

【原著】

The EF SET English Assessment at Hiroshima Bunkyo University

Richard Sugg

広島文教大学での EFSET 英語評価

Richard Sugg

Abstract

This report gives an overview of a five-year trial of the computer-adaptive English First Standard English Test (EF SET). The trial was originally conducted among small groups of first year Global Communication Department (GC) Writing Strategies (WS) course students and has since moved on to include Education Department (ED) Writing Strategies course students, the majority of second year General English (GE) course students from all departments, and third year International Communication Strategies (ICS) course Global Communication Department students at Hiroshima Bunkyo University. This free, online assessment provides students with a Common European Framework of Reference (CEFR) aligned results certificate and claims correlation with both the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS). It was hoped that this free, low stakes, i-Pad friendly assessment could be used to improve generally foreign language test-averse, non-English major students' awareness of the CEFR and their overall level within it, encourage continued study and learning throughout their four years of university, and provide external evidence of students' overall English progress. It was also considered as one way of checking the effectiveness and validity of our own in-house streaming and progress / achievement measuring Bunkyo English Test (BET). The EF SET's methods of administration, effectiveness in achieving these goals, and future feasibility and implementation possibilities will all be discussed.

Introduction

The English as a Foreign or Second Language world in Japan is heavily invested in being able to accurately assess students' ability. However, many teachers are still just training students who will never need to use English for their chosen occupations to memorize for pass / fail tests. Because of this, university educators spend countless hours designing hopefully motivating, increasingly online, in-house assessments for the largely disinterested. At the same time, they are often having to deal with, or outsource, the training of more motivated students to pass expensive, external, high-stakes English language proficiency tests such as the Test of English for International Communication (TOEIC), the 実用英語技能検定 Test in Practical English Proficiency (EIKEN),

the International English Language Testing System (IELTS), or the Cambridge Exam Suite in the traditional four skill paper-based test environment. More recently there has been an attempt to make such ability defining proficiency tests more accessible to all, with the Test of English as a Foreign Language internet-based test (TOEFL iBT) being available both online at test centers and at home. The TOEIC, IELTS and EIKEN (in its two-day S-CBT form) and Cambridge's APTIS test are now also available online, but only at test centers. However, all these tests are prohibitively expensive and are aimed at students with specific English career goals in mind (TOEFL for university entrance, IELTS for study abroad, TOEIC or EIKEN for work with English inside Japan).

While some of our students do take these tests, what we at the Bunkyo English Communication Center (BECC) were looking for was a free, online assessment that would give the majority of our students (non-English major, non-English career path students who still have to study English, or those who are interested in using English but lack the confidence or finances to apply for the aforementioned tests) a reasonably accurate indication of their practical ability without it interfering with their grades or own majors. If such a test could also give students results aligned to the Common European Framework of Reference (CEFR), which we could then also use to link to our own CEFR-based assessments and Self Access Learning Center (SALC) materials and help raise awareness within students of what they 'can-do', then so much the better.

The English First Standard English Test

In late 2016 we were introduced to the Education First Standard English Test (EF SET). An internationally recognized 'study abroad' organization, according to the main Education First (EF) website (Education First, 2021), 'EF is an international education company offering study abroad, language learning, cultural exchange and academic programs around the world'. Whilst also offering studies in Spanish, French, Italian, German and Japanese, by far their largest focus is on the English language learning market. Originally developed between May 2012 and August 2013, with the EF SET assessment

'EF initially set out to develop a free assessment tool for adults that could assess all four language modalities: listening, reading, writing, and speaking. The initial purpose for developing the EF SET was to create a placement/advancement/certification test that would be used internally within the EF environment and then be used as a publicly accessible free standardized test comparable in quality to high-stakes and very expensive tests such as IELTS and TOEFL'. (EF SET, 2014)

The assessment is still free to use for anybody via a Windows or Mac OS computer, tablet or smartphone, and a reliable internet connection. It is currently a 50-minute standardized assessment of only listening and reading skills, designed to classify test-takers' performances in both skills into one of the 6 levels defined by the CEFR: A1, A2, B1, B2, C1 and C2 (COE, 2001). Indeed, the EF SET website claims that 'The EF SET (50 min) is currently the only standardized

English test that reliably measures all skill levels, beginner to proficient, in alignment with the internationally recognized standard, the CEFR'.

A short speech test is now available in BETA form. However, like the fifteen-minute Quick Check that is also available on the EF SET web page, this is seen more as just a 'rough guide' and does not claim to be an accurate assessment of a student's CEFR level.

The EF SET assessment is an 'adaptive test', which means that the levels of the questions change in real-time as test takers demonstrate their ability. This means that if a candidate answers all the first reading section questions correctly, which are 'set at the mid-point of the B1-CEFR level' (EF SET, 2014) then the second section questions will be provided at a higher CEFR level. Accordingly, if a candidate answers all the first reading section questions incorrectly, then the second set of questions will be provided at a lower level (See Figures 1 and 2 for examples of how question complexity changes according to CEFR level).

Figure 1. Example EF SET Task Models for each CEFR Level (Reading) (EF SET, 2014)

A1	A2	B1	B2	C1
SINGLE PASSAGE/MULTIPLE CHOICE OR MULTIPLE OPTION QUESTIONS WITH COMPLETION B2/C1				
100 words	150-200 words	350 words	450-500 words	450-600 words
4 questions	5 questions	6 questions	6 questions + COMP	8 questions + 2 COMP
TWO PAIRED SHORT PASSAGES + 6 FIXED FORMAT QUESTIONS				
50 word each pssg	100 word each pssg	150 words each pssg	150-200 words each pssg	150-200 words each pssg
4 statements/fixed choices	6 statements/4 fixed choices	6 statements/4 fixed choices	6 statements/4 fixed choices	6 statements/4 fixed choices

Figure 2. Example EF SET Task Models for each CEFR Level (Listening) (EF SET, 2014)

A1	A2	B1	B2	C1
MONOLOGUE + M/C OR M/O QUESTIONS + COMP AT B2/C1				
100 words	60 "	60'	90"	240"
	4 M/C questions	4 M/O questions	4 M/O questions + COMP	5 questions + 1 COMP
DIALOGUE WITH MULTIPLE CHOICE				
60 "	60 "	120"	180"	180"
4 M/C questions	7 M/C questions	4 M/O questions	4 M/O questions + 1 COMP	6 FF questions
SPEAKER MATCH seconds by speakers by matches by choices				
90" seconds/4 of 5	120 seconds/4 of 5	180 seconds/5 of 6	220 seconds/5 of 6	400 seconds/5 of 6 x 2

This adjustment continues throughout the assessment, allowing it to measure a candidate’s ability as accurately as is possible. At the end of the assessment, these scores are then combined and placed on the EF scale from 1 to 100, and a final overall grade and CEFR level Certificate is provided instantly to students. These results can then be used to give students an idea of how they might later perform on one of the more popular proficiency tests (Table 1).

Table 1. *EF SET scores comparison (Adapted from EF SET English Levels, (2021))*

EF SET	CEFR	TOEFL-iBT	IELTS	TOEIC #	Cambridge
1–10	Pre-A1	n/a	n/a	n/a	80–99
11–30	A1 Beginner	n/a	n/a	120–220	100–119
31–40	A2 Elementary	n/a	n/a	225–545	120–139
41–50	B1 Intermediate	42–71	4.0–5.0	550–780	140–159
51–60	B2 Upper Intermediate	72–94	5.5–6.0	785–940	160–179
61–70	C1 Advanced	95–120	6.5–7.5	945–990	180–199
71–100	C2 Proficient	n/a	8.0–9.0	n/a	200–230

Combined Reading and Listening Scores only.

The assessment is also self-timing, with both the reading and listening sections having a maximum time limit of 25 minutes each. The student decides when to move to the next question, but having moved on, cannot go back to the previous question page. If a student does not complete a section within the 25-minute limit, then the assessment automatically moves on.

The EF SET assessment is freely available to anyone through its home page. When we first started the trial with our students this is how they accessed the assessment. Designed to provide data for the EF organization and as a way of encouraging students to become interested in further study or study abroad programs, this means that results only go to the student. Originally, this also gave a student their own personal URL from which they could take the assessment at any time. While convenient for the student, an immediate drawback of this for an institution is that teachers are then reliant on the student to provide them with a copy of their results, should they wish to track them. This was both time consuming and prone to students incorrectly reporting their results. In 2018 the BECC joined the EF SET Partner Program, which provided us with our own branded landing page for students to open the assessment and allowed us to get a full report of all student results from EF. In 2019 this page was combined with the landing page of other universities, unfortunately making the initial sign-in process more confusing and leading to some students’ data becoming unavailable. The landing page was again updated in the spring of 2020 restoring our own less confusing access point for students.

Students and Procedure

To ensure as little disruption as possible to students, the EF SET was first trialed at the end of the first semester in July 2017 with four classes of the author’s own students. Two classes were

made up of Global Communication (GC) department first year female students (until 2019, Bunkyo was a women's university) in the Writing Strategies class, aged between 18 and 19. Two more classes were made up of a mixture of Education, Welfare, Psychology and Nutrition first years in our General English (GE) course. One class was streamed as A1–A2 and the other as A2–B1 according to our CEFR-based curriculum. All students were given a brief slide show introduction on how to sign in to and take the assessment on their iPads but were not given any pre-assessment study or practice. While the GC students did take the TOEIC test as part of their studies, most of them, along with the students from the other departments, had not taken any other language proficiency assessments. None of the students had ever taken the EF SET before.

The corresponding groups amongst the first-year students were also chosen the following year, and the same GC students again took the EF SET as second years to start tracking progress. The procedures for taking the assessment remained the same. Although the second-year students had taken it once before, they were not given any extra specific practice during the year.

In the third year of trialing the assessment (2019), the assessment was taken in the second semester. The original GC group (now third year students) was again offered the assessment as part of its International Communication Strategies (ICS) class. All the second-year students who had chosen to take the GE course for a second year were also offered the assessment. First year students were not chosen this time as concerns had been raised about assessment 'overload' for incoming first year students, and due to changes in class make up with the university becoming both co-educational and opening a new separate education department.

In 2020, all first semester BECC lessons were held online due to Covid 19, so the assessment was again run in the second semester when first year GC Writing Strategies (WS) students, and third year GC ICS students were having classes face to face. Thanks to their familiarity with the assessment and working in the online environment, second year GE course students (almost exclusively Education and GC department students) were again offered the EF SET even though their classes were all online.

As of August 2021, the GC first year WS students have taken the assessment. It is currently planned to offer the GE course second years and the third year GC ICS students the assessment in October / November 2021. At this point it should be noted that while the GC students have extra English classes compared to other departments, and some of the Education department students do have an interest in becoming English teachers, none of the BECC students should be considered English majors.

Results

Our original EF SET trial in 2017 was designed to answer three questions: 1) Would the EF SET work on our students' iPads? 2) Would it be feasible for our students in terms of their ability? 3) Could our own streaming process be at all validated by an external assessment? The EF SET results from the first four classes / 72 students are below:

Table 2a. 2017 EF SET CEFR grades for A2–B1 Streamed Students

Class	BECC Stream	A1 Beginner		A2 Elementary		B1 Intermediate	
		Students	%	Students	%	Students	%
GC FE13	A2–B1	4	27	6	40	5	33
FE4	A2–B1	10	45	7	32	5	23

Table 2b. 2017 EF SET CEFR grades for A1–A2 Streamed Students

Class	BECC Stream	A1 Beginner		A2 Elementary		B1 Intermediate	
		Students	%	Students	%	Students	%
GC FE 14	A1–A2	8	53	6	40	1	7
FE8	A1–A2	10	50	10	50	0	0

Table 2c. 2017 EF SET CEFR grades for all 72 students

A1 Beginner		A2 Elementary		B1 Intermediate	
Students	%	Students	%	Students	%
32	45	29	40	11	15

For the first question it did seem that the assessment was compatible with our students' iPads, as only one student across the four classes was not able to complete the assessment. This turned out to be a wi-fi issue rather than a hardware issue. As for question two, all the students who were able to complete the assessment got scores that could be measured, and in-line with what we were expecting after having worked with the students for a period of three months. For question 3, we were actually not too hopeful that we would get any useful correlation between our in-house Bunkyo English Tests (BETs) and the EF SET. This is because the EF SET, much like the Oxford Online Placement Test, could be seen as 'a placement test which is not suitable for measuring achievement of GE Curriculum content and goals' (Bower et al, 2014). The EF SET draws upon grammar, vocabulary, and questions from a much wider range of topics than the students would have encountered in high school and their GE lessons at the BECC. Also, we only have two streams of students, and because of class size and make-up issues (trying to mix relatively even numbers from each department) we often need to split A2 students of similar levels between the two streams.

However, the adaptive nature of the EF SET did result in students answering questions closer to their range of knowledge. In the GE course A1–A2 stream class, all the students were placed between A1 or A2. In the GE A2–B1 stream class, 10 out of 22 students still placed at A1. At first, even accounting for the wider range previously mentioned, this did seem an excessive number. On further review of the data however, only three of these students had an overall score between 20 and 25, meaning that the other seven were closer to the A2 level of 31 points than they were A1. Also, all three of the lower students had one grade (either Reading or Listening) above 25, with two listening scores of 28 and 29, and a reading of 25. In real terms, it seemed that possibly only one or two students had been 'misplaced' by our internal streaming (Table 3).

Table 3. Comparison of EF SET A1 students in the GE A2–B1 Stream

Student	Reading 31 = A2	Listening 31 = A2	Total 31 = A2	Conclusion
1	30	29	30	Ok
2	23	34	29	Ok
3	28	29	29	Ok
4	23	35	29	Ok
5	29	29	29	Ok
6	27	29	28	Acceptable
7	22	30	26	Acceptable
8	21	29	25	Acceptable
9	20	28	24	Misplaced?
10	25	22	24	Misplaced?

Bold and shaded scores denote noticeable deviation.

Looking at the two GC classes we saw similar results. For the A2–B1 streamed class, four students were again placed at A1 by EF SET. Three of these students got a total of 30 or 29, putting them very close to the 31-point A2 threshold. All three had one reading or listening score above 31. The lowest two students had TOEIC scores well above the 379-point average for this class. For the A1–A2 class, only one student was streamed as B1, with a total score of 43, just above the 41-point B1 threshold. This student was absent for the initial BET 1 streaming test resulting in her being put into the ‘lower’ group. Also, the student was given an A2 for reading, which is the BET category score we use to determine the two streams for the writing course these classes were taken from (Table 4).

Table 4. Comparison of EF SET CEFR vs Streamed class deviation students in the GC course

Student	Stream	Reading 31 = A2	Listening 31 = A2	Comments / Overall 31 = A2	Conclusion
1	A1–A2	39	47	Absent for TOEIC and BET 1	Misplaced
2	A2–B1	28	32	30	Ok
3	A2–B1	28	32	30	Ok
4	A2–B1	25	33	29 / TOEIC 465	Ok
5	A2–B1	14	32	23 / TOEIC 410	Ok

Bold and shaded scores denote noticeable deviation.

At the end of the July 2017 trial, it did seem that the EF SET could positively answer our initial three questions, so we decided to continue for a further year. As no further issues were raised in 2018, we expanded the trial in 2019 with a view to begin using the EF SET to both encourage and monitor individual learning beyond our students first year of compulsory English education. To do this we decided to offer the EF SET to the same group of GC students (now taking it for the third time) in their ICS class, and to all second year GE students, including members from two of

the author's 2018 first year classes.

In their first year, the GC students included in the study had two GE classes, one WS, one RS (Reading Studies) and two BEC (Basic English Communication, a basic grammar review course) classes in English with native speakers in the BECC. They also had access to our all English SALC as well as some language related classes in Japanese. They first took the EF SET after three months of study, or around 120 hours. As it can take 100–150 hours for even a reasonably motivated adult student to advance from A1 to A2, and anywhere between 180–230 hours for a less motivated teenager, or 160–240 hours and 200–250 hours to advance from A2 to B1 (Knight, 2018, p10–p11), it was reasonable to assume that taking the assessment at this time would give us an accurate baseline for our students. In their second year, they had the same BECC classes and SALC opportunities minus the BEC course. This time they took the assessment in October at the beginning of the second semester, which meant we could hope to see the effects of a year's teaching. The group took their third and final assessment in January of their third year. In the third year, their only compulsory English course in the BECC was the ICS class, held just once a week. Any other English studies would have been from an elective course or by themselves at home or in the SALC. The results over the three years can be seen in Table 5 below.

Table 5. GC First Semester EF SET Results

Year / (No. of students)	A1 Beginner		A2 Elementary		B1 Intermediate		B2 Upper Intermediate		C1 Advanced	
	Ss	%	Ss	%	Ss	%	Ss	%	Ss	%
1 (28)#	12	43	11	39	5	18				
2 (27)##	3	11	8	30	9	33	7	26		
3 (21)###	5	24	5	24	7	33	3	14	1	5

One student had wi-fi issues and were unable to take the assessment.

Two students had wi-fi issues.

Two students studying abroad, six students declined to take the assessment.

In Table 5 we can see that during their first semester of year 1, all the students fit within our two designated GE course streams, upon which we base our streaming for the WS, RS and BEC classes. Looking deeper at the results, the majority (18) of the students received either a full A2 score, or one A2 score in either reading or listening. Only six of the twelve who were ranked A1 overall actually ranked as A1 for both reading and writing. In addition to the five students who ranked as B1, six students ranked as A2 scored a B1 in either reading or writing.

As hoped, by the time the students took the assessment for the second time, the majority were able to see a marked improvement in their ability. While three A1 students showed no improvement from year 1, four students moved up to A2 and five students improved as far as B1. At the other end of the spectrum seven students placed at B2. Of these, four had moved up from A2 and three from B1.

For this cohort, students had a choice between following an English education and business path,

or a Japanese business path. This was reflected in the number of elective classes students took in their third year, thus resulting in students in the same ICS class receiving very different numbers of hours of native English instruction per week. The EF SET again proved effective in showing which students had progressed overall, which students had progressed in either their reading and listening, and indeed which students had regressed through lack of English practice. Unsurprisingly, when we looked at internal BECC usage and GPA records, the student with the highest score (C1) took the most lessons in the BECC, spent the most time and had the most learning advising sessions in the SALC, and scored the highest amongst the cohort on the TOEIC (885). Furthermore, ten of the top eleven scores (B1–C1) were achieved by students who took the most English lessons within the BECC. At the same time, the five A1 students and the six students who decided not to participate all took the fewest elective classes in the third year, pointing to perhaps an overall lack of motivation to continue studies in English.

Sadly, we could not use the 2019 EF SET data to meaningfully track the progress of the second year GE students who took the assessment in 2018. This was mainly because the second-year course was by now an elective course for all students other than the GC department. Very few of the students in the low stream GE class (Welfare, Nutrition and Psychology students) in 2018 chose to continue with Sophomore English. While the some of the higher stream class did take the course, not all the students were able to take the assessment due to class schedule constraints. Due to the way the landing page was set up that year some students also entered their information incorrectly. Combined with the format the results were sent to us that year, it was impossible to collate results for those students over the two years. We could however analyze the data for most of the second-year students for just that year (Table 6) to see if there was any correlation with our streaming for the second-year classes.

Table 6. *GE Sophomore Semester 2 All Departments Overall CEFR placement*

Total No. of Students	A1 Beginner		A2 Elementary		B1 Intermediate		B2 Upper Intermediate	
	Ss	%	Ss	%	Ss	%	Ss	%
137	73	53	51	37	10	8	3	2

The SE course was divided by our BET 2 (GE end of first year assessment) into seven classes: one A1–A2 stream class, four low A2–B1 stream classes (closer to the A1–A2 threshold) and 2 A2–B1 high stream classes. Again, at first glance, the results were worrying because 73 students were placed at A1, rather than the 16 students we had in that class. However, when we again looked at students’ reading and listening scores, a slightly different picture emerged (Table 7).

Table 7. *GE Sophomore Semester 2 students placed via their highest listening or reading grade*

Total No. of Students	A1 Beginner		A2 Elementary		B1 Intermediate		B2 Upper Intermediate	
	Ss	%	Ss	%	Ss	%	Ss	%
137	31	22	78	57	23	17	5	4

Now we can see that 74% of our students placed in the A2–B1 range, for which we had the most classes. There were still 15 ‘extra’ A1 level students, all in the Education department. Due to scheduling constraints, in 2019 all Education students were in their own classes, while all the other departments were grouped together. Due to lack of teachers, an extra A1–A2 level class could not be provided for the Education course so these 15 students were divided across the A2–B1 low classes.

We also wished to see if the EF SET could motivate students with positive feedback, and to continue with their English studies beyond their compulsory years of study. Sadly, in the first few years of the trial (2017–2019) the general response from students was that it was too difficult and that they probably wouldn’t choose to do it again. Also, they found that doing a longer assessment on an iPad (in our students’ case the iPad mini) was more difficult than doing one on paper.

Ahead of the 2020 trial the EF SET changed its question format, so that the first reading section, while still placed at mid B1 level, consisted of ‘easier’ to read signs or posters to be joined up with a sentence describing their meaning. Previously the first section began with a longer reading text with content-based questions that students had to take time scrolling for to find the answers. This format had been quite off putting for less motivated students, whereas we hoped the updated format would be less intimidating. Also, we noticed that the new layout (which was originally designed for computers or larger mobile devices) worked better if the iPad mini was used in the horizontal position. In the past that is hadn’t been such an issue, but as students’ daily use of mobile smart phones has increased to a point where they are never without them, we have noticed that many students, unless explicitly told, will also use their iPad in a vertical ‘phone’ position. In 2020 we instructed all students to use the iPad in the horizontal position. We also introduced a new wider ranging post-assessment questionnaire that included questions in both English and Japanese asking students how difficult they found the assessment, if they were happy with their results, and how likely they were to take it again. All students who took the EF SET were asked, but not forced, to fill in the questionnaire immediately after finishing the assessment.

Beginning by looking at the students’ responses to the question ‘How difficult was the assessment?’, in October 2020 68 first year writing class (GC and ED) students took the assessment. Of these, six students (9%) said it was too difficult, 25 (37%) that it was very difficult, 28 (41%) that it was difficult, while nine (13%) responded that it was a little difficult. At the same time, 44 second year GC students took the assessment for the second time. Of these, five (11%) said it was too difficult, 12 (27%) that it was very difficult, 20 (46%) that it was difficult, while seven (16%) responded that it was a little difficult. The remaining 59 students who took the assessment were second year GE course students from the remaining departments. Of these, eight (13%) said it was too difficult, 24 (41%) that it was very difficult, 21 (36%) that it was difficult, while six (10%) responded that it was a little difficult (Table 8).

The two sets of students who took the assessment for the first time reported similar levels of difficulty, which perhaps speaks to its adaptive nature. It also shows that the EF SET can be

Table 8. 2020 Students' responses to the question 'How difficult was the assessment?'

Course / No. of Students	Assessment Difficulty							
	Too difficult		Very difficult		Difficult		A little difficult	
	Ss	%	Ss	%	Ss	%	Ss	%
1 st year WS students (68)	6	9	25	37	28	41	9	13
2 nd year GE students (59)	8	13	24	41	21	36	6	10
2 nd year GC students (44)	5	11	12	27	20	46	7	16

considered a 'fair' test as described by Isaacs et al (2013, also in Sneddon and Charles, 2021), in that 'if an assessment is to be fair and equitable, it must give fair access to those individuals engaging with the assessment, as well as be unbiased towards different sub-groups within the cohort'.

The group who took it for the second time (2nd year GC students) reported it as being slightly less difficult. These students had not practiced or prepared but being familiar with an assessment or test's layout and already knowing that it has no effect on grades will always reduce student anxiety and make the assessment seem easier.

Perhaps unsurprisingly for students taking an assessment for the first time, only 12 (18%) out of 68 of the WS course students said that they were happy with their results. However, 67 students replied that they wanted to take the EF SET again. For the second-year students of the 59 taking the assessment for the first time, 15 (25%) said they were happy with their results. This time 45 students, or almost three quarters said that they wanted to take it again. Of the 44 GC students taking the assessment for the second time, only 1 student said they were happy with their result. At the same time, 39 said they wished to take it again (Table 9).

Table 9. 2020 Students' responses to the questions 'Are you happy with your result?' and 'Will you take the assessment again?'

Course / No. of Students	Are you happy with your result?		Will you take the assessment again?	
	Happy	Unhappy	Yes	No
	Students	Students	Students	Students
1 st year WS students (68)	12	56	67	1
2 nd year GE students (59)	15	44	45	14
2 nd year GC students (44)	1	43	39	5

NB: Each year, students were asked for permission to use their response data for research purposes, whilst maintaining their anonymity. Any student who did not give permission, or who did not answer this question in the after assessment survey, had their data removed from the study.

Limitations and Issues

As with any form of assessment, the EF SET does have its drawbacks. The most obvious is to do with technology and the way the assessment is delivered. As an online assessment, and with our students using iPads, we are reliant on a stable wi-fi connection. During the early years of the trial the university was still improving its campus wide wi-fi. This meant that if the assessment was taken by a class of 30 students at the same time as 4 other classes were using the wi-fi in the BECC, students could lose their internet connections. Unlike an assessment on the Moodle system, students' answers are not stored during the assessment. Any student who loses their wi-fi connection immediately loses their answers and must start again. This can be very demoralizing, and if happens late on during a lesson period can mean that a student is unable to finish. Also, if a student does hit refresh or return in their browser, they too lose their responses up to that point. Delivering the assessment through a third-party app such as Showbie (which all BECC GE and WS classes use to deliver lessons to students) has improved the situation by effectively removing the browser limitations, but students are still at the mercy of their wi-fi. Of course, students can easily take the assessment again in their free time, but the situation is less than ideal.

Over the years, the opening landing page and instructions have often changed, sometimes during the same calendar year. Because the assessment instructions are not fully bilingual, and by their very nature are often at a level beyond that of our lowest students, we must remake instructional materials (slide shows, .pdfs or videos) and go through them with the students again. This is time consuming, but necessary to make sure that the assessment is indeed measuring the students' ability to answer the questions, rather than their ability to understand how to do the assessment.

One of the major plusses of the EF SET is that it is free. However, this can be one of its major drawbacks. The first of these is that when signing in to do the EF SET, students must provide a certain amount of personal information that the company will then use to help focus its sales and marketing. We have found ways around this, and our own landing page enables students to sign in using false information in certain areas, but it can be off putting for some students. In the original form of the assessment, along with their emailed results certificate, students received their very own link to use for further attempts. Now, students must sign in and submit their information every time. While helping us collate the student data, this has made it less likely for students to retake the assessment without our prompting.

Another issue is that even with students accessing the assessment through our own landing page, teachers still cannot get immediate access to students' results. We must email English First after each group of assessments to get the full results, and often wait up to a month to get them. We do ask students to provide us with screenshots of their final certificate and to fill in their grades in the post-assessment questionnaire, but this is both very time consuming to collate the data, and open to student error.

As previously mentioned, by purporting to be able to measure students' ability across the whole

range of the CEFR, the EF SET is effectively a placement test drawing its question content from a much wider area than our students cover in our two-year General English curriculum. As our GE curriculum concentrates on basic practical communication which we think our students can use or might need in their everyday lives, it is unlikely that those not taking extra elective courses will ever be able to achieve much more than the B1 level score. In some extreme cases, this leads students to almost 'give up' on the reading section after the mid B1 level reading question 1, before completing the listening section. If a student has a reading score and a listening score two or more CEFR bands apart, then no final overall score is given, which does not help in terms of motivation. We do try to point out to all students that they can see improvement within their CEFR band, but many do just focus on their overall grade.

Because the assessment is free and is designed to assess a student's CEFR level right there and then, there are no practice assessments for students to use. Also, the question items do not change that regularly (which is why we recommend taking the assessment no more than once a semester) so students do not get to see their answers in relation to the correct answers. Not 'practicing' for an assessment is an almost alien concept for Japanese students, and not receiving detailed item feedback does limit the students' reflective process.

Finally, because the assessment can be taken anytime and anywhere without supervision, it is difficult to make a case for it having any kind of security. The results are based solely on a student's honesty, and as such, even though the EF SET home page says that the certificate the student receives can be published on social media and professional sites such as LinkedIn, it is unlikely that many Japanese employers would take this into consideration in the way they do TOEIC or TOEFL scores.

Concluding Comments

Further review of Table 1 and consideration of the make-up of our student body shows why, as an external assessment, the EF SET is a good fit for our students. Upon entry to Bunkyo University, all our first-year students are streamed into one of two levels for their General English course: either A1–A2 or A2–B1. Students at that level would not even register a score with TOEFL or IELTS, and as any TOEIC score below 500 is regarded as largely meaningless in terms of future employment, there is little or no motivation for many of our students to take these expensive tests.

Similarly, one of the criticisms of solely using the CEFR for language learning is that there can be a large ability difference between a student who has just reached the A1 level and who

'Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce themselves and others and can ask and answer questions about personal details such as where someone lives, people they know and things they have. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help' (COE, 2020, p. 175)

and a student who is nearly at the A2 level and thus is almost able to

‘Understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of their background, immediate environment and matters in areas of immediate need’. (COE, 2020, p. 175)

Under general assessment standards however, both students will be classified as A1. To try and address this, particularly for Japanese junior and senior high schools, the CEFR-J has been introduced. The CEFR-J re-divides the six levels of the CEFR 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 (TUFS, 2021). However, this has yet to be widely adopted, and most of our students have reported little or no knowledge of either the CEFR or CEFR-J when they join us.

Through taking the EF SET, even though our students do not receive answer specific feedback, they do receive a score certificate showing how close they are to each CEFR band, and a summary of what they can do. Teachers can also look at this and add extra comments. For example, as a score of 31 equals CEFR A2, a student who scores between 26 and 30 can be said to be more of an A1+ (A1.3 on the CEFR-J scale) and can be encouraged as such to try and improve by those extra 4 or 5 points. This can be more motivating for a student than thinking that they have to try and improve by a whole CEFR level. This reinforcement of how the CEFR works and what it means can be added to the CEFR explanations and experience our SALC Learning Advisors and materials provide. The feedback also dovetails nicely with the CEFR goals in every BECC GE lesson. Within our current in-house curricula, we can only provide CEFR-based feedback for student’s spoken interaction, spoken production and writing abilities (Sugg, 2019), so this extra feedback on their reading and listening skills helps complete a student’s picture of their overall abilities.

Because the EF SET is free, students can also afford to retake it termly, allowing them to see even small improvements within their CEFR band year on year. As a student progresses through their four years of study, they can also use the assessment in a ‘no stress / no loss’ atmosphere to check their English progress, or indeed to see if their abilities have started to decline to a lack of regular practice. The most recent responses from students do indicate that students, although not ‘happy’ with their overall results, are motivated to take the assessment again.

There does seem to be enough of a correlation between the EF SET results and our own streaming for GE classes that we can use it to monitor and possibly help validate our own in-house BETs. From January 2020 we have shortened the end of year 1 BET 2 and the end of year 2 BET 3 and taken them online, with a view to giving students some CEFR based feedback based on these assessments. Having the EF SET to ‘check’ our own test against will hopefully help us attain this goal.

Despite the mentioned limitations, for now it seems a good idea to continue to run the EF SET in the BECC, and perhaps even extend it once again to include all first-year students. We also need to find a way to make sure that students always have access to the BECC EF SET landing page, perhaps via an information station in the SALC. In this way, regardless of how often the EF SET log in procedure changes or whether students are taking elective classes which allow time for the assessment in their curriculum, students will be able to access the assessment throughout their four years at the university.

Hopefully the EF SET can continue to be another tool to help us in making sure that students are aware of even their smallest English language improvements, can relate them to CEFR results in their other English courses, and can use them to motivate and inform their own studies in our Self Access Learning Center.

Final Author's Note: The author and all participating teachers have no affiliation with Education First and have received no compensation for this report or the ongoing use of the EF SET.

Bibliography

- Bower, J., Rutson-Griffiths, A., Sugg, R. (2014). *Setting and Raising Standards-the Rationale for, and the Structure of the Bunkyo English Tests*. Bulletin of Hiroshima Bunkyo Women's University, Volume 49.
- Council of Europe (COE). (2020). *Common European Framework of Reference for Languages (CEFR): Learning, teaching, assessment*. Companion Volume. Retrieved from <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Council of Europe (COE). (2001). *Common European Framework of Reference for Languages (CEFR): Learning, teaching, assessment*. Cambridge University Press.
- Education First Standard English Test. (EF SET). (2014). *EF SET Academic and Technical Development Report*. Retrieved from <https://a.storyblok.com/f/79503/x/6c8302dc10/ef-set-academic-development-report.pdf>
- Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Key concepts in educational assessment*. London, Sage Publications Ltd. Also at: <https://www.doi.org/10.4135/9781473915077>
- Knight, B. (2018). *How Long Does It Take to Learn a Foreign Language?* Research Report, Cambridge University Press. Retrieved from <https://www.cambridge.org/elt/blog/wpcontent/uploads/2018/10/How-long-does-it-take-to-learn-a-foreign-language.pdf>
- Luecht, R. (2015). *EF SET Plus-TOEFL iBT Correlation Study Report*. Retrieved from <https://a.storyblok.com/f/79503/x/8f28ee8ebc/ef-set-toefl-correlation-study-report.pdf>
- Luecht, R. (2015). *EF SET Plus-IELTS Correlation Study Report*. Retrieved from <https://a.storyblok.com/f/79503/x/3044206eea/ef-set-ielts-correlation-study-report.pdf>
- Sneddon, N., Charles, A. (2021). *Designing Effective Assessments and Items*. Online learning course, Cambridge Assessment Network, Cambridge UK.
- Sugg, R. (2019). *CEFR Based Rubrics and Feedback: What you CAN DO!* Bulletin of Hiroshima Bunkyo University, Volume 54.
- Tokyo University of Foreign Studies (TUFS). (2021). *CEFR-J とは/What is CEFR-J?* Available at https://tufs-sgu.com/language_ability/cefr-j/
- The English First Standard English Test is available at <https://www.efset.org/>
- The EF SET English Levels (2021) are available at <https://www.efset.org/english-score/>