# Perceptions of Japanese EFL Student Oral Language Ability: Learner Self-Assessment versus Interviewer Assessment Using CEFR Descriptors

Nicolangelo Becce, Christopher Hennessy

## Abstract

Each April from 2015 to 2017, the authors, in conjunction with other faculty members at a Japanese national university, conducted five-minute oral interviews based on the Common European Framework of Reference for Languages (CEFR) with over 800 first-year students each year for English class placement. This CEFR-based interview system was primarily developed by the authors, utilizing the CEFR descriptors for "Listening," "Spoken Interaction," and "Spoken Production." As an additional part of this interview process, participating students were asked to self-assess their own English language ability through a questionnaire utilizing descriptors from the A1 Level to the C2 level, also in the categories of "Listening," "Spoken Interaction," and "Spoken Production." In the first part of this paper, the authors will explain the rationale for using the CEFR and CEFR-J (Common European Framework of Reference for Languages － Japan) as the framework for developing this interview system, give a brief synopsis of the interview system in general, and describe the CEFR self-assessment questionnaire utilized. In the second part, the authors will give an analysis of Japanese EFL learners' own CEFR self-assessment versus the CEFR profile given by interviewers through the interview process. The results as a whole suggest that, overall, the Japanese EFL learners involved in this study assessed "Listening," "Spoken Interaction," and "Spoken Production" higher than the CEFR profiles they received through the interview process, despite general claims about Japanese language learners' tendency toward "modesty" in describing their own language abilities. In the final section, the paper offers an overview of issues surrounding the development of English interview protocols - including question creation and norming issues - and future research planned by the authors on the oral interview protocol system.

**Keywords**: EFL, CEFR, CEFR-J, oral placement interview, Japanese higher education.

## 1. Introduction

In 2013, the Japanese national university where this research was conducted reformed its English language curriculum after having become one of the funding recipients of the MEXT Global 30 Plus program. This program, supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), aims at promoting the internationalization of the academic environment of Japanese universities. During the academic years 2014-2015, 2015-2016, and 2016-2017, this new language program developed through MEXT Global 30 Plus involved 16 full-time instructors and around 1,600 first- and second-year students of the faculties of Engineering, Medical Sciences, and Education and Regional Studies. The students were divided into 67 English language classes of 24 students each that met twice, on average, a week for 90 minutes. For each of the academic years, the curriculum started with a focus on personal communication, gradually shifting toward English for professional communication and TOEIC (Test of English for International Communication) Listening and Reading test preparation.

Since, in compliance with the MEXT requirements, this English language program had specific goals in terms of average TOEIC scores among the student population, the TOEIC test was also used during the academic years 2013-2014 and 2014-2015 for placement purposes. But, at the same time, given that the first part of the language curriculum started with a focus on communication-based activities, the authors developed a new placement system that could place students according to their spoken English level, as well as an activity capable of producing a positive affective outcome in the interviewees. In order to develop a system that could be valid on an international level, but that at the same time could be flexible enough to be tailored to the specific needs of the student population, the authors decided to follow the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) and its Japanese counterpart, the CEFR-J. Indeed, the CEFR is being integrated into English education in Japan by MEXT from 2020 (MEXT, 2015; "New Standardized College," 2017) and the CEFR-J project is now actively contributing in the updating process of the CEFR (Council of Europe, 2017).

Besides for developing an entire interview system based on CEFR and CEFR-J, the authors also had all 1,600-plus participating students of the interview sessions in April 2016 and 2017 complete language self-assessment questionnaires based on "Listening," "Spoken Interaction," and "Spoken Production." In this paper, the authors will show the results from comparing the CEFR profile given to students by their interviewer versus their own perceived CEFR profile based on their own self-assessment. In order to show this analysis, the authors will first describe the CEFR and CEFR-J and the rationale for using this as a framework, and then describe the CEFR self-assessment questionnaire given to students. They will then lay out the methodology for this comparison, followed by the results and discussion. They will conclude with future directions of their research and the interview process as a whole.

## 1-1 The CEFR and CEFR-J

The Common European Framework of Reference for Languages (CEFR) is a language framework developed by the Council of Europe as a method of learning and assessing language use in Europe. The CEFR divides learners into 6 levels (A1, A2, B1, B2, C1, C2) according to their reading, listening, speaking and writing abilities. A language-independent framework by definition, the CEFR has been adapted to the specific needs of the English language teaching contexts in Japan by Tono and Negishi (2012), who developed a new framework, the CEFR-J (Common European Framework of Reference for Languages － Japan).

The CEFR-J refines the "can do" descriptors developed by the Council of Europe, and further divides the original framework into 12 levels (Pre-A1; A1.1, A1.2, A1.3; A2.1, A2.2; B1.1, B1.2; B2.1, B2.2; C1; C2). Moreover, the CEFR-J introduces a wordlist of 5,639 words, covering the levels from Pre-A1 to B2. This wordlist is based on the analysis of major English textbooks used in Asian regions (Tono & Negishi, 2012). Both the CEFR and CEFR-J "can do" descriptors for speaking were used as a starting point for the creation of a set of new descriptors for this interview protocol.

The authors chose to use the CEFR and CEFR-J as frameworks for creating this interview system for three reasons. First and foremost, as mentioned above, CEFR has been set by MEXT to be integrated into English education in Japan by 2020 (MEXT 2015, "New Standardized College," 2017). Next, with this reform, Tono (2017) has mentioned the timeliness of any research activity connected to CEFR. Third, Tono (投野 2013) further described CEFR as offering a framework to shed light on utilizing both language communication ability and general everyday activities, which was part of the design of the personal communication-based English language program at the university in which this interview system was designed.

## 1-2 The Interview System

In order to conduct over 800-plus interviews over one week with limited personnel, the authors decided that interviews could last no longer than five minutes. Given the time constraints, in terms of both general preparation and norming process for the interviewers as well as in terms of available time to concretely interview the students, the authors decided to develop an interview system using the following "can do" descriptors:

A1 － I can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.

A2 － I can give simple descriptions of things and make comparisons. I can describe past activities and personal experiences.

B1 － I can explain and give reasons for my plans, intentions and actions.

B2 － I can develop an argument well enough to be followed without difficulty most of the

time. I can speculate about causes, consequences and hypothetical situations.

The A1 level is preceded by a Pre-A1 level that does not currently have any "can do" descriptor, given that the authors decided to define it as the result of the interview result in case the interviewee performance was not sufficient to be considered at the A1 level. Apart from the Pre-A1 level, the four "can do" descriptors here reported follow the original CEFR scale model, except that they cover only the first four original levels (A1-B2). This choice has been made for two main reasons: the first being, the authors expected the student population to be composed, in most cases, of A and B-level students, with only a limited number of C-level students based on previous TOEIC scores by the student populations as well as the understanding of the CEFR level system and the level of textbooks used by the students of that specific Japanese national university. Consequently, the authors assumed that, for the expected student population, regrouping the levels B2, C1 and C2 into "B2+" would not have affected, in a determinant way, the grouping of students in different classes. The second reason is directly related to time constraint; as explained above, each interview had to be contained within five minutes. Since the interview protocol workflow (see Figure I) always starts from the lowest levels (from Pre-A1 to B2+), the choice to avoid the assessment of the interviewee performance beyond the B2 level was believed to help the interviewers stay within this five-minute time limit.
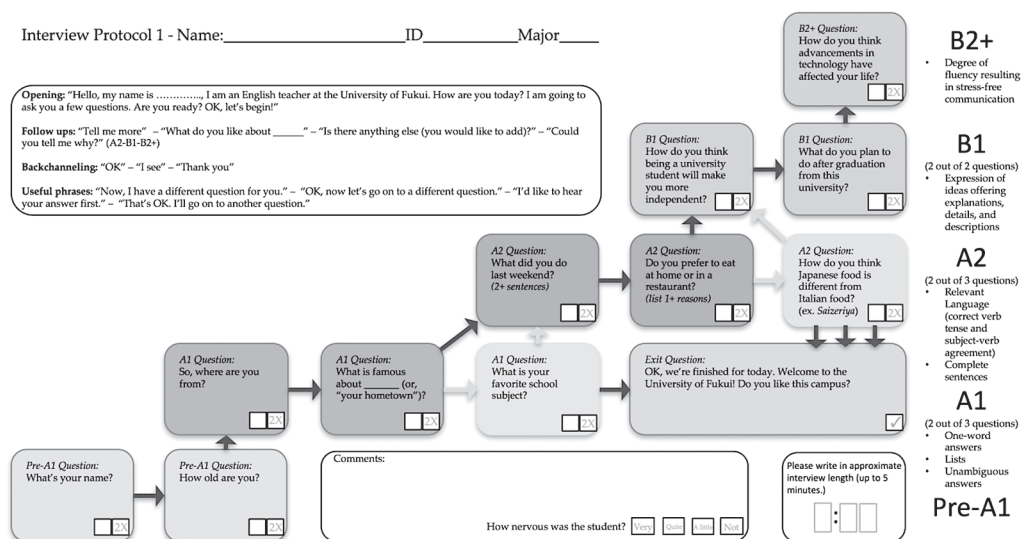


*Figure I.* Sample Interview Protocol

## 1-3 The CEFR Self-Assessment

Before the actual interview, students were explicitly told that they would have a short interview with a Language Center instructor. In addition, they were told that the purpose of the interview was to see how they could speak English, and that during the interviews they should have relaxed, and tried to speak as much as they could. Moreover, students received a self-assessment form in which they were supposed to circle all the descriptors that described their perceived levels of English language proficiency. Figure II represents an adapted version of the self-assessment grid included in the Common European Framework of Reference for Languages and translated by the Goethe-Institut of Tokyo (吉島 2004).

英語の自己評価　　　　　自分に当てはまるもの全てを○で囲んでください。

氏名＿＿＿＿＿＿＿＿　　　　学科＿＿＿＿＿＿＿＿　　　　学籍番号＿＿＿＿＿＿＿

| | | | | | | |
|---|---|---|---|---|---|---|
| 聞くこと | はっきりとゆっくりと話してもらえれば、自分、家族、すぐ周りの具体的なものに関する聞き慣れた語やごく基本的な表現を聞き取れる。 | （ごく基本的な個人や家族の情報、買い物、近所、仕事などの）直接自分につながりのある領域で最も頻繁に使われる語彙や表現を理解することができる。短い、はっきりとした簡単なメッセージやアナウンスの要点を聞き取れる。 | 仕事、学校、娯楽で普段出会うような身近な話題について、明瞭で標準的な話し方の会話なら要点を理解することができる。話し方が比較的ゆっくり、はっきりとしているなら、時事問題や、個人的もしくは仕事上の話題についても、ラジオやテレビ番組の要点を理解することができる。 | 長い会話や講義を理解することができる。また、もし話題がある程度身近な範囲であれば、議論の流れが複雑であっても理解できる。たいていのテレビのニュースや時事問題の番組も分かる。標準語の映画なら大多数は理解できる。 | たとえ構成がはっきりしなくて、関係性が暗示されているにすぎず、明示的でない場合でも、長い話が理解できる。特別の努力なしにテレビ番組や映画を理解できる。 | 生であれ、放送されたものであれ、母語話者の速いスピードで話されても、その話し方の癖に慣れる時間の余裕があれば、どんな種類の言葉も難無く理解できる。 |
| やり取り | 相手がゆっくり話し、繰り返したり、言い換えたりしてくれて、また自分が言いたいことを表現するのに助け舟を出してくれるなら、簡単なやり取りをすることができる。直接必要なことやごく身近な話題についての簡単な質問なら、聞いたり答えたりできる。 | 単純な日常の仕事の中で、情報の直接のやり取りが必要ならば、身近な話題や活動について話し合いができる。通常は会話を続けていくだけの理解力はないのだが、短い社交的なやり取りをすることはできる。 | 当該言語圏の旅行中に最も起こりやすいたいていの状況に対処することができる。例えば、家族や趣味、仕事、旅行、最近の出来事など、日常生活に直接関係のあることや個人的な関心事について、準備なしで会話に入ることができる。 | 流暢に自然に会話をすることができ、母語話者と普通にやり取りができる。身近なコンテクストの議論に積極的に参加し、自分の意見を説明し、弁明できる。 | 言葉をことさら探さずに流暢に社会上、仕事上の目的に合った言葉遣いが、意のままに効果的にできる。自分の考えや意見を精確に表現でき、自分の発言を上手に他の話し手の発言にあわせることができる。 | 慣用表現、口語体表現をよく知っていて、いかなる会話や議論でも努力しないで加わることができる。自分を流暢に表現し、詳細に細かい意味のニュアンスを伝えることができる。表現上の困難に出会っても、周りの人がそれにほとんど気がつかないほどに修正し、うまく繕うことができる。 |
| 表現 | どこに住んでいるか、また、知っている人たちについて、簡単な語句や文を使って表現できる。 | 家族、周囲の人々、居住条件、学歴、職歴を簡単な言葉で一連の語句や文を使って説明できる。 | 簡単な方法で語句をつないで、自分の経験や出来事、夢や希望、野心を語ることができる。意見や計画に対する理由や説明を簡潔に示すことができる。物語を語ったり、本や映画のあらすじを話し、またそれに対する感想・考えを表現できる。 | 自分の興味関心のある分野に関連する限り、幅広い話題について、明瞭で詳細な説明をすることができる。時事問題について、いろいろな可能性の長所、短所を示して自己の見方を説明できる。 | 複雑な話題を、派生的な問題にも立ち入って、詳しく論ずることができ、一定の観点を展開しながら、適切な結論までまとめ上げることができる。 | 状況にあった文体で、はっきりとすらすらと流暢に記述や論述ができる。効果的な論理構成によって聞き手に重要点を把握させ、記憶にとどめさせることができる。 |

*Figure II.* Self-Assessment Form

In this table, only the descriptors related to the categories of "Listening," "Spoken Interaction," and "Spoken Production" are maintained. The purpose of this self-assessment form was to give interviewees an idea of their language abilities that did not focus on the pernicious dichotomy "wrong versus correct answer," but rather, following the principles of the Common European Framework of Reference, it gave them a more empowering overview of the goals that they could achieve in terms of communication in a second language.

Apart from the pedagogical purpose implicit in the activity of self-assessing their language proficiency through CEFR Can-Do descriptors, this self-assessment activity gave the authors a chance to compare the students' perceived language abilities with the interview profiles assessed by the interviewers.

## 2. Methodology

　　The authors sought to compare the results of the interview profiles given to students by the interviewers against the results of students' own CEFR self-assessment. This comparison is based on a simple numerical analysis of interview profiles versus self-assessment profiles, in which A1 is assigned numerical value 1, A2 is 2, B1 is 3, B2 is 4, C1 is 5, and C2 is 6. All student profiles － both the interview and self-assessment － were transferred into numerical values based on this scale through Microsoft Excel. The results will be described in the next section.
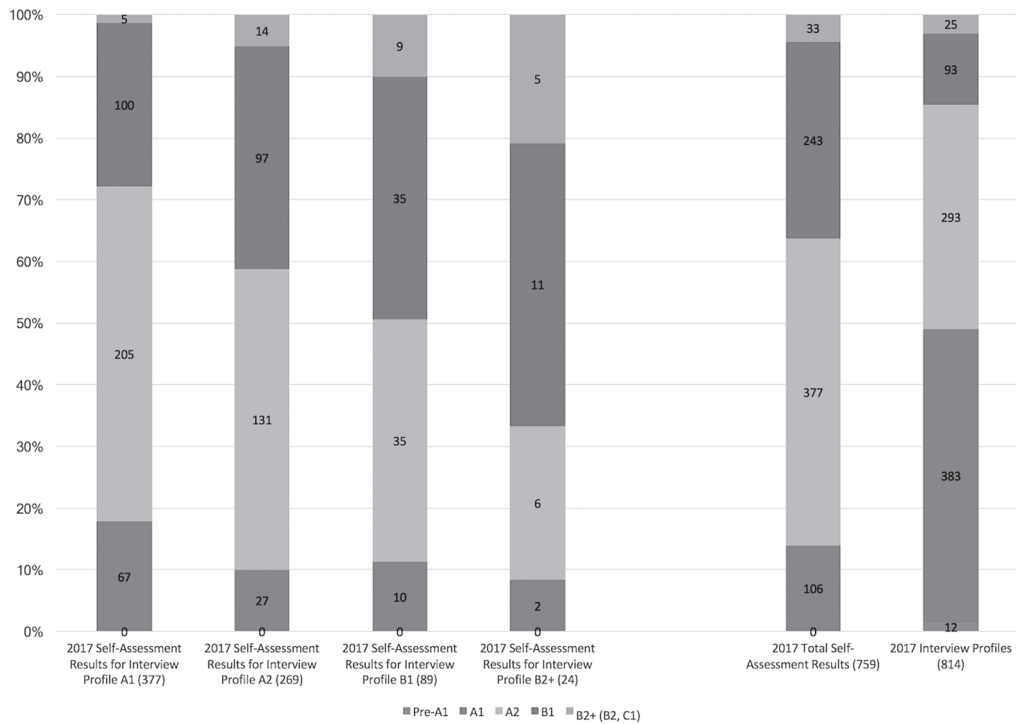


*Figure III.* Comparison Between Interview Profiles and Self-Assessment Results － 2017

## 3. Findings

　　Figure III represents a comparison between the self-assessment results and the interview profiles for the academic year 2017-2018. The first four columns on the left side of the figure show the results of the self-assessment profiles divided into interview profiles. After the authors collected the self-assessment results, they calculated the average between listening, spoken interaction and spoken production of each interviewee. Consequently, they divided all the self-assessment results in four groups, according to the interview results, namely A1, A2, B1, B2+ (which means B2, C1 and C2 combined). Since the original CEFR Self-Assessment Grid does not provide

any descriptors for the Pre-A1 level, the authors decided not to consider all the forms resulting in the Pre-A1 range, because they thought it was possible that some students might have misunderstood the instructions, thus leaving the self-assessment form blank and incomplete.

All the CEFR levels (Pre-A1, A1, A2, B1, B2+) are represented in different shades of gray. The four columns on the left side of the figure seem to show a clear pattern. In fact, it seems that there is a certain degree of consistency between the self-assessment results and the interview profiles, which means that, from the interview profiles A1 to those in the B2+ range, students seemed to gradually self-assess themselves at higher levels of language proficiency.

The right side of the figure shows a clear comparison between the total self-assessment results (second column from the right) and the interview results (first column from the right). What seems to be clear from the interview results is that, in accordance with Tono and Negishi (2012), more than eighty percent of the students were assessed either A1 or A2, and only a limited number of students were assessed at the B1 level or higher (it might be useful to add that very similar results were also obtained in 2015 and 2016). On the other hand, there is an evident discrepancy between the interview results and the self-assessment results, given that less than 65% of the students self-assessed themselves either A1 or A2 level, and roughly 30% of them self-assessed themselves at the B1 level. Considering that this self-assessment procedure was imple-
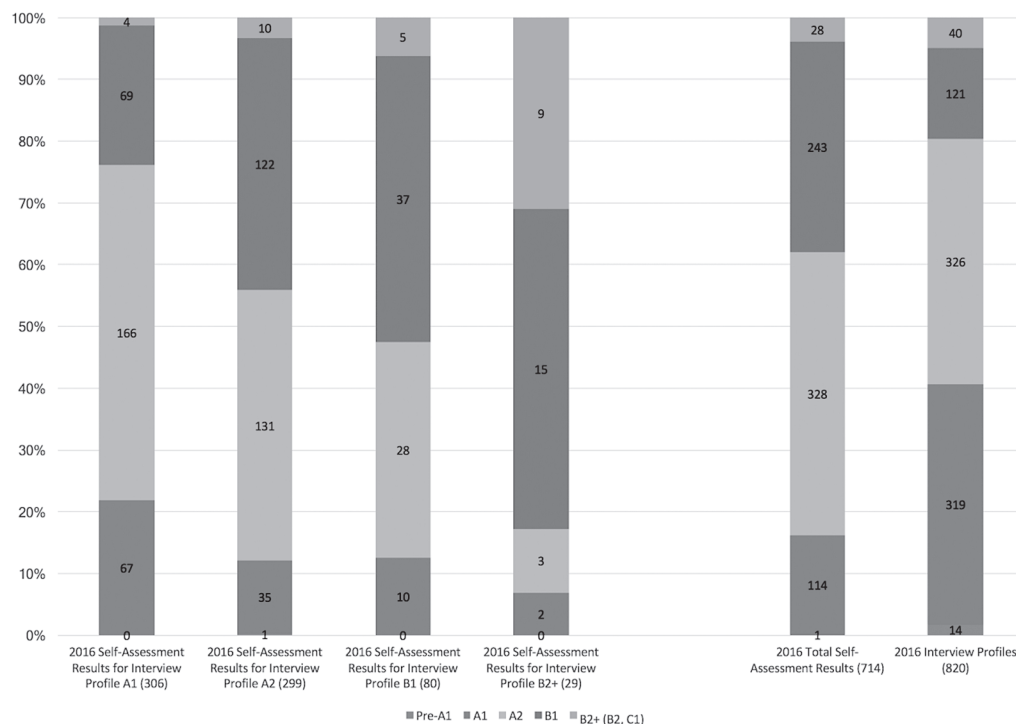


*Figure IV.* Comparison Between Interview Profiles and Self-Assessment Results - 2016

mented in 2016, it is important to note that in both 2016-2017 and 2017-2018 the results of both self-assessment and interviews are quite similar. Figure IV shows the same type of comparison chart between self-assessment results and interview results, this time in relation to the academic year 2016-2017.

## 4. Discussion

The data collected give the opportunity to make a few considerations. As Runnels (2013) pointed out, it seems to be normally assumed that, "for self-assessment surveys in particular, [Japanese survey-takers are] subject to Japanese cultural factors related to modesty" (p. 5; see also Matsuno, 2009). Yet, the data seem to show that students, at least in this specific case, might have been more lenient toward their listening and speaking skills compared to the results of the placement interviews. It is important to point out that the self-assessment form was only related to listening, spoken interaction and spoken production, and it did not have any descriptors for reading and writing skills, meaning that, if students were asked to assess their reading and writing proficiency as well, the self-assessment results could have been different. What seems to be sure from the self-assessment results is that, as can be noticed in figures V-VI, by considering A1
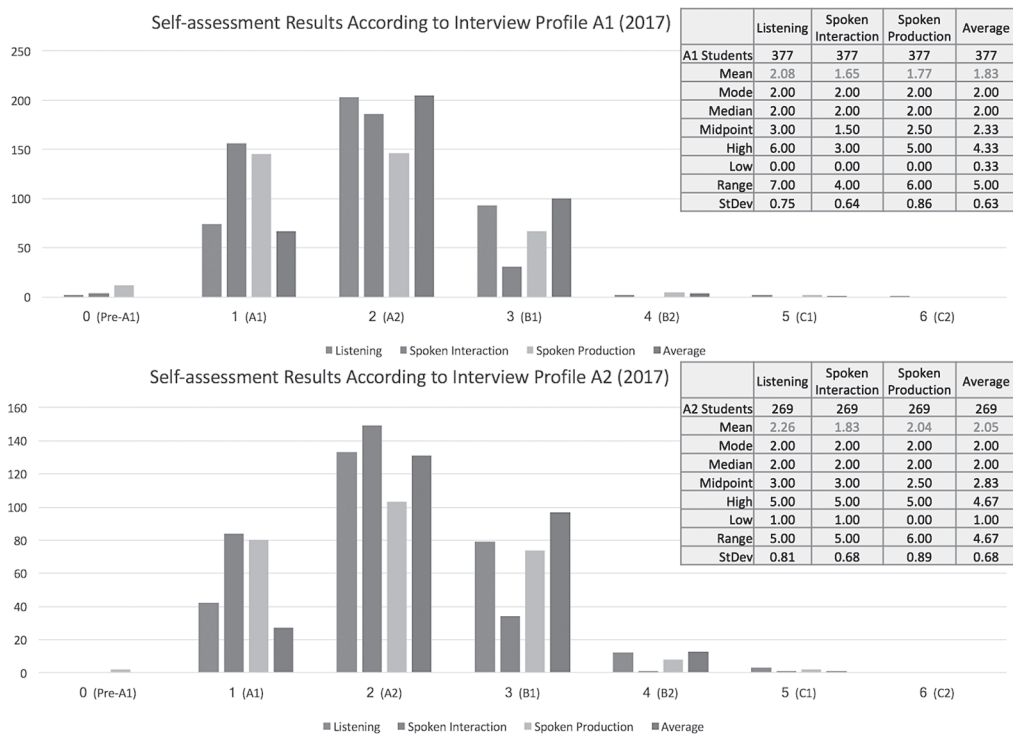


*Figure V.* Self-Assessment Results According to Interview Profiles A1-A2

**Self-assessment Results According to Interview Profile B1 (2017)**

|  | Listening | Spoken Interaction | Spoken Production | Average |
|---|---|---|---|---|
| B1 Students | 89 | 89 | 89 | 89 |
| Mean | 2.36 | 2.36 | 2.21 | 2.21 |
| Mode | 2.00 | 2.00 | 2.00 | 2.00 |
| Median | 2.00 | 2.00 | 2.00 | 2.00 |
| Midpoint | 3.00 | 3.00 | 3.00 | 2.67 |
| High | 5.00 | 5.00 | 5.00 | 4.33 |
| Low | 1.00 | 1.00 | 1.00 | 1.00 |
| Range | 5.00 | 5.00 | 5.00 | 4.33 |
| StDev | 0.86 | 0.86 | 0.98 | 0.80 |

■Listening ■Spoken Interaction ■Spoken Production ■Average

**Self-assessment Results According to Interview Profile B2+ (2017)**

|  | Listening | Spoken Interaction | Spoken Production | Average |
|---|---|---|---|---|
| B2+ Students | 24 | 24 | 24 | 24 |
| Mean | 2.79 | 2.38 | 2.54 | 2.57 |
| Mode | 3.00 | 3.00 | 2.00 | 2.00 |
| Median | 3.00 | 2.00 | 2.00 | 2.50 |
| Midpoint | 3.50 | 2.50 | 3.00 | 2.83 |
| High | 6.00 | 4.00 | 5.00 | 4.67 |
| Low | 1.00 | 1.00 | 1.00 | 1.00 |
| Range | 6.00 | 4.00 | 5.00 | 4.67 |
| StDev | 1.22 | 0.86 | 1.00 | 0.87 |

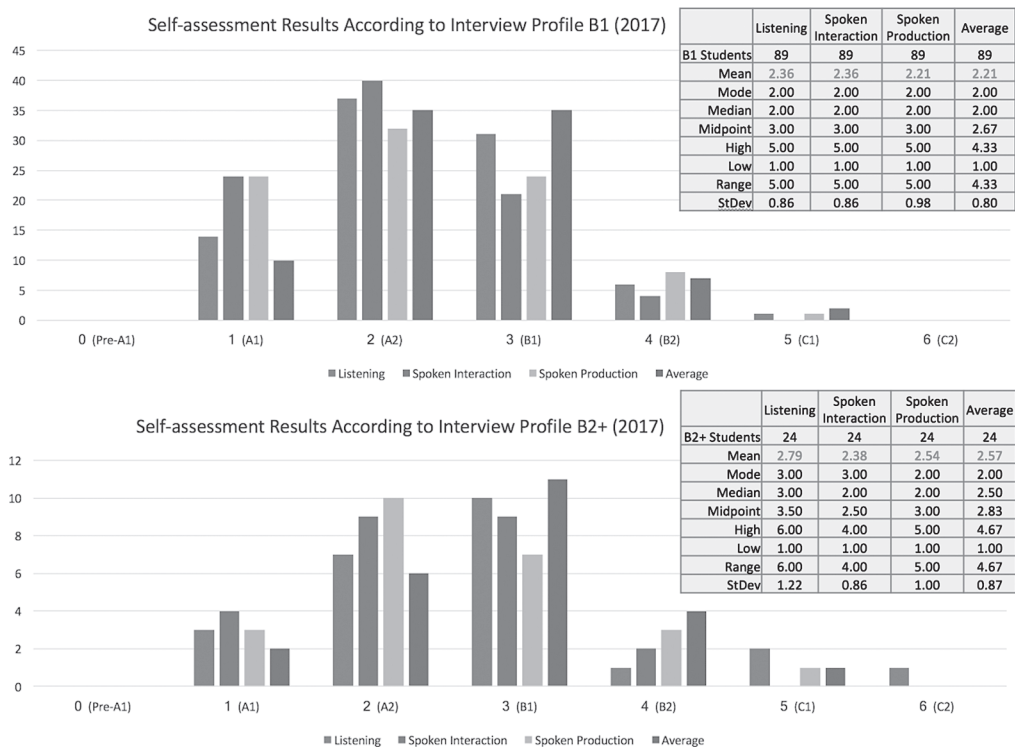■Listening ■Spoken Interaction ■Spoken Production ■Average

*Figure VI.* Self-Assessment Results According to Interview Profiles B1-B2+

=1, A2=2, B1=3 and so on, then the total average self-assessment result becomes 2.16 on a scale that goes from 1 to 6, which means that, at least in this specific case, the idea that Japanese survey-takers normally tend to choose neutral responses seems to be in contradiction with the data collected (see Runnels, 2013; Dörnyei & Taguchi, 2010).

On the other hand, the fact that a relatively limited number of students (at least in comparison with the interview results) chose A1 as well as B2 or higher levels could be related to the fact that low-achieving students might consider their proficiency level as higher than it really is, and, vice versa, high-achieving students might feel that their English proficiency is still too limited compared with their expectations. While this assumption is, at this stage of the research, nothing more than speculation, it might be useful to notice that the same trend was already observed with the self-assessment results of the academic year 2016-2017, and it might be interesting to compare the self-assessment results with other data collected, for example the nervousness level of the interviewees as perceived by the interviewers, or the answers collected through the post-interview questionnaires handed out to the interviewees after the interviews.

### 5. Conclusion

As already mentioned, further research needs to be done on many levels, for example in terms of error variance (Brown, 2005). In this sense, the researchers are trying to collect a number of video interviews from a variety of different student populations (for example, at the University of L'Aquila in Italy) in order to test the flexibility of the protocol system in different social, cultural, and linguistic contexts, and also to use Rasch analysis (Boone, Staver, & Yale, 2015) to test the norming process and study the protocol system from the interviewers' point of view.

Among the other aspects that the authors are planning to investigate, there is how to further streamline and simplify the norming session for the academic year 2018-2019 through a blind assessment based on developed and updated protocols, as well as practice interview sessions with local high schools in order to test the interview system at the secondary level of education. In addition, to further simplify the interview and assessment process for interviewers, a more detailed analysis of the CEFR descriptors will be conducted, and new ways to develop a flexible user interface for various contexts will be explored (for example, through the use of tablets or laptop screens to display the interview protocols rather than single sheets of A4 paper.) To study more in detail the interview system in affective terms for the interviewees, further breakdown and analysis of interviewee self-assessment and interview profile will be conducted, and, based on the questionnaires that interviewees compiled after the interviews conducted during the academic years 2015-2016 and 2016-2017, a qualitative analysis of the interview experience will be conducted.

### References

Boone, W. J., Staver, J. R., & Yale M. S. (2015) *Rasch analysis in the human sciences.* New York, NY: Springer.

Brown, J.D. (2005) *Testing in language programs: a comprehensive guide to English language assessment.* Upple Saddle River, NJ: Prentice Hall.

Council of Europe (2001) *The Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge, UK: Cambridge University Press.

Council of Europe (2017) *Common European Framework of Reference for Languages: learning, teaching, assessment. Companion volume with new descriptors. Provisional edition September 2017.* <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/168074a4e2> (Accessed November 27, 2017).

Dörnyei, Z., & Taguchi, T. (2010) *Questionnaires in second language research: Construction, administration, and processing.* New York, NY: Routledge.

Matsuno, S. (2009) Self-, peer-, and teacher-assessments in Japanese university EFL writing. *Language Testing,* 26 (1), 75-100.

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2002) *Developing a strategic plan to cultivate "Japanese with English abilities": Plan to improve English and Japanese abilities.*

<http://www.mext.go.jp/english/news/2002/07/020901.htm> (Accessed November 27, 2017).

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2015) 生徒の英語力向上推進プラン.

<http://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2015/07/21/1358906_01_1.pdf> (Accessed November 27, 2017).

New standardized college entrance exam to adjust English test format. (2017). *The Mainichi Shimbun.*

<https://mainichi.jp/english/articles/20170516/p2a/00m/0na/020000c> (Accessed November 27, 2017).

Runnels, J. (2013) Student ability, self-assessment, and teacher assessment on the CEFR-J's can-do statements. *The Language Teacher, 37.5, September/October 2013,* 3-7.

Tono, Y. (2017) The CEFR-J and its impact on English language teaching in Japan. *JACET International Convention. Selected Papers Vol. 4.*

<http://www.jacet.org/publication/selected-papers/past-issues/> (Accessed November 27, 2017).

Tono, Y. & Negishi, M. (2012) The CEFR-J: Adapting the CEFR for English language teaching in Japan. *JALT Framework & Language Portfolio SIG Newsletter 8,* 5-12.

投野由紀夫/編 (2013) 『英語到達度指標 CEFR－J ガイドブック CAN－DO リスト作成・活用』東京: 大修館書店.

吉島・大橋 (訳・編) (2004) 『ヨーロッパ言語共通参照枠組み(CEF)の最近の動向』Tokyo, JP: Goethe-Institut, 28-29.

<https://www.jpf.go.jp/j/publish/japanese/euro/pdf/01-3.pdf> (Accessed November 27, 2017).

EFL学生の会話力の認識について：学習者の自己評価対CEFRの基準を用いた面接者による評価

ベッチェ・ニコランジェロ，ヘネシー・クリストファー

　2015年から2017年の３年間、毎年４月に、筆者はその他教員陣と共に、日本の国立大学の800人を超える学部１年生を対象に、英語授業のクラス分けのため、ヨーロッパ共通参照枠（CEFR）に基づく５分間の面接試験を行った。このCEFRに基づく面接試験制度は、同フレームワークの「聴解」「対話」「会話生産」の基準を活用して主に筆者らが開発したものである。この面接試験過程の一環として、参加学生はアンケートへの回答により、自分の英語力について、「聴解」「対話」「会話生産」の各分野についてCEFRのレベルA1からC2までの自己評価を行った。当論文では、はじめに、CEFR及びCEFR－J（ヨーロッパ共通参照枠－日本）を当面接試験制度を開発するうえでのフレームワークとして活用した理論的根拠を説明する。そのうえで、面接試験制度の概要に触れ、また評価において使用したCEFR自己評価のためのアンケートについて紹介する。次に、日本のEFL学習者によるCEFR自己評価、及び面接試験を通じた面接者によるCEFRに基づく評価について比較分析する。全体的な結果として、この研究に参加した日本のEFL学習者は、「聴解」「対話」「会話生産」の全ての分野において、面接者による評価よりも自己評価結果の方が高かった。日本人の外国語学習者は、一般的に、自分の言語能力を「控えめに」表現する傾向にあると言われているにも関わらずである。最後に、設問や評価基準等を含む英語の面接試験を開発するうえでの諸問題について概要を示し、また口頭試験制度に関する将来的な研究計画について記す。