【原著】

Setting and Raising Standards

——the Rationale for, and the Structure of the Bunkyo English Tests——

Jack Bower, Arthur Rutson-Griffiths and Richard Sugg

標準の設定と向上

——広島文教女子大学　英語試験の理論的解釈とその構造——

Jack Bower, Arthur Rutson-Griffiths and Richard Sugg

1．Abstract

This paper outlines the rationale for the Bunkyo English Tests (BETs), which are institutional standardized reading and listening tests administered as part of a General English (GE) curriculum at the Bunkyo English Communication Center (BECC) at Hiroshima Bunkyo Women's University (HBWU). Firstly, a brief history of institutional English reading and listening tests at the BECC is presented. Secondly, an explanation of why the BETs are necessary, and the valuable roles played by the BETs in the evolving BECC GE curriculum is given. Finally, a brief overview of the language skills covered by the BETs, and how these skills are measured is presented.

2．History of Institutional English Proficiency Tests for the BECC's General English Courses

2.a Early History

When the BECC started in 2008, an existing video-based English listening and reading proficiency test developed and used at Kanda University of International Studies (KUIS), known as the Kanda English Proficiency Test (KEPT) was used to assess learner reading and listening proficiency. KEPT results were used for class streaming purposes, and to measure changes in student English language proficiency over two years of the General English program.

However, it was apparent from student reactions during the tests, as well as from the test results, that most students found the KEPT to be too difficult. Different versions of the KEPT were trialled in later years, and a simpler KUIS test, the Kanda Assessment of Communicative English (KACE) was also trialled. Although both tests were suitable for intermediate-advanced English language majors at KUIS, they were too difficult for false-beginner/beginner English non-majors at HBWU. In addition, neither test was able to measure student achievement of GE curriculum content.

Due to other curriculum renewal priorities and a lack of resources, more suitable replacement

in-house tests of listening and reading were not developed until 2011. These tests were loosely based on simplifying KEPT tasks and focused on topics from the GE curriculum materials. These new tests were christened the Bunkyo English Achievement Tests (BEATs). The first BEATs consisted of two versions: one for entering GE students, and one for students at the end of their first year of study.

These two BEATs were subsequently revised in 2012 for their 2013 administrations based on classical testing item analysis, and Rasch item analysis. In addition a third BEAT to assess GE students at the end of their final, second year of study at the BECC was created, resulting in a set of three exams for class streaming and assessment purposes.

2.b Creating the Bunkyo English Tests – Institutional English Reading and Listening Proficiency Tests Aligned to the CEFR level A2

In 2013, as part of a wider Common European Framework of Reference – Japan (CEFR-J) and General English Curriculum alignment project, a decision was made to revise the BEATs to target them at CEFR-J level A2 (see Negishi; 2012 and Negishi, Takada & Tono, 2012 for descriptions of the CEFR-J). Due to its very recent development, there were no sample test tasks or items available for the CEFR-J at the time, so a pragmatic decision was made to model test tasks and item types on Cambridge's Key English Test (KET) (University of Cambridge ESOL Examinations, 2012a), which has a twenty year history as a validated test of listening and reading at the CEFR A2 level. The name of the tests was also shortened to the Bunkyo English Tests (BETs) for the sake of simplicity, and detailed specifications for the BETs were created. For all sections, items were revised by reducing the number of answer choices from four to three, as research has demonstrated that three options are optimal for multiple choice tests (Rodriguez, 2005). This was done by removing the least useful distractor based on Rasch item analysis. In addition, around half of the reading items in each of the BETs were created from scratch based on the new specifications, and the grammar and vocabulary sections of the tests were also revised to match the new specifications. However, the BET listening section items were not updated from the 2012 BEATs in 2013 beyond removing distractors, due to a decision to update the test incrementally. This decision was made because there was a lack of resources for test item piloting, so rather than changing the tests all at once, it was thought to be more practical to keep some statistically well-functioning items from the old tests until the item functioning of the new items had been established.

2.c Broadening the BETs to cover CEFR level B1

In the first semester of 2014 a decision was made to create two separate courses in the GE curriculum, one aiming to bring students from CEFR level A1 to A2 over two years of study, and the other aiming to bring students from A2 to B1 over two years of study (Bower et al, 2014). To ensure that the BETs have a wide enough range of questions to cover the proficiency span of all GE students, one task each was added to the BET reading section and to the BET listening section modelled on the Preliminary English Test (PET) (University of Cambridge ESOL Examinations, 2012b) tasks. These new test tasks aim to assess skills at the CEFR B1 level. In addition, in 2014

each of the BETs will fully match the BET specifications for the first time, with all of the listening section questions being rewritten to match the specifications, and the remaining half of the reading questions also being newly written.

3. The need for a streaming test

In this section the reasons for using the BET to stream GE classes are explained. These reasons are the fairly wide range of ability and motivation of entering students, a need to place students into one of two levelled courses, and students and teachers preferences for streaming.

The first reason for GE class streaming is that students entering the GE program have a rather wide range of communicative English language proficiency. All GE students have at least six years of 'knowledge', but even within the regimented Japanese environment, and within our generally mid- to low level students, there is a range of skill levels. This may be due to three factors: the wide variety of textbooks used in Japanese high schools and junior high schools, inconsistencies in the number of English lessons taken in high school and junior high school classes, and an increase in the use of Admissions Office (AO) entrance exams. These three factors are explained in the following paragraphs.

While most junior high schools use set Monbukagakkusho, or MEXT (文部科学省, The Japanese Ministry of Education, Culture, Sports, Science and Technology), texts such as New Crown and New Horizon, many high schools are not using either of the recommended 'English Expressions' (英語表現) or 'English Communication' (コミュニケーション英語) books (Hiroshima Board of Education, 2013). Often lower level government and technical schools use neither, choosing to pick from the list that each prefecture distributes to schools every year. Langham (2007) states that there are nineteen companies publishing English textbooks for high schools, and that there are twenty one possible oral communication textbooks for high schools to choose from. In Hiroshima prefecture alone, while there is only one choice for each of the lowest level courses of 'Basic English (コミュニケーション 英語基礎) and 'English Conversation' (英語会話), there are twenty three choices for 'English Communication' and twelve for 'English Expressions' in the first year alone, with a total of 74 different books for high schools to choose from over the three years.

Also, there are inconsistencies in the number of English lessons GE students have received in their secondary education. While the government states that schools must provide certain numbers of various English classes over two or three years, the number of classes per week vary, and some schools include their tests within these numbers. While government schools usually have from three to five lessons per week with Japanese teachers and mainly in Japanese, some private schools will have six or seven, one or two of which may be with a native speaker.

A final reason for the varied English proficiency levels of entering students may be the increased use of AO entrance exams, which allow lower ability students, or those who have studied through various forms of cram school or correspondence course and therefore have had little or no

communicative language practice whatsoever, to enter the university without passing a traditional paper test (see Matsutani, 2012).

The second reason for dividing GE students by ability based on BET scores is that the GE curriculum will be split into two levels of courses from 2015. Thus BET results will be needed for course placement purposes. Given that the focus of schools is still on the passing of an essentially grammar based entrance test, rather than any measure of communicative 'can do' ability, it has become clear that as the GE curriculum is to be based on the CEFR, which has the levelling of language ability as one of its core concepts for making curricula, we need to divide our whole curriculum, and our students by ability.

The third reason to stream students is the views of the teaching body concerned. Many foreign and second language programs place students in streamed classes according to their level of proficiency in the target language in order to allow teachers to better tailor lessons to the ability level of their students (Joyce & McMillan, 2010), while "31 EFL teachers at four Japanese universities expressed a preference for teaching streamed classes, citing advantages related to the effectiveness of instruction, appropriateness of materials, student motivation, and the quality of learner-learner interaction" (McMillan & Joyce, 2011, p. 70). These findings seem to back up their idea that by creating relatively homogeneous classes, teachers can design activities that more accurately suit the proficiency level of their students. At our own university, when we asked teachers the question, "Which of the following streaming options would you prefer for 12 Freshman English (FE) or Sophomore English (SE) classes?", 100% of the teachers preferred to have streamed classes, with 62.5% favoring streaming students into three levels, and the second most popular response favoring separating the higher level students from the majority. This in turn matches the breakdown of our students' CEFR ability according to the Oxford Online Placement Test (5% B2 and B1, 42% A2 and 53% A1 and A0), which teachers were not aware of when they completed the questionnaire.

Japanese university students too, seem to be in favour of streaming, with Joyce and McMillan (2010) reporting that streamed students were more able to understand their classroom materials, which indicates further support for the theories quoted previously, and that "while streamed students (of all levels) expressed a clear desire to be part of a streamed class, non-streamed learners were far more ambivalent about their preferred placement method" (2010, p. 224). Again, a survey of our own students seems to support these ideas, with 58.2 % expressing a desire to be in classes with students of similar ability, as opposed to only 2.7% who preferred to have mixed ability classes.

To conclude this section, both our own in house surveys, and the most up to date research carried out in other Japanese universities by the likes of Joyce and McMillan, seem to show that at the tertiary level, in a mixed class where all the students have varying goals, the long held belief that lower students won't challenge the higher ones sufficiently, and that higher level students will be too good for lower level students, still holds true. From the teachers' perspective, streaming helps

to put the more motivated and higher ability students into groups that can improve their communication levels, while at the same time putting lower level students into groups where we can concentrate on improving basic grammar, vocabulary and communication skills.

As classes are already grouped by the university in their second year of study (Early Childhood Education students study with Psychology students, and Nutrition with Welfare), we are not actually searching for a 'number 1' class. The whole aim of streaming based on BET scores is to put students 'like with like' within their own CEFR bands, so as to be able to give them all the highest possibility of increasing their own communicative ability by at least one band of the CEFR over their two years of study. To this end, we currently have three streams (high, middle and low) for our FE and SE Early Childhood Education/Psychology classes, but only two high and three low for our SE Nutrition/Welfare classes.

4. The need to track student progress in English L2 proficiency

In order to refine our curriculum, and to satisfy the demands of parents and an administration who ask us to teach said curriculum, we need to be able to accurately track our students' progress over their two years of instruction. The Vermont Board of Education (1994, p. 171) states that the fifth and "final step in providing effective L2 educational services for students is actually a series of steps for monitoring and guiding student progress", and includes in its guidelines for formal assessments the need to have a "comprehensive proficiency test", which in turn is a "curriculum-referenced test for evaluating mastery of program objectives correlated to the curriculum", and is a "criterion-referenced or standardized, norm-referenced test(s) of language, reading/reading comprehension, or subject area skills needed at the student's grade level" (1994, p.180). The BETs are just such tests.

In addition, we hope that by receiving a numerical score which shows their progress after each year of GE, students will become more motivated to study. Most dictionaries define motivation as one's enthusiasm for doing something, while the Longman English Dictionary Online defines it as being an "eagerness and willingness to do something without needing to be told or forced to do it". The latter point is very important because our GE courses students are indeed being forced to study English for two years. Qashoa (2006, p.2) cites Dornyei (2001) as arguing that motivation is cyclic, going up and down, affecting language achievement and being affected by it. Dornyei (Dornyei & Ushioda 2011, p.138) himself states that "a demotivated learner is someone who was once motivated but has lost their interest for some external reason". By giving students a numerical measure of their progress each year, we can keep the 'I want to study more because I am improving more' cycle constant. Also, it is another way we can show progress to students who are keen to study at first, but who do not generally have everyday opportunities to measure their communicative ability in the real world, thus eliminating one of the biggest negative external factors faced by students living outside of the main cities in Japan.

Finally, by accurately tracking student progress via the application of the BET 2 streaming test we

can further refine our class groupings in the second year, so that students progressing to a higher band of the CEFR can be grouped together and equally challenged, while students who may still lack basic skills can be given more time and opportunities to improve at a pace that is better suited to them.

5．Why not use a commercial test?

After problems with the KEPT and KACE became apparent, and due to the large amount of resources required to produce in-house standardized English language proficiency tests, commercial tests were briefly considered for the purposes of streaming classes and tracking improvements in student English language ability. However, none of the available commercial language tests examined were considered appropriate, partially because they were too expensive, but most importantly because they were not suitable for measuring achievement of the GE curriculum language proficiency goals, which are localized versions of CEFR can do statements based on the syllabus and lesson content. The problems with the three commercial tests considered for use in the GE Curriculum are summarized in the table below.

Table 1: Summary of commercial test considered for use in the BECC

Test	Advantages	Problems
Test of English for International Communication (TOEIC)	• Widely recognized and accepted in Japan • Recognized by all stakeholders • Useful for PR • Valid and reliable	• A norm referenced test "...NRTs are not suited to measuring the specific learning points and skills developed in a particular program" (Brown & Hudson, 2002, p. 7) • A test of business English with no relation to the GE curriculum Not suitable for measuring achievement of the GE Curriculum proficiency goals. • Potential for negative washback (teachers might teach to the test) • Cost
The Oxford Online Placement Test (OOPT)	• Places students at their CEFR level. • Valid and reliable • Cheapest commercial alternative	• A placement test which is not suitable for measuring achievement of GE curriculum content and goals. • Cost
The Key English Test (KET)	• Places students at CEFR levels A1-B1. • Valid and reliable. • Measures all macros skills (speaking, listening, reading and writing)	• Not suitable for measuring achievement of GE curriculum content and goals. • Not administered in Hiroshima. • Cost

Table 2: Cost of Commercial Tests

Test name	Cost for a single test	Approximate yearly cost for GE students (assuming three cohorts of 320 students)
TOEIC	¥5,725 (Educational Testing Service, 2014)	¥5,500,000
OOPT	GBP 3.75 (Oxford University Press, 2014)	¥612,000
Key English Test	¥9,920 (Luna International, 2011)	¥9,500,000

6. Why use listening and speaking tests for course streaming?

Listening and speaking tests are used for class placement purposes in the GE curriculum mainly due to practicality. Multiple choice tests of the receptive skills (i.e. reading and listening) are much easier to administer and grade than direct tests of the productive skills (i.e. speaking and writing) (Brown, 2005). The BETs can be administered to entire cohorts of GE students, which consist of over 300 students, in a single 90 minute classroom period and graded by a bubble reading machine the same day. Converting the bubble sheet reader output to test scores in Microsoft Excel takes only a few more hours. The administrative timeline for the university allows only two days between administering the BET1 to entering GE students, and streaming the students into classes based on their test scores. It would be very difficult if not impossible to administer and grade speaking tests or writing tests within this time period, particularly considering that double grading is needed in order to ensure reliable direct test scores of productive skills (Taylor & Galaczi, 2011).

Speaking tests have been administered at the end of each semester to GE students since 2010. However, these tests were administered in class time, which made it impossible to double grade the tests for logistical reasons. At the time of writing, BECC management is weighing the possibility of administering double graded speaking tests to students beginning from 2015. However, at this time listening and reading tests are still the only viable way to stream students for GE courses, and to track student progress in English proficiency.

7. Future benefits of setting CEFR A2 and B1 cut scores

Cut scores are defined scores on a test that match a performance standard. According to Cizek (2012, p. 4) "performance standards specify what level of performance on a test is required for a test taker to be classified into a given performance category." These performance standards are operationalised in a test by using cut scores, which "function to divide the distribution of test performances into two or more performance categories" (ibid.). Test takers who achieve or surpass the cut score are placed into a higher performance category than those who do not. In the case of the BETs, cut scores would distinguish between students who exhibit B1 performance, A2 performance, and A1 performance in English reading and listening.

Setting CEFR-based cut scores for the BETs is intended to have a number of benefits, primarily through allowing students to receive a CEFR level and a description of their listening and reading performance (see Council of Europe, 2001 for performance descriptors at each level). A CEFR level and its associated performance descriptors are more meaningful than a number alone, providing more transparency in the streaming process. Students will be able to use descriptors to plan their learning, particularly in the Self-Access Learning Center (SALC), where self-study activities are categorized according to CEFR level (Kodate, 2014). They will also be able to track their progress between tests as they progress through their two years of study, and may be more motivated to study to reach the next level.

Students will also receive a CEFR certificate at the end of their study that is understood by other language institutions around the world (Council of Europe, 2001). End-of-course CEFR level certification can also be used for many other purposes. It describes to future employers what students are able to do in English, and provides a measurement of progress that the BECC can use as feedback on the curriculum and the university can use for publicity purposes. In future, setting cut scores based on the CEFR may also allow the BECC to gain accreditation from the Evaluation & Accreditation of Quality in Language Services (EAQUALS), an independent accreditation body that allows institutions to award learners CEFR-based certificates of achievement (EAQUALS, 2014).

8. Positive washback on the General English curriculum

The term 'washback' refers to the influence of testing on classroom instruction, assuming that teachers and learners "do things they would not necessarily otherwise do because of the test" (Alderson & Wall, 1993, p. 5). In the case of large, high-stakes tests this effect has often been taken to be negative, but in the case of smaller, lower-stakes tests, it is hypothesized to have a positive effect on classroom instruction (Cheng, Watanabe & Curtis, 2004). The positive washback effect of the BETs on the General English curriculum can be seen in a number of instances, particularly in curriculum content and the level of classroom exercises.

For a test to have validity, it must cover a broad sample of content from its target domain (Kane, 2013). This is traditionally referred to as 'content validity'. For the BETs the target domain is the lesson handouts for GE courses. To ensure that the BETs had content validity, an analysis of the GE lesson handouts was carried out in 2012 to list curriculum topics, notions, functions and grammar topics. This analysis revealed that the curriculum, particularly in first year, did not cover a wide range of topics, and was lacking explicit grammar instruction that the BETs could test. This analysis was one of the contributing factors to the realization that a revision of the whole curriculum was necessary. The new curriculum, when complete, will cover a wider number of topics, and will include explicit grammar instruction, both of which will improve the content validity of the BETs.

In the writing of the new curriculum, the BETs are exerting positive washback in a number of

areas, in particular in classroom exercises. Messick states that:

> "Ideally, the move from learning exercise to test exercise should be seamless. As a consequence, for optimal positive washback there should be little if any difference between activities involved in learning the language and activities involved in preparing for the test" (1996, p. 1).

Reading and listening texts in the new curriculum are being written according to the specifications for the BETs, such as by adhering to guidelines on word counts, sentence complexity and vocabulary used. Exercises based on these texts are also written according to BET specifications, detailing what information students should extract from the texts and how questions are written (see section 8 for further details).

The writing of reading and listening texts based on BET specifications is anticipated to have a number of benefits. Not only will students be used to the format of reading and listening tasks in the BETs, as suggested by Messick, but also the consistency and quality of reading and listening exercises in the curriculum will be improved. As teachers will be writing exercises for lessons based on common guidelines rather than their intuition, the consistency of exercises between lessons written by different teachers will presumably be improved. The quality of exercises is also expected to improve; as the BET specifications are based on thorough language-learning research carried out by a number of international testing institutions (see section 2), exercises written based on the specifications are more likely to be targeted at appropriate language levels for learners at the BECC.

9. What do the BETs measure and how?

This section gives a brief summary of language proficiency constructs the BETs claim to measure, along with breakdowns of task types in each BET sub-section.

9.a Reading
Construct Definition
The construct definition for reading in the BETs comes from Khalifa and Weir's (2009) model of reading, which lists a hierarchical order of cognitive processes involved in text comprehension. The cognitive processes are briefly summarized in order from simplest to more complex below. For reasons of space only those cognitive processes in the hierarchical model required at CEFR levels A1-B1 are listed.

1. *Word recognition* – involves matching the written word to a mental representation of the word.
2. *Lexical access* – retrieving the mental entry for the word from the mental lexicon. This means accessing information such as orthography, phonology, morphology, syntax and meanings.

3. *Syntactic parsing* – mentally grouping words and phrases into larger units of meaning at the clause and sentence level.

4. *Establishing propositional meaning* – interpreting the literal meaning of a clause or sentence.

5. *Inferencing* – interpreting sentence meaning by utilizing background knowledge to fill in missing information. Also anaphoric resolution to link pronouns and other references to their subject.

6. *Building a mental model* – integrating sentence meanings into a mental representation of the meaning of the text that has been read so far. This involves comprehension across sentences and understanding the meaning of the text as a whole.

According to Khalifa & Weir (ibid.) tasks at the A2 level in the KET elicit mostly levels 1-4 of this model which involves understanding individual clauses and sentences, with a very few tasks requiring understanding meaning across sentences at levels 5 and 6. All test tasks at the B1 level in the PET, however, require processing across all levels 1-6.

Structure

The BET reading section contains four parts, and 30 minutes are allowed for test takers to complete this section. The composition of the parts of the BET reading section is summarized in table 3 below.

Table 3. Summary of reading text and task types contained in the BETs

	Text type	Task type
Part 1	Five simple signs.	Matching of signs to meanings (five questions, eight options).
Part 2	Five single turn conversational exchanges in the form of a question and an answer.	Three option multiple choice (five questions).
Part 3	A short dialogue.	Five matching items selecting from eight possible responses
Part 4	A short informational text adapted from authentic newspaper, magazine articles, or classroom materials	Three option multiple choice (seven questions).
Part 5	A short biography, or description of a group activity or project. This text is more grammatically and lexically complex than parts 1-3.	Four option multiple choice (five questions).

Skills Targeted

BET reading section Parts 1-4 are mainly intended to assess careful reading at the local level of clauses and are supposed to engage stages 1-4 of Khalifa and Weir's Cognitive processing model of reading, with a very few items requiring cognitive processing at levels 5-6. Part 5 of the reading section is intended to assess stages 1-6 of the cognitive processing model, with two items deliberately requiring global comprehension of the text, and most items requiring processing

across multiple sentences.

9.b Vocabulary and Grammar
Construct Definition

For the BET grammar items, students are tested through the use of selected-response type items in the form of passages with gaps targeting the areas of grammatical knowledge to be measured, which are intended to assess the students' knowledge of "grammatical form, or more specifically, syntactic accuracy" (Purpura, 2004, p. 129). The range of testable grammar for the BETs is defined by the grammar list included in the language specifications section of the BET Specifications.

Vocabulary for the BETs is taken from the vocabulary lists for the first year GE course and the second year GE course, which are in turn made up from an overall list of words taken from the CEFR vocabulary for the topic areas we practice in class. In the vocabulary section, the BETs aim to test a student's knowledge of both word meaning and use.

Structure

The BET Grammar/Vocabulary section contains two parts and 15 questions. These are summarized below.

Table 4. Summary of vocabulary and grammar text and task types contained in the BETs

	Text type	Task type
Part 1 (Vocabulary)	Six sentences (including one integrated example) with a connecting link of topic or story line	A five item, three option multiple choice sentence gap fill.
Part 2 (Grammar)	A teacher created text (text length is 110-130 words) based on classroom materials.	A ten item, three option multiple choice sentence gap fill.

Skills Targeted

Part 1 is focussing on reading and identifying appropriate vocabulary, and targets the A1 and A2 levels. Part 2 is focussing on reading and identifying appropriate structural word (auxiliary verbs, modal verbs, determiners, pronouns, prepositions, conjunctions etc.), and again targets the A1 and A2 levels.

9.c Listening
Construct definition

All BET listening items are intended to assess the ability of students on their understanding of local linguistic meanings, defined as including 'the whole of grammatical knowledge – not only phonology, stress and intonation, but also vocabulary and syntax, as well as the ability to use that knowledge automatically in real time' (Buck, 2001, p. 105). Questions intended to assess students at the B1 level may also assess discourse knowledge (such as discourse features), pragmatic knowledge (such as the function of an utterance) or sociolinguistic knowledge (such as idiomatic

expressions).

Structure

The BET listening section contains a total of five parts and 26 questions. The parts are summarized below.

Table 5. Summary of listening text and task types contained in the BETs

	Text type	**Task type**
Part 1	Five short, informal dialogues, e.g. between friends or a customer and assistant in a shop.	Three option multiple choice (one question per dialogue).
Part 2	Longer informal dialogue, e.g. between two friends about a familiar topic.	Matching (five questions, eight options).
Part 3	Longer informal or neutral dialogue, e.g. between two friends about a familiar topic, or a transactional exchange in a shop or restaurant.	Three option multiple choice (five questions).
Part 4	Monologue in a neutral context, e.g. an announcement in a shop or station.	Three option multiple choice (five questions).
Part 5	Informational monologue or interview with one main speaker, e.g. a recorded message or an interview about someone's life and experiences.	Three option multiple choice (six questions).

All listening texts are read from scripted passages and are delivered slowly, but not unnaturally so. All texts are heard twice. Texts are based on authentic situations, but do not generally contain features of natural speech (e.g. hesitation, fillers, interruptions, repair), except those questions that are intended to test at the B1 level. Texts do not contain strong regional accents.

Skills targeted

Parts 1-2 are targeted at level A1. Items at this level do not test any listening skills beyond what may be defined as 'key information', such as listening for explicitly stated times, dates or places. Parts 3-4 are targeted at level A2, and require students to identify factual information from the texts that is not always explicitly stated. Part 5 is targeted at level B1, and requires students to have a detailed understanding of the text in order to identify specific, detailed information or a clearly stated opinion.

10. Concluding Comments

The updated BETs for the 2015 administration fulfil several important functions in the General English curriculum. These functions are class streaming, tracking student progress in reading and listening ability, and fostering positive washback on GE curriculum content and pedagogy. Furthermore the BETs provide objective information about GE student English language proficiency to a range of stakeholders including students, teachers, administrators, parents and employers. In the near future BET scores will also form part of the basis for granting CEFR

proficiency certificates to GE graduates. These certificates will provide stakeholders with detailed descriptors of the English language proficiency of GE graduates, and may prove to be useful for university PR, and also for fostering student motivation. Finally, we believe that the BETs are able to achieve these functions more effectively and economically than commercial alternatives.

References

Alderson, J. C. & Wall, D. (1993). Does washback exist?. *Applied linguistics*, *14*(2), 115–129.

Bower, J., Rutson-Griffiths, A., Kodate, A., Foale, C., Lusk Lehde, L., & Semmelroth, A. (2014). *BECC General English curriculum overview 2014/15*. Unpublished internal document, Bunkyo English Communication Center.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill.

Brown, J. D. & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Routledge.

Cizek, G. J. (2012). An introduction to contemporary standard setting. In Cizek, G. J. (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed.) (pp. 3–14). New York, NY: Routledge.

Council of Europe (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Dornyei, Z. (2001). Teaching and Researching Motivation. Harlow, England: Longman.

Dornyei, Z. & Ushioda, E. (2011). Teaching and Researching Motivation: Second Edition. Harlow, England: Pearson Education.

EAQUALS (2014). *Student certificates*. Retrieved from http://eaquals.org/pages/7255

Educational Testing Service. (2014). *TOEIC 申込方法・受験料*. Retrieved from http://www.toeic.or.jp/toeic/guide01/flow.html

Hiroshima Board of Education. (2013). 平成26年度に県立学校で使用する教科用図書の採択について. Retrieved from http://www.pref.hiroshima.lg.jp/uploaded/attachment/105457.pdf

Joyce, P. & McMillan, B. (2010). Student Perceptions of Their Learning Experience in Streamed and Mixed-Ability Classes. *Language Education in Asia, Volume 1*

Kodate, A. (2014). *Developing SAC Materials to Support a CEFR-Based Syllabus*. Manuscript submitted for publication.

Langham, C. (2007). The writing and screening of a Japanese high school text series. *Proceedings of the 6th Annual JALT Pan-SIG Conference*. Retrieved on 29th of Aug. 2014 from: https://jalt.org/pansig/2007/HTML/Langham.htm

Longman English Dictionary Online http://www.ldoceonline

Luna International. (2011). ケンブリッジ英検 お申込み. Retrieved from http://www.luna-jp004.jp/order/index.html

Matsutani, M. (2012). Student count, knowledge sliding. *The Japan Times* Available at http://www.japantimes.co.jp/news/2012/01/10/reference/student-count-knowledge-sliding/#.VAqrbxYUqSp

McMillan, B. & Joyce, P. (2011). Teacher Perspectives on Student Placement in University EFL Programs. *Nepal English Language Teachers' Association. (NELTA), ISSN: 2091–0487*

Messick, S. (1996). Validity and washback in language testing. *Language testing*, *13*(3), 241–256.

Negishi, M. (2012). The development of the CEFR-J: Where we are, where we are going. 科学研究費補助金 基盤研究 *B* 研究プロジェクト報告書 (pp. 105–116).

Negishi, M., Takada, T. and Tono, Y. (2012). A progress report on the development of the CEFR-J. *Studies in Language Testing* 36: 137–165. Cambridge: Cambridge University Press.

Oxford University Press. (2014). *Browse, try tests and buy tests*. Retrieved from https://www.oxfordenglishtesting.com/browse_tests.aspx?userType=2&menuId=2

Purpura, J. (2004). Assessing Grammar. Cambridge: Cambridge University Press.

Qashoa, S. H. H. (2006). *Motivation among Learners of English in the Secondary Schools in the Eastern Coast of the UAE* (masters dissertation). Retrieved from http://www.asian-efl-journal.com/thesis_Sulaiman_Hasan_Qashoa.pdf

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3–13.

Taylor, L. & Galsczi, E. (2011). Scoring validity. In L, Taylor (Ed.), *Examining speaking: Research and practice in second language speaking.* (171–233). Cambridge: Cambridge University Press.

University of Cambridge ESOL Examinations. (2012a). *Cambridge English Key: Key English Test (KET) CEFR Level A2 Handbook for Teachers.* Retrieved from https://www.teachers.cambridgeesol.org/ts/digitalAssets/117391_Cambridge_English_Key__KET__Handbook.pdf

University of Cambridge ESOL Examinations. (2012b). *Cambridge English Preliminary: Preliminary English Test (PET) CEFR Level B1 Handbook for Teachers.* Retrieved from https://www.teachers.cambridgeesol.org/ts/digitalAssets/117693_Cambridge_English_Preliminary__PET__Handbook.pdf

Vermont Board of Education. (1994). *Serving Students Learning English as a Second language: A Guide for Vermont Educators.* Retrieved from http://education.vermont.gov/documents/guide_94.pdf