CEFR-J Self-assessment with Japanese First-year University Students

Kevin Mueller †

† Tokyo International University　1-13-1 Matoba-kita, Kawagoe, Saitama, 350-1197 Japan
E-mail:　kmueller@tiu.ac.jp †

Abstract

This study, undertaken with 257 Japanese first-year university students at a Japanese university, examines students' self-assessment of their ability based on a portion of the CEFR-J can-do list.　It was conducted at the outset of the year with students who are English Language Communication majors to examine their self-assessment of English ability prior to the beginning of university English courses.　These students are enrolled in a program that has them take nine hours of compulsory English courses each week, primarily with non-Japanese nationality faculty members. The findings reveal that, in general, more students stated they could do the framework proficiency skills at the lower end of the scale (in the A0 to A1 CEFR bands) while fewer self-assessed themselves as able to perform at the higher end of the scale.　This study noted a few outlying outcomes, most notably, that more students responded they could perform at a higher level than at a lower level according to the CEFR-J rubric in two instances, and these apparent discrepancies are explained herein.　Moreover, while the overall study has a high reliability coefficient as measured by the Cronbach alpha and the Mokken scale score, which measure the ranking order that a participant who answered positively to a more difficult question is assumed to answer an easier question appropriately, was rather low.　These reliability factors are elucidated as well.　The paper concludes with further considerations for future can-do self-assessment research employing the complete CEFR-J with Japanese university students.

Keywords　CEFR，CEFR-J，self-assessment，self-evaluation

日本人大学 1 年生の CEFR–J 自己査定

ミューラー　ケビン†

†東京国際大学埼玉県川越市的場北 1–13–1〒350–1197
E-mail:　†kmueller@tiu.ac.jp

概要

　日本で英語を学ぶ 257 名の大学 1 年生を対象に、本研究では、CEFR-J の can-do リストに基づき、学生の自分の能力に対する自己査定に関して調査を行なった。この調査は、英語コミュニケーションを専攻する新入生が、学生が大学での英語の学びを始める前の段階で、自分の英語能力に関する自己査定を調べるために入学時に実施された。これらの学生は主に日本人以外の教員が担当する 1 週間あたり 9 時間の必修授業を受けるプログラムに登録していた。全体的に本調査を通して、多くの学生が　尺度の下位となる CEFR の　A0 から A1 の範囲で文章構造能力に関して出来たと述べる一方で、高い尺度で力を発揮できたと自己査定した者は少なかった。さらに加えると、この研究では、とりわけ CEFR-J のルーブリックにおける 2 つの項目において、多く

の学生が下位のレベルよりも上位のレベルで力が発揮できるという、特異的な通常ではあり得ない結果も示しており、この明らかなる矛盾点を本論文で明らかにしている。さらに本調査全体を通して、より難しい質問に正答できる対象者は、より易しい質問には答えられるであろうとの想定のもとに順位を測定する Cronbach alpha および Mokken 尺度法によって測定された高信頼係数はかなり低かった。これらの信頼度因子も明瞭に説明した。本論文では、一貫した CEFR-J 調査を日本人大学生に用い、今後 can-do 自己査定研究により重点を置くことが肝要であるとの結論を得ることができた。

1. Introduction

In 2001, after more than twenty years of research, the Council of Europe published The Common European Framework of Reference for Languages (CEFR) in both English and French. It is now published in forty languages, including Japanese and was "designed to provide a transparent, coherent and comprehensive basis for the elaboration of language syllabuses and curriculum guidelines, the design of teaching and learning materials, and the assessment of foreign language proficiency. It is used in Europe but also in other continents" [1].

The CEFR outlines foreign language ability at six discrete levels: A1, A2, B1, B2, C1, and C2. In addition, it denotes three "plus" stages (A2+, B1+, B2+). With a foundation based on firsthand experimentation with second language learners and extensive input from linguists, educators and researchers, the framework, according to its writers, makes it possible: "to establish learning and teaching objectives, to review curricula, to design teaching materials and, to provide a basis for recognizing language qualifications thus facilitating educational and occupational mobility" [2].

The CEFR document is far more comprehensive than merely containing a collection of "can-do" statements that outline discernable language skills that a learner is able to proficiently do. This paper will concentrate primarily on these "can-do" aspects of the CEFR in both it original form and the CEFR-J [3] (a framework conceived and organized solely for the English learning and teaching context in Japan) [4].

Within the original CEFR document (Council of Europe), two appendices focus on "can-do" statements. The first, Appendix C, provides self-assessment statements as well as language tests and feedback; this system is in place for learners, primarily those who are independently studying one of the following fourteen European languages: Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish, and Swedish [5]. The second, Appendix D, seeks to outline "a set of performance-related scales, describing what learners can actually do in the foreign language " [6]. There are "can do" descriptors for each of the six levels. These "can do" descriptors were created by the Association of Language Testers in Europe (ALTE).

While the CEFR was designed for the European language learning context, it has expanded throughout the world as an all-purpose, broad construct for understanding language ability from the perspectives of studying, instruction, and evaluation. Tono and Negishi [7] argue that, based on an eight-year period of analyses inside and outside of Japan, implementing the CEFR will be a key instigator in the transformation of English education in Japan. Nagai and O'Dwyer [8] concur with these researchers and claim that the implementation of the CEFR rubric in Japan has been generally beneficial for language education in Japan. Three examples can be shown to illustrate how the CEFR has been applied in Japan at the governmental level. First, Fenelly [9] points out that the December 13, 2013 "English Education Reform Plan corresponding to Globalization" document explicitly mentions CEFR levels noting that junior high

school English education be at the A1-A2 levels and high school instruction be at the B1-B2 levels. Second, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) released a document with five propositions and clear guidelines for improving English for worldwide communication that incorporates measuring students' proficiency with a can-do inventory, as cited by Tono and Negishi [10]. Tono and Negishi maintain that this inventory is inspired by the CEFR. Third, the aforementioned two researchers state the *Nihon Hoso Kyokai* (NHK), the Japan Broadcasting Corporation, which is, in part, funded by the Japanese government, has implemented the CEFR rubric for their foreign language textbooks, and television and radio programs.

At the governmental level, various methods of incorporating CEFR are occurring; however, Negishi's research [11] revealed eighty percent of university students in Japan are in the A or B bands. Thus, the original CEFR A and B levels were altered to more discretely classify learners [12]; A1, A2, B1, B2 in the original were divided up into nine groups (A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2). In addition, a Pre-A1 stage was incorporated; therefore, the CEFR-J has the resulting twelve levels:
[Pre-A1], [A1.1, A1.2, A1.3], [A2.1, A2.2], [B1.1, B1.2], [B2.1, B2.2], [C1] [C2]

2. Self-assessment Validity

From a pedagogical perspective, there are a number of explanations for encouraging language learners to conduct self-assessment. In a language learning context, Oskarsson [13] elucidates six reasons why a learner's self-assessment is helpful. First, it provides the learner with experience in appraising proficiency that then aids in learning development. Second, learners and instructors gain an increased appreciation of distinguishing degrees of proficiency. Third, it stimulates a learner to focus on further learning objectives. Fourth, a teacher can present students with a variety of self-measurement practices. Fifth, students have opportunities to engage in gauging their own proficiency. Finally, students can carry these self-assessment skills with them as they continue studying once a course is completed. As a result of being involved in self-assessment, the language learner is can be an active participant in the learning process.

While being beneficial, self-assessment does have a number of weaknesses. According to Dunning, Heath and Suls [14], a few psychological factors are at work that result in defective self-assessments, and they sort them into two main types. Firstly, inaccurate self-assessment may occur because people usually do not have all the necessary information required to make a correct appraisal, and they do not take into consideration what is unknown. Secondly, inaccurate self-assessment often occurs because people do not pay close attention to significant and advantageous information which they do have. Another related drawback of self-assessments is that students can deliberately lie when assessing their own skills [15].

More pertinent to this study, an L2 learning context, Blanche and Merino [16], in a meta-analysis of the accuracy of self-assessment, found that a learner's self-assessment stretched from being rather accurate to very accurate. In another meta-analysis, Ross [17] corroborated the findings of Blanche and Merino. However, there are a number of caveats that require explanation. Kruger and Dunning [18] recognized that people with higher proficiency tended to rate themselves as being less proficient while less proficient students tended to rate themselves more highly. In research done on Japanese university students' English writing ability, Matsuno [19] evaluated students' self-assessment and instructor assessment and found this tendency that Kruger and Dunning noted as well: more proficient students underestimated their ability. Matsuno posits that this "was

probably caused by the tendency of many Japanese to display a degree of modesty" [20].

In addition to these factors, gender appears to also possibly skew self-assessment results. In a meta-analysis study [21] determined that gender affects self-assessment depending on the subject or content being examined. In the language arts, females were shown to have higher self-efficacy beliefs; however, in the field of language arts when students of comparable proficiency levels were analyzed, males overestimated themselves more than females [22]. There are various methods for students to assess their proficiency themselves in the L2 context, and one study by Bachman and Palmer [23] found that the most accurate question type asked learners about the perceived difficulty with certain facets of English as an L2. They noted that the least effective type of question was the "can-do" question.

Students' self-assessments cannot guarantee objective measurements of true ability. However, to compensate for some of these deficiencies Ross notes that instructors "can strengthen reliability through such strategies as engaging students in rubric construction" [24] and providing learners with training.

The present report is a summary of the findings of a particular self-assessment of skills outlined in the CEFR-J by a first-year Japanese university students. It notes not only the findings but also mentions shortcomings and possible future research along these lines.

3. Method

Participants

257 students in first-year English as a foreign language courses of Tokyo International University voluntarily participated in this study. Participants were from three different departments: Language Communication, International Relations, and Economics. The survey was administered at the beginning of April 2017.

Instrument

The survey was conducted in Japanese and administered online using Google Forms. Participants were required to indicate "yes" or "no" to indicate whether they can or cannot do a particular skill based on 40 CEFR-J can-do statements. The 40 can-do statements were taken from all five skills (understanding: listening, reading; speaking: spoken interaction, spoken production; and writing) from four of the CEFR-J levels: A1.2, A2.1, B1.1, and B2.1. The statements were randomized (see Appendices 1 and 2 for the original Japanese questions and translated English). The rationale for randomizing statements was to have participants be "blind" to the progression of difficulty inherent in the CEFR-J statements. These particular levels were selected as incoming freshmen students at Tokyo International University placed, to varying degrees, in these levels on an in-house placement paper-based test formulated on CEFR standards which was administered at the end of March 2017, prior to the commencement of 2017 spring term classes.

4. Results and Discussion

As Figure 1. and Tables 1 and 2 reveal that there were more positive responses at the A1 and A2 bands. Fewer positive responses were found at the B1 and B2 bands. One notable and seemingly contradictory exception is the spoken production skill category in the A1.2 and A2.2 levels. Here, more positive responses for the more advanced level (A2.2) were given than for the more basic level (A1.2). There may be two reasons for this. One, each question on the questionnaire asks participants to respond in a binary fashion (either "yes" or "no"). Students may have more diffuse or subtle responses to the questions asked and thus a Likert scale type format may have been more appropriate. In addition, the two questions at the A1.2 level (Question 25. I can express simple opinions related to limited, familiar topics, using simple words and

basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance. Question 29. I can give simple descriptions e.g. of everyday object, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.) could be perceived as more challenging as the A2.1 questions (Question 26. I can introduce myself including my hobbies and abilities, using a series of simple phrases and sentences. Question 30. I can give a brief talk about familiar topics (e.g. my school and my neighborhood) supported by visual aids such as photos, pictures, and maps, using a series of simple phrases and sentences.). A possible rationale for the perceived level of difficulty could be that participants concentrated on the phrasing "restricted range of sentence structures" in the A1.2 questions. What is meant by this expression is open to interpretation and each participant may be unclear as to how to gauge the meaning of "restricted range". Thus, these modifying statements in the questions at the A1.2 level may be perceived by participants as being more difficult.

Another apparent anomaly can be seen in the second spoken production question (questions 29 through 32. See Appendix 1 and 2). Here, the largest number of students (161) responded positively to the B1.1 level statement (154 participants stated "yes" for the A1 level statement and 149 stated "yes" to the A2 level statement). Here, two possible explanations can be provided for this seeming contradiction. Firstly, in the A1 and A2 category questions, modifying phrases related to grammatical structure ("restricted range of sentence structures" in question 29 [A1] and "series of simple phrases and sentences" in question 30 [A2]) which do not exist in the B1 question may cause Japanese students who are particularly sensitive to grammatical points when speaking to not respond in the positive. A second reason for the apparent contradiction is the A1 and A2 level questions could be perceived as connoting that the respondent will give a public speech; the two questions explicitly state: "…provided I can prepare my speech in advance" (question 29) and "I can give a brief talk…" (question 30). The B1 category question does not explicitly or implicitly connote a public speech The

B1 category question does not explicitly or implicitly connote a public speech as the language is vague ("I can talk about….").

Table 1: Raw Number of Participants' Positive Responses to CEFR-J Questions According to Skill Group

CEFR-J Level	Listening
A1.2	470
A2.1	405
B1.1	238
B2.1	130

CEFR-J Level	Reading
A1.2	445
A2.1	419
B1.1	230
B2.1	83

CEFR-J Level	Spoken Interaction
A1.2	432
A2.1	379
B1.1	239
B2.1	125

CEFR-J Level	Spoken Production
A1.2	323
A2.1	370
B1.1	256
B2.1	62

CEFR-J Level	Writing
A1.2	449
A2.1	358
B1.1	181
B2.1	80

* Note: The number of students at a given CEFR-J level or above and those who responded "yes" to CEFR-J can-do questions exceeds the total number of participants. The number at students at given CEFR-J level or above is more than the number of participants because students at the A2 or higher level should, in principle, respond "yes" to questions

at the A1.2 level; this would also be true for the A2 and B1 levels for B1 and B2 students respectively.
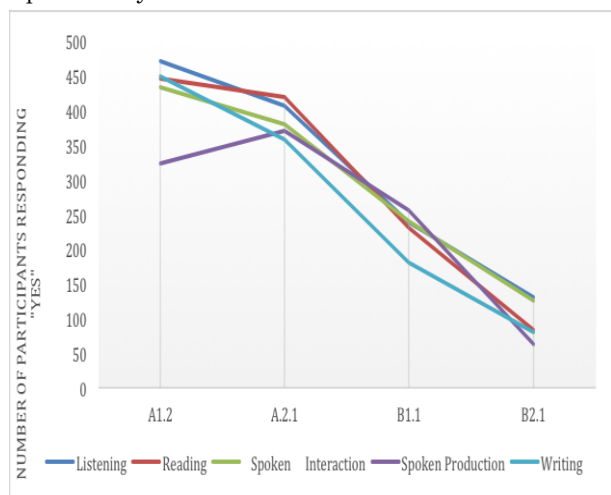

Figure 1: Relationship among raw number of self-ratings in five skill categories

Williams and Andrade, in a study of 243 Japanese university students studying English in Japan, found that "anxiety was often associated with tasks involving speaking in front of others" [25]. Therefore, anxiety related to giving a more formal public speech rather than simply talking about a topic could have skewed responses.

The measure of internal consistency of participants' self-ratings among the five skill groups in this study was statistically reliable when the raw responses of individual students was compared with each skill category as the Cronbach alpha was 0.9062. This scale is from 0 to 1.0 and the higher the number, the more statistical reliability. A reliability coefficient of 0.7000 or higher is deemed "acceptable" in the majority of social science studies. The coefficient in this present study mirrors the finding of internal consistency that Tokeshi and Gao [26] had. In their study that the Cronbach's alpha value among the five skill categories was 0.872 when they calculated "the average self-ratings of

individual respondents for each skill category" [27]. Another study with similar reliability data is Runnels' [28]. When calculating data on the reliability of the entire CEFR-J A level can-do statements with 590 first- and second-year Japanese university students in Runnels' study, across all statements a strong reliability as measured by Cronbach's alpha 0.944 scale was found.

	A1.2	A2.1	B1.1	B2.1
	250 (1)	227 (2)	173 (3)	74 (4)
Listening	220 (5)	178 (6)	65 (7)	56 (8)
	227 (9)	222 (10)	77 (11)	30 (12)
Reading	218 (13)	197 (14)	153 (15)	53 (16)
	216 (17)	188 (18)	142 (19)	63 (20)
Spoken Interaction	216 (21)	191 (22)	97 (23)	62 (24)
	169 (25)	221 (26)	95 (27)	24 (28)
Spoken Production	154 (29)	149 (30)	161 (31)	38 (32)
	228 (33)	170 (34)	98 (35)	46 (36)
Writing	221 (37)	188 (38)	83 (39)	34 (40)

Table 2: Raw Number of Participants' Positive Responses to Each Question Along CEFRJ Skill Groups (question number noted in parentheses)

CEFR-J Level	Cronbach alpha
A1.2	0.744
A2.1	0.684
B1.1	0.799
B2.1	0.853

Table 3. Cronbach Alpha for CEFR-J Levels

As shown in Table 3., the A2.1 level questions had a few anomalies and they may explain the relatively lower Cronbach alpha coefficient in comparison with the other levels. A Mokken Scale analysis for the A1.2, A2.1, B1.1, and B2.1 statements as group was conducted (results displayed in Table 4. below). Runnels' [29] research undertook similar analysis and she pointed out that "Mokken scaling is a statistical technique that assumes the order of difficulty of items is not the same across a population (van Schuur, 2003) and it provides a measure of reliability by identifying items for which Guttman patterning is occurring at higher rates (Molenaar, 1997; Sijtsma & Molenaar, 2002)" [30]. Guttman scaling is utilized in surveys, tests or questionnaires having binary items ("yes" or "no" as in the one employed in the present study). In a Guttman scale, items are arranged in order so that if a participant answers "yes" to an B2.1 level question, they should also respond "yes" to lower proficiency items (e.g. at the A1.2, A2.1, and B1.1 levels). Thus, stating "yes" to an item at a higher level implies that "yes" should be noted at all lower level questions. Runnels [31] writes that Mokken scaling generates a ratio that displays the perceived difficulty of each can-do statement according to the self-rated ability of each participant and the degree to which a larger number of higher proficiency participants perceived the can-do statement more challenging. The statistic that is produced is called the coefficient of homogeneity (H or H-value), and it reveals the structure of answers for each statement by means of item limitations. Thus, this coefficient notes the reliability scale for each can-do statement and exposes the degree that a Guttman model can be seen for each answer. Coefficients of homogeneity range from 0 to 1.0. A higher H-score correlates with an element that measures more in sync with Guttman's proposition (i.e. items are arranged in order so that if a respondent answers "yes" to an B2.1 level question, they should also respond "yes" to lower proficiency items; stating "yes" to an

item at a higher level implies that "yes" should be noted at all lower level questions.). H-values beneath the 0.3 threshold are unacceptable, and scores over 0.6 denotes strong reliability [32].

Table 4. Mokken Scales Coefficients for Each CEFR-J level Examined

CEFR-J Level	H-coefficient
A1.2	0.394
A2.1	0.387
B1.1	0.376
B2.1	0.485

Two hypotheses for relatively lower h-coefficient results (all are on the lower cusp of acceptable results of 0.3 as aforementioned). Firstly, in contrast to Runnels' study [33] where participants responded to the five A levels in the CEFR-J (A1.1, A1.2, A1.3, A2.1, A2.2), the present study took a larger view and examined a range from of levels from A1.2 to B2.1. As previously noted, conspicuous incongruities in participants' answers to two A2.1 can-do statements relative to A1.2 and B1.2 queries may have skewed reliability. Secondly, the present study only had respondents select from binary options ("yes" or "no"). Having a more restricted field of options, while producing more discrete data, constrains and compels participants in an either-or situation which may not truly reflect participants' ability. In a similar study, Runnels' [34] utilized a five category Likert-scale response form (strongly disagree to strong agree). Providing more breadth of responses possibly reduces the error co-efficient and diffuses the data.

A number of weaknesses with this study must be detailed. Firstly, the entire 110 CEFR-J can-do list was not employed and thus while an overview of the participants' self-ratings can be recognized, a complete picture of the group's CEFR-J self-assessment cannot be seen. This would be a step for a further study. The rationale for this limited survey was twofold; first, at the start of an academic year appropriating time for freshmen students to participate in a voluntary-based questionnaire is limited and therefore the researcher decided to use an abbreviated version, and second, at the outset of the academic year, the students in question sat for an in-house

placement test that was developed along CEFR lines. The purpose of this test was to stream students into sections so that students would be with students at similar proficiency levels. This in-house placement test was designed based on the Cambridge KET examination (geared for A1 and A2 learners). Following numerous pilot trials and follow-up analysis and modifications, the in-house test revealed measureable reliability to the KET test (0.753 Cronbach alpha). Thus, administering a questionnaire with queries from the A1, A2, B1, and B2 levels was deemed suitable. A second drawback of this study was the binary response options (yes/no). With more nuanced options in a Likert scale (e.g. "strongly agree, / somewhat agree / somewhat disagree / strongly disagree") may have provided more telling and informative data. Thirdly, certain terms in the CEFR-J can-do list may be unfamiliar to students and thus cause confusion and therefore result in misleading data. As the majority of participants in this study were first-year Japanese university students who entered the university in April, questions pertaining to a work environment would not necessarily pertain to respondents. For example, question 16 states "I can understand in detail specifications, instruction manuals, or reports written for my own field of work, provided I can reread difficult sections." Students in their first year of studies at university in Japan most likely would not be exposed to interacting with English in such a fashion and therefore may be at a loss as to how to respond. A fourth weakness of the study revolves around the term "native speakers" which arises a number of times throughout the questionnaire. The entire notion of "native speaker" and its relevance to English proficiency is a different topic (e.g. English as lingua franca); however, germane to this study is that numerous participants may not have had opportunities to interact with "native speakers". Thus, a question such as number 24, "I can discuss abstract topics, provided they are within my terms of knowledge, my interests, and my experience, although I sometimes cannot contribute to discussions between native speakers" may have

confounded certain respondents.

6. Conclusion

This study offers introductory and narrow results on the CEFR-J can-do statements and measurement methods. While the consistency of participants' self-ratings among the five skill groups was statistically reliable as measured by the Cronbach alpha when the raw responses of individual students was compared with each skill category, there were a number of areas where the CEFR-J scales appeared to require additional investigation as to appropriateness and reliability. The findings reveal, as measured by the Mokken scale h-coefficient, that internal reliability was rather weak; possible explanations for this weakness were commented on and further research on CEFR-J can-do statements can take measured steps to possibly decrease self-rating reliability concerns. Further studies investigating the internal reliability of the CEFR-J statements should be undertaken, noting one particular weakness of this present study of using binary option responses; employing a Likert-scale format may reduce low Mokken scale coefficient results. In addition, the limitations of self-rating, notably with Japanese university students' perceptions of their English skill, and the possible gap with actual proficiency as measured on the CEFR-J requires more examination. One possible method to link participants' self-assessments with more objective proficiency could be analyzing these self-rating data points alongside KET scores (synced with the original CEFR, not the CEFR-J).

文　　献
[1] European Framework of Reference for Languages, Common European Framework of Reference: Learning, teaching, assessment (CEFR). Retrieved August 2, 2017 from https://www.coe.int/en/web/common-european-framework-reference-languages/
[2] Council of Europe, Common European framework of reference for languages: Learning, teaching, assessment. Strasbourg: Council of Europe / Cambridge University Press, 2011.
[3] N. Nagai, and F. O'Dwyer, The actual and potential impacts of the CEFR on language

education in Japan, Synergies Europe 6, pp.141-152, 2011.

[4] Y.Tono,and M. Negishi, M., The CEFR-J: Adapting the CEFR for English Language Teaching in Japan. Framework & Language Portfolio SIG Newsletter, 8, pp.5-12, 2012.

[5] Council of Europe, Common European framework of reference for languages: Learning, teaching, assessment. Strasbourg: Council of Europe / Cambridge University Press, p. 256, 2011.

[6] Council of Europe, Common European framework of reference for languages: Learning, teaching, assessment. Strasbourg: Council of Europe / Cambridge University Press, p. 244, 2011.

[7] Y.Tono and M. Negishi, The CEFR-J: Adapting the CEFR for English Language Teaching in Japan. Framework & Language Portfolio SIG Newsletter, 8, pp.5-12, 2012.

[8] N. Nagai,and F. O'Dwyer, The actual and potential impacts of the CEFR on language education in Japan, Synergies Europe 6, pp.141-152, 2011.

[9] M.G.Fennelly, The Influence of CEFR on English Education in Japan, Bulletin of Shikoku University, 46, pp. 109–122, 2016.

[10] Y.Tono,and M. Negishi, The CEFR-J: Adapting the CEFR for English Language Teaching in Japan. Framework & Language Portfolio SIG Newsletter, 8, pp.5-12, 2012.

[11] M. Negishi, The Development of the CEFR-J: Where we are, where we are going, in New perspectives for foreign language teaching in higher education: Exploring the possibilities of application of CEFR, ed. N. Tomimori, M. Furihatat, K. Haida, N. Kurosawa, & M. Negishi, pp. 105-116, Tokyo WOLSEC, Tokyo University of Foreign Studies, 2012.

[12] M. Negishi, M,. CEFR-J Kaihatsu no Keii [The Development Process of the CEFR-J], ARCLE Review, vol. 5, issue 3, 37-52, 2011.

[13] Oskarsson, M., Approaches to self-assessment in foreign language learning, London: Pergamon, 1980.

[14] D.Dunning, C. Heat and J.M. Suls, Flawed Self-Assessment. Psychological Science in the Public Interest,5 (3), pp. 69-106. doi:10.1111/j.1529-1006.2004.00018.x, 2004.

[15] D.Gardner and L. Miller, L., Establishing Self-Access: From Theory to Practice. Cambridge University Press, Cambridge, 1999.

[16] P. Blanche and B.J. Merino, Self-assessment of foreign- language skills: Implications for teachers and researchers. Language Learning,39 (3), pp. 313-338. doi:10.1111/j.1467-1770.1989.tb00595.x, 1989.

[17] S.Ross, Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing,15*(1), pp.1-20. doi:10.1177/026553229801500101, 1998.

[18] J. Kruger and D.Dunning, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology,77(6), pp. 1121-1134. doi:10.1037//0022-3514.77.6.1121, 1999.

[19] S. Matsuno, Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing,26(1), pp. 75-100. doi:10.1177/0265532208097337, 1999.

[20] S. Matsuno, Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing,26(1), pp. 75-100. doi:10.1177/0265532208097337, 1999.

[21] K.Denies and R. Janssen, Country and gender differences in the functioning of CEFR-based can-do statements as a tool for self-assessing English proficiency. Language Assessment Quarterly,*13* (3), pp. 251-276. doi:10.1080/15434303.2016.1212055, 2016.

[22] F. Parajes, Gender and perceived self-efficacy in self-regulated learning. Theory Into Practice, 41(2), pp. 116-125. http://dx.doi.org/10.1207/s15430421tip410 2_8, 2002.

[23] L.F. Bachman and A.S.Palmer,The construct validation of self-ratings of communicative language ability. Language Testing,6 (1), pp.14-29. doi:10.1177/026553228900600104, 1999.

[24] J.A.Ross, The reliability, validity, and utility of self-assessment. Practical Assessment, Research, and Evaluation, 11(10), pp. 1-13, 2006.

[25] K.Williams and M.Andrade,Foreign Language Learning Anxiety in Japanese EFL University Classes: Causes, Coping, and Locus of Control, Electronic Journal of Foreign Language Teaching, 5(2), pp.181–191, 2008.

[26] M. Tokeshi and L. Gao,CEFR-J Based Survey for Japanese University Students, The European Conference on Education 2015 Official Conference Proceedings, pp. 433-447, 2015.

[27] M. Tokeshi and L. Gao,CEFR-J Based Survey for Japanese University Students, The European Conference on Education 2015 Official Conference Proceedings, pp. 433-447, 2015.

[28] J. Runnels, An exploratory reliability and content analysis of the CEFR-Japan's A-Level can-do statements, JALT Journal 36(1), pp. 69-89, 2014.

Appendix 1

CEFR-J Can-do questionnaire (Japanese version)

(Retrieved from http://www.tufs.ac.jp/ts/personal/tonolab/cefr-j/english/download.html)

1. 趣味やスポーツ、部活動などの身近なトピックにする短い話を、ゆっくりはっきりと話されれば、理解することができる。

2. ゆっくりはっきりと放送されれば、公共の乗り物や駅や空港の短い簡潔なアナウンスを　理解することができる。

3. 外国の行事や習慣などに関する説明の概要を、ゆっくりはっきりと話されれば、理解することができる。

4. 自然な速さの標準的な英語で話されていれば、テレビ番組や映画の母語話者同士の会話の要点を理解できる。

5. 日常生活の身近なトピックについての話を、ゆっくりはっきりと話されれば、場所や時間等の具体的な情報を聞きとることができる。

6. 学校の宿題、旅行の日程などの明確で具体的な事実を、はっきりとなじみのある発音で指示されれば、要点を理解することができる。

7．自分の周りで話されている少し長めの議論でも、はっきりとなじみのある発音であれば、その要点を理解することができる。

8．トピックが身近であれば、長い話や複雑な議論の流れを理解することができる。

9．簡単なポスターや招待状等の日常生活で使われる非常に短い簡単な文章を読み、理解することができる。

10．簡単な語を用いて書かれた人物描写、場所の説明、日常生活や文化の紹介などの、説明文を理解することができる。

11．学習を目的として書かれた新聞や雑誌の記事の要点を理解することができる。

12．現代の問題など一般的関心の高いトピックを扱った文章を、辞書を使わずに読み、複数の視点の相違点や共通点を比較しながら読むことができる。

13．身近な人からの携帯メールなどによる、旅の思い出などが書かれた非常に短い簡単な近況報告を理解することができる。

14．簡単な語を用いて書かれた短い物語や伝記などを理解することができる。

15．ゲームのやり方、申込書の記入のしかた、ものの組み立て方など、簡潔に書かれた手順を理解することができる。

16．難しい部分を読み返すことができれば、自分の専門分野の報告書・仕様書・操作マニュアルなどを、詳細に理解することができる。

17．基本的な語や言い回しを使って日常のやりとり（何ができるかできないかや色についてのやりとりなど）において単純に応答することができる。

18．順序を表す表現である first, then, next などのつなぎ言葉や「右に曲がって」や「まっすぐ行って」などの基本的な表現を使って、単純な道案内をすることができる。

19．身近なトピック（学校・趣味・将来の希望）について、簡単な英語を幅広く使って意見を表明し、情報を交換することができる。

20．ある程度なじみのあるトピックならば、新聞・インターネットで読んだり、テレビで見たニュースの要点について議論することができる。

21．スポーツや食べ物などの好き嫌いなどのとてもなじみのあるトピックに関して、はっきり話されれば限られたレパートリーを使って、簡単な意見交換をすることができる。

22．補助となる絵やものを用いて、基本的な情報を伝え、また、簡単な意見交換をすることができる。

23．個人的に関心のある具体的なトピックについて、簡単な英語を多様に用いて、社交的な会話を続けることができる。

24．母語話者同士の議論に加われないこともあるが、自分が学んだトピックや自分の興味や経験の範囲内のトピックなら、抽象的なトピックであっても、議論できる。

25．前もって発話することを用意した上で、限られた

身近なトピックについて、簡単な語や基礎的な句を限られた構文を用い、簡単な意見を言うことができる。

26．一連の簡単な語句や文を使って、自分の趣味や特技に触れながら自己紹介をすることができる。

27．使える語句や表現を繋いで、自分の経験や夢、希望を順序だて、話しを広げながら、ある程度詳しく語ることができる。

28．ある視点に賛成または反対の理由や代替案などをあげて、事前に用意されたプレゼンテーションを聴衆の前で流暢に行うことができ、一連の質問にもある程度流暢に対応ができる。

29．前もって発話することを用意した上で、日常生活の物事を、簡単な語や基礎的な句を限られた構文を用い、簡単に描写することができる。

30．写真や絵、地図などの視覚的補助を利用しながら、一連の簡単な句や文を使って、身近なトピック（学校や地域など）について短い話をすることができる。

31．自分の考えを事前に準備して、メモの助けがあれば、聞き手を混乱させないように、馴染みのあるトピックや自分に関心のある事柄について語ることができる。

32．ディベートなどで、そのトピックが関心のある分野のものであれば、論拠を並べ自分の主張を明確に述べることができる。

33．簡単な語や基礎的な表現を用いて、身近なこと（好き嫌い、家族、学校生活など）について短い文章を書くことができる。

34．日常的・個人的な内容であれば、招待状、私的な手紙、メモ、メッセージなどを簡単な英語で書くことができる。

35．自分に直接関わりのある環境（学校、職場、地域など）での出来事を、身近な状況で使われる語彙・文法を用いて、ある程度まとまりのあるかたちで、描写することができる。

36．自分の専門分野であれば、メールやファックス、ビジネス・レターなどのビジネス文書を、感情の度合いをある程度含め、かつ用途に合った適切な文体で、書くことができる。

37．簡単な語や基礎的な表現を用いて、メッセージカード（誕生日カードなど）や身近な事柄についての短いメモなどを書ける。

38．文と文を and, but, because などの簡単な接続詞でつなげるような書き方であれば、基礎的・具体的な語彙、簡単な句や文を使った簡単な英語で、日記や写真、事物の説明文などのまとまりのある文章を書くことができる。

39．身近な状況で使われる語彙・文法を用いれば、筋道を立てて、作業の手順などを示す説明文を書くことができる。

40．旅行記や自分史、身近なエピソードなどの物語文について何か自分が知っていれば、多くの情報源から統合して情報や議論を整理しながら、それに対する自分の考えの根拠を示しつつ、ある程度の結束性のあるエッセイやレポートなどを、幅広い語彙や複雑な文構造をある程度使って、書くことができる。

Appendix 2 CEFR-J Can-do questionnaire

1. I can understand short conversations about familiar topics (e.g. hobbies, sports, club activities), provided they are delivered in slow and clear speech.

2. I can understand short, simple announcements e.g. on public transport or in stations or airports, provided they are delivered slowly and clearly.

3. I can understand the gist of explanations of cultural practices and customs that are unfamiliar to me, provided they are delivered in slow and clear speech involving rephrasing and repetition.

4. I can understand the main points of a conversation between native speakers in television programmes and in films, provided they are delivered at normal speed and in standard English.

5. I can catch concrete information (e.g. places and times) on familiar topics encountered in everyday life, provided it is delivered in slow and clear speech.

6. I can understand the main points of straightforward factual messages (e.g. a school assignment, a travel itinerary), provided speech is clearly articulated in a familiar accent.

7. I can understand the main points of extended discussions around me, provided speech is clearly articulated and in a familiar accent.

8. I can follow extended speech and complex lines of argument provided the topic is reasonably familiar.

9. I can understand very short, simple, everyday texts, such as simple posters and invitation cards.

10. I can understand explanatory texts describing people, places, everyday life, and   culture, etc., written in simple words.

11. I can understand the main points of English newspaper and magazine articles adapted for educational purposes.

12. I can read texts dealing with topics of general interest, such as current affairs, without consulting a dictionary, and can compare differences and similarities between multiple points of view.

13. I can understand very short reports of recent events such as text messages from friends' or relatives', describing travel memories, etc.

14. I can understand short narratives and biographies written in simple words.

15. I can understand clearly written instructions (e.g. for playing games, for filling in a form, for assembling things).

16. I can understand in detail specifications, instruction manuals, or reports written   for my own field of work, provided I can reread difficult sections.

17. I can respond simply in basic, everyday interactions such as talking about what I can/cannot do or describing colour, using a limited repertoire of expressions.

18. I can give simple directions from place to place, using basic expressions such as "turn right" and "go straight" along with sequencers such as first, then, and next.

19. I can express opinions and exchange information about familiar topics (e.g. school, hobbies, hopes for the future), using a wide range of simple English.

20. I can discuss the main points of news stories I have read about in the newspapers/ on the internet or watched on TV, provided the topic is reasonably familiar to me.

21. I can exchange simple opinions about very familiar topics such as likes and for sports, foods, etc., using a limited repertoire of expressions, provided that people speak clearly.

22. I can get across basic information and exchange simple opinions, using pictures or objects to help me.

23. I can maintain a social conversation about concrete topics of personal interest, using a wide range of simple English.

24. I can discuss abstract topics, provided they are within my terms of knowledge, interests, and my experience, although I sometimes cannot contribute to discussions between native speakers.

25. I can express simple opinions related to limited, familiar topics, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.

26. I can introduce myself including my hobbies and abilities, using a series of simple phrases and sentences.

27. I can talk in some detail about my experiences, hopes and dreams, expanding on what I say by joining together words, phrases and expressions I can readily use to make longer contributions.

28. I can give a prepared presentation with reasonable fluency, stating reasons for agreement or disagreement or alternative proposals, and can answer a series of questions.

29. I can give simple descriptions e.g. of everyday object, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.

30. I can give a brief talk about familiar topics (e.g. my school and my neighborhood) by visual aids such as photos, pictures, and maps, using a series of simple phrases and sentences.

31. I can talk about familiar topics and other topics of personal interest, without causing confusion to the listeners, provided I can prepare my ideas in advance and use brief notes to help me.

32. I can develop an argument clearly in a debate by

providing evidence, provided the topic is of personal interest.

33. I can write short texts about matters of personal relevance (e.g. likes and dislikes, , and school life), using simple words and basic expressions.

34. I can write invitations, personal letters, memos, and messages, in simple English, provided they are about routine, personal matters.

35. I can write a description of substantial length about events taking place in my immediate environment (e.g. school, workplace, local area), using familiar vocabulary and grammar.

36. I can write business documents (e.g. e‑mail, fax, business letters), conveying degrees of emotion, in a style appropriate to the purpose, provided they are in my professional field.

37. I can write message cards (e.g. birthday cards) and short memos about events of personal relevance, using simple words and basic expressions.

38. I can write texts of some length (e.g. diary entries, explanations of photos and events) in simple English, using basic, concrete vocabulary and simple phrases and sentences, linking sentences with simple connectives like and, but , and because .

39. I can write coherent instructions telling people how to do things, with vocabulary and grammar of immediate relevance.

40. I can write reasonably coherent essays and reports using a wide range of vocabulary and complex sentence structures, synthesising information and arguments from a number of sources, provided I know something about the topics.