

CHAPTER ONE: INTRODUCTION

1.1 Evaluation: An Overview

The key sense of the term 'evaluation' refers to "...the systematic determination of merit, worth or significance of something" (Scriven, 1999, p.1). The Joint Committee on Standards for Educational Evaluation (2000) states that lifelong evaluation is important to individuals, institutions within society and society itself. According to McNamara (2000, p.35), "...making judgments about people is a common feature of everyday life. We are continually evaluating what others say and do, offering criticism and feedback." Payne (1994, p.1) adds, "People are always evaluating. We do it everyday. We buy clothing, a car or refrigerator. We select a movie or subscribe to a magazine. All these decisions require data based judgement." Hence, all work pertaining to evaluation involves collecting and sifting through data, making judgement and inferences from the information gathered. The judgements we make may shape our actions, or we may ignore the implications.

In the field of education, we are particularly concerned to know whether we are doing the correct things and if our actions will lead to what is good or desirable. In order to find this out, we have to establish formal methods of reaching our judgement. Evaluation can aid in making such deductions and judgements. Evaluations that are properly commissioned, conducted, disseminated and used, can help improve education in general and programs, students, teachers and

materials in particular (Joint Committee on Standards for Educational Evaluation, 2000).

Major reforms in education in the past have consistently been accompanied by major reforms in the methods of evaluation. The Joint Committee on Standards for Educational Evaluation (1999) states that throughout the 30s, 40s and 50s, the progress in the field of evaluation was mainly in assessing student performance. But since the 60s, emphasis has also been on the developments related to the assessment of educational programs, projects and course materials. This could be attributed to the outside pressures on educators to explain and justify what they are doing. Evaluation is one way to produce the sort of information required for public accountability. Hence, evaluation has today become an integral and inevitable feature of all aspects of education.

Murphy (1981, p.3) uses a dynamic model (Fig.1.1) to show the importance of evaluation in a curriculum and states that evaluation is equally essential as the methods for conveying content and achieving objectives as each of the components informs and is informed by the other.

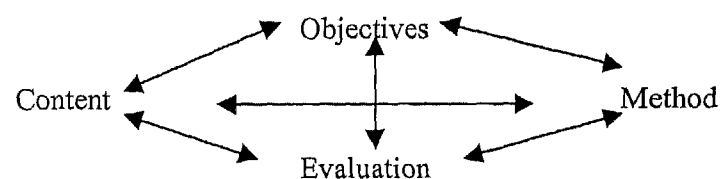


Figure 1.1: Evaluation in a Curriculum

Therefore, evaluation can be viewed as a procedure for making judgements and assessing the value and quality of what is being carried out. Murphy (1981, p.47) states that:

Evaluation should not be seen as a one way process in which learners are judged by teachers, instead, as a multi-way dynamic system in which the learners are involved interactively with others in making judgement about themselves, both as learners and users of the language on the basis of criteria which are defined and negotiated in terms of the learning situation.

Furthermore, it is important that the evaluator makes explicit what is being evaluated and the criteria by which judgements are being made. As stated by Rea-Dickins and Germaine (1992), evaluation in an educational context should be undertaken in a systematic manner, in line with guiding principles based on carefully defined criteria.

Evaluation is carried out for several purposes. Chitravelu, Sithamparam & Teh (2004, p.191) assert that, "One has to bear in mind that the purpose for which evaluation is undertaken affects the way the evaluation is done and how the results are interpreted and used." According to Rea-Dickins and Germaine (1992), the main purpose of undertaking evaluation is to explain or confirm existing procedures. It can help determine if the current classroom practices are effective or otherwise. Information that is gathered is then used to indicate to the teacher the value of trying out new approaches and clarifying what works and what does not.

Another motivation for evaluation is to use the information gathered to initiate change (Rea-Dickins and Germaine, 1992). For example, if the current evaluation shows that the existing textbook is not very comprehensive for the course, the text could then be replaced by a new text in the next term. Through evaluation, teachers can also keep track of a student's relative performance and this could act as an indication of achievement and proficiency for external purposes.

Thus, evaluation is important in the teaching and learning process as it can help to make judgements about pupils' academic performance, provide feedback and incentives to pupils, and also gather information which can be used for the planning of courses or to decide on a possible course of action (Airasian, 1997).

1.2 Evaluating Writing

It is not surprising that extended-response items, typically short essays, are now an integral part of most assessment programs. This is mainly due to the fact that extended response items, such as compositions, provide an opportunity for students to demonstrate a wide variety of skills and knowledge, including higher order thinking skills such as synthesis and analysis (Rudner and Gagne, 2001).

Chitravelu et al. (2004) explain that there are two primary reasons for evaluating composition writing. The first reason for evaluating writing is to motivate

students to write better by enabling them to recognize the progress they have made and appraise them for the good points in their work. Another reason is to highlight their shortcomings or weaknesses and instruct them to do corrections and improve their quality of work.

However, since a written assignment or composition item permits each student to construct a unique response, no one, single answer key can be applied uniformly to all responses. This subjectivity makes scoring a writing assignment extremely problematic. Past research has shown how extremely unreliable markers are, both "...in their own inconsistency and in their failure to agree with colleagues" (Heaton, 1990a, p.78). Scorers may differ enormously in respect of "...spread of marks, strictness and rank order" (Heaton, 1990a, p.144).

Even with rigorous training, differences in the background, training and experience of the scorers can lead to subtle, but important differences in grading (Rudner and Gagne, 2001). Heaton (1990a) also stresses that this unreliability exists because scorers may award marks based on "...what a student has written, what they believe the student meant, handwriting, general impression and previous knowledge of the student" (Heaton, 1990a, p.144). Thus, characteristics such as handwriting, writing style, accuracy, neatness, fatigue, identity of the pupil etc. can influence a teacher's reaction to an essay answer, although, none of these characteristics has anything to do with the actual content of the pupil's response (Airasian, 1997). Hence, little agreement is found between scorers in

assessing writing and little agreement is found as to what constitutes good levels of proficiency in writing.

Therefore, although the field of scoring second language composition has developed in the last three decades, it has proved and continues to prove an issue of controversy and disagreement to researchers, teachers and students.

1.3 Issues to Consider when Scoring Written Work

Chitravelu et al. (2004) has also outlined several issues such as pedagogic usefulness and practicality, to be considered when evaluating student writing and further explains the criteria to be used in considering work for evaluation. There are instances whereby assessing every piece of work might not necessarily help students produce better work. In the case of journal writing for instance, frequent evaluation might deter a student from wanting his or her work to be evaluated. Another concern is that of practicality. If the teacher has large classes of thirty students or more, then, it might not be practical to evaluate every piece of writing. However, in smaller classes, it would be possible to evaluate every piece of written work.

As stated by Chitravelu et al. (2004), in most cases, it is the teacher who evaluates. Therefore, the teacher has to play several roles as evaluator. Firstly, the teacher plays the role of a reader. As a reader, the teacher should tell the students

what the strengths of his or her writing are, what is unclear etc. Next, the teacher is also an advisor who makes suggestions for improvement. However, whatever the suggestions, the teacher must ensure that it is not beyond the student's capacity to rewrite and correct. Another role for the teacher would be that of facilitator or resource person, someone who could inform or direct students to relevant resources such as dictionaries and web sites. Besides just being an evaluator, the teacher needs to encourage pre-evaluation and self-evaluation. According to Heaton (1990b), these types of evaluation can help students examine, evaluate and improve their own writing and become more sensitive to the idea of the audience and recognize their strengths and weaknesses.

According to Reid (1993; p.229), the primary objectives of evaluating written work should be "...long-term improvement and cognitive change." Thus, the results of evaluative intervention should bring about some change in student writing performance.

1.4 Statement of the Problem

Composition writing has become a focal point for critics today, as writing, for obvious reasons, is one of the most visible products of education. Incorrect word usage and grammatical errors have been taken to be indications of inadequacies in the education system. It is obvious that the problem of not being able to write well still persists in our Malaysian schools and institutions of higher learning.

Chitravelu et al. (2004) state that although Malaysian students can communicate their ideas relatively well in English, their writing ability is still inadequate. The scenario is very much the same in the numerous private colleges in Malaysia. Students are still unable to express themselves effectively in composition writing. Although ESL (English as a Second Language) students entering college require training in all language skills, writing is of major importance at this level of education, as students are mostly evaluated through written means such as research papers and essay exams. And while success in college is highly dependent on a student's writing skills, non-native English speaking students show the least proficiency in this area. Hence, teachers have long been concerned about students whose poor written English prevents them from reaching their full potential either at university or in their professional lives. This problem is further complicated by the increase in the number of students with English as a Second Language (ESL), attending college and the changing circumstances and demands of educational institutions.

Time is another important consideration in working with such students. Due to limited financial resources, the students are unable to afford extended language courses. Hence, the pressure on the instructor from both, the student and the institution, is to have the student enter the academic mainstream as quickly as possible. The need then, is to train students as quickly and efficiently as possible in the writing skills for successful college study. Therefore, college instructors of

English as a Second Language have now come to recognize the need for more accurate and efficient methods of scoring the writing of their students.

The researcher is an English lecturer at an established private college – Taylor’s College Subang Jaya (TCSJ), whereby the problems faced by the students and the demands from the institution and students, on the ESL lecturers are very similar to those discussed above. TCSJ has an annual intake of approximately 2500–3000 pre-university students. These students need to obtain a minimum overall band of 6 to 6.5 in the IELTS (International English Language Testing System) exam or 550 in the TOEFL (Test of English as a Foreign Language) / TWE (Test of Written English) exam to gain admission into a foreign university that is recognized by the Malaysian government.

In addition, universities such as the University of Technology Sydney and the University of Sheffield expect a minimum band of 6 in the writing component to guarantee admission. Otherwise, the students are only given a conditional offer. Therefore, the writing component is extremely important. In order to achieve a band of 6 in the IELTS exam or a Band 5 in the TWE test, a student has to be able to write a 250 word academic essay with adequate content, organization, language and vocabulary. Due to time constraints, the students of TCSJ are unable to attend an extended proficiency course. Hence, they enrol for short-term (intensive) IELTS and TOEFL preparatory courses. The pressure is thus, on the lecturers to help students improve their language skills (writing skills especially),

within a very short time frame in order to enable them to obtain the minimum overall band in the IELTS/TOEFL exam.

At TCSJ, the two main methods of scoring students' writing used by the ESL lecturers in short-term (16-hour) language preparatory courses are the analytic and holistic methods. The researcher's interest in investigating these two methods of scoring students' writing in short-term language courses was motivated by a number of factors.

Firstly, previous studies comparing the holistic and analytic methods of scoring writing in short term language preparatory courses are scarce and unable to confirm whether one specific method of scoring (the holistic or analytic) has a greater effect on student writing performance compared to the other (analytic or holistic) approach to scoring compositions.

Secondly, lecturers at TCSJ were uncertain as to which method of scoring would be more effective in improving students' writing performance. Some felt that neither method of scoring could make a difference in short language preparatory courses.

Thirdly, although many lecturers felt that the analytic method was more effective in improving students' writing performance they were doubtful whether it would work in a 16 hour (short term / intensive) language preparatory course.

Lastly, lecturers were divided in their opinions on the overall effect the analytic or holistic method of scoring would have in terms of motivating students to improve on their writing within such a short time span.

These factors, thus, motivated the researcher to investigate the usefulness of these two methods of scoring essays in short-term (16-hour) language preparatory courses and to identify any significant relationship(s) that exist(s) between the method of scoring used and the writing performance of students.

This study aims to answer the following research questions:

1. Do the two methods of scoring (holistic and analytic) affect students' writing performance in short-term (16-hour) language preparatory courses?
2. Do the two methods of scoring (holistic and analytic) affect students' attitude toward their writing?
3. Is there a preference for the holistic or analytic scoring amongst teachers teaching short-term (16-hour) language preparatory courses?

1.5 Significance of the Study

This study will have significant implications for those who devise and use the holistic and analytic scoring methods. ESL (English as a Second Language) writing teachers, especially those involved in short term language preparatory courses, will be able to use these findings to determine whether it is worthwhile for them to spend a lot of time and effort scoring their students' essays. Those who commonly use these methods of scoring will also gain a better understanding of which method of scoring is more useful in helping students improve their writing skills within a short span of time. The findings from the present study would also be able to offer insights to lecturers on the effects each method of scoring has on the attitude of students. Finally, the results should also help to clarify the immediate doubts expressed by the lecturers at TCSJ, as to which is a more effective method of scoring in short-term language courses and prove useful to teachers of other institutions of higher learning or schools who are involved in teaching short-term language preparatory courses.

1.6 Limitations of the Study

The results of this study are limited as it was only carried out in one private institution. Therefore, the findings are not generalizable for all ESL students. Moreover, the study only looked at one type of short term language preparatory

course – IELTS; hence, the findings may not be applicable to students pursuing other short-term English courses.

1.7 Definition of Terms

Evaluation

Evaluation is a systematic process of determining the extent to which pupils achieve instructional objectives. Davies et al. (1999) give evaluation a broader definition. It is viewed as the systematic gathering of information for purposes of decision-making and is stated to compare two components: information and value judgements (Davies et al. 1999). The information gathered could be either qualitative or quantitative. The more reliable and relevant the information gathered, the greater the probability of making correct decisions.

Scoring

The procedure for giving numerical values or scores to the items in a test is called scoring. The steps involved in any type of scoring play a key role in ensuring that the scores are reliable and that they are used for valid reasons.

Holistic Scoring

According to Hughes' (2002) definition, holistic scoring is an assignment of a single score to a piece of writing on the basis of an overall impression of it. Therefore, in this method of scoring, the evaluator reads each essay without marking anything, then rates it as a whole and assigns a single score within a given range.

Analytic Scoring

Hughes (2002) defines analytic scoring as a method of scoring which requires a separate score for each of a number of aspects of a task. Weir (1990) states that it is a method of marking whereby each separate criterion in the marking scheme is awarded a separate mark and the final mark, is a composite of a piece of writing. The marking scheme is usually drawn up by the examiner or a body of examiners.

Band Descriptor Scale

A rating scale, which shows the degree of success or failure of test takers in performing a certain task. It allows us to know how successful or unsuccessful a test taker was. Band descriptor scales are usually contrasted with simple success criterion. A band descriptor scale permits comparisons between test takers as well as giving some idea of where a performance lies in absolute terms.