SWU-SET as a CEFR Standardized English Test

Chuenjit Athiworakun
International College for Sustainability Studies, Srinakharinwirot University, Bangkok, Thailand

Kriangkrai Vathanalaoha
International College for Sustainability Studies, Srinakharinwirot University, Bangkok, Thailand

Thanakorn Thongprayoon
International College for Sustainability Studies, Srinakharinwirot University, Bangkok, Thailand

Krich Rajprasit
International College for Sustainability Studies, Srinakharinwirot University, Bangkok, Thailand

Wachirapong Yaemtui
International College for Sustainability Studies, Srinakharinwirot University, Bangkok, Thailand

Abstract—**Language and Academic Services Centre, International College for Sustainability Studies, Srinakharinwirot University (SWU) understands and foresees the significance of using English in general and academic contexts. As parallel with international stanardized tests that are used to measure test takers' English proficiency, Language and Academic Services Centre attempted to construct Srinakharinwirot University Standardized English Test (SWU-SET) that can measure levels of English proficiency aligned with the concept of Common European Framework of Reference for Languages (CEFR). The aims of this paper were to test quality of SWU-SET based on validity, reliability, and item discrimination and to develop items of SWU-SET aligned with the CEFR. The methodology implemented in this study included a construction of test specification and a 100-MCQ test design under the investigation of test validity, test reliabity, and test item discrimination. As a consequence, two sets of the SWU-SET were launched for a pilot study. Consequently, all test items were revised and re-launched to confirm the test quality. Then, SWU-SETs were under the process of standard mapping which conformed to the CEFR. The results and discussion were presented in this paper to show how the test was designed and spelled out how the test yielded satisfactory results as a standardized test aligned with CEFR.**

Index Terms—**English proficiency test, CEFR, standard mapping, SWU-SET**

I. INTRODUCTION

English is undeniably a significant means of communication in the era of globalization, as well as a key to encounter the academic and professional worlds. Regarding English language proficiency, a variety of standardized test such as TOEFL, IELTS and TOEIC are designed to meet the demand for higher education, organizations, and language learners. The test scores which are used to underline the English proficiency of students as well as employees are used to assess the test takers' English proficiency for pursuing higher education or working in international organizations. Srinakharinwirot University has been adopting a variety of English standardized test as summative assessment for such academic and professional purposes; however, the university aims at developing an in-house test at the present. This leads to the development of Srinakharinwirot University Standardized English Test (SWU-SET) as to assess the test takers' English proficiency.

There are two major challenges for test developers to consider when developing a test. One is how to design the test to meet the standardized quality. These include the test validity, test reliability, and test items discrimination index (Bachman & Palmer, 2010). There are English language proficiency tests which are well-developed nowadays such as TOEFL, IELTS or TOEIC. SWU-SET will be only accredited or reliable if its quality achieves those standardized tests. The other challenge is how SWU-SET will be amalgamated by Common European Framework of Reference (CEFR) as its standardized criteria are accepted among nations in the world. Subsequently, it is vital that SWU-SET scores should comply with the CEFR.

According to the aforementioned, this leads to an urge for Srinakharinwirot University, to develop its own English language proficiency test, despite the major challenges uttered, to reach the standardized level and to be acknowledged internationally.

II. LITERATURE REVIEW

A. English Language Proficiency

The studies concerning the relationship between English language proficiency and academic achievement have been conducted for several decades. Intelligible English language has a direct impact on the learning process in school and is deemed a tool for effective communication contributing to a successful class. It is ascertained that language teachers are supposed to teach and train their learners to master all the four skills of listening, reading, writing and speaking (Oller, 1979). Also, developing proficiency in language enhances the process of learning to read and write (Donalson, 1978). When learners have the problems of learning to speak and learning to write, they tend to have academic failure (Bereiter & Scardamalia, 1982).

In the study of the relationship between bilingual students' level of English fluency and academic achievement, previous studies found that bilingual students who were fluent in English were likely do less well in school, but they generally outperformed their English-speaking peers significantly(Demie et al., 2003). Since then, a strong relationship between the comprehensive assessment and the academic English test has been found and constructively discussed (Kato et al., 2006).

On the contrary, there are arguments against the relationship between language proficiency and academic success. Students failed in school because the tests responding to their skills and knowledge were not grounded on authenticity (Burt & Dulay, 1978). In addition, a group of researchers intended to compare performance on two standardized English language proficiency tests with three measures of success: patient satisfaction; faculty and colleague evaluation; and scores on an objective test of medical knowledge (Eggly & Smulowitz, 1999). They found that only patient satisfaction and faculty and college evaluation were in relation to English language skills and that the scores could not predict students' practical medical knowledge.

B. Validity, Reliability, and Item Discrimination as Test Quality

The SWU-SET aims to test quality of English language proficiency based on validity, reliability, and item discrimination. The following concepts will be discussed because they relate to how the test quality of the SWU-SET was solidated.

Firstly, validity is considered to be a significant principle in language assessment because a test should be valid in order to measure what it is meant to measure. Validity refers to "the extent to which the results of an evaluation procedure serve the particular uses for which they are intended" (Gronlund, 1971, p.142). As validity is complex, there is evidence or types of validity to be considered: content-validity, criterion-related validity, construct validity, consequential validity, and face validity.

Secondly, reliability is also another vital indicator in language assessment because the results of the test should be consistent. Reliability refers to "the consistency of evaluation results" (Gronlund, 1971, p.80). If a test is reliable, its results will be similar over a certain time period with the same or different groups of test takers. For instance, if a test developer administers an achievement test to their test takers on one occasion, they should have the similar scores when taking the same test on different occasions (Brown, 2004). The degree of which test is reliable or consistent can be estimated by calculating a reliability coefficient which can go as high as +1.00 for a perfectly reliable test or as low as 0.00 for an unreliable test. If the test reliability score is, for example, valued 0.70, this means that the test scores are 70% consistent or reliable with 30% measurement error. Split-half method, Kuder-Richardson formula 20 (KR-20) and Kuder-Richardson formula 21 (KR-21). Split-half method is based on the division of odd and even numbers and carries out twice among the same group of test takers. Unlike split-half method, KR-20 and KR-21 are calculated based on mean and standard deviation and carried out once. When compared with the others, KR-20 is an internal-consistency reliability statistic that avoids the problem of understanding the reliability of certain tests and considered a much more accurate estimation of reliability (Kuder & Richardson, 1937).

Finally, item difficulty or item facility (IF) is a representation of statistics used to investigate the percentage of test takers who correctly answer a test item given (Brown, 2004). IF can be calculated by adding up the number of students who correctly answer a particular item and dividing the sum by the total number of the test takers who take the test. The range of IF can be valued from 0.00 (very easy) to 1.00 (very difficult). Although the ideal value falls into 0.50, the acceptable value for IF as to indicate high power of discrimination can be compromised from 0.20 to 0.80. The equation for the item difficulty is as follows:

IF = N correct/N total
N correct = number of test takers answering correctly
N total = total number of test takers taking the test

Item discrimination (r) is represented through statistics that indicates the degree to which an item separates the test takers performing well from the students doing poorly on the given test as a whole (Henning, 1987). The test takers performing well are generally referred as 'high' scorers or of 'upper' proficiency while those performing poorly are defined as 'low' scorers or of "lower" proficiency. A demarcation between high and low scorers helps test developers contrast the performances of the upper test takers on the test with the performances of lower ones. The item discrimination index can range from -1.00 to + 1.00. If the r is of +1.00 value, it means that all test takers in the upper groups answer correctly, while those of the lower groups answer incorrectly. However it is quite difficult that all test items will have item discrimination index of +1.00; it was suggested that the test items range from 0.40-1.00 can be considered 'very good' items, while those of below 0.19 are considered poor items or to be rejected or improved by revision (Ebel, 1979). The equation for the item discrimination is as follows:

r =   IF (upper) - IF (lower)
r =   item discrimination for an individual item
IF (upper) =   item difficulty for the upper group on the whole test
IF (lower) =   item difficulty for the lower group on the whole test

C. CEFR and Standardized Test

Language testing is to "pinpoint strengths and weaknesses in the learned abilities" of language learners (Henning, 1979, p. 1). There are various ways of teaching as well as ways of describing levels of language learning and assessment. At the present, educational institutions including schools and universities use different pedagogic methodologies and different systems to describe proficiency levels. What may be an intermediate level in one country may be an upper intermediate level in another. However, levels may vary in each country because of no global consensus of standardized criteria.

To facilitate teaching and learning, a way to specify what learners are able to do at certain levels is of necessity. As teachers, how these levels can guide teaching and selecting proper course books and resources should be taken into their consideration. In many countries, there is a general agreement that language learning can be organized into three levels: basic/beginner, intermediate, and advanced. In this regard, the Council of Europe developed the Common European Framework of Reference for Languages (CEFR) to establish international standards for learning, teaching, and assessment for all modern European languages. CEFR categorizes the language learners in terms of can-do statements into six specific levels (see Table 1 for holistic rubric): A1 and A2 levels as Basic User; B1 and B2 levels as Independent User; C1 and C2 levels as Proficient User. At this juncture, CEFR has been recognized in English language learning in order to establish guideline for English language educators and learners to abide by.

TABLE 1:
CAN-DO DESCRIPTORS OF COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES (CEFR)

Level	Listening	Reading	Speaking	Writing
C2	Has no difficulty in understanding any kind of spoken language, delivered at fast native speed	Can understand a wide range of long and complex texts	Has a good command of idiomatic expressions	Can write clear, smoothly flowing, complex texts in a logical structure
C1	Can understand enough to follow complex topics, though he/she may need to confirm details	Can understand in detail lengthy, complex texts, provided he/she can reread difficult sections	Can express him/herself fluently and spontaneously	Can express him/herself with clarity and precision
B2	Idiomatic usage influences the ability to understand	Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms	Can interact with a degree of fluency and spontaneity that makes regular interaction	Can express news and views effectively in writing
B1	Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure	Can read straightforward factual texts on subjects related to his/her field and interest	Can exploit a wide range of simple language to deal with most situations	Can write personal letters and notes asking for or conveying simple information of immediate relevance
A2	Can understand enough provided speech is clearly and slowly articulated	Can understand short, simple texts containing the highest frequency vocabulary	Can communicate in simple and routine tasks requiring a simple and direct exchange of information	Can write short, simple formulaic notes relating to matters in areas of immediate need
A1	Can follow speech which is very slow and carefully articulated, with long pauses for him/her to assimilate meaning	Can understand very short, simple texts, a single phrase at a time, picking up familiar names, words and basic phrases	Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech	Can ask for a pass on personal details in written form

D. Standard-setting Process

The standard-setting process applies to standardized tests has been recognized as the Benchmark Method (Faggen, 1994) or the Examinee Paper Selection Method (Hambleton et al., 2000). Despite several types of standard setting used in language testing nowadays, the standard-setting process generally comprises the panelists to score each test item based on individual judgment on percentage basis. On each proficiency level (e.g. B1, A2 and etc.), each panelist has to individually present perceived percentage that each 'just-qualified' test taker (i.e. the lowest proficiency test takers from a particular CEFR level) can do each test item correctly. To illustrate, the question that each panelist must be aware during the standard-setting process is how much percentage that test taker in each level of CEFR can do this test item correctly. As each panelist marks their score, the scores will be collected and publically posted for multiple subjective agreement.

III. RESEARCH METHODOLOGY

A. Research Questions

1. In what way does SWU-SET represent itself as a valid standardized test?

2. In what way does SWU-SET represent itself as a reliable standardized test?

3. How is the SWU-SET aligned with the CEFR?

Regarding the validity of SWU-SET, Index of Item-Objective Congruence (IOC) (Rovinelli & Hambleton, 1976) was used to affirm the test validity. As with the reliability of the test, it is acknowledged that not only a good test should be able to distinguish students' testing ability, ranging from high- and low-ability in a normal distribution (neither negative- nor positive-skewed curve), but the results of the test takers should be similar when they retake the test. To answer the third question, Angoff's multiple subjective agreement was implemented to set the SWU-SET score, based on the descriptors of CEFR (Angoff, 1971). Multiple subjective agreement is a qualitative data by nature, albeit consensus upon quantitative data based upon both percentage from each item from each panelist and the number of peer-review rounds.

B. Production of the SWU-SET and Angoff's Technique

1) Production of the SWU-SET 1 and 2

1.1) Production: Two sets of SWU-SET were developed according to the test specification. The test specification was specifically designed to assess the test taker's English four skills (listening, speaking, writing and reading) of English language as acquainted through conventional standardized tests. It was divided into five parts and each part comprised 20 items. At this juncture, World Englishes were incorporated into the content of SWU-SET to reflect the latest trend of English as an international language (Jenkins, 2009).

1.2) IOC from three English language testing experts was collected to check the content validity of SWU-SET. Each panelist can individually rate each test item according to these values: -1 (unsatisfactory), 0 (neutral), +1 (satisfactory). The IOC scores collected from each expert would average out and any test items would be revised if the IOC was lower than 0.5.

2) 1st piloting the standardized test: purposive sampling: mixed-ability students (1st selection) 300 students (High: Medicine, Moderate: Social Sciences, Low: Physical Education). 150 students were tested by SWU-SET 1 and the other half were tested by SWU-SET 2.

3) 1st item difficulty analysis and the test amendment: This step is used to revise the test items that are too difficult or too easy.

4) 2nd piloting the standardized test: purposive sampling: mixed-ability students (2nd selection) 300 students (High: Dentistry, Moderate: Social Sciences and Education, Low: Fine Arts). In a similar fashion, these students were be divided by half to test SWU-SET 1 (revised) and SWU-SET 2 (revised), respectively.

5) 2nd item difficulty analysis: this step is used to accompany while making a cut-off score for each item via standard-mapping method: Angoff's technique.

6) By using two sets of SWU-SET, we aimed to justify the reliability of SWU-SET 1 and SWU-SET 2 and to calibrate the tests as accordance with CEFR. KR-20 statistics was used to verify the reliability of both SWU-SET 1 and SWU-SET 2 due to the fact that different groups of students were used to compare the results between the try-out and revised tests.

7) Angoff's Technique: standard-mapping method based on multiple subjective agreement (MSA). This step has been recognized by TOEFL and remains the most widely used standard-setting method for selected-response tests (Cizek, 1993; Mehrens, 1995; and Hurtz, 2003). SWU-SET implemented the method via three steps: panelist selection criteria, panelist orientation, and panelist training before the actual standard mapping occurred. The criterion of panelist selection was based on teaching experience. For this pioneering project, all panelists were the lecturers of Language and Academic Services Centre at Srinakharinwirot Univeristy and had at least five years of EFL teaching experience at higher education levels. The panelists were provided with an overview of the purpose of the test and a definition of threshold scores (or cut scores) of "just-qualified" students, as applied to the current purpose. They were consequently trained to get familiarized with the key descriptors of *A2 B1 B2 and above* level of proficiency on the CEFR. Moreover, prior to the meeting, each panelist was individually given an assignment to review the CEFR with its holistic and analytic descriptors and to memorize key indicators for each level. Each level was defined in terms of the English language skills being measured by can-do statements reflecting language skills used in lived experience.

For each test item, each panelist was given charting papers and was asked to mark perceived percentage that just-qualified student could correctly do it each CEFR level. This practice was designed to bring the whole group to a shared agreement by thorough comprehending each of the CEFR levels. Each charting paper was posted and openly discussed so that the whole panel had an opportunity to comment and suggest plausible modifications through their education experience. The whole panel thus underwent the process of locating the most compatible percentage where this contributed to the standard-setting judgment for each test item of the SWU-SET. By doing so, this meant that any of the test items must fall into acceptable percentage range (< 30% range) from all of the panelists. If the range was greater than 30% percentage range, any panelist giving such the highest and the lowest percentage would be asked to clarify their judgement why they particularly rated the item. In a similar fashion, the other panelists would be able to explain and clarify why they marked their ratings. In due course, the percentage was set on the panelists' consent representing the minimum requirement for each CEFR level. The number of test items categorized into each CEFR level was consequently calculated and propelled how the cut-off score was established.

IV. RESULTS AND DISCUSSION

A. Validity and Reliability

Regarding the content validity of SWU-SET, IOC was used to affirm that the test items were specifically designed to serve its purpose for standardized test. The results from three experts in language testing pointed out that only 10 test items from SWU-SET 1 and 8 test items from SWU-SET 2 were below 0.5. However, they were either deleted or revised to improve the test quality. The betterment was apparently shown through higher power of distribution through the revisions of SWU-SET 1 and SWU-SET 2 as they could discriminate high- and low-ability students in accordance with the normal distribution.

As with the reliability of SWU-SET, the analysis of KR-20 could show high reliability of the test. At the outset, it could be seen from the try-out SWU-SET 1 and SWU-SET 2 that the KR-20 values were as high as 0.9359 and 0.9494 respectively. The revisions of SWU-SET 1 and SWU-SET 2 also showed high KR-20 values as 0.75372 and 0.8056 respectively. These values affirmed that SWU-SET was highly reliable. As with the second trial of the SWU-SET, the statistics revealed that the revisions of SWU-SET 1 and SWU-SET 2 could be able to distinguish high- and low-ability test takers. Note that immediate absence of students (271/300 and 273/300) was addressed in this research study and would mark as one of extraneous factors which researchers control by all means. The scores could be illustrated from Tables 2 and 3 below.

1) **Try-out** (271 participants)

TABLE 2:
SCORE OF TRY-OUT SWU-SET 1 AND SWU-SET 2

SWU-SET (1) MAX = 86 MIN = 25 KR-20 = 0.9359	Below A2 = 0 A2 = 74 B1 = 53 B2 and above = 6 No. of students = 133/271
SWU-SET (2) MAX = 91 MIN = 19 KR-20 = 0.9494	Below A2 = 1 A2 = 80 B1 = 34 B2 and above = 24 No. of students = 138/271

2) **Revised Test** (273 participants)

TABLE 3:
SCORE OF REVISED SWU-SET 1 AND SWU-SET 2

SWU-SET (1): revision MAX = 89 MIN = 19 KR-20 = 0.7537	Below A2 = 1 A2 = 53 B1 = 71 B2 and above = 13 No. of students = 138/273
SWU-SET (2): revision MAX = 96 MIN = 18 KR-20 = 0.8056	Below A2 = 1 A2 = 47 B1 = 66 B2 and above = 21 No. of students = 135/273

B. Cut-off Score: Standard Setting (A2, B, and B2 and above)

SWU-SET 1 and 2 could establish a close relationship with CEFR through Angoff's multiple subjective agreement. Apparently, while compared to the try-out examinations (Table 2), the revisions of SWU-SET 1 and SWU-SET 2 (Table 3) showed that the number of test takers was relatively close to the normal distribution. At this juncture, the cut-off score was drawn from the results of multiple subjective agreement and the consensus of three rounds of discussions. The number of A2-level test items was 22 while that of B1-level test items was 50. The last cut-off score was titled *B2 and above* due to its plausibility to assess C1 level. However, due to Angoff's technique, researchers found only 7 test items that fall into C1 level and this meant that any test taker would need to get 93 out of 100 items to reach C1 level. In terms of classifying division, the range from B2 to C1 (25 points) was relatively narrower and unequaled while compared to those of A2 and B1 (28 points) and of B1 and B2 (28 points) respectively. Nevertheless, this classification could inevitably imply test takers, whose scores were closer to 100 points, their English ability encompassing C1 level.

TABLE 4:
CUT-OFF SCORE FOR SWU-SET

A2	= 22
B1	= 50
B2 and above	= 78

C. Limitation and Recommendations

This research was conducted with scrutiny of language testing experts and provided pedagogic implications for standardized test aligned with CEFR; however, there were confounding factors that should be pinpointed for future research studies.

SWU-SET was originally piloted with the aim to develop standardized test as parallel with international standardized tests and mainly focused on content validity. It is rather intriguing to investigate other perspectives of validity. Construct validity may have been questioned as SWU-SET was designed to test only receptive skills, not any of productive skills, especially speaking and writing skills. Although SWU-SET has been designed to test those productive skills in the form of indirect test, it is still questionable whether it could predict authentic ability of test takers. Future testing format, as well as the readiness of test administration, for productive skills is required to measure authentic ability of language learners. It is also intriguing to have post-test interviews from diverse groups of test takers to underline the significance of consequential validity, face validity and backwash of SWU-SET to improve the quality of the test.

Regarding the sample size used in this research, samples through purposive sampling technique was subjective by nature and it could not guarantee that each student received equal background of English language training. Besides, the sample size was considered small by number (271 and 273 students for SWU-SET 1 and SWU-SET 2 respectively). At this juncture, more samples are needed to demonstrate a stronger claim of generalizability of SWU-SET to wider language learners. Most importantly, the morbidity rate found in this research study should be more concerned as this would eventually impact the reliability of SWU-SET.

V. Conclusion

SWU-SET was specifically designed through the acceptable standard-setting framework aligned with CEFR and proven by its validity and reliability. This is one of the pioneering projects in Thailand to measure learners' integrated skills of English through CEFR. SWU-SET is deemed an invaluable asset of Thailand's provision of English proficiency test because it could reduce considerable amount of expense while unnecessarily taking international standardized test for any reason. Last, but not least, SWU-SET provides a viable model for Thai government to see how the trend of World Englishes has been interspersed with one another and how the standardized language testing should be designed to comply with the global changes in general as well as in educational settings.

References

[1] 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.

[2] Bachman, L., Palmer, A. (2010). Language Assessment in Practice. Oxford: Oxford University Press.

[3] Bereiter, C. & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2). Hillsdale, NJ: Erlbaum.

[4] Brown, H. D. (2004). Language Assessment: Principles and Classroom Principles. Longman.

[5] Burt, M. & Dulay, H. (1978). Some guidelines for the assessment of oral language proficiency and dominance. *TESOL quarterly*, *12*(2), 17-192.

[6] Cizek, Gr. J. (1993). Reconsidering Standards and Criteria. *Journal of Educational Measurement*, *30*(2), 93-106.

[7] Council of Europe (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Council of Europe.

[8] Demie, F.; Taplin, A.; & Butler, R. (2003). Stages of English acquisition and attainment of bilingual pupils: implications for pupil performance in schools, *Race Equality Teaching*, *21* (2), 42-48.

[9] Donalson, M. (1978). Children's Minds. London: Fontana.

[10] Ebel, R. L. (1979). Essentials of Education Measurement. New Jersey: Practice Hall.

[11] Eggly, S., Musial, J., & Smulowitz, J. (1999). Research and Discussion Note The Relationship between English Language Proficiency and Success as a Medical Resident. *English for Specific Purposes*, *18*(2), 201–208.

[12] Faggen, J. (1994). Setting standards for constructed response tests: An overview (ETS RM 9419). Princeton, New Jersey: Educational Testing Service.

[13] Gronlund, N. E. (1971). Measurement and Evaluation in Teaching, (2nd ed.). New York: The Macllian Company.

[14] Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. N. (2000). Handbook for setting standards on performance assessments. Washington, DC: Council of Chief State School Officers.

[15] Henning, G. (1987). A guide to language testing. Massachusetts, Newbury House Publishers.

[16] Hurtz, G. M., & Auerbach, M. A. (2003). A meta-analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educational and Psychological Measurement*, *63*(4), 584-601

[17] Jenkins, J. (2009). World Englishes: A resource book for students (2nd Edition). London: Routledge.

[18] Kato, H., Hata, K., Yamamoto, H. and Yoshioka, T. (2006). Effectiveness of the weed risk assessment system for the Bonin Islands. In Koike, F., Clout, M. N., Kawamichi, M., De Poorter, M. and Iwatsuki, K. (eds). *Assessment and Control of Biological Invasion Risk*. Shoukadoh Book Sellers, Kyoto, Japan and IUCN, Gland, Switzerland, pp. 65-72.

[19] Kuder, G. F. and Richardson, M. W. (1937, September). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

[20] Mehrens, W. A. (1995). Legal and professional bases for licensure testing. In J. C. Impara (Ed.), *Licensure testing: Purposes, procedures, and practices* (pp. 33-58). Lincoln, NE: Buros Institute of Mental Measurements.

[21] Oller, J. W. (1979). Language Tests at School. Longman Group: United Kingdom.

[22] Rovinelli, R. J., and Hambleton, R. K. (1976, April). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research*, 2, 49-60.

Chuenjit Athiworakun is a lecturer at Language and Academic Services Center, International College for Sustainability Studies, Srinakharinwirot University. She graduated from her Master Degree in Educational Linguistics from Srinakharinwirot University, Thailand. She is now pursuing a PhD at Thammasat University, Thailand. Her research interest includes linguistics and language testing.

Kriangkrai Vathanalaoha received MA (English Language and Literature) from Thammasat University and Certificate of Postgraduate Research in Applied Linguistics at Lancaster University, UK. He is now a PhD Candidate at Language Institute, Thammasat University, Thailand. He is a lecturer at Language and Academic Services Center, International College for Sustainability Studies, Srinakharinwirot University. His research interests include cognitive poetics, corpus linguistics, genre analysis and pedagogical stylistics.

Thanakorn Thongprayoon holds an M.A. in TEFL and is now pursuing his Ph.D. in Science of Education and Learning Management. He has been a lecturer of English and a teacher trainer. And now he is Associate Dean for Academic Affairs at International College for Sustainability Studies, Srinakharinwirot University. His research interests vary from teaching professional development, English language teaching and learning, and intercultural communicative competence to media literacy.

Krich Rajprasit is a lecturer at a lecturer at Language and Academic Services Center, International College for Sustainability Studies, Srinakharinwirot University. He completed his MA in Business English for International Communication in Thailand. His research interests are English for Specific Purposes (ESP), Intercultural Communication and Communication mobility.

Wachirapong Yaemtui is currently a lecturer at Language and Academic Services Center, International College for Sustainability Studies, Srinakharinwirot University. He received B.A. (English Language and Literature), B. Pol. Sc. (International Relations), and M.A. (Teaching English as a Foreign Language) and is now a PhD Candidate (English Language Teaching) at Thammasat University. His research interest includes World Englishes, English as a Lingua Franca, Intercultural Communication, and English learning and teaching strategies.