Mapping the English Proficiency Test (KU-EXITE) onto the CEFR Framework

Asst. Prof. Dr. Pornsiri Muangsamai

15 March 2017

Test Center of Foreign Languages
Faculty of Humanities Kasetsart University

Mapping the English Proficiency Test (KU-EXITE) onto the CEFR Framework

By Asst. Prof. Dr. Pornsiri Muangsamai

Abstract

The objective of this paper is to map the scores of KU-EXITE, Kasetsart University's English proficiency test, on the global scale of Common European Framework of Reference for Languages (CEFR). To achieve the goal, the modified Angoff method was used for the standard-setting process. It engaged 15 panelists, each with more than five years' experience in English language teaching and testing, and with background knowledge of CEFR and the modified Angoff approach. They were invited to give cut score recommendations for the KU-EXITE English proficiency test. The panelists estimated the percentage of the least able candidates who entered each of the CEFR levels: A2, B1 and B2 in three rounds by using several factors that assured validity and reliability. These were minimally qualified candidates' CEFR descriptors for each level and each test section, feedback, *p*-values, and group discussions. The results of the study include the minimum cut scores for each test section, the total minimum cut scores and the scores in relation to the CEFR levels.

Key Words: English proficiency test, CEFR, KU-EXITE, standard setting, cut scores

Introduction

Nowadays, good English language proficiency is one of the requirements for further study and employment. In order to produce quality graduates who are able to communicate in English effectively, and to enhance their competitiveness, one of the steps Kasetsart University (KU) takes is to make it compulsory for KU students to take an English proficiency test when they are in the third year, starting in 2017. This requirement aims at encouraging KU students to pay more attention to developing their English proficiency. In this regard, KU produces its institutional English proficiency examination called KU-EXITE to measure students' English ability. It consists of three parts. Part I is listening comprehension which is divided into: listening to questions and responses, listening to dialogs and listening to talks. The second part is structure and writing ability which includes sentence completion and text completion. The last part is reading comprehension which contains different types of texts such as letters, announcements, and academic texts. Since a good English examination paper needs to be able to measure test takers' English proficiency meaningfully and convincingly, standard setting, 'the methodology used to define levels of achievement or proficiency and the cut scores corresponding to those levels'(Bejar, 2008, p.1), is applied to serve this purpose.

Among different standard setting methods, the most commonly used for cut score recommendations is the modified Angoff method (MA). Similar to its predecessor, the Angoff method (Angoff, 1971), MA engages subject matter experts (SMEs) in estimating the number of minimally qualified candidates (MQCs) who would answer each test item correctly. The estimates, after being averaged, become the cut scores for the test. However, MA includes more steps to improve agreement among SMEs and to decrease standard errors before giving the final cut score recommendations (Ricker, 2006). For example, SMEs would be presented with descriptors of MQCs so that they share a mental model of MQCs while doing their ratings. Also, they would obtain 'normative feedback' (Cizek & Bunch, 2007, p. 54) --- ratings of the whole panel--- and 'reality feedback' (Cizek & Bunch, 2007, p. 55) --- the *p*-values. The feedback promotes the panel's rating convergence. The standard setting process could be arranged in two (Chang, 1999) to three rounds (Busch & Jaeger, 1990) with discussions between each round. The multiple rounds allow SMEs to exchange their rationales for the given scores and with additional information, SMEs might want to reconsider their own and adjust them. Once these steps have been implemented, the final recommended cut scores are obtained. Because the MA method

involves several factors, namely, SMEs, the training session for their test item rating, descriptors of proficiency at each level, multiple rounds of judgments, normative and reality information, discussions among SMEs, and simple statistics for cut score calculation (Hambleton & Pitoniak, 2006), it has been accepted and followed by many researchers (Baron & Papageorgiou, 2016; Tannenbaum & Baron, 2011; Tannenbaum & Wylie, 2008). This study follows the MA, accordingly.

However, test scores, by themselves, do not sufficiently display what test takers are likely to accomplish. To map them onto a framework of reference is a means to establish socially constructed meaning of test scores among score users, for example, test takers, teachers, and course and material designers. At present, the Common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR) is the most well-recognized and influential language teaching and assessment framework (Papageorgiou et al., 2015). It specifies what test takers can do in three bands: A1-A2 (Basic User), B1-B2 (Independent User), and C1-C2 (Proficient User). Scores of English tests administered internationally such as TOEFL, IELTS and TOEIC have been mapped onto CEFR (Baron & Papageorgiou, 2016; Tannenbaum & Baron, 2011; Tannenbaum & Wylie, 2008).

This study aims at identifying the KU-EXITE cut scores recommended for the A2, B1 and B2 levels and mapping them onto the global CEFR scale. The cut scores will be employed to display the ranges of scores including those below A2, between A2 and B1, between B1 and B2, and above B2. The study results will be used to inform test takers and test users of the levels of test takers' performance and to provide direction for their English development.

Methodology

In this study, the modified Angoff method was implemented to determine cut scores of the English proficiency test, KU-EXITE mapped onto CEFR levels --- A2, B1 and B2. There were two stages in this process: Stage 1: the pre-meeting orientation for SMEs held on March 15, 2017 and Stage 2: the cut score setting process held on March 23-24 and 28, 2017. A week after the orientation, the SMEs examined each test item, consulted information critical for cut score recommendations and their accuracy, and estimated the percentage of the least proficient test takers who could do each test item correctly. The estimates were calculated and the cut scores for each CEFR level were set.

Below are the accounts of SMEs, and the standard setting process.

Subject matter experts (SMEs)

SMEs play a very important role in the standard setting process. This study engaged SMEs knowledgeable of the content and the community of test candidates, the quality of which is necessary for estimating the probability that MQCs would give a correct answer to each test item (Mellone & Faben, 2014). Also, they needed to be competent in processing different sources of information such as descriptors of CEFR levels and *p*-values to identify the estimated cut scores of a test paper. Another major issue related to SMEs and cut score accuracy is their number. Zieky and Perie (2006) recommended 10-15 SMEs on a panel while Wheaton and Parry (2012) stated that 10 or more were encouraged. For best practice of cut score standard setting, the number of SMEs for this study was fifteen. They were selected from English teachers from the Faculty of Humanities who had more than five years' experience in teaching English to students in the university. All had attended seminars or workshops on standard setting methods and CEFR. In short, the SMEs had backgrounds in performing tasks in the standard setting process.

The following are the demographics of the SMEs.

Variable		N
Gender		
	Female	12
	Male	3
Education		
	Master's degree	6
	Doctoral degree	9
Area of Study		
	Teaching English as a foreign language	4
	Curriculum and instruction	1
	Language literacy	1
	Language and communication	2
	Linguistics	1
	Applied linguistics	1
	Language and linguistics	1
	English	2
	English for science and technology	1
	Translation studies	1
Function		
	English lecturers at university level	15
Experience		
	5-10 years	8
	More than 10 years	7

Standard Setting Process

The standard setting process in this study followed a variation of the modified Angoff method conducted in previous research in order to reduce cognitive load (Tannenbaum & Wylie, 2008; Tannenbaum & Baron, 2011). The process involved fifteen content experts to determine the cut scores in a focus-group manner in three rounds with discussions between each round. In the first two rounds, the percentage of MQCs for A2 and B2, the low and the high levels of the study, was estimated and then recorded in spreadsheets. This judgment was item-specific. In the third round, in addition to the descriptors of B1, the panelists used the A2 and B2 averages as anchors for B1 judgment. By this means, the estimate for B1 was established from the holistic judgment in Round 3.

The panelists were reminded of the standard setting process introduced to them in the pre-meeting the previous week, the definition of minimally qualified candidates, steps in the standard setting process, and the top estimate of 95 and the bottom estimate of 25, with increments of five (e.g. 25, 30, 35, 40, 45, 50). They practiced rating the scores that the MQCs

for the A2 and B2 levels would most likely earn by examining five test items similar to those in the KU-EXITE, discussing the panelists' estimates, re-reviewing them with *p*-values, discussing them, and recommending the final cut scores. Finally, the facilitator requested the participants sign in the Standard-Setting Process Agreement to confirm their understanding of it.

After the practice, the panelists made judgment of each test item, in the order of Listening Comprehension (40 items), Structure and Writing Ability (30 points), and Reading Comprehension (30 points). In doing so, they needed to provide estimates of the MQCs who could get each question correct. The estimates for each item were averaged and then, the averages were used to determine the cut score.

In Round 1, the panelists focused on Listening Comprehension. They independently rated the percentage of MQCs for both A2 and B2 levels who could give the correct answer to individual listening comprehension test items. The panelists had with them the descriptors for A2 and B2 for reference. They read a test item, estimated how many MQCs for A2 could answer it correctly, and recorded the number in the test item rating form. Then, they estimated how many MQCs for B2 could give a correct answer to the item, and entered the number in the same form. The task was continued until the last item of the listening comprehension section.

After an individual panelist's estimates for the listening section had been recorded in the expert ratings spreadsheet, summed and divided by 100 for her recommended cut score, the facilitator gave feedback to her for consideration. Then, all of the SMEs' estimates for the test items and their recommended cut scores together with the standard deviation (SD) of recommended cut scores were shared among the experts for discussion.

In addition to the feedback, their own and other estimates for each test item and recommended cut scores, the SMEs obtained the *p*-values of the test items.

In Round 2, the facilitator informed the panelists that if they found the feedback and discussion in the first round useful for their judgments, they could revise their estimates for either A2 or B2 or for both levels.

In Round 3, the panelists considered the recommended cut scores for A2 and B2 of Listening Comprehension by re-reviewing the descriptors of MQCs for each of the three CEFR levels. Discussion for this round was initiated. When they finished the rating activity for A2 and B2, they were instructed to give estimates for B1.

After the first section of the test was completely evaluated, the estimates for the A2, B1 and B2 levels for Structure and Writing Ability, and Reading Comprehension were assigned in the same way as those for Listening Comprehension.

At the end of the standard-setting judgments of all sections, the final panel cut score recommendations were presented.

Results

The standard setting results listed in the following tables are the cut scores of each KU-EXITE section: Listening Comprehension, Structure and Writing Ability, and Reading Comprehension, and their end-of-study-evaluation survey which examined the panelists' perception of the cut score standard setting process.

The recommended cut scores of each test section are presented in three rounds.

Listening comprehension standard setting results

Table 1 summarizes the results of the standard setting for each round of judgments on the Listening Comprehension of KU-EXITE which consists of 40 items.

Table 1 Listening Comprehension Standard-Setting Results

	ROUND 1		ROUND 2		ROUND 3		
	A2	**B2**	**A2**	**B2**	**A2**	**B1**	**B2**
MEAN	31.90	61.58	31.96	63.38	31.49	51.21	68.51
MEDIAN	30	60	30	60	30	50	70
MIN	25	40	25	45	25	30	40
MAX	65	90	60	90	60	80	95

Structure and writing ability standard setting results

Table 2 summarizes the results of the standard setting for each round of judgments on the Structure and Writing Ability of KU-EXITE which consists of 30 items.

Table 2 Structure and Writing Ability Standard-Setting Results

	ROUND 1		ROUND 2		ROUND 3		
	A2	**B2**	**A2**	**B2**	**A2**	**B1**	**B2**
MEAN	28.64	56.24	30.07	64.59	30.02	48.80	67.04
MEDIAN	25	55	30	65	30	50	65
MIN	25	35	25	45	25	25	50
MAX	50	75	50	85	50	75	90

Reading comprehension standard setting results

Table 3 summarizes the results of the standard setting for each round of judgments on the Reading Comprehension of KU-EXITE English proficiency test which consists of 30 items.

Table 3 Reading Comprehension Standard-Setting Results

	ROUND 1		ROUND 2		ROUND 3		
	A2	**B2**	**A2**	**B2**	**A2**	**B1**	**B2**
MEAN	32.16	73.40	32.16	73.30	32.16	53.87	73.94
MEDIAN	30	75	30	75	30	55	75
MIN	25	35	25	50	25	30	50
MAX	65	95	65	95	65	80	95

Round 3 Recommended Cut scores

Table 4 shows the Round 3 cut scores of the three test sections and the standard deviations of the KU-EXITE English proficiency test for three CEFR levels: A2, B1 and B2. The scores were rounded up to the nearest whole number to produce the final recommended cut scores.

Table 4 Round 3 Recommended Cut scores

Levels	A2	B1	B2
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
Listening Comprehension	31 (07.72)	51 (9.92)	69 (9.88)
Structure and Writing Ability	30 (5.72)	49 (9.43)	67 (5.38)
Reading Comprehension	32 (6.65)	54 (9.55)	74 (8.27)

The total cut scores for A2, B1 and B2 were computed by using the recommended cut scores in the three sections in Table 4. By summing the recommended cut scores for each CEFR level and dividing them by three, the total cut scores, after being rounded up to the nearest whole number were 31, 51 and 70 for A2, B1 and B2, respectively.

Correlation Table

Table 5 summarizes the KU-EXITE scores in correlation with the A2, B1 and B2 CEFR levels, including the total minimum test scores, and the minimum scores calculated from the recommended cut scores of each test section.

Table 5 Correlation Table

Total minimum scores	Listening minimum scores	Structure &Writing Ability minimum scores	Reading minimum scores	CEFR levels	CEFR General Description
70	28	20	22	B2 Independent User-Vantage	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
51	20	15	16	B1 Independent User-Threshold	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
31	12	9	10	A2 Basic User-Waystage	Can 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 his/her background, immediate environment and matters in areas of immediate need.

Conclusion

This study attempted to establish the cut scores for KU-EXITE by means of the modified Angoff standard setting process and mapping them on the A2, B1 and B2 levels of CEFR. Fifteen subject matter experts examined each test section of Listening Comprehension, Structure and Writing Ability, and Reading Comprehension, respectively and estimated the percentage of the MQCs who could just enter each CEFR level. The evaluation was conducted in three rounds based on the descriptors and the feedback. The latter included an individual SME's recommended cut scores for each level, those of the others, and the *p*-values of the test items calculated from the scores administered to 215 KU-EXITE test takers. The information allowed the SMEs to reconsider their estimates in the previous rounds. The panel then discussed and possibly changed them so as to reflect test takers' performance as much as possible. Finally, the final recommended cut scores for each section as well as the total minimum scores were obtained and mapped on the CEFR scale.

The study results could be beneficial for test takers, policy makers, teachers and material developers. Test takers will be able to understand what their KU-EXITE test scores mean, how strong their English is and what they need to do in order to develop their English ability. Policy makers can use or adjust the recommended cut scores to serve the institute's purposes such as assigning students to the appropriate level of English class, recruiting new students to international study programs, raising students' awareness of their English proficiency for their study and future career or setting students' English passing score before graduation. Teachers and material developers may use the study results to design English courses and learning materials to strengthen students' English proficiency.

References

Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp.508-600). Washington, DC: American Council on Education.

Baron, P. A. & Papageorgiou, S. (2016). *Setting language proficiency score requirements for English-as-a-second-language placement decisions in secondary education* (Research Report No. RR-16-17). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12102

Baron, P. A., & Tannenbaum, R. J. (2011). *Mapping the TOEFL Junior® test onto the Common European Framework of Reference (Research Memorandum No. RM-11-07).* Princeton, NJ: Educational Testing Service

Bejar,I. (2008). *Standard setting: What is it? Why is it important?* Retrieved from https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf

Busch, J. C., & Jaeger, R. M. (1990). Influence of type of judge, normative information, and discussion on standards recommended for the National Teacher Examinations. *Journal of Educational Measurement*, 27, 145-163.

Chang, L. (1999). Judgmental item analysis of the Nedelsky and Angoff standard-setting methods. *Applied Measurement in Education*, 12, 151-165.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

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

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport, CT: Praeger Publishers.

Mellone, I. & Faben,B. (2014). Are They Mission Ready? Using the Modified Angoff Method To Set Cut Scores. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), 14060.*

Papageorgiou, S. et al. (2015). *The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) level*s (Research Memorandum No. RM-15-06). Princeton, NJ: Educational Testing Service.

Ricker, K.L. (2006). Setting cut scores: A critical review of the Angoff and modified Angoff method. *The Alberta Journal of Educational Research* (52),153-64.

Tannenbaum, R.J., & Baron, P.A. *(2011). Mapping TOEFL ITP Scores onto the Common European Framework of Reference.* Retrieved from https://www.ets.org/s/toefl_itp/pdf/mapping_toefl_itp_scores_onto_the_common _european_ framework_of_reference.pdf

Tannenbaum, R. J., & Wylie, E. C. (2008). *Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology* (TOEFL iBT Series Rep. No. TOEFLibt-06, ETS Research Report No. RR-08-34). Princeton, NJ: ETS.

Zieky, M. & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement.* Princeton, NJ: Educational Testing Service.