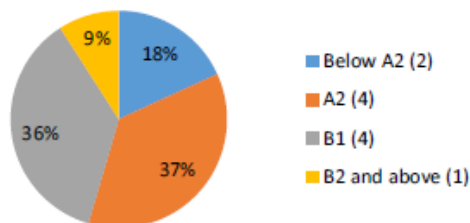
- 1) **First Vote:** individual rating
- 2) Group discussion (**Maximum number of 4 persons for each group**)
- 3) **Second Vote:** individual rating **and whole group consensus**

Question 1
Choose which CEFR level is most appropriate for the question.

- ☐ Below A2
- ☐ A2
- ☐ B1
- ☐ B2 and above

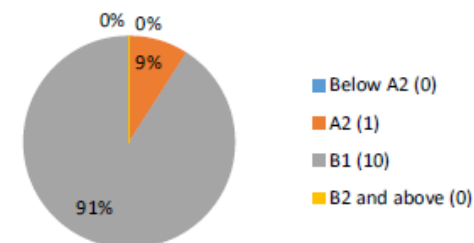
Submit Form

Results: Question 1
(First Vote)



First vote result

Results: Question 1
(Second Vote)



Second vote result

Question form for each
Question/response

Workshop result

Biodata of standard setting panelists

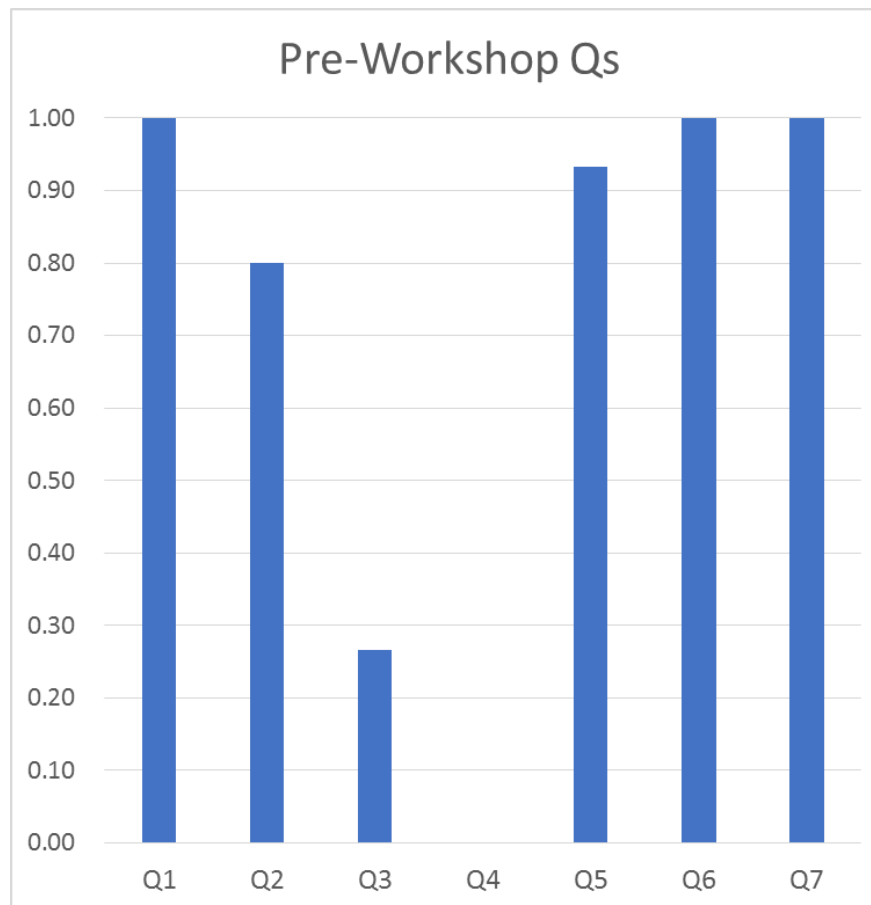
Panelist ID	R	L	W	S	Work experience
1	○	○	○	○	English Teacher/Examiner
2	○	○	○	○	IELTS, TEAP, EIKEN, BULATS, Teach English
3	○	○	○	○	Instructor on University's tutorial English programme
4	○	○		○	English Teacher/Examiner
5	○	○	○	○	English Instructor
6	○	○	○	○	EIKEN/TEAP Proofreader/Rater/Examiner
7	○	○	○	○	English Instructor/Teacher/Trainer
8	○	○	○	○	English Teacher
9	○	○			University Teacher
10	○	○	○	○	Professor
11	○	○	○	○	English Teacher
12	○	○	○		English Instructor, Rikkyo Instructor
13			○	○	Junior High ALT
14			○	○	English teacher
15	○	○	○	○	English Teacher

Other information of panelists (15 in total)

1) Male: 10, Female: 5

2) English native: 12, Japanese native: 3

Workshop result



Q1	I was familiar with the concept of the CEFR before taking part in this project.
Q2	I was familiar with the concept of standard setting before taking part in this project.
Q3	I have had experience acting as a judge/rater on standard setting panels before taking part in this project.
Q4	I have had experience organizing standard setting panels.
Q5	The preparation booklet gave me a clear understanding of the purpose of the project.
Q6	The explanations and tasks in the preparation booklet helped me understand the structure of the CEFR.
Q7	The explanations and tasks in the preparation booklet helped me understand EIKEN and its tasks.

Workshop result

Post-workshop Qs

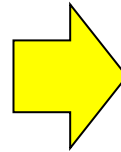
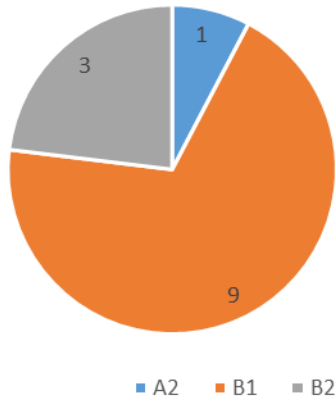
	Questions
Q1	I have a clearer understanding of the concept of the CEFR after taking part in this project.
Q2	I have a clearer understanding of the concept of standard setting after taking part in this project.
Q3	The explanations, training and discussion in today's session helped me understand the structure of the CEFR.
Q4	The explanations, training and discussion in today's session helped me understand TEAP and its tasks.
Q5	The time provided for the discussion of the CEFR was adequate.
Q6	The time provided for the discussion of TEAP and its tasks was adequate.
Q7	There was an equal opportunity for everyone to contribute his/her ideas during the discussion of the CEFR.
Q8	There was an equal opportunity for everyone to contribute his/her ideas during the discussion of TEAP.
Q9	The CEFR training tasks with the items supplied were useful.
Q10	The TEAP training tasks with the items supplied were useful.
Q11	The time provided for the CEFR training was adequate.
Q12	The time provided for the TEAP training was adequate.
Q13	The explanation of the Bookmark Method was adequate and I felt able to undertake the rating tasks for the reading items.
Q14	The time provided for rating the TEAP reading items was adequate.
Q15	The feedback on item difficulty of the TEAP reading items was useful.
Q16	The facilities and food service were adequate and helped create a productive and efficient working environment.
Q17	During the workshop I felt I had adequate opportunities to present my opinions and was able to ask questions when I was not sure of how to proceed.

	R	L	W	S
Q1	0.92	0.92	1.00	1.00
Q2	1.00	1.00	1.00	1.00
Q3	1.00	1.00	1.00	1.00
Q4	1.00	1.00	1.00	1.00
Q5	0.92	1.00	1.00	1.00
Q6	0.92	1.00	1.00	1.00
Q7	1.00	1.00	1.00	1.00
Q8	1.00	1.00	1.00	1.00
Q9	1.00	1.00	1.00	1.00
Q10	1.00	1.00	1.00	1.00
Q11	0.92	1.00	1.00	1.00
Q12	0.92	1.00	1.00	1.00
Q13	1.00	1.00	1.00	1.00
Q14	0.85	0.85	0.85	0.92
Q15	1.00	1.00	1.00	1.00
Q16	1.00	1.00	1.00	1.00
Q17	1.00	1.00	1.00	1.00

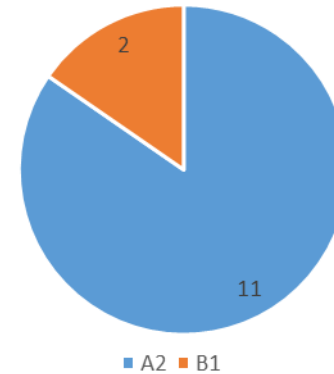
Workshop result

Benchmarking result

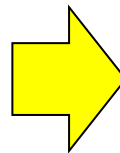
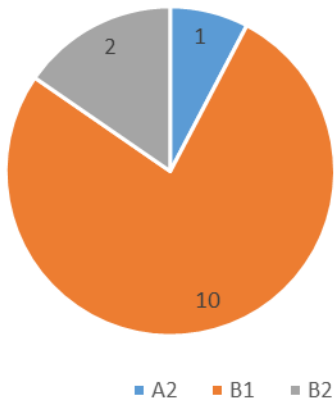
【TEAP】Reading Benchmark 1 Round 1



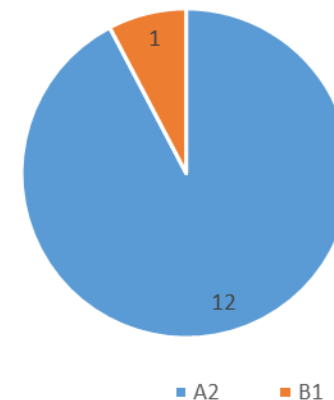
【TEAP】Reading Benchmark 1 Round 2



【TEAP】Reading Benchmark 2 Round 1



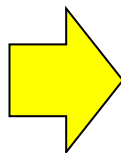
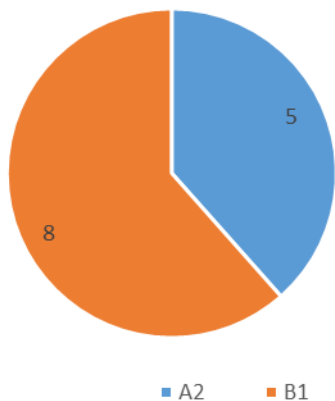
【TEAP】Reading Benchmark 2 Round 2



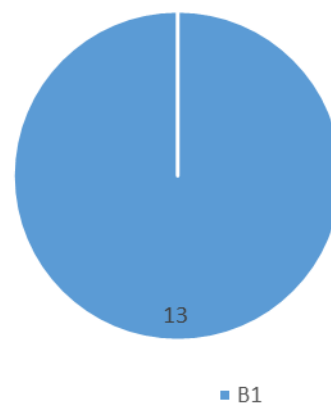
Workshop result

Benchmarking result

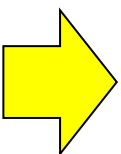
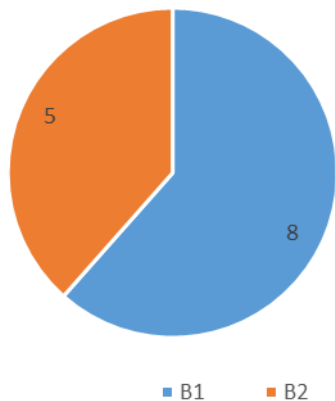
【TEAP】Reading Benchmark 3 Round 1



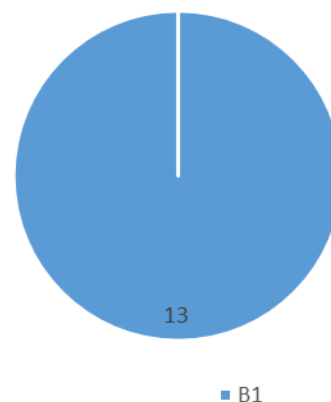
【TEAP】Reading Benchmark 3 Round 2



【TEAP】Reading Benchmark 4 Round 1



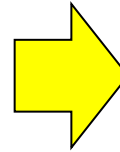
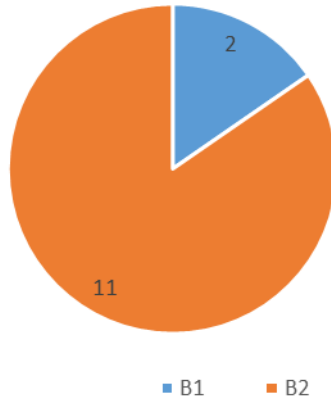
【TEAP】Reading Benchmark 4 Round 2



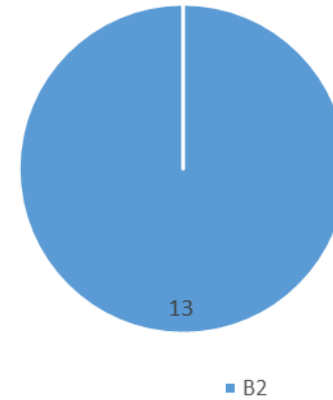
Workshop result

Benchmarking result

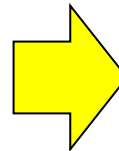
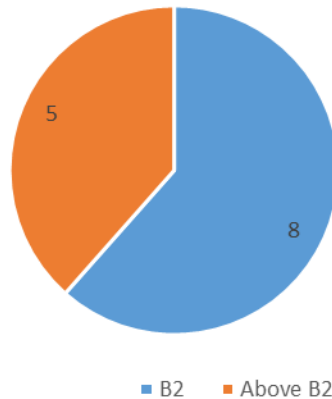
【TEAP】Reading Benchmark 5 Round 1



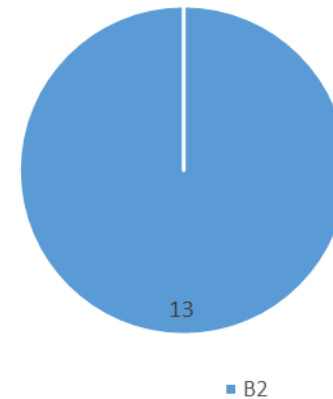
【TEAP】Reading Benchmark 5 Round 2



【TEAP】Reading Benchmark 6 Round 1



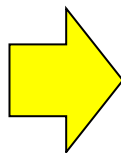
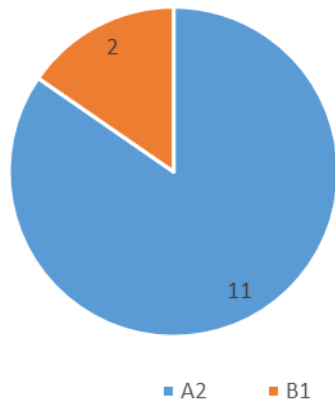
【TEAP】Reading Benchmark 6 Round 2



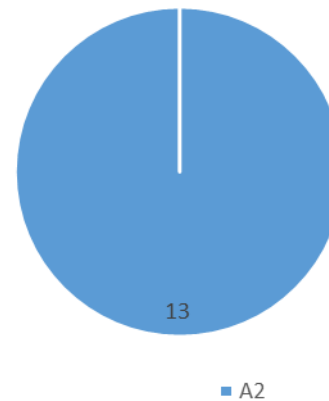
Workshop result

Benchmarking result

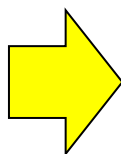
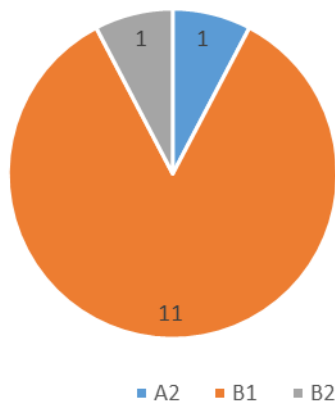
【TEAP】Listening Benchmark 1 Round 1



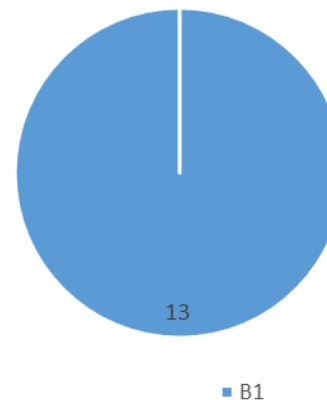
【TEAP】Listening Benchmark 1 Round 2



【TEAP】Listening Benchmark 2 Round 1



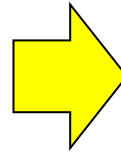
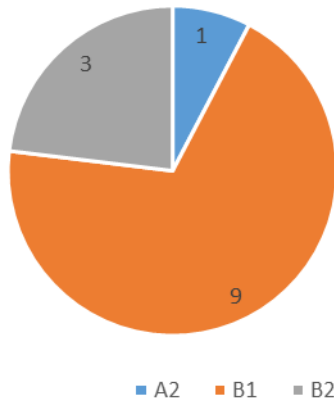
【TEAP】Listening Benchmark 2 Round 2



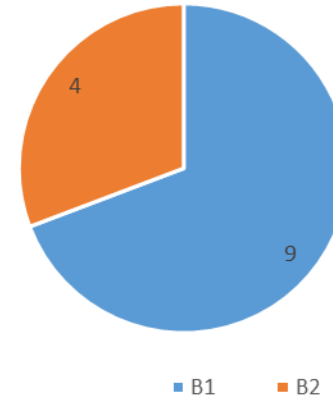
Workshop result

Benchmarking result

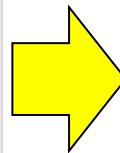
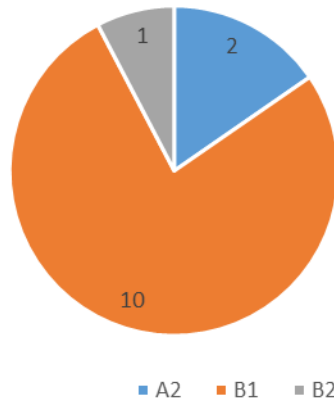
【TEAP】Listening Benchmark 3 Round 1



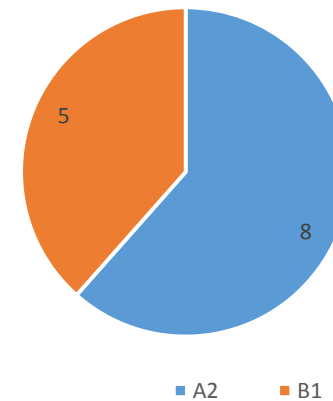
【TEAP】Listening Benchmark 3 Round 2



【TEAP】Listening Benchmark 4 Round 1

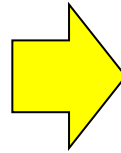
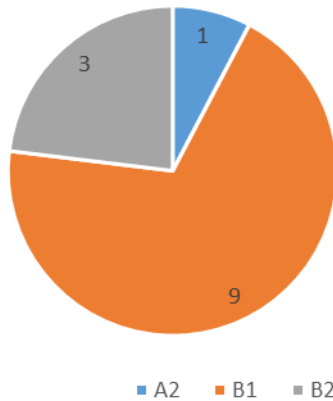


【TEAP】Listening Benchmark 4 Round 2



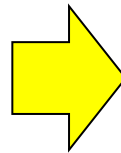
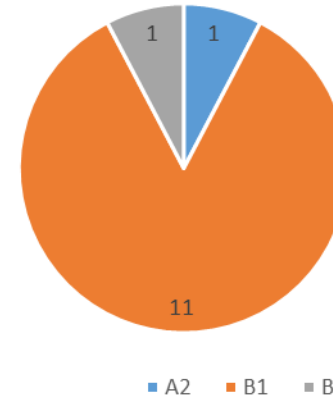
Workshop result

【TEAP】Listening Benchmark 5 Round 1

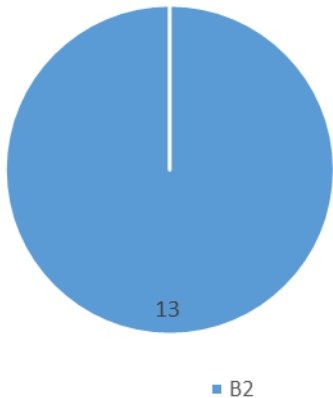


Benchmarking result

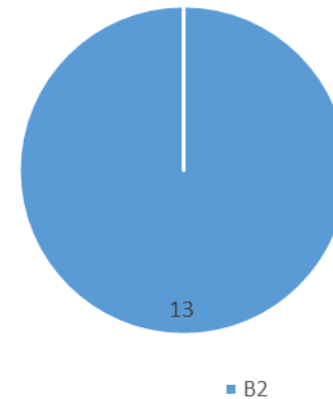
【TEAP】Listening Benchmark 5 Round 2



【TEAP】Listening Benchmark 6 Round 1



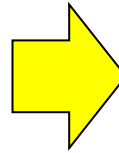
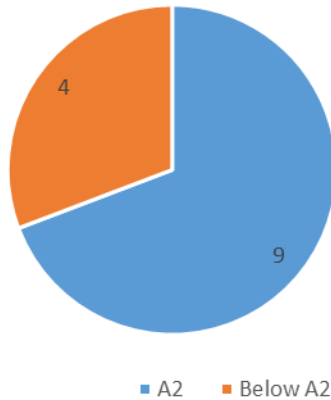
【TEAP】Listening Benchmark 6 Round 2



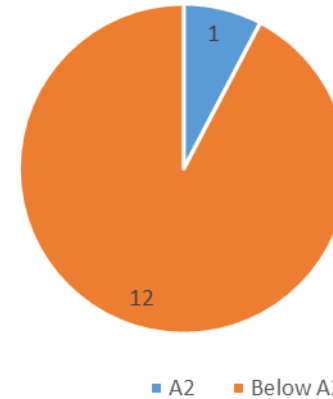
Workshop result

Benchmarking result

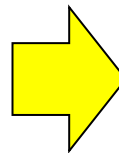
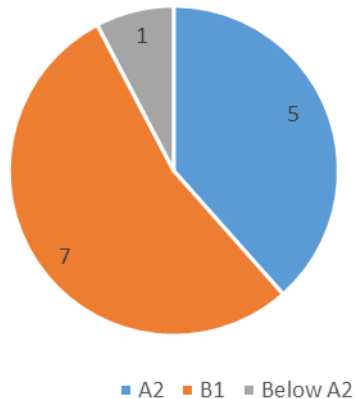
【TEAP】Writing Benchmark 1 Round 1



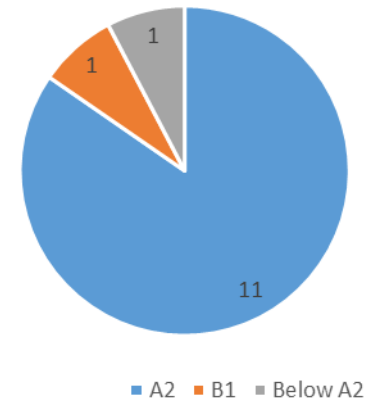
【TEAP】Writing Benchmark 1 Round 2



【TEAP】Writing Benchmark 2 Round 1



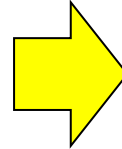
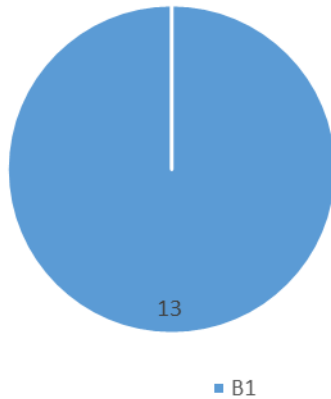
【TEAP】Writing Benchmark 2 Round 2



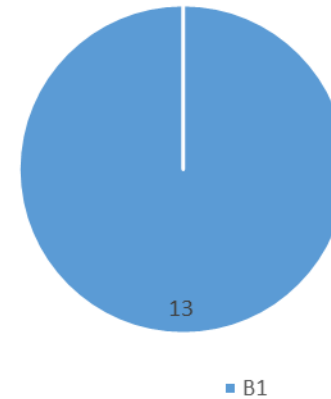
Workshop result

Benchmarking result

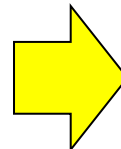
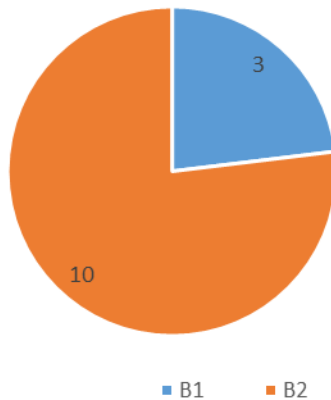
【TEAP】Writing Benchmark 3 Round 1



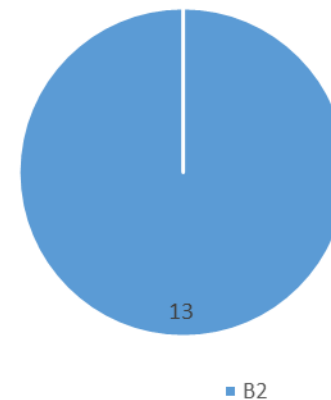
【TEAP】Writing Benchmark 3 Round 2



【TEAP】Writing Benchmark 4 Round 1



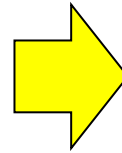
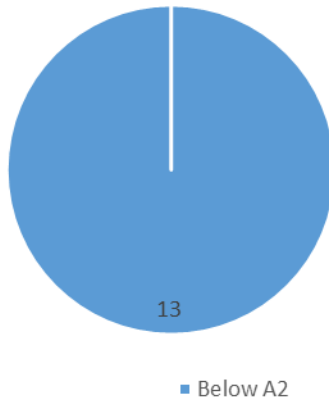
【TEAP】Writing Benchmark 4 Round 2



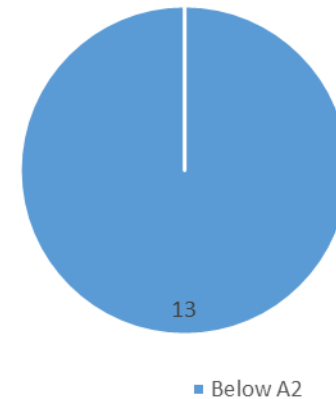
Workshop result

Benchmarking result

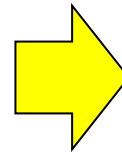
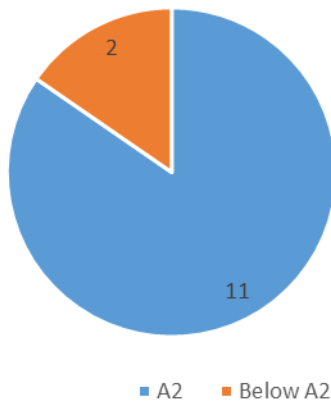
【TEAP】Speaking Benchmark 1 Round 1



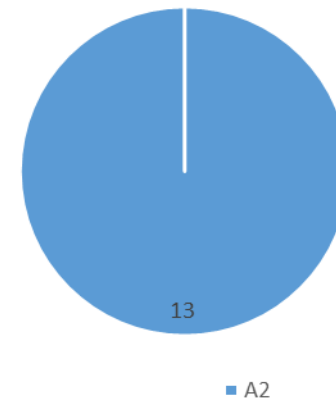
【TEAP】Speaking Benchmark 1 Round 2



【TEAP】Speaking Benchmark 2 Round 1

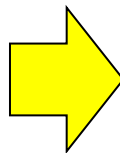
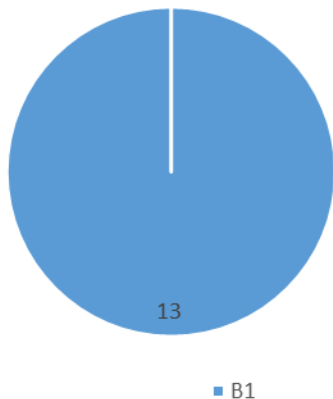


【TEAP】Speaking Benchmark 2 Round 2



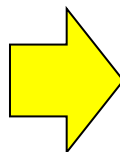
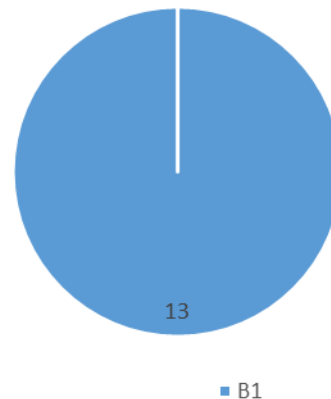
Workshop result

【TEAP】Speaking Benchmark 3 Round 1

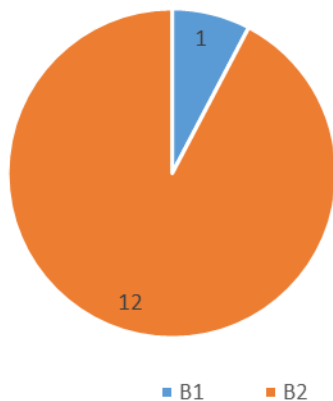


Benchmarking result

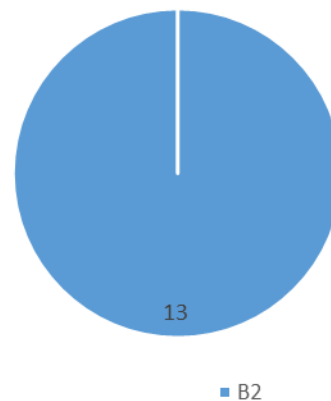
【TEAP】Speaking Benchmark 3 Round 2



【TEAP】Speaking Benchmark 4 Round 1

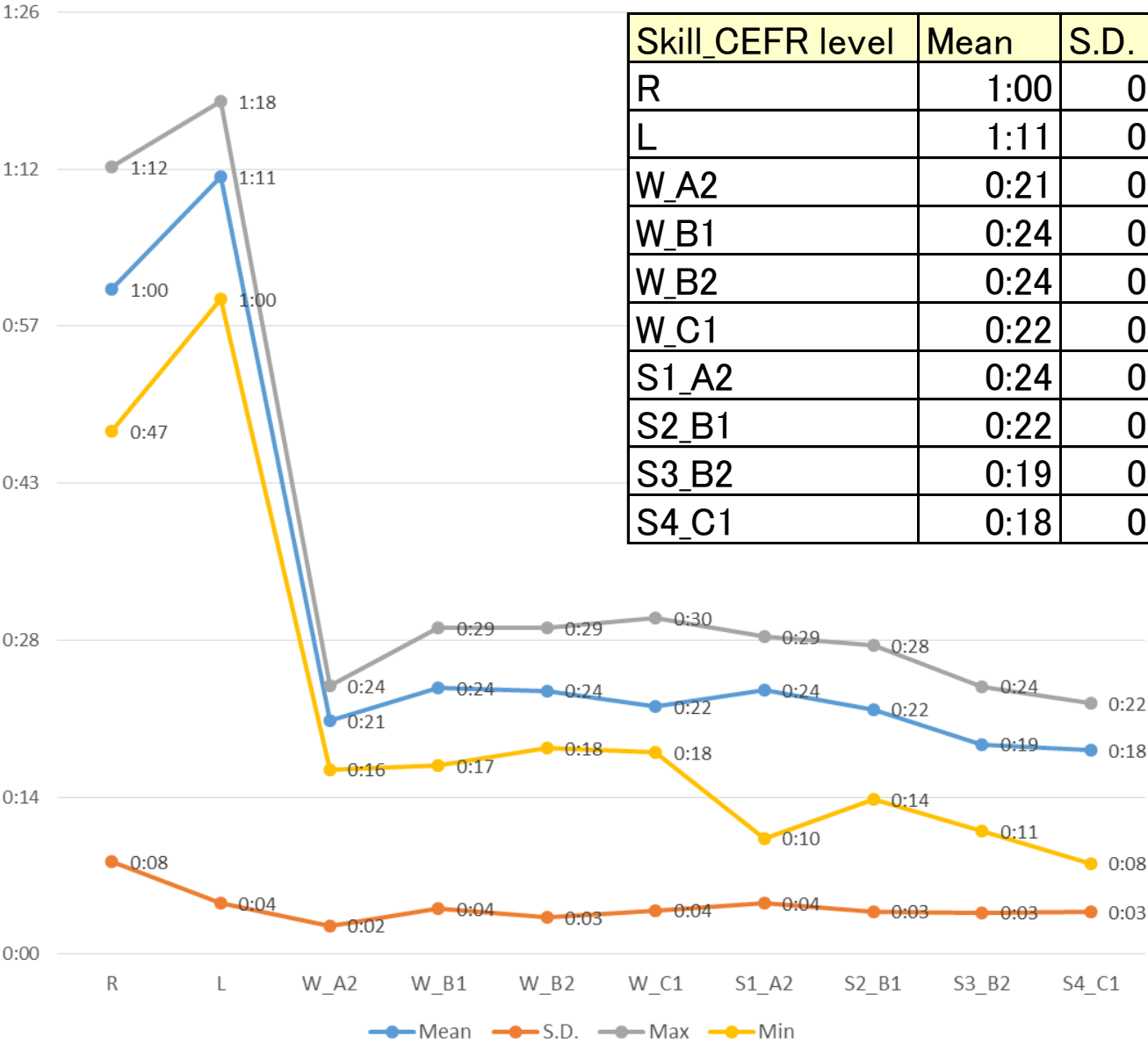


【TEAP】Speaking Benchmark 4 Round 2



Workshop result

Time spent on standard setting (h:mm)



Skill_CEFR level	Mean	S.D.	Max	Min	NoI/P
R	1:00	0:08	1:12	0:47	60
L	1:11	0:04	1:18	1:00	50
W_A2	0:21	0:02	0:24	0:16	10
W_B1	0:24	0:04	0:29	0:17	10
W_B2	0:24	0:03	0:29	0:18	10
W_C1	0:22	0:04	0:30	0:18	10
S1_A2	0:24	0:04	0:29	0:10	10
S2_B1	0:22	0:03	0:28	0:14	10
S3_B2	0:19	0:03	0:24	0:11	10
S4_C1	0:18	0:03	0:22	0:08	10

Workshop result

世界レベルの英語力を

英検

Reading

Panelist ID	Below A2/A2	A2/B1	B1/B2	B2/Above B2
6	Item 1	Item 22	Item 53	Item 60
12	Item 3	Item 13	Item 29	Item 47
7	Item 1	Item 18	Item 38	Item 55
2	Item 1	Item 26	Item 55	Item 60
9	Item 1	Item 27	Item 52	Item 60
11	Item 3	Item 13	Item 29	Item 60
5	Item 1	Item 14	Item 46	Item 60
4	Item 2	Item 21	Item 47	Item 60
10	Item 3	Item 19	Item 43	Item 60
13	Item 6	Item 25	Item 50	Item 60
3	Item 1	Item 19	Item 43	Item 60
8	Item 1	Item 13	Item 45	Item 57
16	Item 2	Item 22	Item 46	Item 59

Panelist ID	Below A2/A2	A2/B1	B1/B2	B2/Above B2
6	31.0681054	54.4690054	79.0513054	91.4874054
12	37.9578054	48.4103054	60.8569054	71.7994054
7	31.0681054	50.8377054	65.3983054	80.5936054
2	31.0681054	56.5790054	80.5936054	91.4874054
9	31.0681054	58.2398054	76.7517054	91.4874054
11	37.9578054	48.4103054	60.8569054	91.4874054
5	31.0681054	48.7949054	71.6821054	91.4874054
4	35.9835054	53.3086054	71.7994054	91.4874054
10	37.9578054	51.1644054	69.3957054	91.4874054
13	42.6085054	56.2589054	73.5129054	91.4874054
3	31.0681054	51.1644054	69.3957054	91.4874054
8	31.0681054	48.4103054	71.4833054	87.5809054
16	35.9835054	54.4690054	71.6821054	89.2062054

Listening

Panelist ID	Below A2/A2	A2/B1	B1/B2	B2/Above B2
11	Item 1	Item 9	Item 28	Item 50
10	Item 1	Item 6	Item 30	Item 49
12	Item 3	Item 18	Item 30	Item 50
2	Item 1	Item 12	Item 27	Item 44
13	Item 3	Item 19	Item 38	Item 50
9	Item 1	Item 20	Item 40	Item 50
7	Item 1	Item 13	Item 28	Item 50
4	Item 1	Item 29	Item 47	Item 50
6	Item 1	Item 17	Item 47	Item 50
5	Item 1	Item 5	Item 19	Item 32
16	Item 1	Item 28	Item 40	Item 50
3	Item 2	Item 17	Item 33	Item 50
8	Item 1	Item 8	Item 40	Item 49

Panelist ID	Below A2/A2	A2/B1	B1/B2	B2/Above B2
11	34.32828618	50.76858618	61.72038618	93.60568618
10	34.32828618	47.18738618	64.03538618	83.28178618
12	38.15768618	57.09018618	64.03538618	93.60568618
2	34.32828618	52.39028618	61.38768618	74.50298618
13	38.15768618	57.45098618	71.73848618	93.60568618
9	34.32828618	57.96468618	71.75188618	93.60568618
7	34.32828618	52.53778618	61.72038618	93.60568618
4	34.32828618	63.75108618	79.31818618	93.60568618
6	34.32828618	56.01578618	79.31818618	93.60568618
5	34.32828618	43.69728618	57.45098618	65.80598618
16	34.32828618	61.72038618	71.75188618	93.60568618
3	36.10958618	56.01578618	66.31828618	93.60568618
8	34.32828618	50.72298618	71.75188618	83.28178618

Writing

A2	36.0	A2	53.8	B1	80.9	Above B2	96.1
Below A2	29.0	B1	65.1	B2	83.9	B2	95.1

Speaking

A2	37.5	A2	54.2	B1	84.7	Above B2	98.7
Below A2	30.7	B1	65.1	B2	91.9	B2	98.4

Workshop result

Before Standard Setting (Until 2017)

CEFR	A2	B1	B2
L	35	51	75
R	33	50	75
S	33	61	85
W	33	61	85
Total	134	223	320

After Standard Setting (Since 2018)

CEFR	A2	B1	B2	Above B2
L	34	54	68	90
R	34	52	71	90
S	34	60	88	99
W	33	59	82	96
Total	135	225	309	375

New CEFR Reference Table as of 2018.3.26

各資格・検定試験とCEFRとの対照表

文部科学省（平成30年3月）

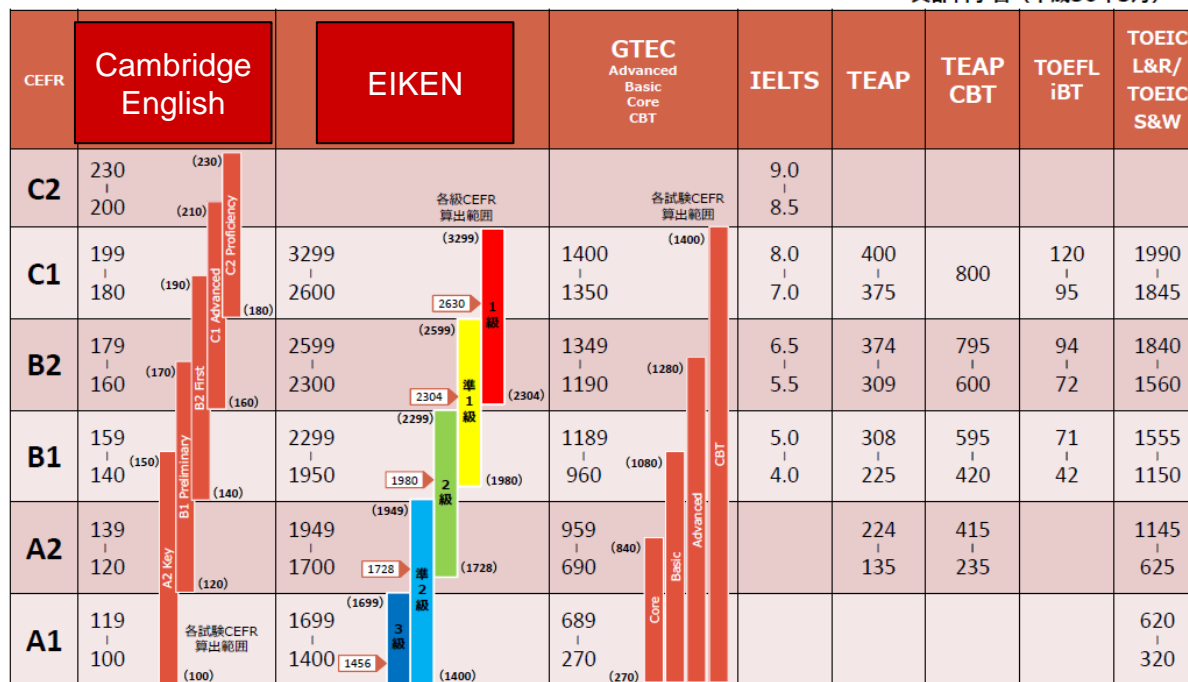
CEFR	Cambridge English	EIKEN	GTEC Advanced Basic Core CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEIC L&R/ TOEIC S&W
C2	230 200 (230) (210)			9.0 8.5				
C1	199 180 (190)	3299 2600 (3299) (2630)	1400 1350 (1400)	8.0 7.0	400 375	800	120 95	1990 1845
B2	179 160 (170)	2599 2300 (2599) (2304)	1349 1190 (1280)	6.5 5.5	374 309	795 600	94 72	1840 1560
B1	159 140 (150)	2299 1950 (2299) (1980)	1189 960 (1080)	5.0 4.0	308 225	595 420	71 42	1555 1150
A2	139 120 (120)	1949 1700 (1949) (1728)	959 690 (840)		224 135	415 235		1145 625
A1	119 100 (100)	1699 1400 (1699) (1456)	689 270 (270)					620 320

→ は各級合格スコア

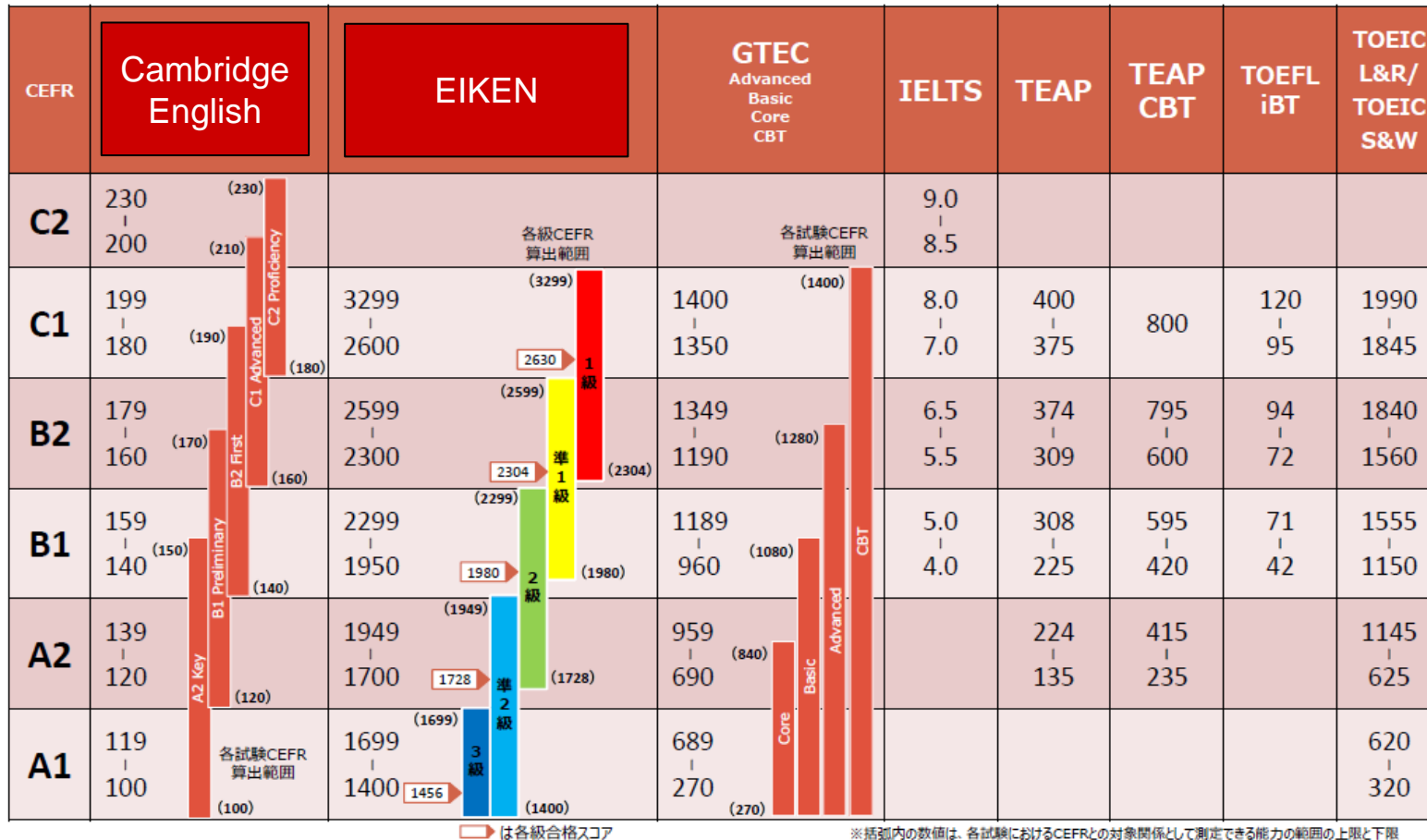
※括弧内の数値は、各試験におけるCEFRとの対象関係として測定できる能力の範囲の上限と下限

New CEFR Reference Table as of 2018.3.26

CEFR	Cambridge English	EIKEN	GTEC	GTEC CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEFL Junior Comprehensive	TOEIC L&R TOEIC S&W
C2	CPE (200+)	—	—	—	8.5–9.0	—	—	—	—	—
C1	CAE (180~199)	Grade 1 (2630~3400)	—	1370–1400	7.0–8.0	400	800	95–120	—	1305–1390 L&R 945~ S&W 360~
B2	FCE (160~179)	Grade Pre 1 (2304~3000)	1190–1280	1160–1369	5.5–6.5	334–399	600–795	72–94	341–352	1095–1300 L&R 785~ S&W 310~
B1	PET (140~159)	Grade 2 (1980~2600)	960–1189	880–1159	4.0–5.0	226–333	420–595	42–71	322–340	790–1090 L&R 550~ S&W 240~
A2	KET (120~139)	Grade Pre 2 (1728~2400)	690–959	510–879	3.0	150–225	235–415	—	300–321	385–785 L&R 225~ S&W 160~
A1	—	Grade 3–5 (419–2200)	—689	—509	2.0	—	—	—	—	200–380 L&R 120~ S&W 80~



Critical issues related to the implementation of the new system



Use of test results: CEFR? Pass/Fail? Scores?

Critical issues related to the implementation of the new system

Which test has the best reliability or validity?

Researchers

Which test can cover our students (B2) ?

Admission officers

What if applicants' CEFR are all B1?

Test takers

Which test is "cheaper" or "easier" to achieve B1?

Which test is "easier" to teach?

Teachers

各資格・検定試験とCEFRとの対照表
文部科学省（平成30年3月）

CEFR	ケンブリッジ 英語検定	実用英語技能検定 1級・2級	GTEC Advanced Basic Core CBT	IELTS	TEAP	TEAP CBT	TOEFL iBT	TOEIC L&R/ TOEIC S&W
C2	230 200			9.0 8.5				
C1	199 180	3299 2600	1400 1350	8.0 7.0	400 375	800	120 95	1990 1845
B2	179 160	2599 2300	1349 1190	6.5 5.5	374 309	795 600	94 72	1840 1560
B1	159 140	2299 1950	1189 960	5.0 4.0	308 225	595 420	71 42	1555 1150
A2	139 120	1949 1700	959 690		224 135	415 235		1145 625
A1	119 100	1699 1400	689 270					620 320

※ 各欄内の数値は、各試験におけるCEFRとの対照値を示し、測定で求められる能力の範囲の上限と下限

Policy maker

Which test has the most test centers across the country?

Which test can the best accommodation for handicapped test takers?

References

- Council of Europe. 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.
- Council of Europe. 2009. Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual. Strasbourg: Language Policy Division.
- Martyniuk, W. (ed.). 2010. Aligning Tests with the CEFR. Cambridge University Press.
- Taylor, L. 2014. A Report on the Review of Test Specifications for the Reading and Listening Papers of the Test of English for Academic Purposes (TEAP) for Japanese University Entrants.
- Green, A. 2014. The Test of English for Academic Purposes (TEAP) Impact Study: Report 1-Preliminary Questionnaires to Japanese High School Students and Teachers.
- Nakamura, K. 2014. Examination of Possible Consequences of new Test within the Context of University Entrance Exam Reform in Japan.
- Weir, C.J. 2014. A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Writing Test for Japanese University Entrants.
- Nakatsuhara, F. 2014. A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants-Study 1&2.
- Nakatsuhara, F. 2014. A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants-Study 3&4.
- Milanovic, M. & Weir, C.J. 2010. Aligning Tests with the CEFR. Cambridge University Press.

Thank you for your attention

ke-nakamura@eiken.or.jp