DEVELOPING AN ENGLISH PERFORMANCE TEST FOR INCOMING INDONESIAN STUDENTS

Bill Glenny Wullur[1]
Universitas Klabat, Manado
email: billwullur@yahoo.com

Abstract: This study constructed, developed and validated an English Performance test designed to provide a valid criterion-based measurement within the placement programs in Indonesian universities. The steps in developing a performance test involved conducting needs analysis, establishing attributes of good performance test, and constructing test specification based on current language testing theory. The test was validated in several ways: (1) face validation, (2) content validation, (3) concurrent validation, and (4) construct validation. The reliability of the test was established using the internal consistency index, and the practicality of the test was established by distribution of score of the pilot tests. In general, the proposed performance test is found valid, reliable and practical. The output of the study is a resultant and valid English Performance Test for Incoming Indonesian College Students. The test is named: Academic English Performance Test or AccEPT.

Keyword: Performance test, Alternative Assessment, Test development, Validation study, communicative testing.

The changes in the theory of language create a new demand of language test—a performance-based assessment in language education. Recently there has been a growing interest in the importance of context in defining performance conditions of students (Weir, 1993). With the advent of the communicative movement in the area of language teaching nowadays, the prominence of language proficiency now is not simply in the competence in linguistic theories, but on the use of language in real communication—the whole embodiment of communicative competence (McNamara, 1996).

The strength of performance tests lies in the fact that they are direct, authentic and highly contextualized. This is because their very design "depends on using stimulus materials and posing tasks to the learners that are based directly on the learners intended (or hypothesized) use of target language" (Bailey, 1998). This communicative/real-life approach in testing is characterized by Weir by the following features: focus in meaning, contextualization, activity processing, genuine stimulus material, authentic operations of text, unpredictable outcomes, interaction based, and the judgment of communicative purposes. Nevertheless, the most dominant characteristic of second language performance test is that assessment will take place when the candidate is engaged 'in an act of communication' (McNamara, 1996). Performance in this sense is not limited to the performance of real world-

---

[1] Full time Teacher & Dean of Faculty of Education, Universitas Klabat, Airmadidi – Manado, Indonesia. He got his MA in Language Education from The University of the Philippines, Diliman,Quezon City in 2004

tasks or the demonstration of skilled execution, but also in the wide sense of the ability to use language in real communication

The nature of any placement test is indeed a proficiency test. A placement test is "designed to assess students' level of language proficiency so that they can be placed in the appropriate course or class" (Alderson, Clapham and Wall, 1995). It is intended to "define a student's language proficiency relative to the levels of a particular program he or she is about to enter" and thus "elicit information that would assign students to one of the levels in the program or exempt them from further ESL instruction" (Bailey, 1998). "In recent years, proficiency tests have increasingly sought to include performance features in their design … because performance is indeed the true evidence of language proficiency" (McNamara, 2000).

A good placement test must take into account the real proficiency—performance of the language in use—of the students. It must be sufficient to contain both the formal system of the language as well as the functions where the language operates. Especially in the context of foreign language teaching, when it is not unusual to notice that there are students with the knowledge of good grammar but lack communicative skills.

This study develops and validates an English Performance test to serve as entrance and placement test in colleges in Indonesia. The test was put side by side with the existing placement test of Universitas Klabat, which has these quality: It is an indirect testing of communication skills, vocabulary and grammar, the format is multiple-choice, while the test's items are discreet and decontextualized; the content is based on grammatical sequence, accuracy oriented, and to conclude, it is a traditional paper and pencil test. Therefore, the proposed test should have the characteristics of: direct testing of academic proficiency, written and oral performance format, authentic student's utterances, and its content is academic task based situation and contextualization, and communicatively oriented. After all, it must be a performance of demonstrated proficiency

The purpose of this research is to design a valid performance test to assess the proficiency of Indonesian students as they enter a college. In the light of communicative competence, the researcher attempts to provide a valid measurement tool to describe the students' proficiency into logical criteria and thus corresponding levels in the placement programs. The study aims to answer the following: 1) How is validity established in the performance assessment? 2) How is reliability established in the scoring of the performance assessment? 3) How is the banding system capable for practical placement procedures? 4) How are the sections of the performance assessment related to each other and to the whole test? 5) What is the strength of relationship of the scores of performance test when put side by side with the scores on the existing placement test? 6) How do the raters interpret the rating system?

This study will benefit Indonesian Higher Education Institutions with a performance test which is developed according to the program objectives. The result this study will help improve the quality of placement decision made in this institution, by reflecting the proficiency of their prospective students in performance-related criteria. In Indonesia, private universities need placement tests to maintain their reputation of the quality of education. These institutions attempt to accommodate every levels of language proficiency by developing its own placement test in their rivalry to compete each other. This study will benefit them since it will give them a working framework in the advancement of institutional placement test.

Furthermore, this study is not designed to be an end for itself, but is geared toward a

standardized proficiency test. The development this test will later provide the private educational institutions in Indonesia a balanced proficiency test: It will fill the gap of research in the integrated proficiency test of Indonesian college students. Specifically, this research will answer the needs of an adequate placement test in Indonesia.

Performance Test

From the early 1970s, a new theory of language and language use began to exert a significant influence on language teaching and potentially on language testing. This was Hymes's theory of communicative competence, which greatly expanded the scope of what was covered by an understanding of language and the ability to use language in context, particularly in terms of the social demands of performance. The feature of communicative language tests ultimately came to have two features: (1) they were performance test, requiring assessment to be carried out when the learner is engaged in an extended act of communication, either receptive or productive, or both; (2) they paid attention to the social roles subjects were likely to assume in real world settings, and offered a means of specifying the demands of such roles in detail (McNamara, 2000). Mc Namara, then concluded that performance assessment found a rationale in the theory of communicative competence.

In the late 70s and early 80s, Cummins claims that language proficiency is not a unitary construct but he proposed a distinction between cognitive academic language proficiency (CALP) and basic interpersonal communicative skills (BICS). He finds that most Non-English background children will acquire superficial BICS in the first two years of learning English. They will take longer, however, to reach grade level CALP achievement. Measurement of BICS may not provide any indication of the CALP skills in children (discussed in Brown, 1984). The later is the dimension of proficiency in which the learner manipulates or reflects upon the surface features of language outside of immediate interpersonal context. BICS shows in what learners often use in classroom exercises and test that focus on form. The former, on the other hand, is the communicative capacity that all children acquire in order to be able to function in daily interpersonal exchanges (discussed in Brown, 2000).

Afterward, many writers have tried to specify the components of communicative competences in second languages and their role in performance. This has been done in order to provide a comprehensive framework for test development and testing research, and a basis for the interpretation of test performance. Various aspects of competence were specified in the early 1980s by Michael Canale and Merrill Swain (1980) as:
1   Grammatical competences, which covered the kind of knowledge of systematic features of grammar, lexis, and phonology familiar from the discrete point tradition of testing;
2   Sociolinguistic competence, or knowledge of rules of language use in terms of what is suitable to different types of interlocutors, in different settings and topics;
3   Discourse competence or the ability to deal with extended use of language in context; and
4   Strategic competence, or the ability to compensate in performance for incomplete or imperfect linguistic resources in a second language.

The elaboration of Canale and Swain's models of abilities underlying performance has been helpful for both mapping research in language testing and classifying language tests, and providing language test developers and researchers with a common language to talk about the focus of their work (McNamara, 2000).

Rooted on Canale and Swain model, Bachman (1990) elaborates and rearranges the categorization of language competence, and propose another model of components of communicative language ability in communicative language use. This framework includes three components of communicative language ability (CLA), i.e. language competence, strategic competence, and psychophysiological mechanism. Language competence comprises, essentially, a set of specific knowledge components that are utilized in communication via language as seen in Figure 1 below.
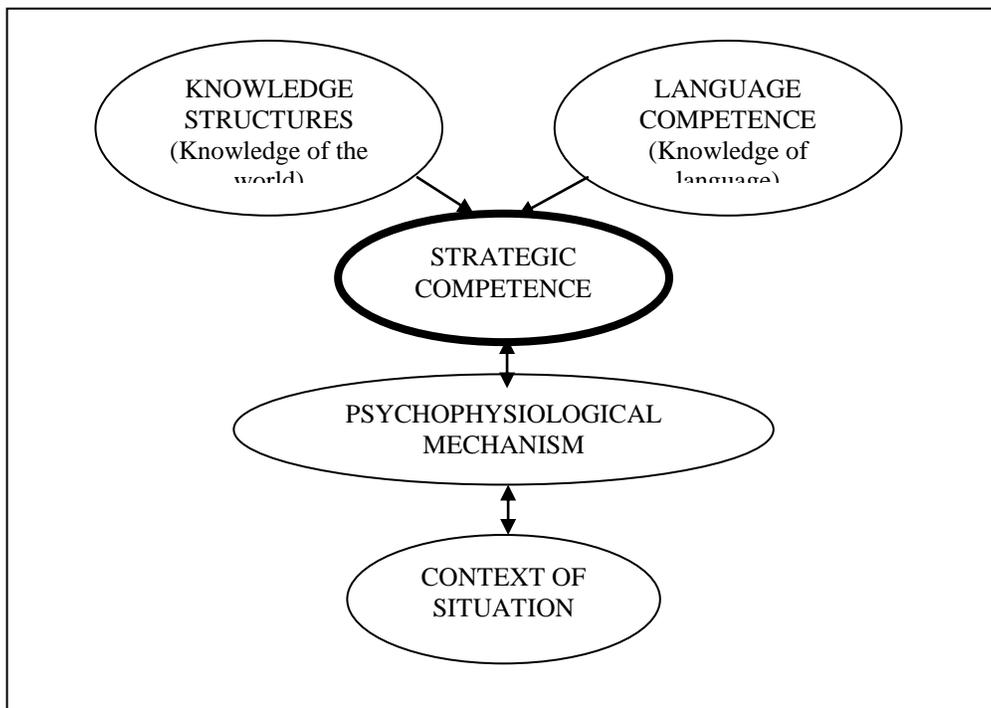

Figure 1.   Components of communicative language ability in communicative language use (Bachman, 1990).

Strategic competence is the mental capacity for implementing the components of language competence in contextualized communicative language use. Strategic competence thus provides the means for relating language competencies to feature the context of situation in which language use takes place and to the language user's knowledge structures. Psychophysiological mechanism refers to the neurological and psychological process involved in the actual execution of language as a physical phenomenon.

Skehan believes that Bachman's model of communicative competence and performance will continue to be of pivotal importance in the years to come as it is theoretically well-founded and comprehensive and it tries to be consistent with a wide range of applied linguistic theorizing (discussed in Kaur, 1999). According to Brown (2000), Bachman schematization of language competence and communicative language ability is integrative, that it captures the modification of the Canale and Swains' model of communicative competence.

Vecchio and Guerrero (in Kaur, 1999) conclude this matter by saying each of these educationally driven concepts of language proficiency shares at least two critical features. First, each definition accommodates the linguistic performance skills. Second, each definition places language within a specific context, in this case the educational setting. Consequently, an English proficiency test should utilize testing procedures that replicate – as nearly as possible- the kinds of contextualized language processing that is used in English speaking classrooms. The study recognizes the specific educational setting in Indonesia as a EFL country and its approach toward English especially in college level—in which an entrance and placement test was administered. The threshold level of student proficiency, which include the minimum requirement of reading ability to understand academic text (i.e. textbook) must first be defined and elaborated in comprehensible criteria.

This research recognizes language proficiency as combination of different constructs of communicative language ability: one of them is Cummins' distinction of CALP and BICS. The former is the construct to be measured in this study. The study recognized the attributes of good performance testing, which include (1) direct testing of academic language proficiency, which requires the students to perform precisely the skills to be measured, (2) integrative testing, which requires combination of language elements in the completion of a task, (3) combination of objective and subjective method of scoring, (4) task-based content, where the whole test is seen as related series of task in a situation, (5) orientation in communication, or assessment of language in use, (6) realistic students' generated text, and (7) authentic academic genre. Table 1 below compares the attributes of a good performance test and those of the existing placement test.

Table 1   Comparison of the Attributes of Existing Placement Test and those of a Good Performance Test

Attributes of the Existing Placement Test	Attributes of a Good Performance Test
Indirect testing of communication skills, vocabulary and grammar: attempts to measure the abilities which underlie skills in which we are interested.	**Direct testing** of academic proficiency, vocabulary and grammar: requires the subject to perform precisely the skill we want to measure.
Discrete point: testing of one element at a time, item by item	**Integrative testing**: Requires the subject to combine many language elements in the completion of a task
Objective method of scoring: no judgment is needed	**Combination of Objective and Subjective** method of scoring: need of judgment of criteria
Grammar based content: item is arranged according to grammatical sequence, and thus decontextualized from the passage.	**Task based** content: task is maintained in a bigger context and situation. The whole test is seen as related series of tasks.
Orientation in **accuracy**: scores obtained from the number of mistakes.	Orientation in **communication**: scores obtain from the degree of communication of task and ability. Assessment of language in use

The format is **multiple-choice**: students only select from tester's made utterance.	The format is **performance** test: students generate authentic utterances.
Genre **varies** from everyday communication to occupational and professional talk.	**Authentic academic** genre

Finally, the performance test was developed in the light of the mission objective and the attributes of good performance test. The developed performance test consists of four sections each assessing a distinctive skill. They are: Formal Letter assessing reading and writing skills, Application Form assessing vocabulary skill, Making Appointment assessing listening skill, and Oral Presentation assessing speaking skill. The whole process of development and validation of the performance test takes place under the theories of communicative language testing, CALP testing, and alternative assessment. The process is shown in Figure 2 below.
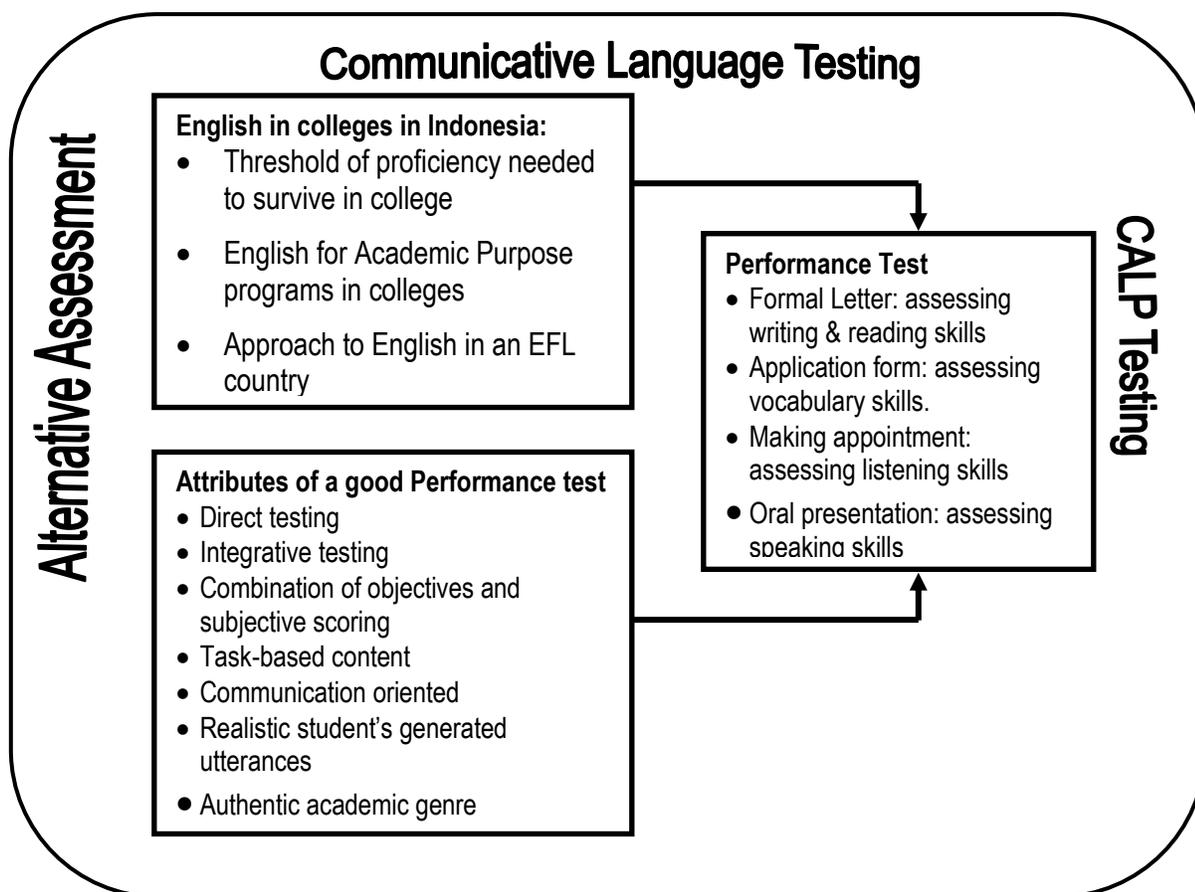

Figure 2.   Conceptual Framework Of The Study

METHODOLOGY

Research methodology in this study uses a combination of a priori and post priori approach for validation study. A priori approach puts an emphasis on the content validation, while post priori approach focuses on empirical validation subsequent to trialing. In this study, the there are six validation analyses identified before, during, and after the trialing, namely, content validation, face validation, concurrent validation, construct validation, internal consistency and practicality of the proposed test. The validation procedures included in research methodological flowchart is seen in Figure 3 below.
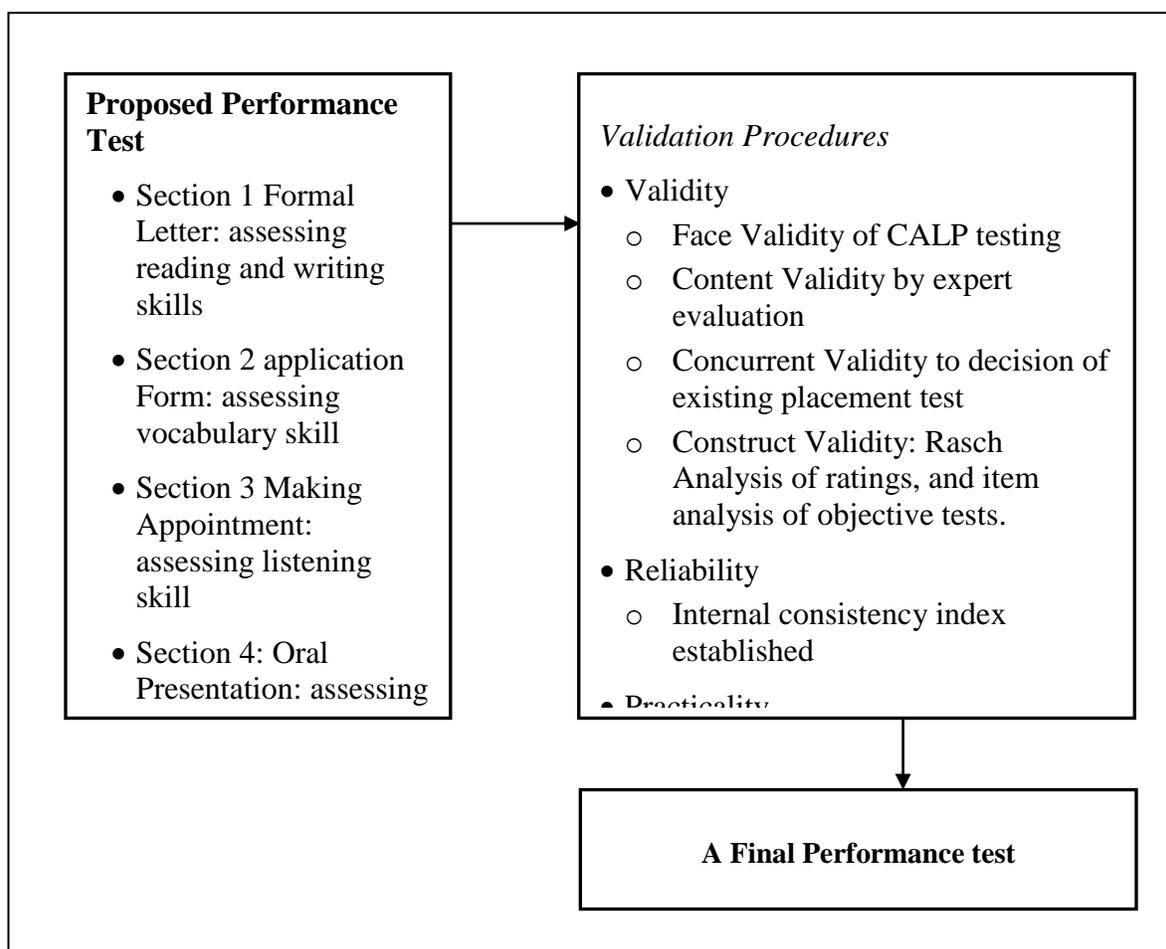

Figure 3   The Methodological Flowchart

The process of data collection procedure is classified under four major steps, namely, item generating, pilot study, item reducing, and field validation of the performance test material. Table 2 below guides the process of developing and validating the performance test.

Table 2 –   Steps of the Process of Developing and Validating the Performance Test

Steps	Procedure	Statistical treatment
1. Item Generating	• Needs analysis • Construction of test specification • Construction of test material • Content validation • Feedback	• Expert judges rating and educational background of judges
2. Pilot Test	• Raters training • Pilot Testing to comparable students • Rating of the comparable students	
3. Items Reducing	• Construct validation • Modification	• Item analysis: (Item facility & Item discriminability) for objective test Rasch Analysis for subjective test
4. Field Validation	• Raters training • Field Testing to subjects • Rating of the subjects • Reliability index • Face validation & Practicability • Concurrent validation • Revision	• Kendall's tau-b, and Kendall's W test. • Spearman's rho test • Gamma test

FINDINGS AND DISCUSSION

This study develops and validates an English Performance test as an entrance and placement test for college students. It is designed to provide a valid criterion-based measurement within the various placement programs in universities. The result of the study is a proposed performance test is categorized in four sections as follows: (1) Formal letter—the subjects initiate business correspondence (letter of inquiry and nomination for the financial assistance), (2) Application form—the subjects complete entries in an application form by using personal data, (3) Making Appointment—the subjects listen to recording and take down key information, and (4) Oral Presentation—the subjects generate a pledge of loyalty, give short

talk on programs to be implemented in their schools, express appropriate feeling, and use social formulae in opening and closing a conversation/speech.

Test Specification

The specification of this proposed test is arranged according to Carroll's 1980 Test Specification of Communicative Performance format. It describes the test in ten features of design, namely: (1) the nature of test subjects; (2) the events and activities; (3) modes of communication, (4) channels of communication, (5) topic areas, (6) test tasks and each corresponding format, (7) language functions, (8) language skills, (9) attitudes and relationships of the communication in the test, and (10) dialect and style. The test specification is seen in Table 3 below.

Table 3 - Test Specification

Features	Test Sections	Description
1. Test Subjects	All	Newly admitted Indonesian students at Universitas Klabat ready to be placed to four levels of an ESL program. The test is designed to assess academic English of the students.
2. Events and Activities Event: A charity is offering several financial assistance plans for the newly admitted students of an institution. Activities: Steps to do to obtain financial assistance.	Section 1: Formal letter	Initially, the subjects are introduced to a brochure containing information on financial assistance. Then, the subjects are required to write a formal letter to ask for more information, and to nominate themselves for the financial assistance plans *Activity: Written correspondence*
	Section 2: Application form	Gaining more information about the financial assistance plan, subjects are required to fill an application form using their personal data. *Activity: Written correspondence (application form)*
	Section 3: Making appointment	Subjects are required to make an appointment for scheduling their presentation by listening to the voice recording on the phone and note down the appointment number, schedule, and the presentation proper. *Activity: Information from a record*.
	Section 4: Oral Presentation	In order to demonstrate their eligibility to obtain different financial assistance plans, the subjects are required to: • Compose and present a pledge of loyalty to the charity • Prepare and present a short speech of ideas on what academic program should be implemented in order to help the teacher/students better • Express appropriate feelings upon hearing the result. *Activity: Oral presentation/interview*

3. **Modes** The test requires students to demonstrate their proficiency in all four modes of communication.	Formal Letter	Writing mode Focus on writing and grammatical skill
	Application Form	Reading and Writing modes, focus on vocabulary skill
	Making Appointment	Listening and Writing modes, focus on listening skill
	Oral Presentation	Speaking mode, focus on academic discourse.
4. **Channels** The channels of communication used here is print and oral presentation.	Formal Letter	Print: brochure, formal letter
	Application Form	Print: Guidelines of the plans, and Application form.
	Making Appointment	Records: Language laboratory, Print: Schedule paper.
	Oral Presentation	Oral Presentation: Pledge of loyalty, short talk, feeling expression.
5. **Topic Areas** Academic: applying And demonstrating eligibility for a financial assistance plan.	Formal Letter	Plans of financial assistance
	Application Form	Personal Information
	Making Appointment	Making an appointment
	Oral Presentation	Demonstration for eligibility
6. **Test Tasks and Formats** Generate academic writing and oral discourse.	Formal Letter	Initiate business correspondence; use information in poster to write a letter of inquiry and to nominate themselves. Format: write a formal letter
	Application Form	Complete an application form, use personal data to match information items. Format: Fill Application form.
	Making Appointment	Listen to recording, Take down key information Format: take down note
	Oral Presentation	Generate a pledge, present a short talk, Put ideas in sequence, extract factual information, express feelings, use social formulae Format: oral presentation

7. Language Functions	Formal Letter	Introductory, formulae, requesting, informing, and showing causation and purpose.
	Application Form	To inform, to state
	Making Appointment	To abbreviate, to show time, purpose and certainty
	Oral Presentation	Greeting, to inform, to state, to explain, to thank, to express regret, to show certainty, causation, and purpose.
8. Language Skills	Formal Letter	(1) Identifying the main point and purpose of text, (2) using appropriate grammatical cohesion devices, (3) using appropriate lexical cohesion devices, (4) deducing meaning of unfamiliar lexical items, (5) understanding and using explicit language
	Application Form	(1) Identifying the main point and purpose of text, (2) using appropriate lexical cohesion devices, (3) reducing information into several words, (4) deducing meaning of unfamiliar lexical items, (5) understanding and using explicit language
	Making Appointment	(1) Identifying the main point and purpose of text, (2) understanding conceptual meaning especially time, location, instrumental, and numbers, (3) reducing discourse to important information, (4) transcoding record into notes and to fill up charts or grids, (5) identifying the important information in a piece of discourse, through: voice underlining, and verbal cues.
	Oral Presentation	(1) Using appropriate grammatical cohesion devices, (2) using appropriate lexical cohesion devices, (3) understanding and using explicit language, (4) understanding and using appropriate sociocultural convention in speaking, (5) pronouncing and articulating of comprehensible communication, (6) giving relevant and adequate response to task.
9. Attitudes and Relationships	Formal Letter	*Attitudes:* formal, certain, and distant. *Relationships:* stranger to stranger, student to administrator
	Application Form	*Attitudes:* formal, certain, and distant. *Relationships*: stranger to stranger, student to administrator
	Making Appointment	*Attitudes:* formal, certain, and distant. *Relationships:* stranger to stranger, student to administrator
	Oral Presentation	Attitudes: formal, certain, and distant. *Relationships:* stranger to stranger, student to administrator
10. Dialect and Style	Formal Letter	Dialect: International English. *Style:* Academic Correspondence
	Application Form	Dialect: International English, some personalization of Indonesian content. *Style:* Academic Form

	Making Appointment	Dialect: International English. The voice of an American (Native Speaker). *Style:* Public announcement
	Oral Presentation	Dialect: International English, Indonesian Speaker accent. *Style:* Formal Presentation

The test was validated in several ways: (1) face validity by comparing the scores of Indonesian studying in the Philippines and those studying in Indonesia. The findings shows that the score is highly correlated at Spearman $\rho = .85$. (2) The content validation relies on the evaluation from four expert informants, and a panel of evaluators, which included of ten professors. The finding shows that the content coverage and relevance of the test is highly satisfactory. (3) The concurrent validation i to existing placement test of Universitas Klabat is also conducted and found that the tests are highly correlated at .964 using gamma test. (4) The construct validation is conducted in two ways: item-per-item analysis in objective section of the test, and Rasch Analysis in the Subjective Section of the test. The findings of item analysis revealed that there are only three items considered "fair items", while the remaining are "good items" and "very good items". The finding of Rasch Analysis showed that the raters could interpret the rating system and that the test is well fit and accurate. In general, the proposed performance test is valid. Table 9 below wraps up the validation of the test.

Table 9 - Summary of the Validation Study of the Test

Validation study	Statistical Treatment / Instrument used	Result	Finding	Interpretation
Face Validation	Percentage of distribution of scores	Spearman rho .85	Comparable distribution between Indonesian pilot test respondents	The test assess CALP
Content Validation	Evaluation sheet, with quantification from 1 to 5	Average point overall is 4.08	The content of the test is highly satisfactory	The content relevancy and coverage of the test is valid
Concurrent Validation	Gamma test	.964	The proposed test is highly correlated to existing placement test	Both test measure the same construct

Construct Validation	Item Analysis for Objective Sections	58% very good items 30% good items 12% fair items	The objective test is well fit for a criterion reference test	The construct of objective test sections of the test is valid
	Rasch Analysis for subjective sections	Item difficulty index ranging from −1.426 to 2.129	The raters leniency and the item difficulty is consistent	The raters interpret the test well
		Q-index ranging from .000 to .025	The error of measurement is very low	Test is accurate in measurement
		Zq index spread	No overfit items	Construct are measured independently
		P index ranging from .097 to .904	No misfit items	Test items are well fit

The reliability of the test was established using the internal consistency index. The findings revealed that the internal consistency of all sections of the test is highly consistent at Kendall's Coefficient of Concordance of .927. The correlation between sections to overall scores is also calculated, and the findings revealed that all sections are highly correlated at Kendall's tau-b correlation coefficient at .942 (Formal Letter), .934 (Application Form), .917 (Making Appointment) and .862 (Oral Presentation). The practicality of the test was established by analyzing the percentage distribution of score of the pilot tests and field test. The findings show that the scores are well distributed in all four levels, and a pattern of distribution exists with correlation at ρ .85. Therefore, the rating system is practical and predictive.

CONCLUSION

One of the useful advances in the area of second language teaching is a synchronization of methods of teaching and testing. The prominence of communicative language teaching and content-based instruction needs a better way of assessing the teaching than the traditional multiple-choice test. Therefore, performance test as well as portfolio assessment is the alternative assessment of product and process. The feature of a good performance-based English test are: (1) Direct testing—directly perform the skill desired to be measured; (2) Integrative testing—mixture of language elements in the completion of tasks; (3) Combination of objective and subjective method of scoring; (4) Task-bases content—task is maintained in a bigger context; (5) Orientation in communication—the scores are obtained from the degree of

communication of task or language in use; (6) Productive—the students elicit and generate their own utterances; and (7) authentic academic genre.

The study also concludes that a performance-based test can be validated through gathering evidence of validity in five ways: (1) face validation of academic tasks through association of students' performance, (2) content validation through expert evaluation, (3) construct validation through Item analysis for objective test and Rasch Analysis for subjective ratings, (4) concurrent validation by comparing the scores to a criteria, and (5) reliability index by establishing the internal consistency of the test sections, and overall.

This study also concludes that the rating system in any valid performance test should have a well-define scoring rubric which is interpretable by the raters. A well define rubric identify and characterize the construct in which the language elements are transparent. This is to ensure that the raters measure the skills in which we are interested, absolutely not the unnecessary skills. In addition, this is to ascertain limited space for raters' leniency over scoring. A set of well-defined level objectives in a language program is a valid criteria in assigning students' placement to the program. In general, the study concludes that a criterion-referenced performance test, which is academic, authentic, and productive, is the answer to the need of synchronizing communicative teaching and communicative testing.

REFERENCES

Alderson, Charles J., Clapham, Caroline & Wall, Dianne (1995). *Language Test Construction and Evaluation*. Cambridge University Press.

Bachman, Lyle F. (1990). *Fundamental Considerations in Language Testing.* Oxford University Press.

Bailey, Kathleen M. (1998). *Learning about Language Assessment: Dilemmas, Decisions, and Directions*. Boston: Heinle & Heinle Publishers.

Brown, H. Douglas (2000). *Principles of Language Learning and Teaching*. (4th Ed.). New York: Addison Wesley Longman Inc.

Canale, Michael. And Swain, Merrill (1980). Approaches to Communicative Competence. *Occasional Paper* No 14. SEAMEO Regional Language Centre.

Carroll, Brendan J. & Hall, Patrick J. (1985). *Make Your Own Language Tests: A practical Guide to Writing Language Performance Tests*. Oxford, UK: Pergamon Press Ltd.

Carroll, Brendan. J. 1980. *Testing Communicative Performance.* Oxford: Pergamon.

Kaur, Sarjit (1999). Problems in Assessing Proficiency in English among Foreign Postgraduate Students. Unpublished paper presented to "Sixth International Literacy and Education Research Network Conference on Learning", Penang, Malaysia, September 1999. [Available FTP: http://ultibase.rmit.edu.au/Articles/may00/kaur1.htm]

McNamara, Tim. (2000). *Language Testing*. Oxford University Press.

McNamara, Tim. (1996). *Measuring Second Language Performance*. New York: Addison Wesley Longman.

Weir, Cyril J. (1993). *Understanding & Developing Language Tests*. UK: Prentice Hall.

Weir, C. J., and Robert, J. (1994). *Evaluation in ELT*. Oxford, UK: Blackwell Publishers.