Title	CEFR Self-Assessment Checklists' Impact on a Japanese National University's English Program
Author(s)	Smith, F. Antonio
Citation	大阪大学英米研究. 2009, 33, p. 125-138
Version Type	VoR
URL	https://hdl.handle.net/11094/99333
rights	
Note	

CEFR Self-Assessment Checklists' Impact on a Japanese National University's English Program

Antonio F. Smith

1   Impact of CEFR Self-Assessment Checklist on the English Program

1.1   Introduction

This paper examines the impact on the English program created by the CEFR Self-Assessment Grid [1] and 'Self-assessment checklists from the Swiss version of the European Language Portfolio' [2] (hereafter, 'Checklist'), both available on the Council of Europe's European Language Portfolio web page. [3]

At the end of the academic year, in January 2008, I collected Checklist results from 2nd year English majors at the former OUFS, now the School of Foreign Studies of Osaka University, to estimate the degree to which CEFR scale targets were being met as well as areas of curriculum that could be improved. Checklist responses were converted to numerical values and processed with Excel. Results revealed students' perceived strengths and weaknesses, quantitatively, in terms of the five CEFR skill areas and individual Can-Do's. The average level was in keeping with the program's 2nd year CEFR goal, essentially B2. Results also indicated how curriculum could be improved with the aim of eventually introducing C1 in all five skill areas as official targets. By converting the Checklist into an automated format (WebCT trials conducted in January, 2009) the English program will be able to obtain Checklist

survey results at the beginning and end of each semester. Moreover, the WebCT program will include a Japanese translation of the Can-Do's that will allow other language programs and English sub-majors to use it easily.

1.2 English Area Studies: Program outline
1.2.1 Students' English education background:

English is part of the national curriculum for junior high and high school, so all students entering the program have at least six years of English education. In addition, many have studied English privately, beginning even in elementary school, and/or have lived abroad.

1.2.2 1st year English curriculum: 5 English classes [4]

1) 2 native-speaker English classes (from April 2008, 3 classes)

2) 3 English classes taught by Japanese (from April, 2008, 4 classes)

1.2.3 2nd year English curriculum: 5 English classes

1) 2 native classes

2) 3 English classes taught by Japanese

3) 730 TOEIC or better required (average score much higher)

1.2.4 3rd and 4th year, in total

1) 2 'content' classes by native speakers

2) Classes in Area Studies: many English texts; Japanese language of instruction

3) Graduation Paper in English

1.2.5 CEFR targets by year, at time of survey (*minimums* to be achieved)

1) 1st year: B1

2) 2nd year: B2 (B2+ for reading) [5]

3) 3rd & 4th year: C1 (hoped for; not officially set, yet)

2 How best to verify that targets are being achieved?

A. Cambridge ESOL tests? Perhaps, eventually. Cambridge tests are desirable because they are tied directly to the CEFR, but they are undesirable because they are not the standard in Japan—TOEIC is, meaning that companies, universities and students accept the necessity of TOEIC. Changing the status quo, and getting Japanese to *replace* their perceived need for TOEIC with a perceived need for Cambridge is very unlikely to happen quickly. Unfortunately, it should be difficult even to get students to *try* Cambridge in addition to TOEIC because of the cost in time and money. Most students cannot afford and/or would not like to take two standardized tests, so the English program cannot require both TOEIC and, for example IELTS, even though our official targets are in terms of the CEFR scales.

B. TOEIC/TOEFL? Not ideal. Equivalencies, though imperfect, do exist. Therefore, TOEIC *can* provide the program with at least some kind of impression of how close its students are coming to its CEFR targets. However, TOEIC does not show *directly* whether a certain CEFR level (A1-C2) has been achieved in terms of the CEFR scale Grid descriptors or in terms of students' abilities to affirm specific 'Can-Do' statements on the Checklist.

C. DIALANG? Not ideal, yet. It is promising because it is free and it is tied a priori to the CEFR. However, some students have reported that it takes a very long time to complete; students need their own computer to use it; its speaking and writing sections measure student abilities indirectly, and students cannot give the English program their raw data to analyze. Students will be encouraged to try it of their

own accord. If and when it is proved effective for the program's students, it can be required.

D. Self-Assessment? Yes, it is the best solution, for now. The Self-Assessment Grid and Checklist break down the CEFR scales into bits that students can understand. As long as students affirm and disaffirm the Can-Do's honestly, the instrunments should provide a picture of students' overall levels and specific language abilities that are accurate enough to be useful for curriculum development. The limit of self-assessment, is, of course, that it is subjective, so eventually, it must be complemented with valid and reliable objective testing.

3 Survey: European Language Portfolio Self-Assessment Checklist

3.1 Objectives:

1) Identify students' perceptions of their abilities and the degree to which they match the English program's official targets.

2) Identify areas of the curriculum that may need adjusting.

3.2 Subjects

The subjects are 2nd year English majors with ages ranging between 19 and 21. 57 surveys were returned, 53 with TOEIC scores.

3.3 Method

1 . January 2008, distribute Self-Assessment Checklist in class with instructions in Japanese; instructions include a request for students' most recent TOEIC or other standardized test score.

2 . Convert CEFR scales to numbers: A1=1, A2=2, B1=3, B2=4, C1=5, C2=6.

3 . Number the Can-Do's within each skill area of each section.

4 . Transfer student data to Excel.

5 . Identify most 'difficult' Can-Do's for students.

6 . Identify students' strengths and weaknesses by CEFR skill area.

7 . Identify overall level.

3.4 Results

3.4.1 Can-Do's most and least often affirmed

The students marked each Can-Do with a:

- 0 to disaffirm it
- 1 checkmark to affirm it （under normal circumstances）
- 2 checkmarks to indicate it is easy. Sums show which abilities seemed 'easy' or 'difficult' within a skill area and level. See Listening and Spoken Interaction examples below.
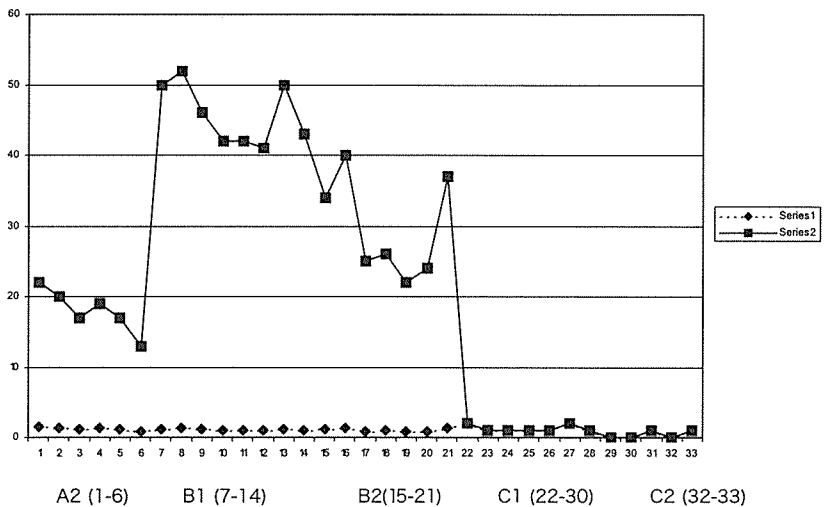

A2 (1-6)   B1 (7-14)   B2(15-21)   C1 (22-30)   C2 (32-33)

Figure 1. Listening Abilities

This Graph indicates that within level B1 in the Self-Assessment Checklist, 10-12 (numbering my own and not part of original Checklist) were most difficult and that within level B2, 17-20 were most difficult.

10. I can understand the main points of radio news bulletins and simpler recorded material on topics of personal interest delivered relatively slowly and clearly.

11. I can grasp the main points in TV programmes on familiar topics when the delivery is relatively slow and clear.

12. I can understand simple technical information, such as operating instructions for everyday equipment.

17. I can understand most radio documentaries delivered in standard language and can identify the speaker's mood, tone, etc.

18. I can understand TV documentaries, live interviews, talk shows, plays and the majority of films in standard dialect.

19. I can understand the main ideas of complex speech on both concrete and abstract topics delivered in a standard dialect, including technical discussions in my field of specialization.

20. I can use a variety of strategies to achieve comprehension, including listening for main points and checking comprehension by using contextual clues.

In Spoken Interaction (see Figure 2.), B1, 21 was by far the most difficult; in B2, 22, 25, 26, 29 and 30 were most difficult (numbering not part of original Checklist).

21. I can manage most discussions involved in the organization of my studies,
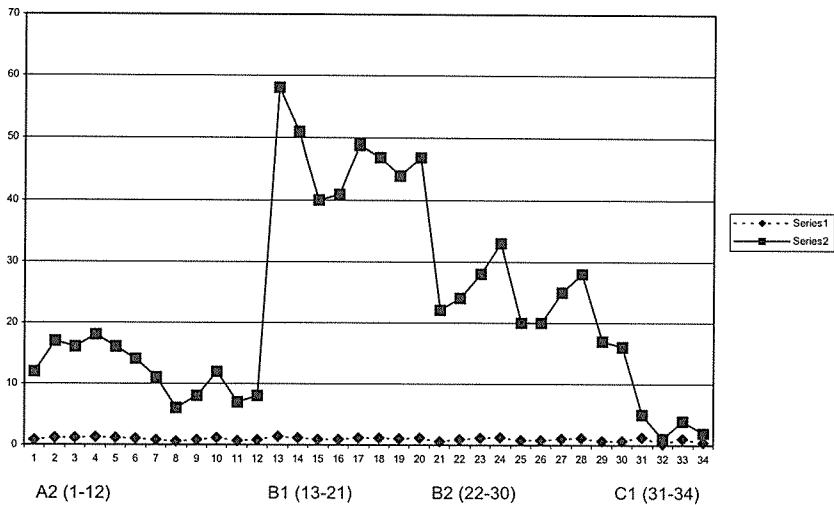
Figure 2. Spoken Interaction Abilities

either, face to face or by telephone.

22. I can initiate, maintain and end discourse naturally with effective turn-taking.

25. I can engage in extended conversation in a clearly participatory fashion on most general topics.

26. I can account for and sustain my opinions in discussion by providing relevant explanations, arguments and comments.

29. I can actively participate in conversations on specialized or cultural topics, whether during or outside of courses.

30. I can efficiently solve problems arising from the organization of my studies, for example, with teachers and the administration.

Such concrete evidence of the Can-Do's students find most difficult can inform curriculum development. In the development process, a teacher or program can

131

prioritize the Can-Do's according to student wants/needs. For example, as Dr Brian North commented after a presentation of these results at ALTE (Association of Language Teachers of Europe) 2008, it is possible that numbers 21 and 30 may not be especially relevant to students studying English in their home countries. Therefore, although about half of the English program's students plan to study abroad at some point, it may be advisable to give priority to the other difficult Can-Do's listed above, given limited time and resources. If 29 and 30 had been eliminated from the English program's version of the Checklist, then students' average spoken interaction scores would have been slightly higher.

3.4.2   Average self-assessed level

Students first read through the Self-Assessment Grid and circled the level for each skill area that they felt most closely matched their actual level (for a snapshot of students' perceived levels, this would be enough). Then they filled out the relevant Checklists. Almost always, the Self-Assessment Checklist results corroborated the Grid results. However, a few subjects went to the Checklist indicated by the Self-Assessment Grid, found they could affirm virtually every Can-Do, and completed the next higher level without revising the circled boxes on the Grid. By averaging students' Checklist responses (approximately 80% or better to qualify for a level) I obtained the following results.

Table 1. Average student level by skill area

Skill	Average	CEFR level
Listening	3.29	Low B2
Reading	3.73	High B2
Spoken Interaction	3.29	Low B2
Spoken Production	3.36	Low B2
Writing	3.82	High B2

Overall, students judged themselves to be in the B2 range, which was the English target for 2nd year. It is not surprising that they are more confident about their reading ability than their listening and speaking ability. A pleasant surprise, however, was their confidence in their writing ability.

Based on these results, it is clear that means of improving listening and speaking should be tried. It would also appear that the writing curriculum has been successful and probably should not be changed without persuasive counterevidence.

3.4.3   Gap in Self-Assessment vs. ETS published equivalence for TOEIC

One ETS web page with the heading, 'TOEIC scores and Common European Framework level descriptors' claims the following TOEIC sum-score ranges per CEF level: 'B1 = from 479 to 619; B2 = from 619 to 803; C1 = from 803 to 943' (Tannenbaum & Wylie 2004).

Interestingly, however, in my survey, of the 53 TOEIC scores returned, 26 are 800 or higher and 11 are above 900, but only one 900+ student claimed C1 ability in all skill areas and two claimed C1 ability in three skill areas. The other subjects claimed mainly B2 and some even B1! Why? (1) Does it have to do with Japanese culture? (2) Are the TOEIC-CEFR equivalencies questionable?

With respect to the first question, the answer would be 'Yes', at least according to linguist Tomoko Yamashita Smith [6], also a Japanese national and Japanese language teacher, who predicted such a result years ago, before I conducted the survey (personal communication). She points out that even many highly advanced Japanese speakers of English, including herself, would feel uncomfortable claiming many C1 and even some B2 abilities without hedging. This she said, was because of the value Japanese put on *kenkyo*, 'modesty' or 'humility', which holds that it is better to underestimate one's abilities than to overestimate them. She intends to conduct a study to ascertain whether or not Japanese and European students of English with comparable standardized test scores rank themselves differently using the Self-

Assessment Checklists.

If they do, the consequences may be that the Grid and Checklist instructions for Japanese—or even the various Can-Do statements themselves—need to be 'adjusted' for a Japanese audience. Preserving the original meaning of the statements while at the same time taking into account cultural differences related to language/speech acts is a delicate matter, one that will no doubt require considerable research.

As for the second question, the answer may also be, 'Yes.' As ETS *TOEFL iBT Research Report, TOEFLiBT-06*, June 2008 explains, 'Asking panelists to create an interpretative bridge between the CEFR and a test, particularly a test that is not designed a priori to measure the CEFR, should not be taken for granted, and should appropriately be considered and treated as a research-based question.' (Tannenbaum & Wylie 2008: 4).

The Scaled-Score Cutscore Results for the TOEIC, in this report were, for C1, 200/200 in writing, 200/200 in speaking, 490/495 in listening, and in reading no cutscore could be obtained. These figures represent a level of difficulty that is categorically different from that suggested by ETS in 2004, leaving one to wonder which estimate is best and whether either is accurate. Further confounding matters is a 2007 ETS standard setting report that suggests a C1 reading cutscore of 455/490 TOEIC, based on 45% of the panelists. [7] Such variations would seem to corroborate the 2008 report's position on mapping. That is, 'It is more likely that tests developed specifically to map to the CEFR would pose less of a linking challenge than tests relying only on a post hoc approach, as was the present case' (Tannenbaum & Wylie 2008:29).

If the latest ETS cutscores are accurate, then my subjects were not underestimating their abilities, and only students with virtually perfect TOEIC scores should be thought of as making the cut for C1. It is notable, however, that these latest results, even if proved 100% accurate, would not make TOEIC an ideal test for gauging C1 level because they do not allow for discrimination *within* C1.

The TOEFL *iBT* results, on the other hand, according to the 2008 ETS report, do at least leave some room for discrimination within C1: writing 28/30, speaking 28/30, listening 26/30 and reading 28/30, thereby making TOEFL a more attractive choice, in at least this one respect. Nevertheless, as the preceding quotation explains, there is at least one very good reason why tests developed specifically to map to the CEFR, such as Cambridge ESOL tests, should be best at doing just that.

4 Conclusion

CEFR Self Assessment Checklists can produce valuable data for understanding students' abilities/levels, for curriculum development and for setting achievement goals.

4.1 Near-term goals based on survey results:

1) Introduce more listening and speaking: 135 hrs. and ~ 100 hrs., respectively, in terms of class time and homework, in 2008-9 thanks to the additional native class starting this year, and the agreement of other native teachers to hit these targets, per my request.

2) Develop speaking test based on the *Manual for Relating language examinations to the CEFR*[8], in particular *the CEFR Grid for Speaking, developed by ALTE members (input)*[9] and the DVD, *Samples of oral production, illustrating for **English**, the levels of the Common European Framework of Reference for Languages*, from the Council of Europe's Language Policy Division, familiarize students with the criteria, and start speaking tests at the end of first and second year to focus and encourage speaking practice.

3) Ensure that the most difficult B1 and B2 Can-Do's are covered in the native-speaker curriculum, mainly in first year.

4) Verify that 2nd year curriculum targeting C1 ability, including speech and debate, with its writing component, contributes to high self-assessment in writing.

5) Automate the Checklist for use twice a semester with all students in the English program and potential use by other language programs.[10]

6) Include a section in the automated version in which students can describe abilities they want/need, but that do not appear in the Checklist.

7) Update the Checklist based on more recent versions and more recent research as needed (Regarding this and the preceding point, I would like to thank Professor John Trim, who at ALTE 2008, commented about the particular version of the Checklist I was using, 'It is not written in stone, you know.')

4.2　Long-term goals related to CEFR:

1) Confirm that Osaka University officially approves the CEFR-based language education achievement system created by the former OUFS (this has been done).

2) Help the CEFR-research group secure further government funding to facilitate implementation of the CEFR-based achievement system.

3) Collaborate with all English teachers in the program, both Japanese and native, to establish *curriculum that officially and efficiently targets C1 or better before graduation for all English majors.*

4) Work to institute valid and reliable testing to prove achievement of CEFR targets for 1st, 2nd, 3rd and 4th year English majors. This could—perhaps—be done indirectly via four-skills TOEIC/TOEFL testing correlated to CEFR. However, most students may find the 2008 cutscores unrealistic goals to attempt, and formally adopting them would no-doubt result in undesirable washback. Therefore, a better possibility may be for Cambridge ESOL to

136

allow the program to participate in pilot testing and provide us test results directly linked to CEFR. This could be a stepping-stone on the way to becoming an official testing center someday, which would allow students to get valuable feedback on their test results (My thanks to Dr. Neil Jones of Cambridge ESOL for clarifying this last point). In addition to letting the English program prove C1 ability, the scores from the pilot-testing program would make it possible to find a correlation between the English program's Checklist results and the Cambridge test we eventually use.[11]

5) If Self-Assessment and testing show that students do not reach C1 before graduation, advocate more English classes in 3rd and, perhaps 4th, year.

6) Proclaim, as soon as it is proved, that our English program consistently produces graduates with C1 or higher level. Perhaps we will be the 1st in Japan to do so.

1) This site was last referenced on 31 August 2008.

2) These descriptors were developed for the Common European Framework and the Portfolio in the Swiss National Science Foundation project by Schneider, Günther & North, Brian (2000): *Fremdsprachen können – was heisst das?* Chur/Zürich, Rüegger.

3) This site is last referenced on 31 August 2008.

4) From April, 2008, it became 7 classes.

5) According to a freshmen CEFR self-assessment survey conducted by Smith in April, 2008, the average freshmen starting level is now B1; this, together with the increase in the number of 1st year classes and end of term TOEIC results, justifies the program's intention to raise the 1st year goal to B2 and the 2nd year goal to a "hedged" C1, with C1 solidified in 3rd and 4th year.

6) PhD in Linguistics, UC Berkeley 2005.

7) This standard setting was conducted by Tannenbaum & Wylie in 2006.

8) This site is last referenced on 31 August 2008.

9) This site is last referenced on 31 August 2008.

10) A project to automate a more recent version of the Checklist using WebCT, was begun in May, 2008.

11) Recently, STEP and the English program have been discussing a research project using STEP BULATS, which is linked to CEFR and is administered by STEP through an agreement with Cambridge ESOL.

References

CEFR Self-Assessment Grid, The Council of Europe 31 August 2008 <http://www.coe.int/T/DG4/ Portfolio/documents/assessment_grid_english.doc>

Mapping the TOEIC and TOEIC Bridge Tests on the Common European Framework of Reference for Languages, 2007 Educational Testing Service. 31 August 2008 <http://www.ets.org/ Media/Tests/TOEIC/pdf/toeic_cef_mapping_flyer.pdf>

Self-assessment checklists from the Swiss version of the European Language Portfolio, The Council of Europe, 31 August 2008<http://www.coe.int/T/DG4/Portfolio/documents/ appendix2.pdf>

The CEFR Grid for Speaking, developed by ALTE members (input), 9 December 2005, The Council of Europe, 31 August 2008 <http://www.coe.int/T/DG4/Portfolio/documents/ ALTE%20CEFR%20Speaking%20Grid%20INput51.pdf>

The Council of Europe's European Language Portfolio, The Council of Europe 31 August 2008 <http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/levels.html>

The Manual for Relating language examinations to the CEFR, The Council of Europe 31 August 2008 <http://www.coe.int/T/DG4/Portfolio/?L=E&M=/documents_intro/Manual.html>

TOEIC scores and Common European Framework level descriptors 31 August 2008 <http:// www.eu.toeic.eu/toeic-sites/toeic-europe/table-toeic-cecr/>

Tannenbaum, R.J., & Wylie E.C. (2004) *Mapping test scores onto the Common European Framework: Setting standards of language proficiency on the Test of English as a Foreign Language (TOEFL), The Test of Spoken English (TSE), The Test of Written English (TWE), and The Test of English for International Communication* (TOEIC). Princeton, NJ: Educational Testing Service. 31 August 2008 <http://www.eu.toeic.eu/toeic-sites/toeic-europe/table-toeic-cecr/>

_____ (2008). *Linking English-Language Test Scores Onto the Common European Framework of Reference: An Application of Standard-Setting Methodology*, Educational Testing Service, Princeton New Jersey. 31 August 2008, <http://www.ets.org/Media/Research/pdf/RR-08-34. pdf>