

## Performance Assessment: Use of Self-assessment(SA) for the Communicative College English Program<sup>1)</sup>

Jihyeon Jeon

(Catholic University)

Jeon, Jihyeon. (1998). Performance assessment: Use of self-assessment (SA) for the communicative college English program. *English Teaching*, 53(3), 179-197.

This study examines the possibility of using self-assessment (SA) in college EFL programs. The major purpose of SA is to provide the opportunity for learners to recognize the performance goals and to reflect on their progress toward established goals, using their metacognition. SA seems to be useful since it is time and cost efficient; it yields a positive washback effect; it gives learners greater control over their learning; it enhances learner awareness of the learning process; and it encourages autonomous learning. An exploratory SA instrument has been developed for the group of college learners under the study. The developed SA instrument is then assessed in terms of test/retest reliability and internal consistency. The validity of the SA instrument is checked by concurrency with the other proficiency measures. Scores on the SA test are compared to the learners' ability scores obtained through IRT (Item Response Theory) model on a pseudo-TOEIC test, a teacher-made achievement test, and a cloze test. Then, affective variables that perverse in learners' performance measured through SA are investigated. Results are discussed in light of practical issues in the application of the SA test into college EFL programs.

---

1) The author wishes to acknowledge the financial support of KRF (Korea Research Foundation) made in the program year of 1997. The author is greatly indebted to Professor Namshik Park in writing this manuscript for his inspiring suggestions and generous encouragement.

## I. INTRODUCTION

As English becomes an important medium for international communication, college English curriculum has been evolving gradually toward a more performance-based approach. College English, a mandatory course for college freshman, often characterized by its long standing reading-centered material, has been criticised by students and instructors alike because it does not foster the English communication skills needed in today's society. Recognizing the ever growing demand for English, many universities and colleges have currently started to add functional or practical English courses taught by native English speakers. Despite the well-intentioned efforts, however, many concerned college teachers feel that students' English proficiency is not improving dramatically. In the past, a lack of opportunity to use English was blamed for students' lack of English communicative ability. Why hasn't student proficiency improved with the advent of native speaker classes? Students are given better opportunities to learn English than ever with these resourceful native speakers. Although students seem to enjoy their native instructors' English class, students do not seem to exert further effort to utilize their great opportunities of having native speaker instructors available to them. Why aren't students utilizing native speaker instructors' office hours?

What is missing in our system of teaching/learning English seems to be a lack of direction. Providing opportunities is one of the necessary conditions for a successful language program, yet it is not a sufficient condition. Learners in EFL learning environments recognize the need to learn to communicate in English but often their specific needs are not identified and thus they cannot sustain their motivation to learn English. For a successful English education, we not only need to provide students with opportunities to learn and use English, we also need to heighten their motivation throughout their learning cycle. Learners' motivation is heightened and sustained when they clearly recognize the skills they need and reflect on their progress toward established goals. In this sense, initial curriculum planning should be based on needs analysis and learners must be regularly consulted on course contents and instructional methods. Furthermore, they should be given regular and specific feedback on their skill attainment. The assessment of learners' attainment should be formative rather than summative, so that it helps learners redirect themselves.

In order to establish a successful performance-based college English program,

we must reevaluate current assessment tools. Assessment is an important component of the educational process and as such it has a powerful influence on any curricula or instructional change. Unless assessment is in concert with curriculum goals, changes in classroom practice would not likely happen. It should be highlighted at this point that assessment is to be understood differently from testing: While tests are given at a particular point in time of student learning and the result is often given as a single score or grade, assessment is an ongoing strategy through which student learning is not only monitored but by which students are involved in making decisions about the degree to which their performance matches their ability (Hancock, 1994; Wiggins, 1994). If students' performance is tested summatively as an end-product, it does not allow students to get formative feedback. Assessment should take place as an important process that gives students the opportunity to improve the outcomes of their endeavor while there is still time to do so. Through the assessment procedure, students should be able to diagnose problems and monitor their progress.

Self-assessment (SA) is a way to assess students' knowledge or ability using their metacognition. The major purpose of SA is to provide the opportunity for learners to develop an understanding of their own level of English skill in relation to their goals. There are several advantages of including SA in a performance-based college English curriculum. First, SA is advantageous to both learners and instructors. SA can ensure continuous and meaningful feedback if the course objectives are explicitly defined by careful consideration of the learning environment and are clearly understood by students in the learning process. Secondly, SA gives learners greater control over their learning. Through SA, students can realize what they should be able to do, can diagnose what they can currently do, and can be aware of what they need to develop more. SA helps learners by defining the goals to achieve and monitor their performance in light of the established goals (Ellis & Sinclair, 1989). By doing this, learners can increase confidence, self-esteem, and motivation (Dickinson, 1987; Jafarpur, 1991). Thirdly, SA encourages autonomous learning through enhancing learner awareness of the learning process by making students understand the process of learning (Cram, 1995). Students are able to establish their own criteria for judging the accuracy and appropriateness of their performance and diagnose their performance against established criteria. Consequently, they can make decisions about their future direction. Fourthly, SA has a positive

*washback* effect. Since SA gives constant feedback on what learners can do or cannot do, it is easily reflected in instructional practice. Teachers and learners would make collaborative decisions about the contents and methods of instruction and thus make a program more learner-centered and needs-oriented. Finally, cumulative records through SA help to develop functional and affective learner profiles, which are essential for a well-organized language program. In addition to these advantages, SA is time and cost efficient. SA can be administered within a short time, usually less than 20 minutes. Furthermore, if it is used appropriately, there are no cheating nor scheduling problems.

There are some potential problems with using SA in a program. To begin with, teachers may be reluctant to use SA because they are not familiar with SA procedure and thus may feel it less accurate than other conventional tests. By the same token, if students have no experience in self-monitored learning, and if they are not familiar with assessment standards or procedures, they may not feel comfortable assessing their own ability. Finally, the results from SA should be carefully interpreted. The results of SA varies according to the degree of training learners have received in self-assessment and the type of assessment carried out, which is usually developed for particular learners in a particular context utilizing learners' language proficiency, learning context, available resources, and affective factors.

Although it is well recognized that SA is beneficial to second language learners, studies on SA are still scarce. Studies have reported that learners can accurately evaluate their performance through SA: Learners' self-assessment is highly correlated with learners' actual performance except for grammar (Von Elek, 1982); SA can assess learners' own knowledge of the second language (LeBlanc & Painchaud, 1985); Learners' self-evaluations on their language ability were accurate overall (Blanche, 1990). Self-assessment has been also studied as an alternative to formal testing either for course placement or for determining language proficiency levels: Considerable variation in SA is reported (Rolfe, 1990); factors that threaten the validity of SA are discussed (Blanche & Merino, 1989); Conclusions are made that SA instrument should not be used in place of an objective performance test (Ready-Morfit, 1991). Cohen (1994) concludes that SA by learners could be one of the means for multiple assessment, while the use of SA for large-scale assessment might be problematic, especially when administrative decisions are involved. In this light, application of SA and gains in adult ESL program are reported (Cram, 1995).

It seems clear that SA can be beneficial to college English programs, if used appropriately. Considering the current movement toward the proficiency-based approach, an ongoing assessment tool that can provide constant feedback and guide students' learning seems to be urgently needed. The present study explores the possibility of including SA in a college English program. If used for non-administrative purposes, would students reliably respond to the SA instrument? Would the students have a fair idea of evaluating their performance? Would the result of SA be comparable to the result of other forms of assessment? Are there any perverse affective factors related to their responses on SA? In order to find answers to these questions, the present study aims to do the following:

1. Develop an exploratory SA instrument for the group of college learners under the study.
2. Assess the reliability of the developed SA instrument in terms of test/retest reliability and unidimensionality.
3. Assess the validity of the SA instrument in terms of the concurrency with other proficiency measures.
4. Investigate affective variables that perverse in responding to SA.

## II. METHOD

### 1. Participants

The participants were 147 college students enrolled in the freshman English course at a university in Seoul. Participants were all freshmen mostly aged from 18 to 19. In addition to responding to the SA instrument developed for the study, they also took the achievement test, the proficiency test (TOEIC), and the cloze test.

### 2. Instrumentation

#### 1) Self-assessment

A self assessment(SA) instrument was developed for the present study. The

SA instrument contained four parts covering listening, speaking, reading, and writing skills. Based on the RSA profile objectives (Royal Society of Arts, 1987), items were derived from a specification of the tasks which Korean EFL learners typically have to carry out. As a result, the following goals were established for the listening, speaking, reading, and writing skills.

### **Listening**

- understand instructions
- understand main idea and purpose
- extract specific information
- select relevant from irrelevant

### **Speaking**

- express oneself intelligibly
- convey intended meaning accurately with sufficient command of vocabulary
- use language appropriate to context
- interact with other speakers fluently

### **Reading**

- read for overall comprehension
- skim for gist and scan to extract specific information
- decode meaning within reasonable time
- interpret text for attitude and style

### **Writing**

- convey exact meaning accurately and clearly
- organize ideas in a clear logical structure
- use register appropriate for task/situation

The items had the following form:

- I can follow simple instructions. (Example from the listening section)
- I can introduce myself and greet people. (Example from the speaking section)
- I can understand short and simple written communications such as advertisements, announcements, and posters. (Example from the reading section)
- I can fill in a form for self and family. (Example from the writing section)

The students were asked to read each of the statements (20 for the listening, 20 for the speaking, 18 for the reading, and 15 for the writing) and to give themselves a score on a scale ranging from 1 ("I cannot do this at all") to 4 ("I can do this all the time"). Students were asked to respond to the self-assessment measure developed for the study during their regular class meeting.

## **2) The Proficiency Test**

The criterion-referenced proficiency test, a pseudo-TOEIC test, contained 2 parts - listening comprehension and reading comprehension. The listening comprehension part consisted of 4 sub parts - picture description, question and answer, conversation, and explanation. The reading comprehension part was composed of 3 sub parts containing grammar-vocabulary, usage, and reading test.

## **3) The Achievement Test**

The teacher-made achievement test was also composed of 2 parts - listening comprehension and reading comprehension. Since the English course that the participants were enrolled in used functional textbooks and the test was constructed based on the functional use of English language, the test was similar to the proficiency test in its form and content. Although the achievement test highlights certain functions of the language as they were taught in the course, they were still parts of the proficiency that the other proficiency test (the pseudo-TOEIC test) attempted to measure. Both the proficiency test and the achievement test required the students to respond to questions in multiple-choice format on a computerized answer sheet.

# **III. DATA ANALYSIS**

The reliability of the SA was assessed first. The SA scores were then compared to the scores on the other formal tests - the proficiency test (TOEIC), the achievement test, and the cloze test. In addition, learner variables were factor analyzed and predicting ability of the emerged factors on SA was assessed using multiple regression.

The statistical analyses - reliability assessment, factor analysis, multiple regression - were done using SPSS (95 Windows version). Students' ability scores on the proficiency test were analyzed through BILOG (DOS version 3.04), an IRT<sup>2)</sup> (Item Response Theory) application software.

## 1. Reliability of the SA Instrument

The SA was subject to the reliability analysis. Reliability was checked in terms of stability and internal consistency. In order to check the stability of the SA instrument, the test-retest reliability was assessed. The SA instrument was given twice to 32 students and then two set of scores were computed. In an attempt to minimize differential practice effects and differential changes in ability, the second administration of the SA instrument was conducted 2 weeks after the first administration without prior notice. As seen in Table 1, the test-retest reliability ranged from .72 to .87.

Each component of the SA instrument (Listening, Speaking, Reading, Writing) was then examined in terms of its unidimensionality (internal consistency). In order to check the unidimensionality of each component of the SA instrument, Cronbach alphas<sup>3)</sup> were assessed. Cronbach alphas were obtained from 147 students who took all the tests (the SA instrument, the proficiency test, the achievement test, and the cloze test). Cronbach alphas ranged from .90 to .94, indicating that the items in each component of the SA instrument indeed measure a single or unidimensional ability or trait.

---

2) IRT (Item Response Theory) is different from CTS (Classical True Score) in estimating learners' ability on a test. While CTS determines learners' ability based on the number of items they get right, IRT estimates learners' ability based on the item characteristic curves. Item characteristic curves are determined by guessing parameter (the probability that an individual of low ability can answer the item correctly), discrimination parameter (the degree to which the item discriminates among individuals of differing levels of ability), and the difficulty parameter (the level of difficulty of the item) (See Bachman, 1990). For further information on IRT and its application, look at Van der Linden and Hambleton (1997) and /or Hambleton and Swaminathan (1985).

3) A method for assessing internal consistency of a measure. It is also referred to as coefficient alpha or simply as alpha. For a good discussion of reliability and validity assessment, consult Huck and Cormier (1996).



**TABLE 1**  
**The Reliability of the SA**

Reliability	Listening	Reading	Speaking	Writing
Stability (n = 32)	.8245	.7214	.8722	.7959
Unidimensionality (n = 147)	.9185	.9027	.9355	.9073

Items which had low or high test-retest reliability coefficients were further analyzed. It was found that items which had low reliability (below .50) were those with unspecified situations and those with inexperienced situations. Examples of the items which had low reliability coefficients were:

- I can understand native speaker over the phone even when he talks fast, using colloquial expressions and idioms. (Example from the listening section)
- I can describe an action, scene, or an event. (Example from the speaking section)
- I can get information from simple diagrams, timetables, time sheet, and street plans. (Example from the reading section)
- I can change writing style when writing to different people or for different purposes. (Example from the writing section)

On the other hand, items with high reliability coefficient (over .70) were concrete questions with detailed situations or frequently met situations.

- I can understand simple learned phrases such as "Hello," "How are you?" "What's your name?" "Where do you live?" (Example from the listening section)
- I can ask directions. (Example from the speaking section)
- I can find errors while reading a simple sentence. (Example from the reading section)
- I can write a simple informal letter. (Example from the writing section)

## 2. Concurrent Validity of the SA Instrument

Concurrent validity of the SA instrument was assessed by examining the

students' scores on the standardized proficiency test, the achievement test, and the cloze test. Table 2 shows the correlations between the overall SA scores and the total scores of the other formal tests. Except for the cloze test, the correlation between the SA score and the other test scores were relatively high: the correlation between the overall SA score and the total proficiency test score was .64; the correlation between the overall SA score and the total achievement score was .61. On the other hand, correlations among other tests were relatively low: the correlation between the proficiency test and the achievement test was .37; the correlation between the achievement test and the cloze test was .03; and the correlation between the achievement and the cloze test was .50.

There are many forms of English tests we use in order to assess students' English proficiency. It is plausible to think that the scores students get from an achievement test and a more general proficiency test would be highly related, especially when the language program is proficiency-oriented. As found in other studies, however, their correlations are not necessarily high. Given these, the result that SA related to more than two forms of other tests indicates that students can assess their own skills of the second language to some extent. Although the results are not of the highest level, it is in accordance with the .53 result by Leblanc and Painchaud (1985).

**TABLE 2**  
Correlation Coefficients Between Self-assessment and Total Test Scores

	Proficiency	Achievement	Cloze
Self Assessment			
Listening	.54	.57	.27
Reading	.77	.37	.01
Speaking	.57	.65	.34
Writing	.49	.63	.36
Total	.64	.61	.28

\* Proficiency vs. Achievement: .37

\* Proficiency vs. Cloze: .03

\* Achievement vs. Cloze: .50

\* Correlations whose values are greater than .15 are all significant at  $\alpha = .05$ .

**TABLE 3**  
Correlation Coefficients Among Subtests

	SL	SR	SS	SW	ST	PL	PR	θ <sub>L</sub>	θ <sub>R</sub>	AL	AR
Self Assessment											
Listening (SL)	1										
Reading (SR)	.83	1									
Speaking (SS)	.80	.72	1								
Writing (SW)	.77	.73	.88	1							
Total (ST)	.92	.89	.93	.92	1						
Proficiency											
Listening (PL)	.42	.68	.35	.33	.48	1					
Reading (PR)	.49	.64	.60	.51	.61	.48	1				
Listen-Ability (θ <sub>L</sub> )	.34	.60	.34	.23	.41	.96	.51	1			
Read-Ability (θ <sub>R</sub> )	.52	.65	.57	.47	.60	.44	.95	.50	1		
Achievement											
Listening (AL)	.37	.20	.58	.34	.42	.15	.19	.23	.18	1	
Reading (AR)	.57	.39	.60	.66	.61	.47	.21	.45	.21	.58	1
Cloze	.27	.01	.34	.36	.28	-.02	.06	.07	.05	.23	.54

\*Ability scores (θ<sub>L</sub>, θ<sub>R</sub>) were estimated from 2-PL IRT (Item Response Theory) Model (N = 1002).

\*Correlations whose values are greater than .15 are all significant at  $\alpha = .05$ .

A close look at the correlations between the SA subscores and the subscores of the other tests are presented in Table 3. The correlation between the SA listening subtests and the other listening subtests were: .42 between the SA listening score and the proficiency listening score; .37 the SA listening and the achievement listening score. The correlation between the SA reading subtest and the other reading subtests were: .64 between the SA reading score and the proficiency reading score; .39 between the SA reading score and the achievement reading score. Additionally, since the scores on the SA represent students' perceived ability, the ability scores adopting Item Response Theory (IRT) seemed to be more meaningful than simple scores following Classical True Score (CTS). Therefore, students' ability scores on the proficiency test (a pseudo-TOEIC test) were obtained using 2 parameter (2PL) model<sup>4)</sup>, one of the IRT

4) The ability score on the 2 PL (two parameter) model is based on the discrimination parameter and the difficulty parameter. The 2 PL model assumes that there is no probability of getting a correct response by guessing. The 3 PL (three parameter) model assumes the guessing parameter. The 1PL model, also known as Rasch model, assumes that the discrimination of all the items are equal and there is no

procedure. In order to estimate the ability scores, the scores of 1002 students, who took the same pseudo-TOEIC test were evaluated. These ability scores were matched to the other test scores and the correlation to the SA subtest scores were then assessed. The correlation using the ability scores following IRT were more or less the same as using classical scores. The correlation between the SA scores and the proficiency ability scores were: .34 between the SA listening score and the proficiency listening ability score; .65 between the SA reading score and the proficiency reading ability score.

The correlations among the listening subtests were substantially lower than those between the listening subscores and the reading subscores. These heterogeneous correlations are probably because students have not had much experience in English listening, and thus they assume their listening ability partly based on their reading ability.

### 3. Affective Variables on Self-assessment

Out of the wide variety of affective factors that might be expected to influence learners' self-measured performance, the affective variables investigated for the present study were selected based on the saliency among adult foreign language learners (Park, 1995). In an attempt to find out meaningful clusterings of affective variables, factor analysis was conducted. Four factors were extracted using the Principal Component Analysis followed by Varimax rotation. The cutoff criterion was .30. As can be seen in Table 4, factor 1 receives appreciable loadings from five items. Features of factor 1 pertain learners' boredom, enjoyment toward the class, voluntary and conscious involvement in learning situation, and yearning for group cohesiveness. It seems best labeled as an *Attitude toward learning situation*.

Factor 2 is defined by five items including anxiety and past contact with spoken and written English. This factor was labeled *Past Contact* and *Anxiety*. Since the two variables involved in factor 2 were distinctly classified in a number of other studies (Park, 1995), they were analyzed as two factors - past contact and anxiety for multiple regression analysis.

Factor 3 evidences appreciable loadings from three items, the features of which were: giving priority, pursuit of learning opportunities, and exertion of effort. It therefore seems to be best labeled as a *Motivation*.

Factor 4 is defined by 2 items concerned with learners' current contact with English, and thus labeled as a *Current contact*.

**TABLE 4**  
Factor Analysis of Affective Variables

Items	Factors			
	Factor 1 Attitude	Factor 2 Anxiety/ Past Contact	Factor 3 Motivation	Factor 4 Current Contact
13. Boredom (reversed)	-.7206	-.0113	-.0904	-.1112
11. Enjoyment	.6772	.0739	.3808	-.0366
4. Voluntary involvement	.6418	.2543	.0935	.1638
5. Conscious involvement	.4645	.3988	.4236	.1580
12. Group cohesiveness	.4192	-.3992	-.0454	.3747
10. Embarrassment	-.1295	-.7227	-.0062	-.0111
15. Alienation	-.5096	-.6370	.0163	-.0136
8. Past contact (Spoken)	-.0460	.6336	.2840	.4196
14. Lack of relaxation	-.5390	-.5731	.2589	-.1354
9. Past contact (Written)	-.0173	.4868	.3797	.4431
3. Pursuit of opportunity	.1814	.0662	.8452	.0532
2. Priority	.0358	-.0875	.7926	.1415
1. Exerting effort	.1733	.3743	.4674	.1941
7. Current contact (Written)	.0442	.0816	.1049	.8493
6. Current contact (Spoken)	.3982	.1450	.1670	.6871
Eigen Value	4.7195	1.7691	1.4525	1.1856
% of Variance	31.5	11.8	9.7	7.9
Cumulative % of Variance	31.5	43.3	52.9	60.8

Factor 1 = Attitude : Factor 2 = Anxiety/Past contact : Factor 3 = Motivation :  
Factor 4 = Current contact

In order to see which factors are perverse in relation to the SA scores, multiple regression analysis was conducted. Mean scores of the variables were used rather than the weighted sums.  $R^2$  was adjusted to reduce errors by increasing the number of variables. As can be seen in Table 5, factors related to the SA listening component were *current contact* and *anxiety*, which accounted for 46% of the total variance. The SA reading component was related to *attitude* and *anxiety*, which explained 39.5% of the total variance. The SA speaking component was related to *current contact*, *anxiety*, and *past contact*: they explained 43.9% of the total variance. The SA writing component was related to *anxiety*, *current contact*, *past contact*, and *motivation*: these four factors accounted for 78.9% of the total variance.

**TABLE 5**  
**Stepwise Multiple Regression Result of Affective Variables on SA**

	Self Assessment			
	Listening	Reading	Speaking	Writing
Attitude	excluded (.158)	step 1 (.002)	excluded (.225)	excluded (.884)
Anxiety	step 2 (.009)	step 2 (.032)	step 2 (.007)	step 1 (.000)
Motivation	excluded (.913)	excluded (.575)	excluded (.161)	step 4 (.036)
Current Contact	step 1 (.004)	excluded (.413)	step 1 (.001)	step 2 (.000)
Past Contact	excluded (.533)	excluded (.562)	step 3 (.048)	step 3 (.007)
Adjusted R2	46.0%	39.5%	43.9%	78.9%

\*The numbers in parenthesis represent p-values from the t-tests of the corresponding factors.

## IV. DISCUSSION

The primary purpose of the present study is to examine the possibility of using self-assessment in a college English program. An exploratory SA instrument was developed for the group of college learners under the study. The developed SA instrument was then assessed in terms of test/retest reliability and unidimensionality. The validity of the SA instrument was checked by concurrency with the other proficiency measures. Then, affective variables that perverse in learners' performance measured through SA were investigated.

The SA used for the present study was a focused form of assesment for particular learners in a particular context thus, it is not necessarily suitable for other groups of learners. The study does not have any intention to generalize the results: the results may be meaningful only in this context.

SA, if used appropriately, seemed to reliably measure students' performance of English. Students' performance measured through SA were related to more than two forms of other tests, indicating that students can assess their own skills of English to some extent. Items which had low or high test-retest reliability coefficients were further analyzed. It was found that items which had

low reliability were those with unspecified situations and those with inexperienced situations. On the other hand, items with high reliability coefficients were concrete questions with detailed situations or frequently met situations. In developing a reliable SA instrument, items should be made concrete and detailed. This can be done by analyzing what tasks the learners may have experienced or typically have to carry out using English.

Learners' performance estimated through SA revealed that for all four skills, anxiety is highly related. This may indicate that in developing performance-based English programs, learners' anxiety should be considered. For the listening, speaking, and writing, current contact was highly related to students' performance in addition to anxiety. This indicates that in developing these skills, learners need to be provided an ample amount of opportunities for using the language. For the productive skills - speaking and writing - not only current contact with English but also past contact with English seems to play a role. This may indicate that in order to help students perform better in their speaking and writing of English, learners should be provided continuous opportunities for using English. Past opportunities of using English seem to be carried over to current productive skills. Interestingly, both current and past contact with English was not related to learners' performance in their reading. Only anxiety and attitudes were related to their reading performance. This indicates that in order to develop reading proficiency, it is important to raise positive attitude toward learning.

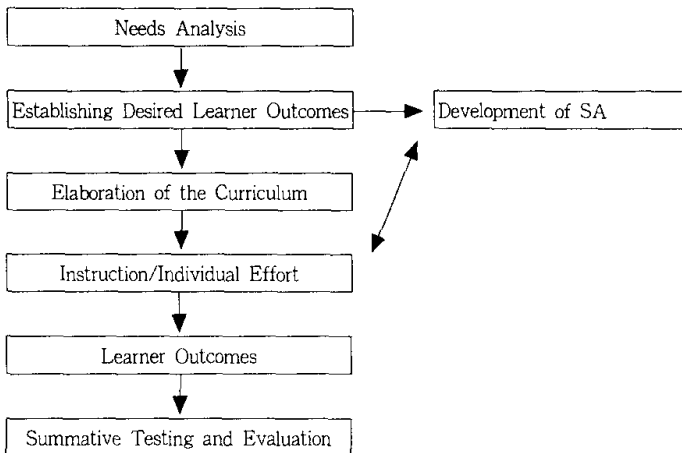
Based on the results of the present study, SA seems to be a useful tool to introduce in a performance-based college English program. As Gimenez (1996) pointed out, students' performance is tested only as end-products, so that it is often too late for students to get formative feedback. SA is useful because it is a process assessment and thus allows formative feedback. Students need the opportunity to improve the outcomes of their endeavor while there is still time to do so. For a successful performance-based English education, there should be coherence among the educational goals, processes, and evaluations (Valette, 1994). To tightly connect the educational goals, instruction, and the assessment, learners' performance standards for the program must be clearly stated and a measure needs to be developed for learners to diagnose and check their progress toward the stated goals. SA is most desirable when it is used as part of formative assessment, where SA outcomes can provide feedback to learners (Brindley, 1989). In order to make students able to diagnose and continuously check their achievement, the criteria for SA needs to fit the local needs. The program objectives and accurate descriptions of what will be achieved, should be derived and specified from needs analysis. These specified objectives (derived

from needs analysis) should guide the course content as well as provide the criteria against which learners' performance can be judged.

Figure 1 shows schematic representation of a program model suitable for the performance-based and learner-centered English program including the SA procedure. Different from the widely used top-down approach in establishing learner outcomes, this program starts with the analysis of learner needs. In addition to the formal evaluation steps at the end of the learning cycle, this model suggests the development of SA instrument at the same time as the establishment of the learner outcomes, so that the SA procedure can have positive washback effects on learners' effort or the instructional practice.

**FIGURE 1**

**Schematic Representation of a Performance-based College English Program**



## V. CONCLUSION

In order to have a successful journey towards becoming a competent English speaker, learners need accurate guidance. Just as travellers need a map and compass to check how far they have to go or if they are on the right track, individual English learners need clear instructions/learning goals to find the path to their goal. As a part of formative assessment, SA can be a valuable tool in a



performance-based college English program. If learners are given enough explanation and practice, and if the result of SA is not for administrative purpose, it encourages learners to be more involved in planning their own learning and reflecting on their progress. It can give valuable feedback to both the teacher and the student. It can develop progress profiles that can be used as a form of continuous goal-setting and as descriptive and reflective tools for improving the program to suit learners' needs.

Although there is no single method that provides a solution to all problems in teaching English, as De Jong(1992) pointed out, successful English learning/teaching at any level, may require two steps urgently. The first is the clear definition of practical and attainable goals in the particular learning context. The second is learners' active cooperation heightened by a criterion referenced definition of goals and regular feedback on progress. In order to check learners' progress toward the established objectives, we need more assessment and less testing. Up to date, as Wiggins(1994) illustratively put it, "we have been so accustomed to testing learners rather than assessing learners and thus, those tests have served the interests of psychometricians and policy makers and not those of students and teachers. Often sacrificing validity for reliability, students have not been the primary client of assessment information."

Learners ability to perform in English can be assessed by having them perform the language in responding to particular tasks and contexts. However, measuring learners' performance usually means a lot of work and high cost. "authentic assessment," "alternative assessment," and sometimes "performance assessment," have a common goal of guiding instruction so that all students can achieve high levels of mastery. If used appropriately as a part of formative assessment, self-assessment can do the job with minimum work and cost. Among other performance assessment such as portfolio, oral proficiency measures, collaborative assessment, learning logs, and journals, SA is the most practical and feasible.

SA can also be used in the evolving cyber/distance education environment, where learners' motivation and performance evaluation are two problem areas. Further research on the use of SA in the assessment process is needed. It would be interesting to see motivational changes in learners upon introduction of SA in the learning process.

## REFERENCES

- Bachmann, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313-340.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing*, 7(2), 202-229.
- Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum*. Sydney: National Centre for English Language Teaching and Research (NCELTR), Macquarie University.
- Brindley, G. (1990). *The second language curriculum in action*. Sydney, Australia: National Centre for English Language Teaching and Research (NCELTR), Macquarie University.
- Brindley, G. (Ed.). (1995). *Language assessment in action*. Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. Boston, MA.: Heinle & Heinle Publishers.
- Cram, B. (1995). Self-assessment: from theory to practice: Developing a workshop guide for teachers. In G. Brindley (Ed.), *Language assessment in action* (pp. 271-305). Sydney: National Centre for English Language Teaching and Research (NCELTR).
- De Jong, J. H. A. L. (1992). Assessment of language proficiency in the perspective of the 21st century. *AILA-Review: Language Teaching in the Twenty-first Century. Problems and Prospects*, 39-45.
- Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge University Press.
- Ellis, G., & Sinclair, B. (1989). *Learning to learn English: a course in learner training*. Cambridge: Cambridge University Press.
- Gimenez, J. C. (1996). Process assessment in ESP: Input, throughput and output. *English for Specific Purposes* 15, 233-241.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer · Nijhoff Publishing.
- Hancock, C. R. (1994). Alternative assessment and second language study: What and why? *ERIC Digest*. [On-line]. <http://eriac2.educ.cua.edu/>

- db/edo/ED376695.htm.
- Hancock, C. R. (1994). *Teaching, testing, and assessment: Making the connection*. Lincolnwood, IL: NTC.
- Huck, S. W., & Cormier, W. H. (1996). *Reading statistics and research*. New York, NY: Harper Collins College Publishers.
- Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation and Research in Education*, 5, 145-157.
- Lafayette, R. C. (1996). *National standards: a catalyst for reform*. Lincolnwood, IL: NTC.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19, 673-687.
- Mislevy, R., & Bock, R. D. (1990). *BILOG 3: Item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software, Inc.
- Park, J. J. (1995). *Adult learners' motivation in learning a non-cognate foreign language*. Unpublished doctoral dissertation. University of Minnesota, Twin Cities.
- Ready-Morritt, D. (1991). *The role and limitations of self-assessment in testing and research*. Ottawa, Canada: Second Language Institute, University of Ottawa. Unpublished manuscript.
- Rolfe, T. (1990). Self - and peer - assessment in the ESL curriculum. In G. Brindley (Ed.), *The second language curriculum in action* (pp.163-186). Sydney, Australia: National Center for English Language Teaching and Research (NCELTR), Macquarie University.
- Royal Society of Arts. (1987). *Practical skills profile scheme*. London: Royal Society of Arts.
- Valette, R. M. (1996). Teaching, testing, and assessment: conceptualizing the relationship. In C. R. Hancock (Ed.), *Teaching, testing, and assessment: Making the connection* (pp. 1-42). Lincolnwood, IL: NTC.
- Van der Linden, W. J., & Hambleton R. K. (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.
- Von Elek, T. (1982). *Test of Swedish as a second language: An experiment in self-assessment*. (Work Papers 31) Molndal, Sweden: Language Teaching Research Center.
- Wiggins, G. (1994). Toward more authentic assessment of language performances. In C. R. Hancock (Ed.), *Teaching, testing, and assessment: Making the connection* (pp. 69-85). Lincolnwood, IL: NTC.