Learner perspectives: familiarization, knowledge, and perceptions of the CEFR

Gary Cook, Hiroshima Bunkyo University

Yukari Rutson-Griffiths, Hiroshima Bunkyo University

Since 2012, various aspects of the Common European Framework of Reference for Languages (CEFR) have been utilized in language curricula at Hiroshima Bunkyo University's language learning center: the Bunkyo English Communication Center (BECC). Whilst teacher training regarding the CEFR had taken place, students had not received any direct CEFR-education. A decision was taken in 2017 to involve one cohort of students from the Global Communication Department in activities over three years to a) determine their ability to sort Can Do descriptors into their respective levels on the self-assessment grid (A1-C2), and b) establish their knowledge and perceptions of the CEFR through questionnaires. The sorting activity, termed the CEFR shuffle in this paper, was also intended to raise students' awareness of the CEFR. Results of this research comparing two groups of students streamed into low (A1-A2) and high (A2-B1) classes show that 1) their ability to correctly sort descriptors as a cohort showed no change over a period of 2 years, 2) the high class performed better than the low class in terms of making fewer sorting mistakes in the first and third years of the study but these results were not found to be statistically significant, and 3) students' knowledge of the CEFR improved slightly while perceptions received mixed results. Further results are discussed and suggestions made for the improvement of this study in future editions.

Keywords: Awareness, Self-assessment, Curriculum, Training, Knowledge, Perceptions, CEFR, Can Do descriptors

1 Introduction

The Bunkyo English Communication Center (BECC) was established in 2008, to provide English language education for students of Hiroshima Bunkyo Women's University University, currently known as Hiroshima Bunkyo University since the admission of male students in the 2019 academic year. Since 2012, staff at the BECC have been involved in major curriculum projects which have seen the employment of the Common European Framework of Reference for Languages (CEFR) (Council of Europe [CoE] 2001) as a guide for the creation of its English language courses. One example is a compulsory first-year English Communication course created in-house which has materials based on 13 language themes provided by the *Waystage* document (Ek and Trim 1991), CEFR-informed goals from the European Association for Quality Language Services (Eaquals) bank of descriptors (2015), and the Association of Language Testers in Europe (ALTE) Can Do project (2002), and assessments modeled on the CEFR aligned *Cambridge English: Key* (University of Cambridge ESOL Examinations 2012a) and *Cambridge English: Preliminary* (University of Cambridge ESOL Examinations 2012b) tests. For detailed information on this project see Bower, Runnels, Rutson-Griffiths, Schmidt, Cook, Lehde and Kodate (2017) and Bower, Rutson-Griffiths, Cook, Schmidt, Lehde, Kodate and Runnels (2017). The project required a great deal of collaboration amongst staff, including the creation and adaptation of a plan by management, organization of staff into various committees to create the curriculum, and ultimately numerous rounds of feedback from teachers as a result of the implementation of lessons. This collaborative effort inspired further projects to coordinate other BECC language courses with CEFR-informed goals and assessments (Cook 2019).

There were various motivations behind this study. Learner autonomy is synonymous with the CEFR. Indeed, one of the main functions of the CEFR-linked European Language Portfolio (Council of Europe 2000) is to support the development of learner autonomy. Due to one of the authors being a learning advisor in the BECC's Self-Access Language Center (SALC), one of the motivating factors was to empower students with knowledge regarding their own language learning. The SALC is an English-only space which primarily exists to encourage students to develop their autonomous language learning skills. Goal setting and reflection are two such skills that are fostered through advising sessions with learning advisors and in-house materials including SALC activities, which are a core component of the aforementioned first-year English communication course. Students select these activities based on what they believe their CEFR-levels to be in the skills of reading, listening, speaking or writing. Pre- and post Can Do descriptors within activities help guide these choices (for more information on the SALC activities, see Kodate 2017). Can Do descriptors are also utilized as individual lesson and overall curriculum goals in a variety of the BECC's language courses. Therefore, we thought it beneficial for students to engage them in a Can Do descriptor familiarization task followed by goal setting and reflection, based on interacting with the CEFR self-assessment grid at all levels.

Furthermore, with BECC staff having undergone various training regarding the CEFR, such as workshops on the history of the use of the CEFR in Japan, how to write Can Do descriptors, and more recently the introduction of the CEFR Companion Volume, familiarizing students considering their role as important stakeholders was a practical step in the continued implementation of the CEFR at the BECC. As stated above, the BECC provides its students with English language education based on the CEFR. Students are presented with CEFR levels and Can Do descriptors in class and the SALC, and are given chances to reflect on their language proficiency before and after each task and assessment. However, those tasks and assessments often give a reference to their performances within a limited range, and the students do not receive many opportunities to engage with Can Do descriptors at all levels. Having stated that one of the intended uses of the CEFR is to support self-directed learning including "raising the learner's awareness of his or her present state of knowledge" and "self-assessment" (Council of Europe [CoE] 2001: 6), it is argued "the main potential for self-assessment . . . is in its use as a tool for motivation and awareness raising: helping learners to appreciate their strengths, recognise their weakness and orient their learning more effectively" (CoE 2001: 192).

As it is stated that the accuracy of self-assessment is enhanced if some training is given to the learners (CoE 2001), the study had first-year students engage in a sorting activity with Can Do descriptors from the CEFR self-assessment grid, termed the CEFR shuffle in this paper. The activity had students in groups read and closely study the CEFR Can Do descriptors from all levels and place them in order from the lowest level, A1, to the highest level, C2, by analyzing what qualities or components differentiate the Can Do descriptors. This was repeated in students' second and third years of study.

Another motivation factor for this research came from the growing popularity of the CEFR in Japan. This is evident from the development of a Japanese version of the CEFR; the CEFR-J (see Tono and Negishi 2012), in addition to the Ministry of Education, Culture, Sports, Science and Technology's (MEXT) decision in 2013 to propose CEFR-referenced goals for all junior and senior high schools (MEXT 2013). Most of the students in this study were attending high schools in Japan from the 2014 academic year, so we were curious as to what CEFR-knowledge our students had obtained before admission to the university in 2017, and how that knowledge may have changed in their first three years of education at the BECC.

The integration of the CEFR in Japan has been well documented (Nagai and O'Dwyer 2011; O'Dwyer 2015; Fennelly 2016; Schmidt et al. 2017); however, there is a lack of research that has a focus on what Japanese learners of English know about the CEFR. Therefore, this article has a dual purpose: firstly, to show the results of our attempt to shine light on the extent of our learners' knowledge of the CEFR, and secondly, to publish research with a hope that other educators may follow suit and carry out similar studies with a focus on the language learner and the CEFR in practice. In addition to the CEFR shuffle,

students were required to answer a questionnaire to establish the extent of students' knowledge and perceptions of the CEFR. The questionnaire was only conducted in the first and third years of the study. The following three research questions were investigated:

1. How well can students sort Can Do descriptors into their respective levels on the self-assessment grid?
2. What effect does language ability have on the Can Do descriptors sorting exercise?
3. How has students' knowledge and perceptions of the CEFR changed in three years of English language education?

2 Methodology

2.1 Participants

This study was conducted from the academic year of 2017 until the end of the 2019 academic year among 31 students enrolled in the Global Communication Department at Hiroshima Bunkyo University. The students received compulsory English lessons every year: six lessons per week in the first year, four lessons per week in the second year, and one lesson per week in the third year, with extra elective classes also available in the third year. Each lesson was 90 minutes. The students' TOEIC scores varied from 180 to 515 in the first year of the study. At the beginning of each year, all the students were streamed into two classes. The overall TOEIC score averages across the three years of the study are shown in Table 1.

Table 1. *Overall TOEIC score averages across three years*

	April, 2017	July, 2018	July, 2019
High-streamed	370.4	477.1	574.1
Low-streamed	282.9	379.3	398.8

   It should be noted here that there were a few changes to the number of participants across the three years due to absences on the days of the study and leaves for long-term study abroad programmes. In addition, three students initially placed in the high-streamed class in 2017 exchanged places with three students from the low stream in the third year of this study.

2.2 CEFR shuffle

The Manual provided by the Council of Europe for language educators who wish to incorporate the CEFR into examinations suggests that they prepare for the implementation through some familiarization tasks (Council of Europe 2009). One of the activities proposed in the document entails reconstructing the CEFR self-assessment grid. For this activity, the participants place the provided Can Do descriptors in the correct empty cells of the CEFR table. This study adapted a similar task as a learner training exercise. Every year, the students in each class were separated into groups of three or four and were given a set of the CEFR self-assessment grid table and 30 Can Do descriptor cards: five skills and six levels from A1 to C2. After receiving a brief explanation on what the CEFR is and how it is utilized in the BECC, they were asked to place the Can Do descriptors onto the correct places of the table. The descriptors were color-coded according to the skills, so students only needed to determine the levels of the descriptors. The descriptors were written in Japanese, and students collaborated to complete the task speaking Japanese. After checking the answers with the researchers, photos were taken of the results, and the students were asked to map three TOEIC scores (225, 550, and 785) and three Eiken grades (pre-2, 2, and pre-1) onto equivalent CEFR levels. The TOEIC score average of each group is shown in Tables 2 and 3[1].

---

1.   It is noted here that although the groups in the first and second years consisted of the same members, new groups were formed in the third year due to the exchange of students in the high- and low-streamed classes.

Table 2. *TOEIC score average of each CEFR shuffle group in 2017 and 2018*

	Groups	Number of students	April, 2017	July, 2018
High	1	4	347.5	400.0
	2	3	401.7	493.3
	3	3	398.3	563.3
	4	4	348.8	477.5
Low	5	4	300.0	373.8
	6	4	231.7	355.0
	7	4	315.0	356.3
	8	3	268.3	441.7

Table 3. *TOEIC score average of each CEFR shuffle group in 2019*

	Groups	Number of Students	July, 2019
High	1	4	536.3
	2	4	633.8
	3	4	587.5
	4	4	538.8
Low	5	3	357.5
	6	3	425.0
	7	4	406.3
	8	4	392.5

After the CEFR Shuffle in the first and third years, students were given time to reflect on their English learning and set learning goals based on the CEFR self-assessment grid.

2.3 Survey

Approximately three months after the CEFR shuffle in 2017 and 2019, the participants were asked to take a short survey to find out their knowledge and perceptions of the CEFR (see Appendix). In the following section, the results of the survey from 2019 will be compared with those of 2017, reported in Cook and Rutson-Griffiths (2018).

3 Results
3.1 CEFR shuffle results

The following results are reported from one cohort of students across three years (2017-2019). The low-streamed class results can be seen in Table 4. Students made 128 mistakes in total across three consecutive years with 90 mistakes coming from the upper half of the self-assessment grid (B2 to C2), contrasting with 38 from the lower half (A1 to B1). Students made mistakes in all five skills with reading receiving the highest number of mistakes (32), and spoken production the fewest (14). The level that was least accurately categorized was C1, which drew 29% of total mistakes. This can be seen in contrast with A1 which only drew 8% of total mistakes.

Table 4. *CEFR self-assessment grid mistakes of lower-streamed class for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	1	1	1	8	7	5	23
Reading	1	3	3	3	11	11	32
Spoken Interaction	2	3	2	9	8	5	29
Spoken Production	1	1	0	0	6	6	14
Writing	5	7	7	5	5	1	30
Total	10	15	13	25	37	28	128

Whilst the results from the high-streamed class in Table 5 show students made 34 fewer total mistakes than the low-streamed class, similar patterns can be seen. The upper-level descriptors (B2-C2) drew more mistakes, 67 in total, whereas 27 mistakes occurred with the lower-level descriptors (A1-B1). All skills received mistakes with reading once again topping the count (25 mistakes), and spoken production incurring the fewest mistakes (12). The C1 level was again categorized least accurately with 31% of total mistakes, and the A1 level the most accurate with only 4% of the total mistakes.

Table 5. *CEFR self-assessment grid mistakes of higher-streamed class for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	0	1	1	5	7	3	17
Reading	2	4	1	4	8	6	25
Spoken Interaction	0	1	2	7	7	5	22
Spoken Production	0	0	1	1	5	5	12
Writing	2	6	6	2	2	0	18
Total	4	12	11	19	29	19	94

In Table 6 the low- and high-streamed class results are combined to show how this cohort fared with classification errors in each category across the three years of this study. As noted previously with Tables 4 and 5, the upper-level descriptors (B2-C2) were more problematic for students to categorize than the lower-level descriptors (A1-B1). 71% of all mistakes came from the upper-level of the self-assessment grid. In particular, the C1 level proved most challenging to categorize correctly with 30% of all mistakes, followed by its neighboring grids, C2 (21%) and B2 (20%). The A1 level was the most accurately categorized for this cohort of students with only 6% of total mistakes coming from the lowest level of the CEFR self-assessment grid. In terms of mistakes according to the five separate skills, reading incurred most mistakes (26%), in contrast with spoken production incurring the fewest mistakes (12%). Indeed, the spoken production skill only had one mistake per level from A1 to B2. There were only two other Can Do descriptors that received just one mistake across three years, with the A1 listening and C2 writing descriptors being the most accurately categorized, which are interestingly at opposite ends of the scale of levels.

Table 6. *CEFR self-assessment grid mistakes of combined high- and low-streamed classes for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	1	2	2	13	14	8	40
Reading	3	7	4	7	19	17	57
Spoken Interaction	2	4	4	16	15	10	51
Spoken Production	1	1	1	1	11	11	26
Writing	7	13	13	7	7	1	48
Total	14	27	24	44	66	47	222

To determine how each class performed on a yearly basis, we can see in Figure 1 that the low-streamed class made 41 mistakes on their first attempt at sorting Can Do descriptors on the self-assessment grid. They improved slightly in 2018 by collectively making 2 fewer mistakes, however, in 2019 their overall performance deteriorated to make a total of 48 mistakes, which was worse than their initial attempt in 2017. Conversely, the high-streamed class started in 2017 with 31 mistakes, 10 fewer than the low stream, but then matched the low-streamed class in 2018 with 39 mistakes, and finally improved with their best result in 2019 of 24 mistakes, exactly half the number of mistakes of their low-streamed counterparts. A point of interest here is the corresponding average TOEIC scores per class for each year. In 2017, the low-streamed average TOEIC score was 282, versus the high-streamed class' average of 370. Both classes improved their average by approximately 100 points in 2018, however, in 2019 whereas the high stream continued to improve their average by approximately a further 100 points to 574, the low-streamed class improved only slightly by 20 points, to an average of 399 points. The role of language ability in sorting Can Do descriptors has piqued the interest of the authors, especially due to the fact that this CEFR shuffle task was completed in the students' native language of Japanese.
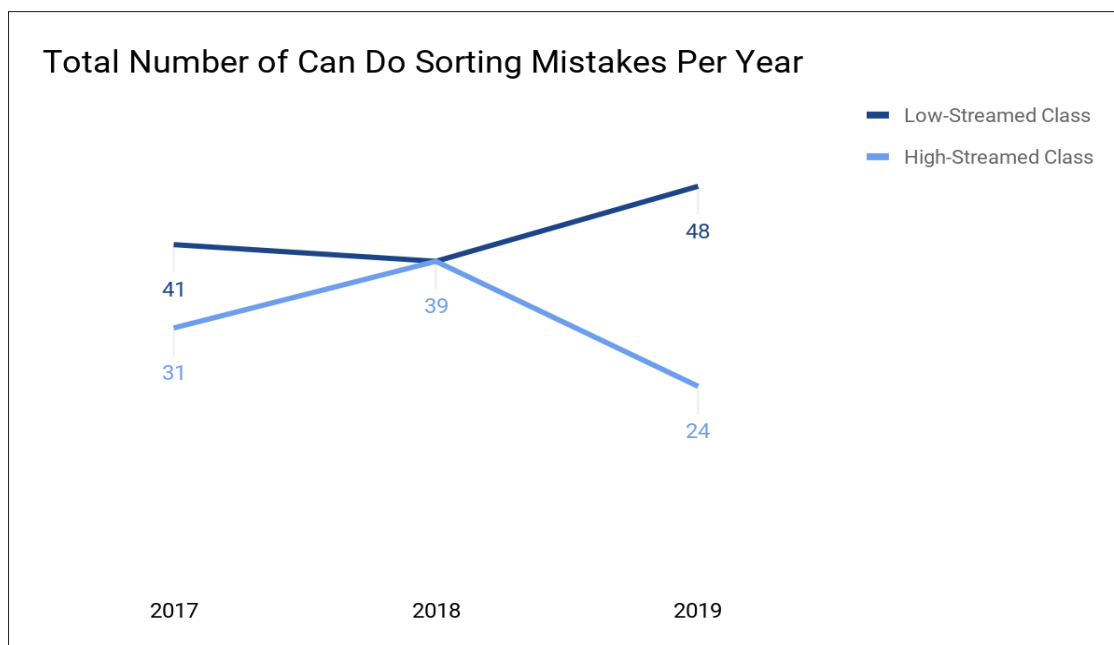

Figure 1. Can Do descriptor sorting mistakes according to class and year.

In order to examine the correlation between language proficiency and ability to correctly sort Can Do descriptors, each group's TOEIC average was compared with the number of mistakes they made each year, making 24 cases in total: four low-streamed groups and four high-streamed groups across

three years. According to the regression analysis on this data set, there was no statistically significant correlation between the students' TOEIC scores and the number of sorting mistakes (r^2=.11, p=.11).

3.2 Survey results

As can be seen in Figure 2, in 2019, when asked whether they can assess their language proficiency based on the CEFR scale, just over 30% of the students reported that they can, which showed some improvement from approximately 13% in the first year of the study. Also, no students reported that they couldn't remember the scales this time, while about 13% of the participants in the first year couldn't remember them.
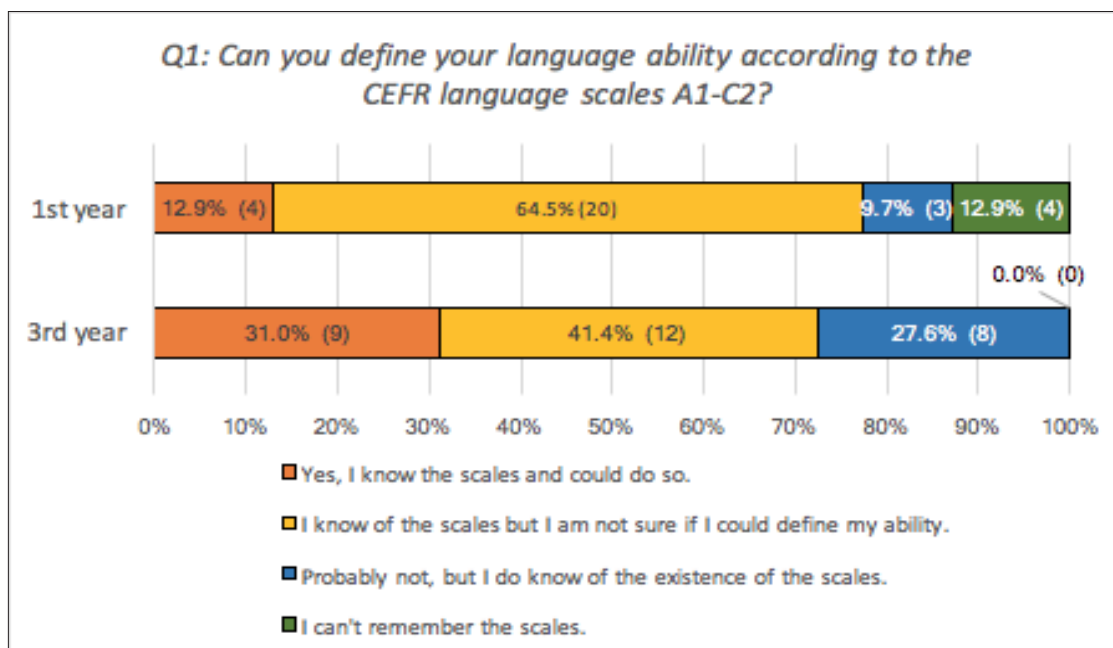
Figure 2. Students' ability in self-assessing their CEFR levels.

In 2019, more participants reported that they think the CEFR levels are useful to some degree. Although fewer participants chose "absolutely," when the two positive answer choices "absolutely" and "somewhat" are combined, the percentage of the participants who recognized the usefulness of the CEFR levels went up to 93.1% in the third year from 74.2% in the first year (Figure 3). In a similar vein, more students felt that Can Do descriptors are useful for them in the third year when comparing the two positive responses "absolutely" and "somewhat" in 2017 and 2019 (61.3% and 72.4% respectively). 22.6% and 38.7% of students did not know whether the CEFR levels and Can Do descriptors were useful for them in the first year, and these percentages came down to 3.5% and 24.1% of students in the third year.
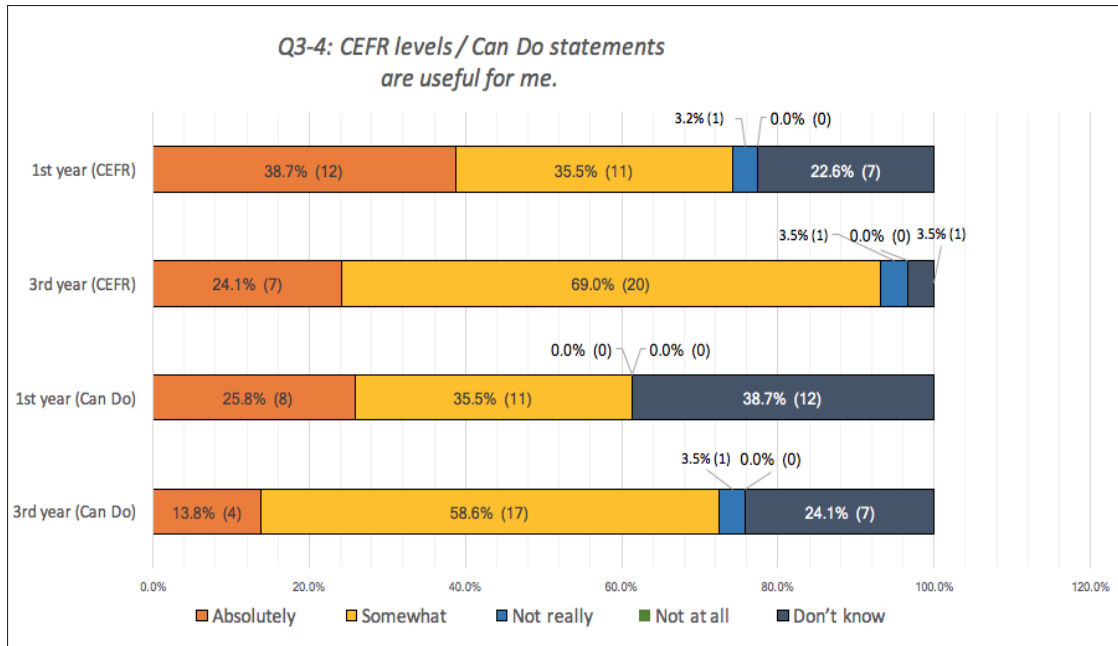
Figure 3. Usefulness of CEFR levels and Can Do descriptors.

When asked about the usefulness of the CEFR shuffle, Figure 4[2] shows that 82.7% of students reported positively in the first year of this study, and 75.9% in the third year. Three students in the first year said it was useful because it helped them set goals in their language learning. Eight students in the first year and seven students in the third year said it helped them better understand the CEFR levels. However, the percentage of the positive responses reduced from 82.7% to 75.9%, with three participants reporting that it was not really useful to them in 2019. Two of these three students' reasons were related to forgetting the descriptors soon after the activity. The other student did not give a specific reason for their choice.
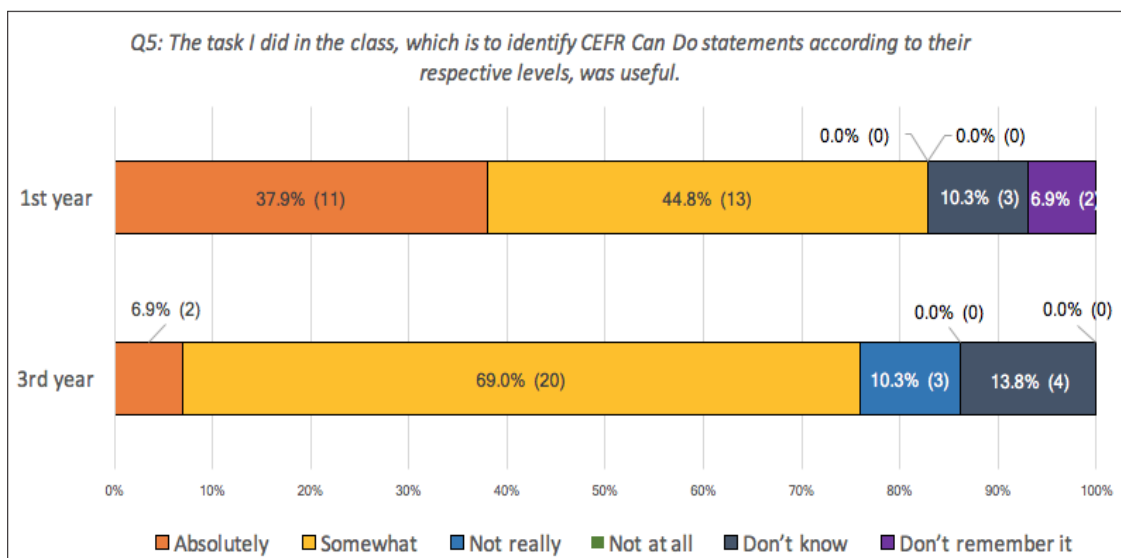

Figure 4. Usefulness of the CEFR shuffle.

---

2. It is noted here that two students who were absent on the CEFR shuffle day in the first year are excluded from the first year data.

Figure 5 shows the students' preference among three references to English proficiency. TOEIC was the most popular answer, followed by CEFR levels and Eiken grades in both the first and third years of the study. Slightly more students reported that CEFR levels are important to them in the third year while fewer students chose TOEIC. The Eiken grade received a similar score in 2019 compared to 2017.
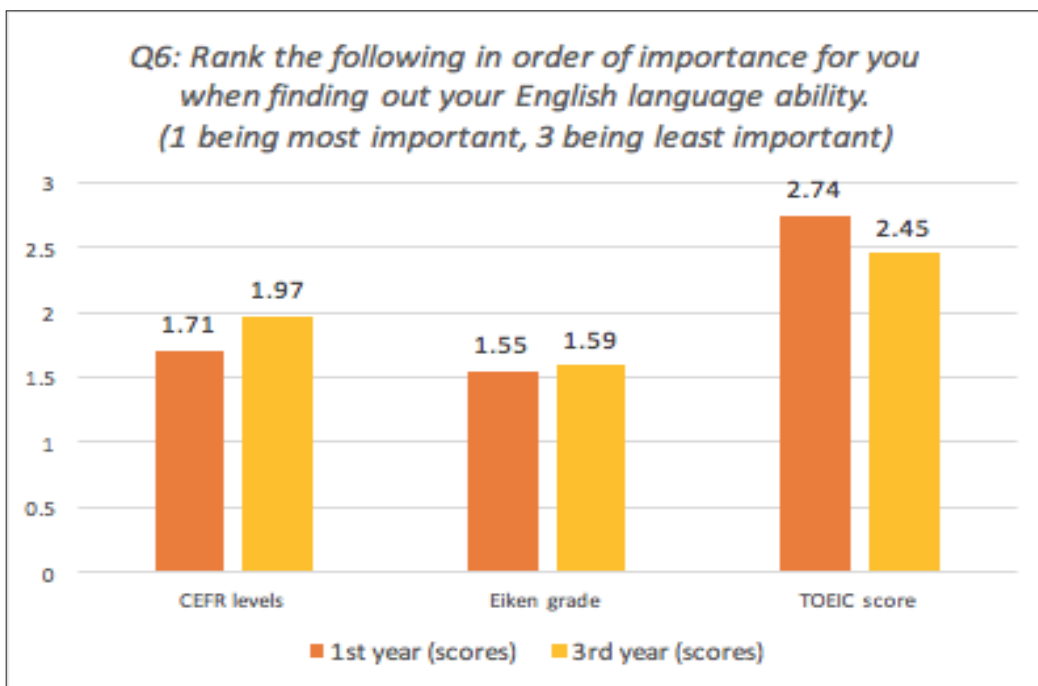

Figure 5. Importance of three language measurements.

Question 7 investigated how familiar the participants are with CEFR levels, TOEIC scores, and Eiken grades. Compared to the first year, the participants' knowledge of these language proficiency levels increased in the third year; nearly 80% of the students were able to place all the items in the correct order while 58.1% of them were able to do so in their first year (Figure 6).
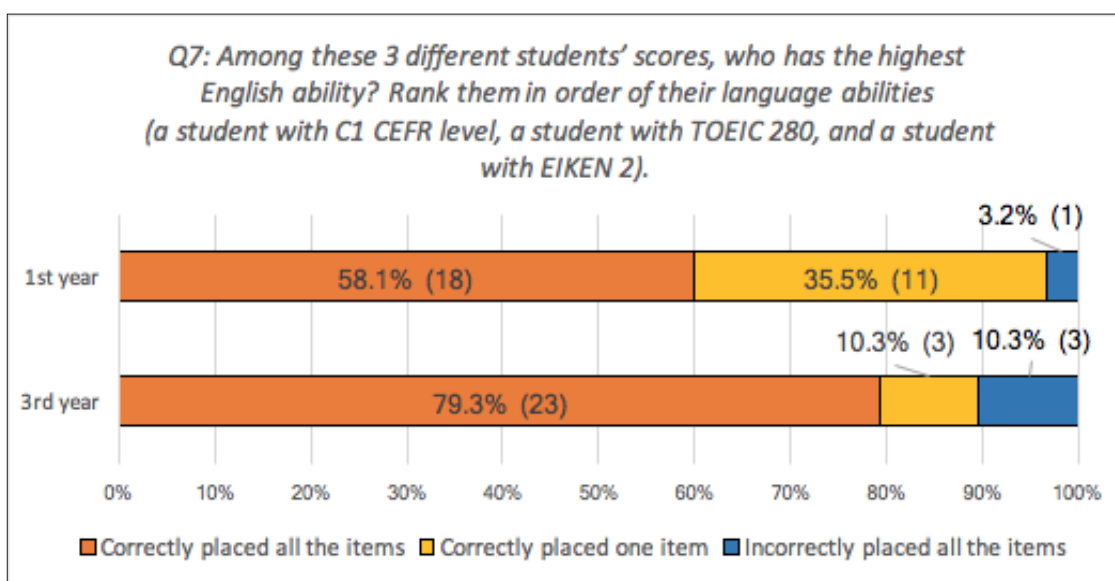

Figure 6. Students' knowledge of ranking CEFR, TOEIC and Eiken language ability.

For Questions 9 to 13, students were asked to self-assess their five language skills based on the CEFR scale (A1-C2). Displayed in Table 7 are the results of reading and listening, along with whether their self-assessment CEFR levels are supported by the results of an external English test conducted by an international education company which gives CEFR feedback: EF Education First. The results of this online test only give students an indication of reading and listening levels, hence we have not included spoken production, spoken interaction, and writing here. In terms of accuracy, more than half of the students who self-assessed their reading were correct in year 1 (58.3%), versus only 15% of students being correct in year 3. Conversely for listening, more students were correctly able to self-assess their levels in year 3 (42.1%) when compared with year 1 (16.7%). Students who responded that they could self-assess their reading and/or listening skills, but did not receive test results from EF Education First, were excluded from this table.

Table 7. *Students' accuracy in self-assessing their CEFR reading and listening levels*

	Reading		Listening	
	1st year	3rd year	1st year	3rd year
Students who assessed correctly	7	3	2	8
Students who assessed incorrectly	5	17	10	11
Total number of students	12	20	12	19

Questions 14 to 16 asked to what extent knowing their own language levels based on the three references: Eiken, CEFR, and TOEIC was useful when they graduate university. It is observable from these questions that more students appreciated the importance of these language assessments in the first year compared to the third year. Although TOEIC received the most positive responses among the three items in both years, fewer students were certain about its usefulness in the third year. Four students (approximately 14%) each reported in the third year that neither the CEFR nor Eiken results were useful for their future career or studies.
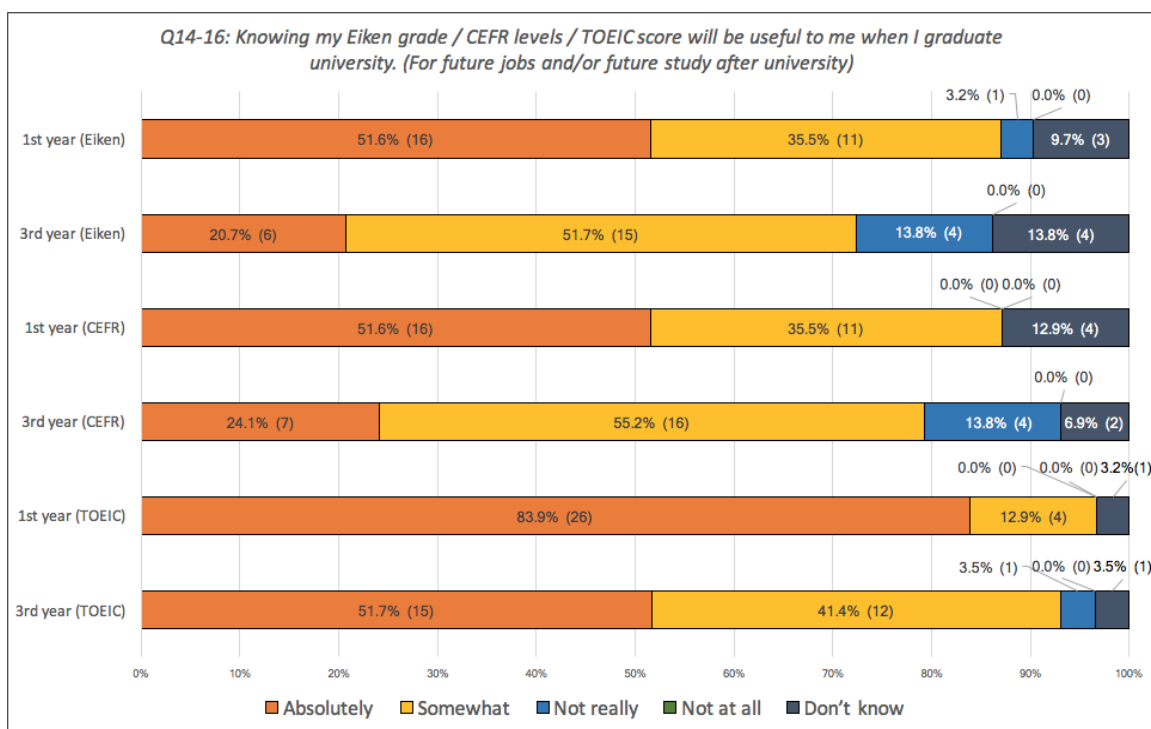

Figure 7. Popularity of three language measurements for future career and studies.

4 Discussion

The purpose of this research was to investigate how well students could sort Can Do descriptors into their respective levels of the self-assessment grid, examine what effect language ability may have had on the Can Do descriptor sorting exercise, and gain an understanding as to how students' knowledge and perceptions of the CEFR changed in their first three years of English language education at the BECC. The results from three consecutive years of conducting the CEFR shuffle show that as a cohort, students made no improvement in being able to correctly place Can Do descriptors onto the self-assessment grid. In fact, the total number of mistakes in years 1 and 3 were exactly the same. The results gave an indication that the levels of the CEFR had some effect as to where mistakes were more likely to occur. In particular, the upper half of the grid, levels B2 to C2, were most often incorrectly categorized with C1 consistently topping the mistake count each year. We suggest two reasons for this occurrence. First, the average level of our students in their third year of study is on the border of A2/B1, which is below where most mistakes occurred. Students may have had more problems conceptualising tasks above their level of English compared to their current or previously attained levels, hence confusion with which descriptor to place in B2, C1 or C2. However, we note here that this was not the case for all skills at these higher levels, with the exception being writing which had more mistakes in the A2-B1 range than the B2-C2 range. In fact, the C2 level writing descriptor was one of the most accurately placed descriptors with only one mistake in three years. The second reason we offer as to why our students were more consistent with accurately placing lower-level descriptors over higher ones is in the wording of the descriptors and abstractness of tasks. Lower-level descriptors often describe more concrete tasks using simple words, such as understanding familiar words on posters (reading A1), whereas higher-level descriptors utilize more difficult vocabulary, for example "contemporary literary prose" (reading B2), and abstract ideas: "appreciating distinctions of style" (reading C1) (CoE 2001: 27).

At first, language ability appeared to have some effect on students' ability to sort descriptors accurately. In both years 1 and 3 the high-streamed class outperformed the low stream, with year three being particularly of interest due not only to the high stream making the fewest mistakes of the study (24), but the result from the low stream making twice as many mistakes (48) as their counterparts. When comparing each group's average TOEIC scores, it is apparent the high-streamed class continued to improve year-by-year, with an average gain of 100 points per year, whereas the low group only managed to improve their third year TOEIC average by 20 points. However, as discussed earlier, the results of a regression analysis on the correlation between each groups' TOEIC average scores and the number of mistakes showed that there is no statistically significant correlation between those two factors. These mixed results may warrant a further study as students worked in groups and individual opinions may not have been reflected in the results. One possible change to the CEFR shuffle in the future is to add a step after the initial group discussion whereby students are individually asked to determine where Can Do descriptors are placed and examine the correlation between the results and their individual language proficiency.

The post-activity surveys conducted in the first and third years of the study suggest that the students' knowledge of the CEFR improved to some extent. When asked to define their CEFR levels, more students reported that they were able to do so (Question 1). However, when accuracy is considered with the results of an externally validated English test, students were not very successful or consistent at estimating their CEFR levels by skill. Reading levels were self-assessed more accurately in Year 1 (58.3%) than Year 3 (15%). Conversely, listening levels were self-assessed more accurately in Year 3 (42.1%) than Year 1 (16.7%) (Table 7). We consider the result that more students gained confidence to self-assess their language ability by skills in Year 3 a positive, however, and are currently trying to determine ways of improving accuracy of self-assessment for future students. The results of Question 7 showed that more students were aware of the CEFR levels in relation to the two major language proficiency tests in Japan: TOEIC and Eiken. As stated earlier, CEFR levels and Can Do descriptors are used in various forms in the BECC, from lesson materials and self-access resources to assessment and feedback. It is natural

to assume that the students had many chances to be exposed to the CEFR's level system in the last three years. Cook and Rutson-Griffiths (2018: 78) argue "students should become more aware of their CEFR levels as teachers give feedback on assignments utilizing CEFR-informed rubrics." This is probably the main reason for the improvement of their CEFR-knowledge. Although it is difficult to argue with certainty as there is no supporting data, it is hoped that our annual CEFR shuffle also contributed to this positive change.

Regarding the students' perceptions, the results of Questions 3 and 4 show the number of students who find CEFR levels and Can Do descriptors useful to some degree increased in two years. Additionally, Question 5 saw the majority of students agreeing that the CEFR shuffle task was useful to them with reasons being attributed to its contribution towards goal-setting and understanding of the CEFR levels. We believe the results of these three questions align favourably with a promotion of self-directed learning as was originally intended with the foundation of the CEFR (CoE 2001: 6), and one of motivating factors behind this study. However, this was not the case for all students. There were still several students who did not recognize the usefulness of CEFR levels or Can Do descriptors in the third year. This combined with two students' comments that the CEFR shuffle was not really useful to them due to the descriptors being quickly forgotten highlight the need for continued training to promote a more effective learning experience.

According to the results of Question 6, which was to find out the popularity of three major language proficiency measurements in Japan, TOEIC was the most popular, although its popularity slightly declined, and in turn, the CEFR attracted a little more favor in the third year. Our interpretation of this result is students' knowledge of the CEFR increased as they saw more examples of its use in the BECC over two years. The outcome was a better awareness of how the CEFR may be beneficial for their futures. A possible reason for the minimal change in the language measurement popularity is that these students are required to take TOEIC by their department at regular intervals, and it is still the most widely used language test for job hunting in Japan. It is interesting, however, to see that the popularity for all three language measurements declined in the third year (Figure 7). This could be the result of some of the students realizing or deciding that they would not need English in their future career or studies, leading to the somewhat negative perception towards these measurements of language ability.

5 Conclusion

We carried out this study to investigate students' familiarity, knowledge, and perceptions towards the CEFR and to raise their awareness of this framework upon which their English language education is based. Through our observation we discovered that students did not show improvement sorting Can Do descriptors over two years and that most mistakes were consistently made in the upper-level of the self-assessment grid, from B2 to C2 levels. However, this study is not without its limitations. Although the high-streamed group made half the number of mistakes as the low-streamed group in the third year of this study, the results of analysis on the correlation between students' language proficiency and descriptor sorting ability suggest there is no statistically significant correlation between those two factors. We suggest a further study that would have students engaging in the CEFR shuffle in small groups but making their final decisions individually to investigate the role between language proficiency and descriptor sorting ability. Feedback is one area that can be improved on in future editions. There was little discussion as to why Can Do descriptors were incorrectly placed. With better feedback, students may increase their accuracy of the placement of descriptors, which could contribute to improved results with students' ability to self-assess their language levels. Further limitations to this study were the small number of participants and the use of a single language test as a reference to compare the participants' task performances. A possible change to future research may be involving a larger sample size and employing a different language proficiency measurement. The contrasting results of students' CEFR knowledge improvement and a small decline in their interest in language proficiency references may

suggest that there are other factors to consider such as students' motivation and needs for English studies for their future. Nevertheless, this three-year study helped us to better understand our students' prior knowledge as well as initial perceptions of the CEFR and how these may change over a few years. The CEFR shuffle was viewed positively by most of our students, and although we hope this study also served to contribute towards self-directed learning we suggest continuous education about the CEFR and training on self-assessment.

6 References

Association of Language Testers in Europe. 2002. *The ALTE Can Do project: Articles and Can Do statements produced by the members of ALTE 1992-2002.* https://www.alte.org/Materials (accessed 19 February 2020)

Bower, Jack, Judith Runnels, Arthur Rutson-Griffiths, Rebecca Schmidt, Gary Cook, Lyndon Lusk Lehde & Azusa Kodate. 2017. Aligning a Japanese university's English language curriculum and lesson plans to the CEFR-J. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 176-225. Cambridge: Cambridge University Press.

Bower, Jack, Arthur Rutson-Griffiths, Gary Cook, Rebecca Schmidt, Lyndon Lusk Lehde, Azusa Kodate & Judith Runnels. 2017. The key questions in Bunkyo. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 247-266. Cambridge: Cambridge University Press.

Cook, Gary. 2019. Working in silos: A report on the coordination of course collaboration at a Japanese university. *The IAFOR International Conference on Education-Hawaii 2019 Official Conference Proceedings*. https://papers.iafor.org/submission42228/ (accessed 12 February 2020)

Cook, Gary & Yukari Rutson-Griffiths. 2018. Investigating first-year students' perceptions and knowledge of the CEFR at a Japanese university. *Hiroshima Bunkyo Joshi Daigaku Kiyo* [Bulletin of Hiroshima Bunkyo Women's University] 53. 67-83.

Council of Europe. 2000. *European Language Portfolio (ELP): Principles and guidelines.* Strasbourg: Language Policy Division. https://rm.coe.int/16804586ba (accessed 19 February 2020)

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. https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr (accessed 19 February 2020)

Ek, Jan Ate van & John Leslie Melville Trim. 1991. *Waystage 1990*. Cambridge: Cambridge University Press.

European Association for Quality Language Services. 2015. *Eaquals bank of descriptors.* https://www.eaquals.org/resources/revision-and-refinement-of-cefr-descriptors/ (accessed 19 February 2020)

Fennelly, Mark Graham. 2016. The influence of CEFR on English language education in Japan. *Shikoku Daigaku Kiyo* [Bulletin of Shikoku University] 46. 109-122. http://id.nii.ac.jp/1456/00000378/ (accessed 19 February 2020)

Kodate, Azusa. 2017. Developing ELP-informed self-access centre learning materials to support a curriculum aligned to the CEFR. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 226-247. Cambridge: Cambridge University Press.

Ministry of Education, Culture, Sports, Science and Technology. 2013. *Gurobaru ka ni taio shita eigo kyoiku kaikaku jisshi keikaku* [English education reform plan corresponding to globalization]. https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1343704.htm (accessed 19 February 2020)

Nagai, Noriko & Fergus O'Dwyer. 2011. The actual and potential impacts of the CEFR on language education in Japan. *Synergies Europe* 6. 141-152. https://gerflint.fr/Base/Europe6/noriko.pdf (accessed 19 February 2020)

O'Dwyer, Fergus. 2015. Toward critical, constructive assessments of CEFR-based language teaching in Japan and beyond. *Gengo bunka kenkyu* [Studies in Language and Culture] 41. 191-204. https://ir.library.osaka-u.ac.jp/repo/ouka/all/51427/slc_41-191.pdf (accessed 19 February 2020)

Schmidt, Maria Gabriela, Judith Runnels & Noriko Nagai. 2017. The past, present and future of the CEFR in Japan. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 18-48. Cambridge: Cambridge University Press.

Tono, Yukio & Masashi Negishi. 2012. The CEFR-J: Adapting the CEFR for English language teaching in Japan. *JALT Framework & Language Portfolio SIG Newsletter.* 5-12.

University of Cambridge ESOL Examinations. 2012a. *Cambridge English Key: Key English Test (KET) CEFR level A2 handbook for teachers*. Cambridge: UCLES.

University of Cambridge ESOL Examinations. 2012b. *Cambridge English Preliminary: Preliminary English Test (PET) CEFR level B1 handbook for teachers*. Cambridge: UCLES.

7 Biographies

Gary Cook has been a lecturer at the Bunkyo English Communication Center in Hiroshima Bunkyo University since 2011, and is currently the Global Communication Department coordinator. He has previous teaching experience in Spain, France, England, Georgia, and his native New Zealand. His research interests are in the areas of curriculum development, analysing the needs of language learners, and the CEFR.

Yukari Rutson-Griffiths has been a learning advisor at Hiroshima Bunkyo University since 2015 and is the coordinator at the self-access learning center at the university. She has also been involved in creating self-access learning materials using the CEFR and CEFR-J. Her research interests include self-access learning, learner autonomy, and the CEFR.

Appendix
Survey regarding students' knowledge and perceptions of the CEFR

Q1: Can you define your language ability according to the CEFR language scales A1-C2?

- *Yes, I know the scales and could do so.*
- *I know of the scales but I am not sure if I could define my ability.*
- *Probably not, but I do know of the existence of the scales.*
- *I can't remember the scales.*

Q2: Where have you seen Can Do statements and/or CEFR levels? Write as many ideas as you can.

Q3: CEFR levels are useful for me.

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q4: Can Do statements are useful for me.

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q5: The task I did in the class, which is to identify CEFR Can Do statements according to their respective levels, was useful.

- *Absolutely / Somewhat / Not really / Not at all / Don't know / Don't remember it*

Q6: Rank the following in order of importance for you when finding out your English language ability. (1 being most important, 3 being least important)

- *CEFR levels / Eiken grade / TOEIC score*

Q7: Among these 3 different students' scores, who has the highest English ability? Rank them in order of their language abilities.

- *A student with C1 CEFR level / A student with TOEIC 280 / A student with Eiken 2*

Q8: What is your most recent TOEIC score? If you can't remember it please write "I don't remember."

Q9: What is your reading CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q10: What is your listening CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q11: What is your writing CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q12: What is your spoken interaction CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q13: What is your spoken production CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q14: Knowing my Eiken grade will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q15: Knowing my CEFR level will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q16: Knowing my TOEIC score will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*