CHAPTER I

INTRODUCTION

Assessment plays an integral role in the improvement of instructional practice for teaching as well as student learning. It serves a number of functions such as diagnosing students' strengths and weaknesses, monitoring their progress through the curriculum and providing instruction tailored to instructional objectives therefore enhancing student achievement.

However, although diagnosis is a major concern throughout the instructional process and is central to its success, teachers do not optimally use diagnostic tools such as tests to obtain feedback. If the main purpose of testing is to assess a student's status with respect to a domain of learning objectives, teachers need to be able to pin point why students perform as they do and prescribe instruction accordingly. Thus there is a need for teachers to be able to link assessment with teaching, and to make instructional decisions rationally.

One rational approach would be to use diagnostic test data to help make instructional decisions. If the teacher is to link assessment information to teaching, the assumption is that there exists a relationship between the diagnosis of learning abilities and the treatment to be administered to the students. The diagnostic tests, which are the instruments of diagnosis, used in the diagnostic assessment model are based on the assumptions that relevant abilities exist and can be measured, the tests are reliable, the tests are valid and that there is a positive impact of diagnosis on achievement.

Significance of Study

Assessment is the process of identifying, gathering and interpreting information about progress in students' learning. Assessment information for students will provide feedback to monitor and evaluate their learning progress as well as to facilitate self assessment. Effective teaching practices too rely on diagnostic assessment. This means that teachers need to use their professional judgment to make decisions about student learning progress, develop assessment strategies and plans as part of the teaching and learning programme, identify aspects of the content and skills to be assessed within the unit of instruction as well as helping students progress towards the achievement of personal learning goals and curriculum objectives.

Traditionally, assessment in the Malaysian education system is based on norm-referenced measurements. Public examinations of academic achievement or cognitive ability such as the *UPSR*, *PMR*, *SPM* and *STPM* can be classified as norm-referenced measures. Norm-referenced measures ascertain a student's performance in relationship to the performance of other students on the same measuring instrument. The meaningfulness of the individual student's score emerges from the comparison. Since norm-referenced measures are devised to facilitate comparisons among students, their primary purpose is to make decisions about students based on their relative standing of the students in a group. Norm-referenced measures of achievement outcomes, however, do not indicate which instructional objectives of a subject curriculum a student has mastered. Noting the importance of tests as tools for monitoring student progress, the Ministry of Education has planned for wider use of criterion-referenced testing in the school curriculum. This is as stated in the Seventh Malaysian Plan (1996-2000):

> With the full implementation of 11 years basic education, there will be continuous progression of students from Standard 1 to Form V. In this regard, extensive monitoring and evaluation of students' performance at all levels will be undertaken, including greater use of the Criterion-Referenced Test, which will further develop new teaching and learning approaches, as well as enhance students' academic performance, particularly in the early years of schooling. This will enable early detection of slow learners and low achievers and remedial measures be implemented to improve their performance. These measures will also motivate students to strive for excellence and contribute towards increasing the success rate at various levels (p.326).

To achieve excellence in educational outcomes as stated above, new learning strategies that possess objectives-based characteristics are needed. These learning strategies should lead to instructional programmes being set up and executed according to well-defined, specific learning objectives. Different approaches may be used to realise these objectives. The focus of this study is the development of a diagnostic assessment model and implementing this model in the classroom. Diagnostic assessment is a form of assessment that uses criterion-referenced tests for formative assessment purposes. It is hoped that the interpretation of the feedback obtained from student test scores will allow teachers to provide remedial assistance to individual students. Moreover, the teacher can use the feedback from students for teaching and curriculum evaluation.

Research Questions

This study specifically wish to answer the following six research questions:

(1) What are the effects of an instructional strategy based on diagnostic assessment on physics achievement scores in the classroom?

(2) How are students with different achievement scores in physics affected by the mastery versus non-mastery strategy employed in diagnostic assessment?

(3) What are the perceptions of students toward the use of diagnostic tests in the classroom?

(4) Is there a significant difference in the perceptions of different physics ability groups towards the classroom diagnostic test?

(5) Is there a significant difference in students' perceptions of multiple-choice versus essay type diagnostic tests?

(6) Do different item discrimination indices differ in their effectiveness in discriminating between mastery and non-mastery students on a classroom diagnostic test?

Limitations of the Study

The pretest-posttest design used in this study does not involve a control-group. There is also the absence of randomisation that is required for a true experimental design. Hence the quasi-experimental design used may fail to account for uncontrolled rival hypotheses or extraneous variables that could also explain any obtained findings. In order to determine conclusively that any observed difference is caused by the treatment effect, a control-group for comparison should have been selected. This control-group's performance can then be compared with the performance of the experimental-group that received the treatment of diagnostic

assessment. However this is not possible in this study because there is only one physics class of students in the Degree Foundation Programme at the Sepang Institute of Technology. The random selection of students to be placed into experimental and control groups cannot be carried out. In this study, the means of the physics achievement scores before and after the treatment are compared for each of the three ability groups: low, medium and high. A one directional matched-group t – test is used to test whether there is any significant improvement in the scores for each of the ability groups. Due to the small sample size of each group (n = 11), the stability of the t-test results will be low.

In the absence of random selection and a control-group, and realising the need for eliminating some of the extraneous variables that may exist in the pretest-posttest design, an interrupted time-series single subject design is included. The procedure involves studying the effect of diagnostic assessment on the mean test scores of the three achievement groups. The series of test scores in the first semester are used as pre-treatment scores while the criterion-referenced test scores in the second semester represent the post-treatment scores. The pre-and post-treatment test scores for each group are plotted on a graph and visually inspected to determine whether the pattern of behaviour shows a real effect of diagnostic assessment in the classroom. The assumption made in this design is that the pattern of pretreatment scores for each group would have continued to be stable if the treatment had not been given. This limitation of the design can be reduced if a third phase involving the withdrawal of diagnostic assessment is undertaken. If the pattern of test scores reverts back to its pretreatment level, then this will demonstrate that diagnostic assessment produced the change in achievement scores observed and not some other extraneous variables or

rival hypotheses. However this threat to internal validity cannot be removed completely using the time-series design in this study. This is because of the fact that the physics course covers only two semesters and not three. The third phase of reverting back to the pretreatment instructional strategy could not be carried out.

Item analysis has many purposes. It is used to inform teachers about item performance in the construction of tests as well as to detect flaws or ambiguity in an item so that it can be revised. The focus of this study is not on item selection prior to teacher-made diagnostic tests since this is highly impractical. The study is limited to the inspection of raw observed item data as well as the application of discrimination and difficulty indices to identify items that discriminate between a group of 'masters' and a group of 'non-masters' after a test has been administered. However, the results of the study will help the researcher to build up a bank of carefully calibrated test items. This 'item bank' will be of great practical value to the researcher in future test development.