From Classroom to Real World: Application of Outcomes-Based Assessment in English Courses

Supalak Nakhornsri

Abstract— Quality Assurance (QA) has become one of the current concerns in higher education worldwide. Learning outcomes are seen as central QA processes since the outcomes provide key benchmarks for maintaining standards and enhancing teaching and learning. Outcomes-based assessment (OBA) necessitates a change in educational practices, to a focus on what students have to learn rather than what educators have to teach. Moreover, it is connected with purposeful planning for the delivery and evaluation of intended outcomes. According to this, the purposes of this study were to investigate the expected learning outcomes of English courses for undergraduate students so that grading criteria can be established. In addition, the concurrent validity or investigation to prove whether students achieved the expected outcomes established from the expected language standards was implemented. Hence, the students' study results and the scores obtained from an English language standard (TOEIC) were computed. The findings can be significant in several ways. Theoretically, the findings can contribute to a better understanding of expected learning outcomes. This information can assist English language instructors in developing their teaching and course evaluation. Finally, the assessment of the concurrent validity can be useful to the interpretation of the study results due to the measurements performed by the evaluating instruments and by the standard instruments. Therefore, the scores obtained from the study results can be compared with the standardized scores. This allows classroom evaluation to be connected with more meaningful standards outside the classroom.

Keywords—concurrent validity; English courses; expected learning outcomes; outcomes-based assessment

I. INTRODUCTION

The QA processes indicate that higher education institutions see learning outcomes as key QA processes as their results may be key criteria for standards, performance evaluation and enhancement of teaching and learning[1]. In Thailand, the Office of National Education Standards and Quality Assessment (ONESQA) in the higher education sector has identified policy objectives and suggested a commitment to improving QA, as well as the need for QA to be created by administrators in higher education institutions[2]. Barkey and Majory[1] identified three interrelated components for successful pedagogy: (1) to define and document clear objectives and outcomes for the study; (2) to support students in achieving those outcomes through actively / devoted learning; Implementing these interrelated elements of higher education involves a change in training strategies with the goal of learning more from educators.. The implementation of outcomes-based learning (OBL) requires a shift from educational inputs to outputs in the form of direct and/or indirect evidence of student achievement in order to judge educational quality.

From the above, assessment is a critical aspect of providing evidence of student achievement and may be used as part of outcome-oriented QA [3].

However, the method of assessment commonly implemented is a norm-reference (NR). This is defined as a solution to problems in interpreting raw scores by indicating the relative standing of different test takers within a given group [4]. This method possesses an undeniable disadvantage in that it cannot measure the progress of an entire population—only where individuals fall within the whole [5]. In addition, NR can minimize educational standards because NR embraces the notion that education suits a curve of bells.

The main purpose of NR is to identify and sort students, and not to assess whether students have learnt the material they have been taught. [6].

Therefore, the inclination to use a normal curve as the basis to determine individual performance should be shifted away. Outcome-based evaluation is preferred, since it is a method of evaluation in which students are evaluated on their results against standards that are independent of the performance of the other participants. The degrees or study results of the students are indicated whether the requirements or predicted outcomes for evaluation have been fulfilled. At the start of the course, the results should be determined by time and effort, by focusing on and defining the expected results and making these explicitly available to students.

This means that the outcomes or criteria need to be established so that the performance of students can be judged according to the set outcomes. The students' study results or grades given then will be assigned and judged as to whether they have satisfied the criteria or the outcomes expected [7].

Since students from vocational institutions are expected to be prepared for the workplaces or their particular professions, the English instruction for these students should be designed to meet the needs of the learners and other stakeholders so that the students can communicate in their work or for study purposes in specific disciplines [8]. Khunmontri, Trakulkasemsuk, and Bunsom [9] stated the problems of students at the vocational level, indicating that they possess rather a low level of English proficiency. This can reduce their readiness to enter the working world.

At King Mongkut's University of Technology North Bangkok (KMUTNB) in Thailand, there are a lot of programs offered to those that have received a higher vocational diploma degree which is equivalent to the second-year university level. These students will take 2 more years in a university (a 2-year curriculum) to get a bachelor degree.

 Supalak Nakhornsri, Department of Languages, Faculty of Applied Arts, KMUTNB, Thailand (e-mail: supalak.n@ arts.kmutnb.ac.th).

For compulsory English courses, they are required to take Practical English I and Practical English II in the first and second semester respectively.

As mentioned, there is a necessity to implement QA in the teaching and assessment of these courses in order to be certain that these courses are connected to real world tasks and that the students' English abilities can meet the needs of society.

For the purposes of this study, the expected learning outcomes of the English courses were studied so that the grading criteria could be established. Once the OBL assessment had been implemented, the students' study results and the scores obtained from the English language standard were computed for measuring the concurrent validity, and as indicated earlier, the concurrent validity can investigate whether the students achieved the expected outcomes established from the expected language standard.

II. LITERATURE REVIEW

Outcomes-based assessment

The principle of continuous improvement is often a common feature of results-based assessment[10]. This ensures that performance-based assessments are related to tailored scheduling for the implementation and assessment of expected results. The evaluation process is also designed to allow the knowledge to learn how to achieve the desired results[10].

Because the evaluation process is an essential component of a traditional learning phase, some important aspects are included. The course process and assessment include planning, training, review, analysis and development, as shown in Figure 1. [11]
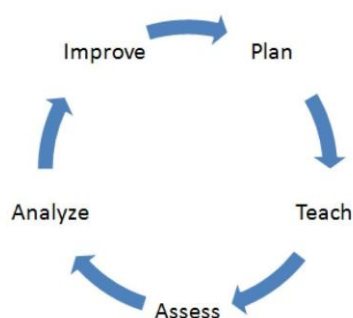

Fig. 1. Instructional Cycle and Course Assessment [11]

This process allows different players to make informed decisions to boost the quality and value of student learning experiences.

Evaluation is included from the process. Anwar, Ahmed and Al Ameen[11] claimed that they can make the educational model a success with the requisite evaluation data for various courses offered. It depends on the precise and timely availability to the approved evaluation committees of universities / departments of the assessment data collected by course instructors.

Framework of outcomes-based assessment

The assessment of student learning is central in this process Because the success of the students in the assessment activities can be used to determine what they have learned and done in a constructive way with expected learning outcome. El-Maaddawy and Deneen[1] proposed the following method to carry out results-based evaluation on courses..

Fig. 2. Outcomes-based learning framework [1]

The framework above was used as a guide for applying outcomes-based assessment to the English courses.

III. RESEARCH METHOD

This research is implemented as R&D so that the findings can be used to connect the course contents and the expectations of the real world. In order to collect and analyze the important data, the instruments, consisting of an open-ended questionnaire asking about expected learning outcomes and the TOEIC, were employed. Moreover, the course evaluation tools, including 1. evaluation forms for attendance and participation, 2. marking rubrics for an interview project, and 3. achievement tests (a listening test, a midterm test, and a final test), were used to measure and evaluate the students' study results.

The purposive sampling technique was applied. Ten English teachers teaching Practical English I and II participated in this study to specify the expected learning outcomes of the courses and to establish the grading criteria by means of Angoff method. The Angoff method is a process used for determining if a minimally-qualified performer would achieve the expected ability. A panel of experts is to review the test items and estimate the probability that a minimally-qualified performer would answer the item correctly. Then the average of the estimates are used to establish the cutting score [12]. For concurrent validity analysis, Dornyei [13] suggested 30 as an optimum number for a correlational research. Thus, 34 students that were enrolled in Practical English I and II and that had received at least a B grade in these courses were purposively selected to take the TOEIC.

Three stages were involved in the study. First, the expected outcomes of Practical English I and I courses were studied. The curriculum, the English proficiency standard, i.e. CEFR and other learning theories, were reviewed. Then, the cut-off score for each grade was assigned by means of the Angoff method. After that, the courses were implemented according to the excepted outcomes and the students' performances were evaluated according to the established grading criteria. Next, the concurrent validity was measured. Due to the fact that there were about 300 students taking Practical English I and II, and only 13 students received A grades from these two courses, the students that had received at least a B grade were included in the study. The students' total scores from Practical English I and II were then correlated with TOEIC scores.

Although the expected outcomes were established according to the CEFR descriptors, the TOEIC was used since this standardized test score is widely used in occupational settings and its scores are comparable to the CEFR. According to this, the students' TOEIC scores were compared with their CEFR levels so that it could be determined if their outcomes were equivalent to the expected English language standard.

IV. FINDINGS

The findings of this study are reported according to the 3 main research questions: 1. What are the expected learning outcomes of the English courses?; 2. How are the grading criteria established?; and 3. What is the concurrent validity of the students' study results and English language standard (TOEIC)?

The first research question asked: What are the expected learning outcomes of the English courses?

The open-ended questionnaire was distributed to the English teachers that taught Practical English I and II. The questions included: 1. What aspects should be included in the expected learning outcomes?; 2. What are appropriate assessment tools?; and 3.What is the weighted percent for each learning outcome?

TABLE 1 shows the findings obtained from the questionnaire and they were used to establish the course contents and its evaluation.

The teachers reported that 3 aspects should be included in the course contents and evaluation: 1. language ability, 2. the affective domain, and 3. The psychomotor domain.

Regarding language ability, the teachers agreed that students should be able to achieve CEFR levels of A2+ to B1. Moreover, they suggested that the students should grow in terms of feeling and emotion as well, so the affective domain was included [14]. Lastly, to provide the students with practical skills, the psychomotor domain or manual and physical skills must unavoidably be taken into account [14]. Table I below summarizes the expected learning outcomes reported by the English teachers.

TABLE I: WEIGHED PERCENT OF EXPECTED OUTCOMES

Expected Outcomes	Behavioral Objective	Assessment Tool	Weighed Percent
Language Ability (CEFR: A2+ - B1)	Listening: -Provided speech is clearly articulated in a general recognizable voice, identify simple factual information on specific everyday or work-related topics, which distinguishes both general and special messages. -Specify certain specific basic points of discourse for common matters regularly found in the fields of work, university, leisure etc.. Reading & writing: -Express comprehension of the main points that are obviously common ingredients on familial topics, which happen on a regular basis in work, university, leisure, etc. -Describe events and experience, desires, aspirations, goals and briefly describe circumstances that are likely to arise when traveling in an area where the language speaks.s	Listening Test	10
		Midterm Test	35
		Final Test	35
Affective Domain (adapted from Bloom's revised taxonomy)	-Follow the classroom's regulations in order to be autonomous learners through language activities and practice	Self-evaluation form: 1. class attendance 2. participation	5 5
Psychomotor Domain (adapted from Bloom's revised taxonomy)	-Communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field -Exchange, verify, and validate information, deal with less standard circumstances, and explain why something is a problem — express thoughts on more abstract subjects, cultural issues such as movies, books, music, etc. Exploit a variety of easy-to-understand languages to deal with most situations when traveling.	Interview foreigners (relevant to the learned topics: city living, food, shopping, sports, at home and relationships)	10

The second research question asked: How are the grading criteria established?

Ten English teachers teaching Practical English I and II were asked to consider the cut-off score for each grade by means of the Angoff method. They used the descriptions of the expected outcomes as a guide. After that, their cut-off score for each grade was collected and calculated by arithmetic mean.

TABLE II illustrates the final cutting score for each grade obtained from the teachers.

TABLE II: GRADING CRITERIA

Evaluation Point	Full Score	A	B+	B	C+	C	D+	D
Attendance	**5**	4.5	4.5	4	4	4	3.5	3
Participation	**5**	4.5	4.5	4	4	4	3.5	3
Listening	**10**	8.5	8.5	8	7.5	7	6.5	6
Interview Project	**10**	8.5	8.5	8	7.5	7	6.5	6
Midterm Test	**35**	28	25	22	18	17	14	12
Final Test	**35**	28	25	22	18	17	14	12
Cutting Point		82	76	68	59	56	48	42
Grade Range		82-100	76-81	68-75	59-67	56-58	48-55	42-47

The third research question asked: What is the concurrent validity of the students' study results and English language standard?

Thirty-four students took the TOEIC. Initially, the students were subcategorized according to their grades from the 2 courses. TABLE III presents the number of students in each subgroup according to the TOEIC compared to the CEFR.

Table III: Comparison Of English Results And Standardized Scores

TOEIC*	CEFR	A/A	A/B+	A/B	B+/A	B+/B+	B+/B	B/A	B/B	C+/A	C/B	D+/B	Total
945	C1												0
785-943	B2												0
550-784	B1	2		1									3
225-549	A2	11	2		4	2	1	3	2	2	1	2	30
120-224	A1								1				1
Total		**13**	**2**	**1**	**4**	**2**	**1**	**3**	**3**	**2**	**1**	**2**	**34**

*[15]

The table clearly shows that the majority of the students achieved a CEFR level of A2. Three students achieved B1.

As concurrent validity measures how well a new test correlates with a well-established test, it focuses on the extent to which the scores on a new measure are related to the scores from that well-established test administered at the same time [16].

The scores obtained from the English courses and the TOEIC scores were analyzed for their relationship by means of Pearson product-moment, and the prediction of the TOEIC scores by the scores obtained from the English courses was implemented by regression analysis.

In order to examine the relationship between the English courses and the TOEIC scores, hypothesis 1 was set.

Hypothesis 1: There is a significant relationship between the scores from the English courses and the TOEIC scores at the 0.05 level. (H1: $r_{xy} \neq 0$)

In order to test hypothesis 1, Pearson product-moment correlation coefficients were calculated and are presented in the table below.

TABLE IV: CORRELATION COEFFICIENTS

Pearson Correlation	TOEIC	English Courses
TOEIC	1	.50**
Sig. (2-tailed)		.005
N		34

** Correlation is significant at the 0.01 level (2-talied).

The findings from Table IV show that the correlation coefficient between the scores from the English courses and the TOEIC scores was 0.50. This coefficient value was considered to be a positively-moderate correlation. It can be said that the scores from the English courses had a moderate positive correlation with the TOEIC scores. That means that if one variable positively increases, there is a positive increase in another valuable at the moderate level.

Regarding predictability, hypothesis 2 was established to examine if the scores from the English courses could predict the TOEIC score.

Hypothesis 2: The scores from the English courses can significantly predict the TOEIC scores at the 0.05 level. (H2: B≠0)

A simple linear regression analysis was employed to assess the second hypothesis. Table V shows the model summary of the simple linear regression analysis.

TABLE V: MODEL SUMMARY OF THE SIMPLE LINEAR REGRESSION ANALYSIS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.50	.23	.20	88.23

Predictors: (Constant), English Courses
Dependent variable: TOEIC

The R coefficient of 0.50 is shown in Table V. The R-Square is 0.23. This means that the independent variable was able to explain 23% of the variance in the TOEIC score. The standard error of the estimate was 88.23.

Table VI illustrates the coefficients of the regression model.

TABLE VI: COEFFICIENTS OF THE REGRESSION MODEL

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	81.75	143.50		-.57	.573
	Eng Courses	2.65	.87	.475	3.05	.005

Dependent Variable: TOEIC score

The unstandardized coefficient B is the increase or decrease in the dependent variable when an independent variable increases by one unit. The regression equation used to predict the TOEIC scores can then be written as follows:

$$y = a + bx$$

TOEIC score = 81.75 + 2.65 (scores from English courses)

According to the above equation, the scores from the English courses could be used to predict the TOEIC scores. With a unit increase in the score, the TOEIC score will increase by 2.65 with the standard error of .87.

In summary, hypothesis 2, which states that the independent variable can individually or in combination significantly predict the TOEIC score at the 0.05 level (H2: B≠0), was accepted. If test takers know their scores from the English courses, it is possible that they will know their TOEIC scores.

CONCLUSION AND RECOMMENDATIONS

A. Conclusions

In order to assure that students who pass these English courses will be able to apply their gained knowledge to real world requirements, the expected learning outcomes of the English courses determined by the English teachers included CEFR levels of A2 + to B1. The other 2 aspects from Bloom's taxonomy (the affective domain and the psychomotor domain) were also integrated. It can be said that the courses' goals were to nurture the students—not only their language ability but also other necessary aspects.

In terms of the evaluation of the study results, the grading criteria established using the Angoff method allowed the English teachers teaching these courses to consider the cut-off score for each grade according to the descriptions of the expected outcomes as a guide. The grading criteria therefore are not only used to assign the student's study results, but the criteria can provide descriptions of the students' ability. Students receive descriptions for the grades that they get, and this means that they will get to know their strengths and weakness, which can be beneficial for the improvement of their abilities. Importantly, the established criteria can demonstrate if the course objectives or expected outcomes can be achieved.

In order to make inferences concerning the study results and real world standards, concurrent validity or measuring how well the scores correlated with a well-established test. This focuses on the extent to which the scores are related to the scores from the standard test was calculated.

Initially, it explicitly showed that the majority of the students achieved a CEFR level of A2. Three students achieved a B1 level. This means that 34 students that passed the 2 courses with at least one course with a B grade were able to achieve the expected outcomes regarding the expected language ability.

Regarding the analysis of concurrent validity, the correlation coefficient between the scores from the English courses and the TOEIC scores was 0.50. It can be summarized that the scores from the English courses had a moderate positive correlation with the TOEIC scores. That means that concurrent validity was detected, although the correlation coefficient was at a moderate level. The cause of this might have been the fact that the scores for the English courses were a result not only of the English tests themselves but were influenced by affective and psychomotor domain effects.

Since there was concurrent validity with the TOEIC scores, the regression analysis could be implemented in order to examine if the English scores could be a predictor of the TOEIC scores. An equation for predicting the TOEIC scores (TOEIC score = 81.75 + 2.65 (scores from English courses) was obtained. Hence, if a student's score from the English courses was known, it could be used to estimate his or her TOEIC score.

B. Recommendations

Once teachers and students are aware that learned contents are connected to the real world, what they have mastered in the classroom must be very meaningful. The measurement and evaluation of the learning achievements can be a way to signify the connection of classroom practices and real-world requirements. Consequently, outcomes-based assessment is highly recommended so that information is not only provided about the results of the students' study, but also about other aspects of their learning experience. The students' improvement and development can be implemented during the instructional processes.

In order to qualitatively evaluate concurrent validity, stakeholders such as employers, business owners, and teachers of other subjects should get involved in defining expected learning outcomes and the investigation of the students' learning achievements.

Last, for educators, the connection of real-world expectations and what is taught teach should be clearly relevant. Additionally, the contents of the classroom should be updated and more flexible. Once students achieve the course goals, they should feel or be certain that they are ready for real-world tasks.

ACKNOWLEDGMENTS

REFERENCES

1. El-Maaddawy, T., & Deneen, C. (2017). Outcomes-Based Assessment and Learning: Trialling Change in a Postgraduate Civil Engineering Course. *Journal of University Teaching & Learning Practice*, 1-19.
2. Rattananuntapat, M. (2015). *Quality Assurance Policies in Thai Higher Education*. Pittsburgh: University of Pittsburgh.
3. Carless, D. (2015). Excellent in university assessment: Learning from award-winning practice. Abingdon, UK: Routledge.
4. Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
5. k12academics. (2019, May 5). *Advantages and Limitations*. Retrieved from k12academics: https://www.k12academics.com
6. FairTest. (2019, May 20). *Norm-referenced Achievement Tests*. Retrieved from Fairtest: http://www.fairtest.org
7. Centre for Higher Education Research and Teaching. (2012). *A Brief Guide to Outcome-based Assessment*. Grahamstown: Rhodes University.
8. Mahbub, M. A. (2018). English Teaching in Vocational High School: A Need Analysis. *JEELS*, 229-258.

Retrieval Number: A4469119119/2019©BEIESP
DOI: 10.35940/ijitee.A4469.119119

4864

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

9. Khunmontri, S., Trakulkasemsuk, W., & Bunsom, T. (2014). Vocational College English Teachers' Awareness of Establishment of the ASEAN Community and Its Impact on English Language Teaching. *Journal of Liberal Arts, Ubon Ratchathani University*, 77-106.

10. Bresciani, M. J. (2011). Identifying Barriers in Implementing Outcomes-based Assessment Program Review: A Grounded Theory Analysis. *Research & Practice in Assessment*, 5-16.

11. Anwar, M. A., Ahmed, N., & Al Ameen, A. (2012). An Outcome-Based Assessment and Improvement System for Measuring Student Performance and Course Effectiveness. *Contemporary Issues in Education Research*, 279-294.

12. Wheaton, A., & Parry, J. (2012, March 20-23). *Using the Angoff Method to Set Cut Scores*. Retrieved from questionmark: www.questionmark.com

13. Dornyei, Z. (2011). *Researcch Methods in Applied Linguistics*. New York: Oxford University.

14. Clark, D. (2015, January 12). *Bloom's Taxonomy of Learning Domains*. Retrieved from Big Dog & Little Dog's Performance Juxtaposition: www.nwlink.com/donclark/hrd/bloom.html

15. ETS. (2011). *Correlation Table TOEIC Listening and Reading Scores Descriptors and the CEFR levels*. Retrieved from ETS: GLOBAL, A subsidiary of Educational Testing Service: www.etsglobal.org

16. Lin-L, W., & Yao, G. (2014). *Encyclopedia of Quality of Life and Well-Being Research*. Retrieved from Springer Link: www.link_springer.com

AUTHORS PROFILE

Supalak Nakhornsri was born in Thailand and earned a Ph. D. degree at Chulalongkorn University in Thailand in 1998. She is an Assistant Professor in the Department of Languages in the Faculty of Applied Arts at King Mongkut's University of Technology North Bangkok (KMUTNB). She has also been the Chair of the M.A. program in English for Business and Industry Communication for ten years. Her research areas include English language assessment and TEFL. Regarding the area of language assessment, her research paper entitled "The Development and Validity of Authentic English Reading Comprehension Computer-Adaptive Online Test" was published in the Selected Papers of CamTESOL 2008 and a paper entitled "The Varieties of English and Other Factors Affecting the Online Academic Listening Ability of EFL Students" was published in a journal in 2011. Her latest published article relevant to TEFL is entitled "Comparative Exploration of Learning Styles and Teaching Techniques between Thai and Vietnamese EFL Students and Instructors" (2016).