
EVALUATING THE STUDENTS' LANGUAGE PROFICIENCY IN THE ENGLISH DEPARTMENT, COLLEGE OF BASIC EDUCATION IN KUWAIT

Dr. Hanan A. Taqi & Dr. Khaled M. Shuqair

English Department, College of Basic Education, Kuwait

ABSTRACT: *In the context of English language programs and the evaluation of such programs, this study analyzes the usefulness of the program taught to English-teachers-to-be students in the English Department, College of Basic Education in Kuwait. Based on the calculation of grades obtained in the initial placement test and a replica test conducted four years later, the change in the students' language proficiency was measured. The paper reviews the results of the 50 participant students in both tests in five main testing categories based on the four language skills. The findings reveal very little improvement in language proficiency, which also seems to be very weak initially for an English teacher. This finding implies a major problem with the current program. Finally, a number of recommendations for program and student improvement for the College of Basic Education, and English programs in other colleges and universities, are considered.*

KEYWORDS: Evaluation of program, higher education, English Language, Curriculum design, testing language programs, evaluating language skills.

INTRODUCTION

A corollary premise that has long been acknowledged in educational research and practice is that "improvement" cannot be achieved without evaluation. Evaluation is the corner stone of any educational practice. In distinguishing evaluation from research, Stufflebean (1981) says: "The purpose of evaluation is to improve, not to prove." The statement *to improve* means that a judgment must be made with regard to what constitutes worth or value. In other words, improvement is directly linked to evaluation and the term evaluation is associated with making judgment about the value or the effectiveness of a certain action.

This study was concerned with evaluating the students' language proficiency in the English Department, College of Basic Education (CBE) at the Public Authority for Applied Education and Training (PAAET) in Kuwait, by measuring students' achievement of the language skills they have acquired after the completion of the requirements of the program. We assessed their performance at this stage in order to gain some insights into the English language proficiency levels they have achieved. Participants in the study are students in their graduating year, who sat for the department placement test twice: upon joining the department four years earlier and at the time of this study when they were asked to take it again by the researchers. Their initial results on the placements test (pretest) were compared with those results they achieved when they took the same test upon graduation in order to gauge their improvements. In doing so, we hope to be able to contemplate on the likelihood of student' attainment of the goals of these courses, and hence we will be able to gain valuable preliminary assessments of the impact of the language courses, and, indirectly, their ability to fulfill the goals they are set for.

Purpose of the Study

The main purpose of the study is to evaluate the effectiveness of the English program in the College of Basic Education in Kuwait. More specifically, the students' competencies in the four skills are examined in their senior year upon finishing all the courses offered by the program. The researcher's ultimate aim is to contribute to the improvement of the primary school English education program taught at the CBE.

Significance of the Study

It is in this context and in any English program context that the present study can make significant contribution. The study attempts an evaluation of the English language proficiency upon the completion of the requirements of the English Department. One of the objectives of the English program taught in the English Department is the improvement of the English language skills of students joining the Department and the development of their performance in the subsequent specialized courses. It is hoped that this study will provide policy makers and educators involved in this English program with pertinent information that will help them evaluate this program and take informed decisions about it in particular and decisions with regard to the problem of the deficiency in the English language skills that Department students suffer from. The significance of the study arises from the fact that it is the first study of its kind that attempts an evaluation of these courses in terms of language learning.

More specifically, the four skills development plays an important role in the language learning process. Ensuring that the students possess the previously aimed competencies in four skills before graduating is one of the preliminary goals of any English education program. Therefore, it is vital to evaluate the students' language skills, which they will need later when they graduate and teach English for the primary stage in the Kuwaiti public schools. In the recent years, evaluation of the students' skills joining the English program has become urgent, as the feedback reports from the Ministry of Education about the language skills of teachers who graduate from the program are not favorable.

Furthermore, the results will provide information regarding the materials, teaching methods, assessment and communication opportunities: all of which will definitely add up to the suggestions to improve the identified deficiencies in primary school English education in general. Another significant aspect of this study is that it will contribute to the scant body of literature on primary school program evaluation in Kuwait. As such, the results of the study may be of assistance to other universities in understanding the deficiencies in their programs.

Limitations of the Study

The main limitation of this study is that it relies on only students' achievements in the same placement test before they join the program (pretest) and in their senior year when they complete the requirements of the program (posttest). It may be more informative to support the findings of such a study with students' self-reported (how students perceive their own competencies) and other varieties of measurement tools, such as direct observation and feedbacks from the instructors. Another limitation is related to the gender of the participants in the study. As the department is an all-female department, the findings are limited, in terms of gender, to females and cannot be generalized to both sexes.

Research Questions

This study was guided by one major research question:

How effective is the English program offered in CBE in improving the students' language competencies?

The sub-questions related to the major question are:

1. Are there statistically significant differences between the students' results of the pretest and posttest in the skill of reading to the advantage of the posttest?
2. Are there statistically significant differences between the students' results of the pretest and posttest in the skill of language use to the advantage of the posttest?
3. Are there statistically significant differences between the students' results of the pretest and posttest in the skill of structure to the advantage of the posttest?
4. Are there statistically significant differences between the students' results of the pretest and posttest in the skill vocabulary to the advantage of the posttest?
5. Are there statistically significant differences between the students' results of the pretest and posttest in the skill of writing to the advantage of the posttest?

Overview of the English Program and the Placement Test

Goals and Outcomes

The Department of English offers a four-year program focusing on a set of teaching interests within language arts and education. The program consists of 130 credit hours designed to prepare students and develop their instructional abilities to be teachers of English in the primary schools. For this end, courses in education and psychology are integrated in the program, in addition to hours of practical and field training.

The aim of the program is to develop the teaching and English language skills of student teachers. Students who complete the courses successfully will be skilled in analysing their pupils' needs and in planning and teaching English lessons. They will have an understanding of the theory of teaching, learning and related topics, and they will be able to describe and use English well enough to teach it and to read about teaching and learning in English.

The English Department has stated its program outcomes as follows;

1. To give students the special skills required to be able to instruct children properly instruct.
2. To help students understand a child on all levels (physical, emotional, intellectual, and others).
3. To show students how to interact well with young learners in a classroom.
4. To let students explore the most suitable methods in teaching English to young learners.
5. To teach students how to create a good classroom atmosphere for learning English.
6. To help students properly manage children's behavior.
7. To demonstrate to students how to use teaching resources easily and efficiently.
8. To allow students to acquire proper instructive strategies using flashcards, stories, music, drama, crafts, games, projects, and pair work.
9. To further develop students teaching career and to increase students level of professionalism.

The program content is a reflection of the program objectives. The program has four strands, the language-oriented modules, the linguistic-oriented modules, the theory and methodology-oriented modules and the literature-orientated modules.

1. The language-oriented modules include those aiming to develop students' own language skills and communication skills: listening, speaking reading and writing and the basics of grammar.
2. The linguistic-oriented modules include courses which teach the student teachers about linguistics and how it will best serve them in the classroom.
3. The theory and methodology-oriented modules look at theories of language learning in the field of young learners, at curriculum development issues and at the changing role of technology. The methodology-oriented modules deal with both the teaching of young learners and with the teaching of different aspects of language competence (Teaching Speaking and Listening, Teaching Reading and Writing, Teaching Grammar and Vocabulary).
4. The literature-orientated modules give students an insight into English literature written specifically for children, to help them see its importance for complete language learning.

Admission Requirements

The minimum requirements for admission to the English Department are the following:

1. Applicants must hold the General Secondary Certificate of Education (GSCE) or its equivalent to be considered for admission to the Department.
2. Applicants must earn at least 70% in the GSCE to be considered for admission to the Department.
3. Because of the large number of applicants, admission is highly competitive and is mainly based on a placement examination and an interview.

In order for an applicant to be considered for admission in the department they must attain in both the placement test and interview a score no less than 60%. While the placement test evaluates applicants' performance in the skills of reading, structure, vocabulary and writing, the interview evaluates applicants' performance in the oral skills. The placement test consists of 100 multiple-choice questions distributed on the language skills as follows:

- | | |
|--------------------------|--------------|
| 1. Reading Comprehension | 24 question |
| 2. Language Use | 19 questions |
| 3. Structure | 15 questions |
| 4. Vocabulary | 27 questions |
| 5. Written expressions | 15 questions |

The time allowed for the test is two hours, during which applicants are asked to provide their answers on a separate answer sheet by darkening the bubble that represents the right choice. Results are attained by processing the answer sheets through a scanning machine.

LITERATURE REVIEW ON CURRICULUM EVALUATION

It is a fact that evaluation may be conducted for a wide range of reasons in every part of our life. In terms of education, it can be stated that the main purpose of evaluation is to obtain

information about student performance. In the same way, the goals might also be the identification of strengths and weaknesses of particular activities in a program.

There is no consensus on the definition of evaluation. While some educators connect evaluation with measurement, others define it as the assessment of the extent to which specific objectives have been attained. Some view evaluation as primarily scientific inquiry, whereas others argue that it is essentially the act of collecting and providing information to enable decision-makers to function effectively (Worthen and Sanders, 1998). Despite this lack of consensus, Murphy (2000) defines evaluation as a way to determine the degree in which a program attains its objectives. In this sense, it gives support to stakeholders in decision making for program improvement through careful analysis of information gathered. Kiely (2009 , p.99) asserts that “ evaluation has emerged from studies on teaching methods which were stimulated by theories on language learning with a focus on quality assurance and enhancement .”Talmage (1982) defines evaluation as the act of rendering judgments to determine value-worth and merit without questioning or diminishing the important roles evaluation plays in decision making. Moreover, “evaluations can differ in terms of many dimensions, among which the design (experimental, quasi-experimental, regression discontinuity) intent (advocacy versus objective assessment), philosophical underpinnings (quantitative versus qualitative), and others” (Frechtling, 2007 p. 104).

Cronbach (1991) makes a distinction among three types of decisions that requires evaluation. He starts first with “course improvement”; deciding what instructional materials and methods are satisfactory and where change is needed. The second type of evaluated decisions is “decisions about individuals, which is based on identifying the needs of the pupil for the sake of planning his/her instruction, judging pupil merit for purposes of selection and grouping, acquainting the pupil with his own progress and deficiencies. Finally, Cronbach states that the third type was “administrative regulation”; judging how good the school system is, how good individual teachers are, etc.

According to Ralph Tyler (1991), evaluation is a process essential to curriculum development. The purpose of evaluation was stated as to determine the extent to which the curriculum had achieved its stated goals. Evaluation was the basis for the identification of strengths and weaknesses in the curriculum, followed by replanning, implementation and evaluation (Gredler, 1996). Similarly, Worthen and Sanders (1998) stated that evaluation is the formal determination of the quality, effectiveness or value of a program, product, project, process, objective or curriculum. In addition, there are several judgment methods that are used for evaluation during this determination process. These are mainly determining standards for judging quality and deciding whether those standards should be relative or absolute. Secondly, collecting relevant information, and finally applying the standards to determine quality. Hence, in the light of these definitions related to evaluation, it can be concluded that program evaluation is a systematic inquiry designed to provide information to decision makers and/or groups interested in a particular program, policy or other intervention. This inquiry might be exemplified as ‘How does the program work?’, ‘Does the program produce unintended side effects and so on?’ (Cronbach, 1980, p. 87) Program Evaluation generally involves assessment of one or more of five program domains. a) the need for the program b) the design of the program c) the program implementation and service delivery

d) the program impact or outcomes and e) program efficiency (cost effectiveness) (Payne, 1994, p. 15).

Mackay (1994) states that in the field of foreign language teaching, the term 'program evaluation' is applied to a wide variety of activities, ranging from academic, theory - driven research to informal enquiries carried out by a single classroom. Thus, evaluation may focus on many different aspects of a language program such as curriculum design, classroom processes, the teachers and students.

Evaluation is a central component of the educational process. Thus, it is certainly a critical and challenging mission. Kelly (1999) defines curriculum evaluation as the process by which we attempt to gauge the value and effectiveness of any particular piece of educational activity. The two common goals of program evaluation, as stated by Lynch (1996) are evaluating a program's effectiveness in absolute terms and/or assessing its quality against that of comparable programs. Program evaluation not only provides useful information to insiders on how the current work can be improved but also offers accountability to outside stakeholders. It aims to discover whether the curriculum designed, developed and implemented is producing or can produce the desired results. The strengths and the weaknesses of the curriculum before implementation and the effectiveness of its implementation can be highlighted by the help of evaluation (Ornstein and Hunkins, 1998). Thus, a systematic and continuous evaluation of a program is significant for its improvement, which ultimately leads to the need for curriculum evaluation.

A different way of analyzing curriculum evaluation is in terms of the timing of the evaluation, the ways in which it is made, the instruments used and the purpose for which the results are used. Scriven (1991) introduced into the literature of evaluation the concept of *Formative* and *Summative Evaluation*. Formative evaluation requires collecting and sharing information for program improvement. While a program is being prepared, the formative evaluator works to provide the program planners and staff with information to help adjust it to the setting and improve it (Morris and Fitz-Gibbon, 1978). Formative evaluation is typically conducted during the development or improvement of a program or product or person and so on and it is conducted often more than once (Scriven, 1991). The purpose of formative evaluation is to validate or ensure that the goals of the instruction are being achieved and to improve the instruction if necessary by means of identification and subsequent remediation of problematic aspects (Weston, Mc Alpine and Bordonaro, 1995). Therefore, it is apparent that formative evaluation provides data to enable on-the-spot changes to be made where necessary. Students' learning activities can be refocused and redirected and the range and depth of instructional activities of a curriculum can be revised in 'mid-stream' (Tunstall and Gipps, 1996). Hence, it applies to both course improvement and students' growth, although some writers tend to concentrate only upon the former (Pryor and Torrance, 1996). In brief, formative evaluation is conducted during the operation of a program to provide program directors evaluate information useful in improving the program. Summative evaluation, on the other hand, is conducted at the end of a program to provide potential consumers with judgments about that program's worth or merit. For example, after the curriculum package is completely developed, a summative evaluation might be conducted to determine how effective the package is with a national sample of typical schools, teachers

and students at the level for which it was developed (Worthen and Sanders, 1998). The summative evaluator's function is not to work with the staff and suggest improvements while the program is running but rather to collect data and write a summary report showing what the program looks like and what has been achieved. Summative Evaluation is the final goal of an educational activity. Thus, summative evaluation provides the data from which decisions can be made. It provides information on the product's efficacy. For example, finding out whether the learners have learnt what they were supposed to learn after using the instructional module. Summative evaluation generally uses numeric scores or letter grades to assess learner achievement.

While formative evaluation leads to decisions about program development including modification, revision and the like, summative evaluation leads to decisions concerning program continuation, termination, expansion, adoption and so on. This study conducts a summative evaluation of the English program in the College of Basic Education since the program has been going on for fifteen years now, and the researchers are concerned with finding out whether the learners have learned what they are supposed to learn. That is, the study seeks to find out whether the outcomes of the programs relevant to the language proficiency have been attained after a long time of implementation.

METHOD

Participants

This research was conducted towards the end of the academic year 2012/2013. The population of the study is 200 students, which is the number of students expected to graduate that year after finishing the requirements of the program. The sample of the study consisted of 50 students chosen randomly from the body of graduating students. Every care was taken to make sure that the students chosen are in their graduating semester by checking with the registration office in CBE.

Procedure

Once the students were chosen, the researchers checked their scores in the placement test they sat for when they joined the Department four years earlier. The scores were classified for each student based on the five sections of the test, so that each student has five scores reflecting her achievements in each of the four skills in question. Then the researchers met with the sample of students and explained to them the purpose and procedures of the experiment. Students were eager to participate and take the placement test for the second time because they themselves wanted to gauge their improvement in the language skills before they graduate and embark on the teaching career. This attitude on the part of students gave assurances to the researchers that the students will take the test seriously.

On the assigned day of the test, every care was taken by researchers to simulate the circumstances of the pretest in terms of procedures and timing. The researchers proctored the test and made sure the test was administered according to the followed procedures. When the students finished the tests, the researchers collected the answer sheets and fed them to the scanning machine. The results of the posttest were classified like those of the pretests, so that for each student in the sample there appeared two scores on each of the five main parts, one

represented her achievement upon joining the department, and the other in her graduating year.

RESULTS

The results obtained from the pretest and the posttest were entered in SPSS. They are compared for statistic significance, and the mean for each score is obtained. Moreover, the researchers looked at minimum and maximum scores to view the level of difference between students, and the level of increase/decrease.

Analysis of the placement test

The results of the pretest and the posttest are analyzed from two different angles. The first angle looks at the results of both tests individually, sighting overall differences between students. In addition, the grades are compared between students in the different parts of the exam. This allows the researchers to find points of weaknesses and strengths students had prior to entering CBE. The next step depended on the comparison of results between the two tests, in a general overview and a part-by-part analysis. The results of the pretest showed significant variation in the students' abilities. While the average percentage is %47.4, the minimum is only %21 and the maximum grade is %78. This reflects a highly significant difference in the abilities of the participants. As for the reading exam, the mean is 12.38 out of 24 that is % 51.58. The lowest percentage obtained in the reading exam is %8.3, and the highest grade is %87.5. The difference in the reading comprehension section is quite high, yet the difference in the language use section is even higher; where some students have scored a full mark (19/19), others have only scored a two.

The average grade in the language use section is 9.7. The exam contained 15 questions, and the highest grade in this section is 13 (%86.6), and the lowest is 1 (%6.6). The vocabulary section contained 27 questions and the highest score was %81.4. The lowest grade obtained was %18.5, which is the highest low in comparison to other parts. This difference is the least between the groups as the average grade was %49.9. The final part in the exam is the 'written expressions' part, which contained 15 questions. The lowest grade in this section was 0, and the highest grade was 12 (%80). Thus, the average grade obtained would be (%42.2).

The difference between the students' abilities in every section individually has shown to be highly significant. By conducting a T- Test (one sample test), it is found that only the 'written expressions' test in the Placement test did not show significance. The significance between grades in the 'language use' part is 0.011 (where f is significant at $f \leq 0.05$). The grammar part also shows significance of $f=0.006$, while the vocabulary part is significant at $f<0.001$. The reading comprehension part also shows significance of $f=0.016$.

Analysis of the posttest

The second exam was given after four years of studying in CBE. It is assumed that the grades, at this stage, would improve. As expected, the posttest shows higher grades. However, in some cases the grades were not significantly higher, and sometimes a little less. At first, the mean score is collected for each section, with the highest and lowest grades also obtained. The best score in the repeated test is %87 and the minimum is %31, while the mean

is %54.3. A one sample test shows that the difference between the two results is significant where $f < 0.001$.

\

In the reading part, the highest grade is 24 (%100) while the lowest is only %20.8, making the average %57. When compared to the first test, the difference is found to be insignificant ($f=0.23$). This might be due to the fact that the grades in the reading comprehension section seem to be quite high in comparison to other sections in the exam. However, significance is found in comparing the 'language use' pretest results where $f=0.001$. The highest grade obtained in the repeated exam is 18 out of a possible 19, while the lowest is a mere 4. The average score obtained is 11 in comparison to a 9.7 in the original Placement test.

In the grammar part, participants scored an average of %56.9 (as opposed to %46 obtained in the pretest test). The highest grade obtained in the repeated test is 14 out of a possible 15 and the lowest grade obtained is 2. In conducting a paired T-Test, it is found that the difference between the two grammar parts is significant (0.05).

The exam also includes 27 vocabulary questions. The highest grade in this section is 24 (%88.8), the lowest grade is 4 (%14.8). The average score is 13.68 correct answers out of 27, while in the pretest the average is 12.38. No significance is found in this section as the difference between the grades is not high. Finally, the written section has shown a lower score of 3 out of a possible 15, which is %20. The highest grade obtained is 11 (%73.3). Thus, the average grade is %50.4. Significance is found to be quite high when comparing the two parts of 'written expression' in a paired T-Test, where $f < 0.001$. This is due to the fact that the average score gained in the pretest is %37.5.

It is worth noting that significance is found within each part as the difference between students still kept its place. The overall exam shows significance of $f < 0.001$. A closer look within the different parts of the exam shows the reading exam to be the only part that has no significance and vocabulary too (0.171). There is high significance in the language use part (0.011), the grammar part (0.006), the vocabulary part (0.00) and the written expressions part (0.016). In addition, the part with the lowest grade in both exams is the 'written expressions' part, while the highest score in the pre test is found in the reading exam. In the repeated test, the highest score is obtained in the 'language use' section. The highest difference between the two exams is found in the 'grammar' section where the students have shown improvement of %10.93. On the other hand, the 'reading comprehension' section shows the least improvement (%5.58). This might be due to the fact that the achievement in the reading section was better to start off than the other sections.

DISCUSSION

An initial look at the results of both the pretest and the posttest reveals that there is an overall improvement. However, it is not the degree of improvement expected after four years of studying English as a major. To be teachers of English, the program is supposed to bring about better improvement in the students' competence of the four skills. The mean of the scores of the pretest was %47.4 while the mean of the scores of the posttest was %54.3. Of course, we need to bear in mind that the placement was originally designed to measure students' mastery of the skills at an intermediate level. That is to say, the test takers are those

who studied English during their school years as one subject among other subjects, so it was expected that after four years of studying English as a concentrated discipline the results should be much better. It is even more disappointing considering the fact that the posttest, used in this study is the same as the pretest designed to measure the abilities of those students who joined the department. This is also true when we look at the students' performance on each section of the test.

In the "Reading" section the mean score obtained in the posttest was %57, compared to %51.58 in the pretest. In the "Language Use" section, the mean score obtained on the posttest was %57.89 in comparison with %51.05. In the "Grammar" section, participants obtained a mean score of %56.9 in the posttest as opposed to %46 in the pretest. The "Vocabulary" did not show better results as the mean score in the posttest was %57.00 while the mean score in the pretest was %51.58. In the "Writing" section though the mean score obtained in the posttest is significantly higher (%50.40) than that obtained in the pretest (%37.5), still both scores fall below the level required of students majoring in English. Consequently, in response to the questions posed by this study, the differences between the results of posttest on all sections (representing the English skills) and those obtained in the pretest are not statistically significant, whereas, according to the goals of the English program in the College of Basic Education we should expect statistically significant differences to the advantage of the results of the posttest. This is to say, it seems that the English department graduate students are not competent in the skills of the English language. The picture becomes gloomier when we come to know that those students graduate to be teachers of English in public school.

In light of the foregoing discussion in the introduction about the goals and outcomes of the English program, the weak performance of the students on the posttest, which is a replica of the pretest, suggests quite clearly that the skills needed to achieve the goals of the English program have not been attained by the students of the English Department in the College of Basic Education. The language ability, or rather the lack of it, which the participant of this study exhibited, indicates that they have not come even close to the desired language skills expected of students who major in English. They didn't exhibit satisfactory abilities, befitting English graduates, in reading, language use, grammar, vocabulary or writing. Are the results obtained by the students on the posttest true reflection of the poor mastery of the language skills required by the program? We can say with a high degree of certainty that the students of the English Department in the College of Basic Education, for some reasons, have not been able to develop their abilities in the English language proficiency to a level required of students graduating from English departments. The question to be raised here is: what are the reasons behind their poor attainment of the language proficiency? We believe that in order to arrive at a satisfactory answer for the question and identify the possible reasons, other aspects of the English program in the College of Basic Education, and perhaps in other universities in the country, must be thoroughly investigated. Careful studies devoted to investigating the implementation and the process phases of the program as well as the factors surrounding its implementation must be considered (see Isaac and Michael (1981) for discussion on the phases of program evaluation). However, as we mentioned under the limitations of the study these aspects of the program (the implementation and the process

phases, the teaching methods, the attitudes of both instructors and students) are not within the scope of this study.

An important issue that, we believe, has serious consequences for the students' poor proficiency is that the English program under consideration has ignored a very crucial element in its implementation, pivotal to any program, which is the process of evaluation. As we pointed out in the background of the study, improvement and effectiveness of any program are closely associated with evaluation. As Webb (1997) points out, to develop sound programs, it is important to ensure alignment between expectations and assessment. He defines alignment as " the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do" (p. 3). This simply means that achievement of outcomes and goals is very much intertwined with evaluation. Alignment is not achieved by one time evaluation attempt or single test, but, as Webb points out, through more comprehensive collection of assessments that is in full congruence with the curriculum. That is, there should be a coherent system of evaluation designed to ensure that students are learning what they are supposed to learn, and continuously ensuring that they are moving towards the attainment of the outcomes of the program.

As such, this important aspect was missing in the implementation of the English program at the College of Basic Education. There has not been any form of evaluating what instructors are doing or what students are learning, let alone having a continuous coherent system of evaluation. Indeed, this is a major deficiency in all educational programs in our institutions and other universities, whether these are large-scale educational programs, or intervention programs, or even frequent amendments on existing programs. In our institutions, once a reform, a program, or an initiative is developed, no stone will ever be turned on such a change. No evaluation procedure will ever take place to find out whether a proposed program or change works or not. The status quo could stay forever the way it is, and oftentimes someone may drop the idea or make new changes on the premise that the first proposed initiative failed. Did it really fail? Why did fail? How did it fail? Questions that nobody, including those who sponsor the new change, could answer. This is simply because program evaluation is not part of our educational policies and practices. Lack of systematic process of evaluation, we believe, has directly influenced the students' poor achievement and attainment of the program outcomes. The English program has been implemented for more than fifteen years now, but has not incorporated any system of evaluation as part of its implementation, so the students' low achievement in the English language proficiency would seem to be an inevitable result.

Pedagogical Implications and Recommendations

On basis of the foregoing discussion, we can attempt to specify the conditions that must be available in order for the English program at the College of Basic Education to achieve its goals and outcomes. In what follows, we present our recommendations that specify such conditions:

1. For any educational program to yield fruits, it must incorporate the process of assessment and evaluation as part of its implementation. We, therefore, recommend that step be taken right

away to develop a coherent continuous system of evaluation that is in alignment with the objective of the teaching program. This is necessary in order to provide very much needed feedback to all who are involved in the teaching process so that the program can be evaluated and improved. Information technology can be utilized to create databases and baseline information on students, teachers, courses and testing practices, so that evaluation processes and research projects can be conducted easily.

2. To get reliable information about students' abilities before they start the program and after taking each of the courses the program offers, testing and grading procedures that are currently followed must be completely revised. They must be done in professional manners, and the students' results must reflect their true and accurate achievements. Course evaluation of students' achievements must be kept away from questionable evaluation practices where students' results don't reflect their achievements.
3. It is important that a review process of the English program takes place sooner than later in order to revise the program and make the necessary amendments of the teaching and learning experiences. Some of the important aspects that experience of the College of Basic Education shows through this study include, but not restricted to, the following. First, the outcomes and objectives of the program must be defined in clear measurable terms. Second, the evaluation process must be included in the program design and its procedures must be specified. Finally, the course plans and components must be revised in order to establish a link between the curriculum and the implementation of the program and ensure alignment or congruence between the objectives and the assessment.
4. It is recommended that further studies be conducted on the course plan and curriculum of the program, as well as on the attitudes of teachers and students. In doing so, we can obtain comprehensive and sound evaluation of the English program.
5. It is recommended that students who graduate from the English department be subjected to standardized tests that are recognized internationally in assessing the English language proficiency. Such tests would enable us to measure the students' abilities in the English skills against standardized levels, and hence would help us place students on an international scale.
6. One or two remedial courses could be offered to students who score lower than a pass score (60%) in the placement test prior to their college acceptance.

REFERENCES

- Cronbach, L.J. (1991). Course improvement through evaluation. *Teachers' College Record*, 64, 672-683.
- Fitz-Gibbon, C.T., and Morris, L.L. (1987). *How to design a program evaluation*. LA: Sage Publications, Inc.
- Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. San Francisco: Jossey-Bass.
- Gredler, M. E. (1996). *Program evaluation*. NJ: Prentice Hall.
- Isaac, S. and Michael, William (1981). *Handbook in research and evaluation* (2nd ed.). San Diego: EdITS Publishers.
- Kelly, A.V. (1999). *The curriculum: Theory and practice*. London: Paul Chapman.
- Kiely, Richard. (2009). Small answers to the big question: learning from language program evaluation. *Language Teaching Research*, 13(1), 99-116. doi:10.1177/1362168808095525
- Lynch, B.K. (1996). *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.
- Mackay, R. (1994). Understanding ESL/EFL program review for accountability and improvement. *ELT Journal* Volume 48/2. Oxford University Press.
- Ornstein, Allan C., and Hunkins Francis P. (2004). *Curriculum: Foundations, principles and issues*. Englawood Cliffs, NJ, Prentice Hall.
- Payne, A. (1994). *Designing educational project and program evaluations*. London: Norwell.
- Pryor, J., and Torrance, H. (1996) Teacher-pupil interaction in formative assessment: assessing the work or protecting the child. *The Curriculum Journal*, 7.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Stufflebeam, D. L., and Shinkfield, A. J. (1985). *Systematic evaluation*. Boston: Kluwe-Nijhoff.
- Stufflebeam, D.L. (2000). The CIPP model for evaluation. In T.Kellaghan and D.L Talmage, H. (1982). *Evaluation of programs*. New York: Free Press.
- Tunstall, P., and Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Association*.
- Tyler, R.W. (1949). *Basic principles of curriculum and instruction*. Chicago: The University of Chicago Press.
- Webb, Norman (1997). *Criteria for alignment of expectations and assessments in mechanics and science education*. Washington, DC: Council of chief state school officers.
- Weston, C., McAlpine, L., and Bordonaro, T. (1995). A model for understanding evaluation in instructional design. *Educational Technology Research and Development*, 43(3), 29-46.
- Worthen R., and Sanders, R. (1998). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.

Appendix**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
overall posttest	50	31	87	54.30	12.364
reading posttest	50	5	24	13.70	3.412
language use posttest	50	4	18	11.18	3.160
grammar posttest	50	2	14	8.54	3.265
vocabulary posttest	50	4	24	13.68	4.897
writing expressions posttest	50	3	11	7.56	2.022
overall pretest	50	21	78	47.40	13.265
reading pretest	50	5	21	12.38	3.374
language use pretest	50	2	19	9.72	3.742
grammar pretest	50	1	13	6.90	3.125
vocabulary pretest	50	5	22	12.00	4.518
written expressions pretest	50	0	12	6.34	2.504
Valid N (listwise)	50				

ANOVA

		Sum of Squares	Df	Mean Square	F	Sig.
reading posttest	Between Groups	413.833	31	13.349	1.534	.171
	Within Groups	156.667	18	8.704		
	Total	570.500	49			
language use posttest	Between Groups	407.297	31	13.139	2.881	.011
	Within Groups	82.083	18	4.560		
	Total	489.380	49			
grammar posttest	Between Groups	441.753	31	14.250	3.180	.006

One-Sample Test

	Within Groups	80.667	18	4.481		
	Total	522.420	49			
vocabulary posttest	Between Groups	1111.047	31	35.840	10.106	.000
	Within Groups	63.833	18	3.546		
	Total	1174.880	49			
writing expressions posttest	Between Groups	164.403	31	5.303	2.658	.016
	Within Groups	35.917	18	1.995		
	Total	200.320	49			

ANOVA

		Sum Squares	of Df	Mean Square	F	Sig.
reading pretest	Between Groups	474.280	29	16.354	3.917	.001
	Within Groups	83.500	20	4.175		
	Total	557.780	49			
Language use pretest	Between Groups	596.163	29	20.557	4.573	.000
	Within Groups	89.917	20	4.496		
	Total	686.080	49			
grammar pretest	Between Groups	384.667	29	13.264	2.827	.009
	Within Groups	93.833	20	4.692		
	Total	478.500	49			
vocabulary pretest	Between Groups	864.500	29	29.810	4.400	.001
	Within Groups	135.500	20	6.775		
	Total	1000.000	49			
written expressions pretest	Between Groups	211.970	29	7.309	1.535	.161
	Within Groups	95.250	20	4.762		
	Total	307.220	49			

Significance in overall grades

	Test Value = 0					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
overall posttest	31.055	49	.000	54.300	50.79	57.81
overall pretest	25.267	49	.000	47.400	43.63	51.17

Significance between all parts:

Reading comprehension:

One-Sample Test

	Test Value = 0					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
reading posttest	28.391	49	.160	13.700	12.73	14.67
reading pretest	25.946	49	.160	12.380	11.42	13.34

Language use:

One-Sample Test

	Test Value = 0					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
language use posttest	25.015	49	.001	11.180	10.28	12.08
language use pretest	18.368	49	.001	9.720	8.66	10.78

Structure:**One-Sample Test**

	Test Value = 0					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
grammar posttest	18.494	49	.050	8.540	7.61	9.47
grammar pretest	15.613	49	.050	6.900	6.01	7.79

Vocabulary:**One-Sample Test**

	Test Value = 0					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
vocabulary posttest	19.755	49	.065	13.680	12.29	15.07
vocabulary pretest	18.783	49	.065	12.000	10.72	13.28

Written expressions:**One-Sample Test**

	Test Value = 0					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
writing expressions posttest	26.439	49	.000	7.560	6.99	8.13
written expressions pretest	17.904	49	.000	6.340	5.63	7.05

Average percentage

	%
overall posttest	%54.30
reading posttest	%57.08
language use posttest	%58.84
grammar posttest	%56.93
vocabulary posttest	%50.66
writing expressions posttest	%50.40
overall pretest	%47.40
reading pretest	%51.50
language use pretest	%51.15
grammar pretest	%46.00
vocabulary pretest	placement
written expressions pretest	%42.26
Valid N (listwise)	