THE PROBLEMS OF THE "NON-FAILURE" POLICY IN ELEMENTARY SCHOOLING

THE PROBLEMS OF THE "NON-FAILURE" POLICY IN ELEMENTARY SCHOOLING

By

PAULA MARTELLI, B.A.

A Project

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts (Teaching)

McMaster University

September 1985

MASTER OF ARTS (TEACHING) (1985) McMASTER UNIVERSITY

(Philosophy) Hamilton, Ontario

TITLE: THE PROBLEMS OF THE "NON-FAILURE" POLICY IN ELEMENTARY SCHOOLING

AUTHOR: PAULA MARTELLI, B.A. (McMASTER UNIVERSITY)

SUPERVISORS: Dr. S.M. Najm, Dr. C. Beattie

NUMBER OF PAGES: 75

ABSTRACT

In recent years the widely publicized criticism that a great number of students lack the basic skills have deepened the concerns of teachers, parents and employers about the adequacy of student promotion policies. Guidelines introduced by the Ontario Ministry of Education since the mid-seventies have failed to address this issue by allowing local school boards to set and implement their own policies for student promotion.

At present, many elementary schools are committed to the "non-failure" policy claiming that "continuous" promotion is beneficial to students. However, there is strong evidence to support the claim that such policy fails to promote effective learning.

The topic was selected to examine the validity of the assumptions on which such policy is based and to discuss the implications for the elementary schools. Adequate promotion policies are crucial to the effectiveness of education and teachers must question the appropriateness and validity of evaluation procedures.

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my project supervisors: Dr. S.M. Najm, Chairman of the M.A.(T)-M.Sc.(T) Programme at McMaster University, who has always assisted me with expert advice and encouragement and Dr. C. Beattie of the Department of Philosophy, for her valuable insight. I would also like to thank my family for its continuous support throughout my study.

TABLE OF CONTENTS

INTRODUCTION

In the last decade education in Canada and in the United States has been subjected to a new wave of criticism. The charge of the informed critics and of the public at large has been that a great number of students lack the basic skills.[1] The prevailing assumption seems to be that this state of affairs has been the result of widespread adoption of improper promotional policies, and of over-extended curricula both at the elementary and at the secondary level of instruction.[2] Consequently, in the late 1970's in both countries a "back to the basics" movement has been gathering momentum, advocating core curricula and the use of standardized tests as the means to improve the standards of student achievement.[3]

In response to public demands for efficient delivery of educational services, federal and state agencies in the United States, and provincial ministries in Canada have undertaken the task to bring about educational reforms. Interestingly, however, while the "back to basics" movement has fostered a great impetus for the formulation of core curricula, the use of competency testing, and the adoption of more efficient student promotion policies in the United States[4], here in Canada it has resulted mainly in the adoption of core curricula.[5] This is not to say that questions about the testing of student achievement, and about the appropriateness of promotion policies have gone undebated in this country. In fact, both issues are well documented and even acknowledged

by some provincial ministries.[6] Nevertheless, the main thrust of educational reforms in Ontario has been - and still is - on the curriculum.[7]

The question arises as to why standardized and/or competency testing, and promotion policies have seemingly received little attention at the time when educational reforms were in the making. The decision to focus on core curricula was made by the Ministry whose main preoccupation seemed to be that of trying to "...rebuild and maintain the confidence of the public, particularly parents, in our school system."[8], and whose fear was the danger of a "gradual erosion of support" from the public. In addition the responsibilities for testing and for promotion policies had already been handed out to local educational agencies, and the revoking of these arrangements would have been politically counterproductive.

Therefore, while the "considerable study" in evaluation and testing undertaken by the Ministry in 1976[9] resulted in publications of evaluation documents such as the Evaluation of Student Achievement: A Resource Guide for Teachers (Ontario Ministry of Education, 1976), the Ontario Assessment Instrument Pool: A Curriculum Based Aid to Evaluation (Ontario Ministry of Education, 1979) and Evaluation and the English Program (Ontario Ministry of Education, 1979), their use in schools was not made mandatory. Furthermore, even the most recent Ministry document, Ontario Schools: Intermediate and Senior Division (1984), with respect to student promotion policies prescribes only that "Parents should also be made aware of the school's promotion policy..."[10]. Thus, once again, the Ministry is allwing local education agencies to set out

and implement their own policies.

Unfortunately, the evaluation of student achievement, as outlined in OSIS,[11] is still vague. This document, in fact, seems to continue to reflect some of the ambiguous rhetoric characteristic of Ministry guidelines in the past ten years. For instance, in OSIS the Ministry says, "these circulars reflect the policy of the province of Ontario that the program in the publicly supported educational system should be designed to provide the greatest possible opportunity for every student to develop as completely as possible his/her abilities and interests and to meet each student's special needs."[12]. This closely resembles the one found in the Formative Years: "It is the policy of the Government of Ontario that every child have the opportunity to develop as completely as possible in the direction of his or her talents and needs."[13]

In light of the above, the promise that "...curriculum guidelines will describe appropriate evaluation techniques,"[14] does not inspire great confidence that the near future will bring the needed reforms in the evaluation of student achievement because what is needed is not a description of evaluative techniques, but a mandatory prescription of them.

Therefore, as long as the Ministry continues to delegate the authority and the responsibility for formulating and implementing evaluation policies to local boards, there will be no uniformity of standards of student achievement. This represents a serious threat to the effectiveness of elementary education in Ontario, as it will perpetuate the less than satisfactory present state of affairs wherein schools adopt

one of two alternative policies for student promotion: the "non-failure" policy which endorses "social promotion" or "continuous promotion", and the merit policy which requires that students demonstrate the attainment of some standards of achievement.

The aim of this paper is to present and defend the thesis that the "non-failure" policy in elementary schooling engenders serious instructional and educational problems: low standards of achievement; progressive accumulation of learning deficiencies; functional illiteracy and a false sense of self-worth.

At first my intention was to conduct an empirical investigation to study the effect of these two policies within the context of the classroom situation at the elementary level. Regrettably, this has not been possible due to lack of empirical evidence at the local education agencies. In fact, it has not been possible to even obtain the documents which outline the policies. Therefore, this project will only present a theoretical approach.

I will first examine the theoretical bases of each policy and outline the probable effects that each would have in the classroom situation. Secondly, I will present the significant developments which have influenced the evolution of modern evaluation of student achievement, discuss the various measurement instruments, and examine the validity of the claims of psychological and social theories which have shaped the basic assumptions of the "non-failure" policy. Thirdly, I will present an analysis of the concept of evaluation in order to establish the type of criteria and process which ought to be employed in the

evaluation of student achievement. Lastly, based on the conclusion of this investigation, I will offer some thoughts for the possible improvement of current practices.

CHAPTER I

AN ANALYSIS OF CURRENT EVALUATION POLICIES

i) The "Non-Failure" Policy

The "non-failure" policy in elementary schools is an offspring of the philosophy of non-graded education popular in North America in the 1960's. In Ontario this philosophy was officially introduced into the school system by the Ministry document known as the Hall-Dennis Report which advocated "...the complete abolition of the graded system;..." and that "...the concepts of promotion and failure should be removed from the schools..."[1]. Although, in practice, the theory as a whole has been rejected[2], its underlying concepts such as "child-centred" education and "continuous progress" have been retained. These concepts have played an important role in the formulation of the above mentioned policy because their implicit egalitarianism seemed to respond to the demands for democratization of education and to the need to create a more skilled society. Child-centred education, in fact, assumes that it is the function of the school to provide for each child a learning situation tailored to his/her needs.[3] Hence its endorsement was seen as the means to bring about an upgrading of the educational achievements of disadvantaged minority groups. The continuous progress concept, on the other hand, assumes that the child's learning process is an ongoing developmental process determined by his/her intellectual and affective status.[4] It postulates that a child's performance in school

is affected by the way a child perceives himself and the expectation of others, hence this postulation offers the possibility of altering a pupil's performance by appropriate manipulation of the school setting and the feedback on the student's performance.[5]

Moreover, the conceptual framework of the "non-failure" policy is ultimately grounded on psychological theories, such as Piaget's cognitive development theory which describes the mechanism and processes of learning,[6] and sociological theories, such as Brookover's theory of self-concept which describes the optimum affective states of the learner.[7] Therefore, while Piaget's "equilibration" justifies the tailoring of curricula,[8] Brookover's "self-concept" justifies the promotion practices.[9]

In a school system which is committed to the "non-failure" policy, evaluation of the learner's achievement focuses on his/her developmental status rather than on how his/her achievement measures up to pre-set standards. Since it is claimed that the traditional summative evaluation of the child's achievement would be inconsistent with the notion of continuous progress, evaluation, in this context, is formative as it represents a means to improve both teaching and learning. The assessment of the learner's achievement is used, in fact, diagnostically in order to plot the next teaching/learning experience at the end of a unit of study, or at the end of a school year. In the latter case, the evaluation will enable the teacher to place the student in the next grade, regardless of the quantity and/or quality of his/her achievement because, theoretically, the tailored curriculum will ensure his/her continuous progress the following year. This "promotion", it is

claimed, gives the learner self-esteem which, in turn, will enhance his/her subsequent motivation and performance, thus enabling the under-achiever to aspire to and attain a higher level of education.

ii) The Effects of the "Non-Failure" Policy in the Classroom

How does the "non-failure" policy affect the actual classroom situation? The obvious long range effect is that if students do not fail, given the current mainstreaming policy,[10] the same group of learners will find themselves together year after year, from kindergarten to graduation from elementary school. In most cases this is what happens in fact. However, the situation is much more complex, as it causes many structural and instructional problems which merit closer scrutiny.

Let us suppose that there is a group of children entering kindergarten in a school system with the "non-failure" policy, and let us see how this policy will affect them in the course of their elementary education. We may safely assume that this kindergarten class, like any other class, reflects a variety of backgrounds and abilities and that, therefore, it necessitates a tailoring of the curriculum in order to meet the needs of each child. At the end of the school year all the students in this class will be assessed to establish their individual developmental status, but, they will all start grade one in the Fall.

The following year this class will still present a variety of abilities and necessitate tailoring of the curriculum. It may be argued that some of the learners who received remedial instruction the previous year, could have matured enough not to require that special care any longer. True, but some will still require it. Furthermore, it is

possible that, due to the increased difficulty of a higher level of instruction, other students who did not require remedial instruction the previous year might require it now. In other words, the grade one teacher is also faced with the same instructional and structural problems that the kindergarten teacher faced the previous year, but at a little more complex level. At the end of this school year the developmental level of these children will be assessed again and, understandably, the evaluation will show again a variety of achievement levels.

Although this will have been the second year in school for these students, it may be found that some learners, in some areas, might not have acquired learning beyond the level of the first year. Educational research shows that it is normal for some learners to require more time to master certain things than others.[11] That is true, but is the "non-failure" policy really beneficial to these and other kinds of learners in the classroom situation?

From the short analysis presented above we can claim that some kinds of learners will continue to require modified curricula and approaches throughout their elementary schooling, and that teachers will have to implement modified curricula at each grade level.[12] A further implication is that some learners, as they advance through the grades, will find their rate of progress farther and farther behind the level of others in the class, because the progressive complexity of the subject matter may compound their difficulties in learning. Therefore, it is possible in the senior grades to find students whose performance is only at the junior level, but who will, nevertheless, graduate with an incomplete elementary programme.

There are other aspects that we should consider, such as the teacher's task in the classroom and the effects of the teaching/learning routine on all kinds of learners. What is the teacher's task in a senior elementary grade with a class of multilevel abilities? First of all, he/she has a given curriculum which (at this level) is quite diversified and relatively sophisticated in subject matter. Secondly he/she has a set amount of time in which to "cover" this curriculum. However, in this case, due to the multilevels class, the teacher will have to modify the curriculum in order to meet the specific needs of individual students. Does that mean individualized instruction? If the number of students in the class is relatively small, possibly so, but in a regular public school class, where the number of students is quite large (sometimes as many as 35 or 40 pupils) this is not possible.

13

The next best alternative would be to group the students, as much as possible according to their abilities, and to teach them in groups, modifying the curriculum accordingly. This means that the teacher would have to prepare and teach a different lesson for each ability group in the class. Obviously, this teacher's task would be to teach two or three curricula within the same time allotted for the teaching of one.

It may be argued that over the years teachers have learned to rotate teaching time with independant activity (seatwork, followup, creative writing), thus affording extra teaching time to be devoted to the teacher's interaction with groups within the class. But is this routinely possible in a senior grade and in subjects such as science or

history? Would this not affect the quality and/or quantity of teaching and/or learning? In other words, would all kinds of learners receive the optimum curriculum for their level of ability, or would some of them be taught a diluted version of the subject matter even though they are capable of learning more?

Is the situation not stressful for both the teacher, who is responsible for multicurricula, and the learners, who, because of the multilevels, have a limited time to interact with the teacher? It may be argued that there are provisions made to relieve the classroom stress by periodic withdrawal of exceptional learners, whether they be slow or gifted, so that the teacher is allowed to concentrate on the remaining students. But are these withdrawals sufficient to compensate for the difficulties created by the multilevels situation?

Another possible alternative to the one discussed above is that rather than regrouping the learners according to their abilities during the teaching of the lesson, the teacher will provide seatwork geared to the different abilities of the learners. Hence the lesson is presented to the whole class at the same time. This would partly eliminate the stressful task of a multilevels curriculum, but would the particular needs of all kinds of learners be met?

It would seem reasonable that in this situation the teacher should not adopt a level of teaching which is too low for most of the pupils in the class, nor a level which is too difficult for the slow learners. It would seem probable, in this case, that the teacher would adopt a level of teaching somewhere in the middle. However, one has to ask whether this situation would, in fact, provide both a challenge for

gifted learners and a basic level of learning for the slow learners.

I believe that the effect of the "non-failure" policy on the instructional organization of the school (rather than improve the teaching/learning experience in the classroom) is to create a complex situation[14] in which efficiency of teaching and learning is questionable[15] and which is stressful to both teacher and learners. Furthermore, this policy seems to militate not only against academic achievement, but also against the affective growth which it claims to promote.

The concept of continuous progress does not justify the practice of "non-failure", because the continuity of the learning process does not depend on the student being "promoted" to subsequent levels of instruction, but rather, on the continuous acquisition of knowledge. Failure to acquire a given knowledge does not imply that a person's learning process has been interrupted, hence the assumption that the traditional summative evaluation is incompatible with the notion of continuous progress is not valid.

The concept of self-esteem which purportedly would motivate the slow learners to better school performance[16] seems to have had a reverse effect because it gives such students a false sense of achievement at first. But as the years go by, the same students acquire a more realistic sense of self-awareness which has a negative effect on self-concept. In fact, sooner or later, the slow learner who has been placed in a mainstreamed class, will realize that he cannot do the same tasks that other students do, and that his/her literacy and numeracy levels are far behind others. For this reason the slow learner who has been

placed in a mainstream class will start feeling frustrated and begin to act out this frustration in such a way as to disrupt the rest of the class.[17]

On the other hand, students who are capable of working toward a high level of attainment may feel frustrated by the instructional situation in the classroom which results from the presence of the slow learners. This may lead to a lack of motivation and effort from these students and consequently to a lowering of standards.[18]

It seems then that the adoption of the "non-failure" policy in school, rather than being beneficial to all kinds of learners, generates many educational problems in the classroom situation: the difficulty of adjusting the curriculum to the multilevels; the apathy of capable students;[19] the sense of ineptitude of the slow learners; the discipline problems. It also seems probable that if the practice of "social" promotion[20] is carried out throughout the elementary system, at some point, one would end up with senior classes in which students cannot function because the learning/teaching situation of certain subjects such as science, geography, history, etc., requires reading and comprehension skills beyond the level of some of the students.[21] Consequently the value of an elementary school certificate becomes questionable and the implications for high school education become quite serious.

iii) The Alternative Policy

The alternative policy in elementary school is the traditional promotion/failure policy which assumes that the school function is that of imparting "knowledge" to all students. Its underlying concept is that there is a "recognized" body of knowledge which all students must

acquire and which is organized in terms of a year of work. The learner is expected to "cover" the curriculum of any given year to the best of his/her ability and his/her performance is evaluated at the end of each unit and, most importantly, at the end of the school year.

This evaluation is intended to measure the student's progress "...in terms of his ability to cover a given amount of content..."[22] In other words, the student's performance is taken as an index of what he/she has learned in terms of quantity and quality. For this reason the performance is graded, either using percentage symbols (100, 90, etc.) or letters (A, B, C), and each numerical or alphabetical symbol is taken as an indication of a level of performance. This mode of evaluation assumes that there is a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance. In this case, it is also assumed that along this continuum there is a point which constitutes the minimum accepted proficiency level which the student must attain in order to be allowed to take the next level of instruction.

This implies that all learners must meet a certain standard and that the student who fails to do so will have to remain at the same level of instruction in order to improve his/her proficiency level at least to minimum standards. This policy does not assume that all students progress at the same pace, nor that all students have the same interests and needs, but it is grounded on the concept of the essential value of the chosen body of knowledge for all learners. In fact it is for this reason that those students whose proficiency indicates an accepted level of knowledge acquisition are promoted to the next one,

and those who have attained below the minimum accepted level are not promoted.

iv) Effects of the Alternative Policy in the Classroom

A child entering a school system with the "failure" policy will find a given curriculum which he/she will have to "cover". If this child is not "ready" to meet the standard he/she will have a difficult first year in school. On the other hand, the kindergarten teacher will be confronted with a class of multilevels of ability and a given curriculum to teach.

At the kindergarten level the curriculum is neither "sophisticated" nor highly diversified. It is possible, however, that even at this level some students might require remediation programmes. At the end of the school year those students whose level of proficiency has been proven acceptable will be promoted to grade one, while the others will repeat the year, or be placed in a double grade situation with the possibility of advancing to the next level during the course of the school year if their proficiency level indicates their readiness.

Those students entering the second year of study can be said to be a group whose ability range is more uniform than that in the previous year. Therefore, having all met the minimum proficiency requirements, they can all be taught the same curriculum at the same time. Surely, even in the course of the second year of school there will be some students who will fall behind and, because of the increased difficulty of the subject matter, might not meet the required standard for advancing to the next unit of work or to the next grade. In either case these students will have to spend more time mastering the required skills

before proceeding to the next level.

It does not follow that in the senior grades one would have classes with only very good students because the slow learners have been left behind. What does follow is that all students in the senior grades will have demonstrated the required competency which will enable them to function at that level. Hence, the teacher at each level is enabled to teach the subject to the class as a whole because, in this case, the weak learners are not those who have fallen two or three levels behind the given grade, but those whose competency, though not very good, has been at an acceptable level to ensure at least minimum success in the level. Hence, the teacher can "cover" the given curriculum without lowering the standards. Fewer students may graduate and, according to some, this may lead to academic elitism. But it ensures that graduation from elementary schooling designates a minimum level of academic competency.

SUMMARY

Thus far I have presented two alternative policies for evaluating student achievement, examined their theoretical foundations, and considered the effects that each has in the classroom. It appears that the "non-failure" policy which focuses on the learner's developmental status, rather than on his/her achievement claims that this mode of evaluation is consistent with the concepts of "continuous" progress and "self-worth". However, since this claim seems to be invalid, a further discussion seems to be necessary.

For this reason, in the next section I will examine the recent changes in evaluation, in order to discuss the way in which assessment of student achievement has been influenced by social and psychological theories about the concepts of "failure", "success", and "self-worth" in relation to student achievement.

CHAPTER II

EVALUATION AND EDUCATION

i) Some Developments in Educational Evaluation

Evaluation is a systematic process of determining the effectiveness of every aspect of education, from teaching and learning to curricula and programmes, from individual student or teacher to groups of students or teachers. In the last two decades, in fact, the purposes and uses of educational evaluation have been broadened and refined mainly due to three factors: the innovations in the field of education, the growth of the body of knowledge, the increased responsibility placed upon the educational sector by a consumeristic public aware of their rights and their power, and therefore demanding accountability.

Modern educational evaluation began, at the turn of the century, with the work of Joseph Rice and Alfred Binet.[1] However, their tests and those developed afterwards were, for many years, only concerned with measuring either the pupil's achievement (be it in a specific subject matter, or in general knowledge), or his/her mental ability.

The cognitive domain continued to be the main issue for educational psychologists for a great many years until a new area of interest developed: that of the affective domain. Benjamin Bloom's 1964 publication, Taxonomy of Educational Objectives, Affective Domain,[2] is a landmark in the development of educational evaluation. In fact, since then, many theorists have devoted their studies not only to the development of mental abilities, but also to that of feelings and emotions, claiming

that the affective domain is closely linked to the learning of cognitive skills.[3]

The recognition of the affective domain as one of the coefficients of the learning process brought many changes in the field of educational evaluation. These changes have affected evaluation in three main areas: the identification of new educational objectives, the implementation of new instruments of measurement, and the use of evaluation to provide feedback and data on both the curriculum and teaching strategies. In fact, the increased interest in the creative and emotional side of the pupil meant that his/her achievement could no longer be evaluated in terms of cognitive skills alone. Instead, a more comprehensive evaluation was recommended, one which would include affective objectives as well.

The old paper and pencil standard tests became almost obsolete as their modes, structures, and methods could not be employed in the non-curricular areas. New tests were developed to encompass the affective domain, and new models were developed which would be best suited to measure achievement in these areas. Hence, while some types of paper and pencil test, modified and refined according to taxonomies, still remained in use, a new array of non-written and, in many cases, non-verbal tests began to be used. However, the changes had also affected the purposes and uses of educational evaluation. In fact, it was no longer sufficient to determine what and how well a pupil had learned. The new development in educational psychology had pointed out that, due to individual differences among students and teachers, some teaching approaches might yield more effective learning than others.

Hence, the assessment of teaching strategies became very important.[4]

Furthermore, new technological advances and new political and social developments made new demands on the individual, which the old knowledge could not satisfy. In response to these changes, educational evaluation became concerned not only with the pupil's intellectual and affective achievement, but also with the assessment of teachers, methods and programmes.[5] Moreover, evaluation began to be used not only as an instrument to measure final outcomes (i.e. student achievement)[6], but also as an instrument for diagnosing and programming learning remediation and for improving teacher/programme effectiveness.[7]

ii) Evaluation as a Measurement of Student's Achievement

Ahmans and Glock define pupil evaluation as "...a process in which a teacher commonly uses information derived from many sources to arrive at a value judgement."[8] about what has been learned. They note that such sources may be classified into two groups: testing procedures and non-testing procedures. Testing procedures were once the only means of evaluation in the classroom. The purpose of tests was to determine the pupil's success or lack of it in the attainment of pre-set standards of achievement, with no reference to individual differences among the testees. In this case, the level of achievement was expressed as a numerical quantity which was objectively scored. D.G. Lewis calls this type of evaluation "backward-looking",[9] because it is used to assess the efficiency of performance at the end of a learning experience (be it a unit, a course, or a school year).

Testing procedures use two main types of measurement: the

criterion-referenced and the norm-referenced. The former is founded on certain underlying concepts which R. Glasser defines as an absolute standard of quality and a continuum of knowledge acquisition.[10] This means that the criterion-referenced measurement presupposes standard objectives of achievement and a scale which covers the two extremes from zero, indicating no proficiency at all, to perfect score indicating a perfect performance.

This continuum of acquisition of knowledge is a necessary presupposition for determining what an individual has or has not learned, but as R.L. Ebel points out[11] it is not sufficient to determine quantitatively and/or qualitatively the level of knowledge acquired. In order to do so one must assume that the criterion-referenced measurement also presupposes the concept of a continuum of knowledge. W.M. Gray gives, in fact, a more satisfactory definition asserting that "...Criterion-referenced tests are those designed to produce measurements directly interpretable in terms of specified performance standards where the standards form a continuum of knowledge that is dependent on the prerequisite relations among the various levels of the continuum.[12]

Thus the CRM[13] presupposes a continuum of knowledge which enables the tester not only to determine what has or has not been learned, but it also enables him to determine if an individual has attained the minimum competency for progressing to a higher level of learning be it a unit, a course, or a grade. In this sense the CRM is not only summative, but also formative as it provides information related to future performance. Hence the CRM is an instrument, not only useful to determine whether an individual has acquired the abilities for

which he/she has been tested; but also an instrument for defining the remedial and/or advancement levels of learning for the selection of "streams" within a classroom or a school.

As the CRM ratings rest upon an "absolute" and axiomatic interpretation of the learner's performance, independently of that of other learners, the assignation of rewards (pass-fail) based on this type of evaluation would appear to have a high degree of validity which cannot be said of the norm-referenced measurement which determines individual performance in relation to the performance of others.

The difference between the two types of measurements can best be seen in a comparison of the two. In a testing situation involving the CRM, the individual's performance would be rated in relation to a predetermined perfect performance or perfect score; therefore, the testee's performance would be considered high or low according to how closely it approximates such a score. On the other hand, in a testing situation involving the NRM,[14] the individual's performance is rated in relation to that of others, which means that the testee's performance is considered high or low accoring to how closely it approximates the highest performance in his/her group of testees.

Clearly the difference in rating methods has serious implications for the reward system. In the first situation, for instance, an individual would be rated an A student because his/her test performance was 80%, thus coming only 20% short of a perfect score. In the second situation, an individual would be rated an A student because his/her performance was the highest in a group of students and yet his/her score

might be only 50% of the perfect score.

Obviously the NRM type of evaluation needs to be interpreted and, for this reason, may give rise to ambiguity in terms of performance rating. The possibility that, in practice, these two types of ratings could be confused and held to be indicative of the same kind of achievement opens the field to speculation on the dangers of misusing achievement rating. One has only to think of the consequences of mixing students whose ratings have been obtained through different types of evaluation and who are expected to perform at the same level.

Let us suppose that there are two groups of grade four students. Group A is evaluated according to the CRM and group B according to the NRM. In the former case the cognitive objectives have been set according to an "absolute" standard of quality, which means that objectives have been set according to what has been deemed to be necessary knowledge at that level. When the individual scores are calculated the number or letter assigned to each individual will signify how competent the student is in the knowledge of the specific area tested, without comparison to his peers.

Group B will be tested on cognitive objectives which have been set according to a norm, which means that those objectives have been proven to be attainable, in various degrees of efficiency, by a sampling of grade four pupils. In this case, the individual scores are calculated in relation to the performance of everyone else in the group. Now, whereas a testee with an A rating in the CRM has, at least in principle, a good knowledge, the testee who has been rated with an A in the NRM has a good competency only in relation to those who had a lower

level of performance. What would happen if these students, because of their raings, were put in the same class? It is possible that, while the first group could go on without difficulty to the next level of learning, the second group instead might find it very difficult to cope because they had not been tested for the minimum level of proficiency required for the next level.

In fact, the frameworks of the CRM and the NRM are quite different: each type is constructed to yield a different type of information about the examinee. The NRM is constructed to measure the individual's performance in relation to the performance of others in a well-defined comparison group (e.g. a class, a school, a system); the CRM is constructed to measure the individual's mastery of well-defined educational objectives. Therefore, while the information provided by the NRM is for the purpose of rank-ordering the individual, that provided by the CRM is for the purpose of assessing the individual's mastery of a specific area of instruction.

Even though both the CRM and the NRM are types of testing procedures for evaluating cognitive achievement, the framework of the latter makes it more adaptable for evaluating affective objectives.[15] However, it is doubtful if any form of standardized measure can be successfully applied to the affective domain. A.L. Costa claims that only a few aspects of this domain can be assessed with some degree of validity through conventional testing procedures.[16] He further claims that "Because affective behaviours are idiosyncratic, spontaneous and unpredictable, evaluative procedures that rely solely on predetermined

outcomes are not satisfactory."[17]

Another type of evaluation is that which makes use of non-testing procedures. These procedures may be classified as observational assessment and involve three main modes or techniques: the checklists, the rating scales and the anecdotal records. These techniques may be used to evaluate a process (i.e. a step by step learning situation) and/or a product (i.e. an individual's achievement). While they all share the same mode of assessment, viz. observation, they may differ in the method of scoring the responses, since scoring varies according to the technique used and according to the type of behaviour being assessed. In fact, while checklists (quantitative and qualitative) usually require the observer to choose between two opposite responses (yes/no, right/wrong, correct/incorrect, etc.), rating scales (which are used to assess the degree of the dimension in question) make use of numerical, graphic and descriptive graphic ratings. In this case the observer has a multiple choice of responses on a continuum between opposite poles.

Both the checklists and the rating scales can be used obtrusively, when the specific behaviour is elicited by the assessment situation, or unobtrusively when the behaviour being assessed occurs normally. The anecdotal records instead are written descriptions of an individual's natural or idiosyncratic behaviour and there is no method for scoring them. M. Priestly, in fact, says that "The notion of scoring implies judgement and anecdotal records themselves should not be judged or evaluated; they are used only to describe particular incidents."[18] He further suggests that the usefulness of anecdotal

records consists in the fact that "...a teacher...may become aware of significant patterns of behaviour that would be important for making diagnoses and recommendations."[19] Cumulative anecdotal records of student profiles are held to give a great range of information from which one can derive, as MacIntosh states, a "more rounded picture of the student,"[20]. These profiles consist of written comments on observed behaviour which includes the affective domain.

Since Bloom's Taxonomy, many theorists have provided descriptions of observable behaviours as indicators of affective responses. Yet the problem with these descriptions is the fact that they are open to different interpretations by different people. Hence even when the written comments in a student profile are made in accordance with the descriptions of the theorist, there is still the difficulty of ensuring that their interpretation will remain the same for different teachers. The main problem with observational assessments is that they put too much in the hands of the observer and the interpreter, thus incurring the risk of subjectivity.

As Costa notes, observational assessments "...are subject, of course, to misinterpretation due to low inter-observer reliability and faulty inferences about the observed behaviour's value antecedents in a particular situation."[21] This is particularly true if the observational assessment is used for the purpose of deciding whether or not a student should go to a higher grade.

In fact, how would the teacher determine if the student had sufficient mastery of cognitive skills to enable him to function at a

higher level of learning? Obviously, the logical way to determine that would be to combine the scores of all the observational assessments taken of that student during the course of the school year, or give a final assessment of the instructional objectives taken during the year. But then how is the final rating calculated?

Checklists and rating scales normally use verbal ratings such as yes, no, poor, good, excellent, etc. Priestly suggests that "In most cases, ratings on all dimensions can be converted to numerical values and added together to derive total scores."[22] This may be true, but what does the total score really mean when it has been arrived at by adding numerals assigned to different verbal ratings such as "yes", "poor", "seldom", etc.?

It may be argued that all verbal ratings are expressed either by opposite choices (e.g. right/wrong, yes/no, etc.) or by a continuum between negative and positive (e.g. never, seldom, sometimes, always), hence one could combine the ratings according to positive and negative scores. For instance if a student's rating shows a greater number of "no" and "never" than of "yes" and "always" (ratings), his/her final assessment would be negative and therefore, the student would remain at the present level of instruction. Conversely, if a student's rating shows a greater number of "yes" and "always", his/her final assessment would be positive and the student would go to the next grade. What would happen to a student whose rating shows a greater number of "yes" than "no", but a lower number of "often" than "never"? In the former case, one might justify the totalling of quantitative and qualitative scores by virtue of their being all extremes (either positive or nega-

tive), but this would not work for the above situation. Furthermore, how can one total different kinds (qualitative, quantitative) of test results together if each dimension has been achieved in various degrees?

Clearly this shows how important and, at the same time, how extensive the interpretative role of the observer is. Interpretation of course, is often a subjective way in which we ascribe meanings to words, facts, actions. What is interpreted to be "good" by one person might not be so interpreted by others. In other words, a value judgement which is the result of an interpretation of what the case is, is always a relative judgement because it is coloured by subjectivity. Hence, we must conclude that observational assessment does not appear to be a reliable mode to evaluate cognitive skills and it is questionable as to whether the interpretation of observed behaviour really reflects the characteristics which it is assumed to define and identify from the affective domain.

iii) The Concepts of Success and Failure

The concepts of success and failure were clearly defined in the traditional school evaluation policies: success meant that the student's performance indicated, at lest, a stipulated minimum attainment of the learning objectives in a course of studies: failure meant that the student's performance indicated a lack of minimum attainment. Furthermore, the grading of such performance indicated to the student (as well as to the teacher) the extent to which his/her attainment approximated the perfect score, or complete mastery of the given objectives. Success and failure were considered to represent two

different levels of achievement, arrived at quantitatively through grades. Hence, success implied that the student's performance indicated lack of minimum attainment.[23]

In the 1960's, however, this state of affairs[24] changed. At that time, education was swept by multisided changes. Its focus shifted from subject-centred to child-centred, and with this shift came a new pedagogy. This pedagogy required that education should be concerned with the total development of the child, hence education expanded its field of concern to include the affective domain. The advent of affective objectives into the educational process brought about many changes, including changes in evaluation processes and policies.

It was argued that, while the acquisition of knowledge was the aim of the cognitive domain, the development of a positive self-image was the aim of the affective domain. Hence, self-conception became the most important affective objective to be nurtured in school. The work of Wilbur Brookover pioneered a new field of research in self-conception and its relationship to student performance. He wrote, "It is hypothesized that the child learns what he perceives he is able to learn. It is further hypothesized that his self-perception is acquired during interaction with significant others who hold expectations of the student as a learner."[25] Furthermore, he postulated that "If self-concept is subject to modification,...and if modification in the images and expectations which others hold for the student takes place, then significant enhancement of achievement may be possible."[26]

Thus, the way was open for educators to study, devise and apply a new avenue of intervention to improve students' performance, and this

shifted the focus of the educational setting from the student's ability to learn to the student's self-concept. Hence, schools modified their programmes and policies to foster a positive self-concept in order to improve student performance. In fact, research data collected by Brookover and an increasing number of theorists, seemed to show a positive correlation between self-concept of ability (i.e. the student's perception of his/her ability) and achievement. These findings became the theoretical basis for the implementation and the justification of new programmes and policies in education.

Self-concept theorists claimed that the meritocratic system of the past fostered low performance by students (especially those coming from poor socio-economic background) because it enforced the success/failure standards of evaluation which undermined the self-worth of students. These theorists viewed "success" and "failure" quite differently. They argued that these terms were, in fact, labels which hindered students, thus seriously affecting the efficiency of education.[27] In 1969 William Glasser's book, Schools Without Failure, expounded and popularized this thesis. He claims that educational writers have not "...looked deeply enough into the role education itself has played in causing students to fail...."[28]

Glasser argues that the determining factor in student achievement is the sense of self-worth. In fact, while a positive self-conception leads to success, a negative feeling about one's worth leads to failure. He also claims that traditional education generally fails to foster self-worth by placing roadblocks "in the path of students

attempting to achieve a successful identity..."[29] and that the inadequacy of traditional education resides in its phlosophy,[30] based on what he calls "the certainty principle" and the "measurement principle".[31] In his opinion the former emphasizes memorization of facts -- often irrelevant to the student's future -- the latter assigns a numerical value to the ability of recalling these facts. Hence, while memorization and irrelevance of facts lead to lack of motivation, measurement procedures (tests, ratings, etc.) label a student either a success or a failure.[32]

Clearly, Glasser does not perceive academic success and failure as denoting a valid measure of student achievement; rather, he seems to hold them as psychological and social labels with serious implications for the student's future. In fact, Glasser feels so strongly about his convictions that he develops his own theory of failure. His entire book is an argument to abolish human failure. However, Glasser's theory must be closely investigated because it makes serious claims concerning traditional academic evaluation, and because many of his problematic assumptions and conclusions have already been accepted by educators and widely applied in evaluation practices and policies which endorse "social" promotion.[33]

Glasser derives his concept of failure from his clinical experiences with delinquent girls. These girls -- who are all failures in the common sense of the word -- display a lack of self-worth and an inability to relate to others in a positive way. Hence he concludes that failure is "failure to love and failure to achieve self-worth."[34] This thesis seems to be confirmed by the attitudes and behaviour displayed by those people outside the correctional facilities "whose common

denominator is failure."[35]

Glasser's definition of failure raises many serious questions. First of all, although the definition is intended to include all cases of failures[36] it's formulation is only given in terms of one dimension, the affects. Secondly, the two psychological components (lack of love and self-worth) given as a constant pair are not necessarily present as such in all cases of failure, nor do they necessarily produce "failure". In fact, one needs only to point out that people commonly held to be "successes", could conceivably have a psychological complex such as either the inability to love, or the feeling of inferiority (poor self-worth).[37]

Another difficulty with Glasser is with respect to his inferences. Since he finds that all cases of failure are cases of inability to love and to achieve self-worth, he infers that there is a causal relationship between the latter and the former. While one might concede that inability to love and to achieve self-worth entails failure, in some respect, we are, nevertheless, not warranted to infer that this inability and this lack are the causes of failure. What we might infer is that the relationship, in this context, is one of entailment between the definiens (i.e., inability to love and to achieve self-worth) and the definiendum (i.e. failure).

Glasser's argument is an example of the causal fallacy which has been traditionally called "*post hoc ergo propter hoc*". This fallacy concludes that B was caused by A just because B followed A. In fact, he concludes that (B) lack of love and lack of self-respect are the causes

of (A) failure, because these elements are present (after the fact one might say) in cases of failure. His argument for causality is invalid not only in terms of deductive reasoning, but also in terms of scientific inquiry. He never completely proves that the identified elements of failure are indeed the only relevant ones. In fact, he introduces other elements commonly associated with failure, but he dismisses them without giving us an acceptable reason for doing so.[38]

Another disturbing aspect of Glasser's writings is the fact that he holds education responsible for the "failures" in society.[39] He argues that when a child receives failing marks he is unable to have a sense of self-worth, and that each negative evaluation undermines more and more the possibility of success, until the child reaches "The age beyond which failure is difficult to reverse...."[40] He claims, therefore; that it is also the school's responsibility to correct this state of affairs by eliminating failure in school.[41] Since he believes that "the school practice that most produces failure is grading",[42] and that "grades are destructive",[43] and "objective tests are killers of kids",[44] he purposes to eliminate failure by eliminating these practices.[45] The major difficulty with Glasser's claim is that he equates total failure with a person who fails in a specific area. Hence he transfers the characteristics ascribed to a class of people (failures) to the effects of a specific event (learning evaluation).

The most serious problem with Glasser's view of success and failure is its implications for the educational setting. If the evaluation of learning must always be positive,[46] is the concept of success still meaningful? In fact, success and failure are two opposite

terms indicating the directions of a dimension, that of evaluation; if one of them is eliminated, does it make sense to speak about the other?

Furthermore, traditional evaluation enforced the success/failure standards because of the pre-set limiting factors of the educational setting: the body of knowledge, the length of compulsory attendance. Glasser's alternative evaluation would also have to operate within this setting. Hence, if a child's learning is always accepted as successful regardless of the quantity of his/her achievement, there is the possibility that a child will find him/herself having completed the compulsory time in school without having learned much. In this case, the student's lack of knowledge would be attributable to the educational system, which accepted the child's standards rather than setting standards for him/her. Furthermore, in this case, what would be the individual's chances for a good future? What will happen to his/her ability to love and to feel self-worth when he/she finds out that the education he/she has received and believed to be "good", is inadequate for employment?

Hence, an evaluation that eliminates the concept of failure in the measurement of achievement might foster self-worth, but the question still remains as to whether this self-worth might thereby be falsely conceived and nurtured.

iv) Self-Concept and Academic Achievement

Since self-concept has become a major issue in many educational policies, especially in evaluation, it is important that it should be discussed here. In practice self-concept is taken as a synonym of self-

worth, hence when it is said that a child's self-concept is low, it is understood that the child feels little self-worth. In addition, educators use terms such as "self-image", "self-awareness" and "self-esteem" interchangeably, without the distinction between descriptive denotation of the component image - which refers to how an individual perceives himself -- and the evaluative denotation of the component esteem -- which refers to how an individual rates himself. This may be due to the fact that self-hood denotes reflexivity and, as such, it involves the individual as the subject as well as the object of either a description or an evaluation.

Education has been concerned with the development of self-concept as an integral part of the learning process and has assumed the responsibility to foster it. In the elementary sector the development of a positive self-hood (in the sense of self-worth) is considered one of the most desirable objectives of education, since self-hood is deemed to be a strong determinant of the individual's future.[47] However, educational policies seem to have taken self-concept not only as a synonym of self-worth, but also as a term denoting a global rather than a specific view of self, thus failing to recognize that self-concept consists of "...three broad regions: the extant self (how the individual sees himself); the desired self (how he would like to see himself); and the presenting self (how he shows himself to others)."[48] Also, they have apparently assumed that whatever affects the idea of the self negatively should be avoided. Hence school evaluation practices and policies have been restructured to exclude failure and failing marks which are deemed to undermine the ongoing development of a positive

self-hood. In this respect, educational policies have failed to recognize the importance of the various components of self-hood, much in the same way as Glasser did.

In other words, the exclusion of failure and failing marks as a means to improve the child's global self-concept (in the sense of self-esteem) is a measure intended to insure the improvement of his academic performance. The underlying assumption is that a child's poor achievement is attributable to his belief that he is not intelligent. However this assumption is not valid. In fact, Rosenberg points out that "...this logic assumes that the general is transferable to the specific and that a change in the global self-attitude will produce a corresponding change in the specific self-attitude.... The assessment of one's academic ability and the view of one's general self-worth are two separate attitudes whose relationship must be investigated, not assumed."[49]

Researchers, including Brookover,[50] also point out that self-hood consists of many dimensions and each dimension might behave differently. For instance, a student might view him/herself as very capable in mathematics, but very poor athletically, or he/she might view him/herself as having poor academic abilities, but as very good at making friends. As early as 1962, Borislov conducted an experiment which included both the specific and the general measures of self-concept, and concluded that "the findings...do not support general self-evaluation, based on the global concept of personality adjustment, as a factor in investigating non-intellectual factors in academic achieve-

ment."[51] Furthermore, many social scientists agree with Brookover's assertion that "...when used as a composite score, these global self-concept measures are not nearly as predictive as are the more task-specific self-concept instruments."[52] Therefore, the view that student self-concept affects achievement raises the issue of a causal relationship between them. Empirical evidence to support this claim of causality, however, reveals difficulties in three major areas: a) legitimacy of causal inferences, b) direction of influences, and c) degree of influences.

Thus, the position taken in education[53] regarding self-conception and its implications for schooling does not seem to be supported by research.[54] In fact, "educators" seem to have assumed that the correlation found by researchers between what they call "self-concept of ability" and achievement holds true when self-concept is taken in its global notion. Furthermore they have interpreted this correlation as a causal relationship directed from self-concept to academic achievement.[55] While this causality has also been claimed by researchers in the field, their opinion as to its direction has been polarized: some claim that self-concept is a determinant of academic achievement; others believe that academic achievement is a determinant of self-concept.[56] Thus the difference of opinions among researchers weakens the argument for a causal relationship.

Although we may concede that a positive conception of one's ability is necessary before one tries a given task, this alone is not sufficient to ensure achievement. More simply, even though a person needs to feel that he is capable of jumping high before he will attempt

to make a high jump, this feeling alone will not make him succeed. In order to attain success he also, and foremost, needs the athletic ability. In fact, Wilson and Portes claim that "self-assessment appears to be only a by-product of the ability variables, interesting perhaps in its own right, but not a systematic part of the causal process of status attainment."[57]

Therefore, we must conclude that, while empirical evidence speaks of a correlation between self-concept and achievement, this cannot be taken as proof of causality. Statistically correlation is a mathematical relationship between two or more variables, which carries no implication of causality.

Another serious problem with self-concept is that of measurement. This difficulty resides in the scientists' working tool, which Carlton calls a "self-referent construct".[58] This construct requires the subject to evaluate himself in relation to his performance by identifying what he perceives to be the causes of a specific outcome. The scientist derives measurements of dimensions of self-concept from measuring self-attribution responses of the subject. Hence self-attribution is the crucial issue. In fact, its validity impinges on the validity of the construct and on the implications derived from it. This self-attribution implies serious considerations for both the self-concept and its application in the educational setting (i.e., self-concept as a determinant of achievement).

Self-attribution theory has usually focused on a single internal-external dimension.[59] That is, researchers have measured

attributions only in terms of "locus of control", where "internal" refers to a person's belief that outcomes depend on one's own characteristics (such as effort, ability, etc.), and "external" refers to a person's belief that outcomes depend on external causes (such as luck, task difficulty, etc.). Within this context, however, research has two different approaches: the dispositional and the situational. The former investigates "individual differences or dispositional tendencies in use of self-attributions by the same subject in different situations,"[60] in order to show that a person's perception across different situations is systematically different from that of another person. The latter manipulates "components of the situation...to determine their effect on attributions...to show that situational manipulations produce systematic differences that generalize across subjects."[61]

It is important to realize that these approaches represent two different ways of investigating self-attribution, hence one should be careful not to assume "that findings derived from the situational approach can be applied to questions about dispositional differences, though this practice is common in attributional literature."[62] If this is, in fact, what happens in attributional literature, then we must question the validity of what they say about attribution. Furthermore, if such literature has misinterpreted research findings we must question whether this misinterpretation has influenced any theoretical or practical areas of education.

Research findings,[63] mostly of the situational approach type, demonstrate that a person's self-attribution differs in systematic ways

for success and failure. A person tends to identify his ability as the cause of his success, but tends to blame the task difficulty for his failure.[64] In other words, a person seems more likely to internalize the causes of his success, but to externalize those of his failure. This seems to suggest that the subject might be using a self-serving bias as "an attempt to protect or enhance self-esteem."[65] However, further research into this phenomenon has resulted in a polarization of opinions.[66] The uncertainty regarding the significance of the self-serving bias is due to the fact that the research in this area was conducted with the situational approach which, as previously described, is not designed to analyze dispositional differences but rather to show that situations can influence self-ascription in a systematic way.

In addition, a study conducted by Fennema which includes some dispositional data, also demonstrates that "Success and failure scales show little correlation, except for ability/success and ability/failure, which are negatively correlated (i.e. if I attribute success to ability, I do not attribute failure to lack of ability)."[67] However, these findings do not represent a validation of other results since "the poor quality of measurement instruments and the lack of comparability of measurements procedures used by different researchers makes the comparison of results from different studies a problematic exercise."[68]

A recent study, which combines the dispositional and situational approaches, and which is, therefore designed to measure distinct components, concludes that the self-serving bias[69] is used largely by students who are the most able and have the highest self-concept, and

suggests that "the self-serving bias is clearly not "a simple response bias that might be expected if a student were merely trying to distort their public image, but rather it shows a clear and logical pattern of relationship to other variables."[70] (e.g. math/abiltity, reading/ability). The study also points out that a major fault of attributional research is the forced-choice technique used to measure attributions. In fact, this technique forces the subject to opt for the least self-damaging response, thus compromising the validity of the findings. For example, if a child with a high self-concept must choose between ability and effort as causes for his failure, he will certainly choose the latter because the alternative choice is self-threatening. Similarly a child with low self-concept, who is given the choice between internal (ability/effort) and external (luck/task difficulty), will opt for an internal, but less threatening variable such as effort, as the cause of his failure.

The above mentioned study emphasizes the need for caution in making inferences from research findings. This is especially true concerning the relationship between self-attribution and self-concept. A large body of research in this field has produced much evidence, but this should not be taken as conclusive. In the light of what has been previously discussed these findings are far from conclusive. Attribution researchers have yet to convince us that the generalizations are correct, since their measurement devices appear to be inefficient and inaccurate, and their findings have produced divergent results. Therefore, we must also conclude that since "Attributional research, both situational and dispositional, is plagued by a disregard for

important measurement issues and a lack of construct validation of the measurement devices that are employed",[71] any implication for the educational setting must also take this into account.

SUMMARY

In this chapter I have presented some of the social and psychological theories which have affected education and consequently the processes and policies of evaluation. I have examined the changes which have occurred in the area of measurement, and discussed how the introduction of the affective domain in education has resulted in redefining the concept of "failure" as a consequence of the prominent attention given to the development of "self-worth".

In addition, I have examined some of the research on self-concept as it relates to student achievement and found little evidence that improving student self-conception results in improvement in achievement. The question, therefore, arises: what form of evaluation of student achievement would be adequate to meet contemporary needs? For this purpose, I shall try to offer a brief analysis of the concept of evaluation, and relate it to the assessment of student achievement.

CHAPTER III

AN ANALYSIS OF EVALUATION

i) Some Thoughts on the Concept of Evaluation

The term evaluation denotes a judgement of some kind. For instance, one talks about "evaluation of performance", "evaluation of achievement", "evaluation of efficiency", etc. In all these cases evaluation conveys a sense of judging the worth of something specific about either a person -- such as in the case of evaluating somebody's performance -- or about a process -- such as evaluating the efficiency of learning. However, in ordinary life, what is meant by evaluation is often interpreted and applied with some measure of confusion. In fact, evaluation is of two distinct types: that which reflects a personal opinion, and that which represents an objective appraisal.

It is important for the purpose of this paper to discuss the two types of evaluation in order to clarify the necessary elements and the characteristics of each type, so that we might define which is appropriate to use in a given set of circumstances. Let us first examine what is meant by evaluation in the sense of a personal opinion.

In ordinary life, one often makes value judgements regarding many things: art, theatre, literature, etc. In these cases, expressions such as "The performance was good", "That painting is bad", represent judgements of worth. Yet, it is not uncommon that the same performance or the same work of art could be judged "good" by some and "bad" by others. This is because what is meant by "good" or "bad" in

relation to something varies according to people, since these criteria cannot be defined for everyone by pointing to the empirical world. In fact, to say that something is "good" is equivalent to comparing it to one's own individual idea of how that something ought to be. Hence criteria such as "good", "bad", "mediocre", are derived from the individual's likes or dislikes and are, for this reason, personal: in other words, subjective.

Clearly then, value judgements of this type cannot be taken to assert an undisputable value, they must only be taken to indicate an opinion/or a preference. Hence, while the agent of evaluation has the right to hold his judgement as valid for himself, he may not expect it to be so for others. While it is legitimate to make value judgements based on one's own opinion and preference (e.g., the purchase of a pair of shoes), it is not legitimate to hold that all value judgements should be made this way. In fact, there are circumstances which require that the judging should be based on criteria acceptable to others (e.g., the hiring of a person). Therefore the type of evaluation required by any given set of circumstances will depend on the purpose of that particular evaluation. Judgements which must be valid for not only the evaluator but also for others must be arrived at through an objective evaluation.

An objective evaluation denotes a judgement which expresses an objective appraisal of worth. This presupposes, first of all, that what is meant by "worth" in a class of circumstances, is objectively defined. But how can value concepts, which are normative in nature, be objectively defined? Usually we explain our ideas by describing them in

empirical terms. This description enables others to understand what we mean. If everybody were to agree on a certain description of a concept, then this description would, by agreement, become the standard interpretation of the meaning of that concept. Therefore, an objective evaluation must begin with objective descriptions of the values which it purports to appraise.

Secondly, an objective evaluation implies that the appraisal has been done in a certain way, namely according to objective standards. Hence it is necessary that the criteria, and the method used in the evaluation, should be properly defined, by agreeing on a precise description of which set of empirical evidence counts as "good", "bad", "poor", etc. in a particular set of circumstances. This, in fact, would ensure not only that the evaluative symbols reflect an objective measure of worth which is supported by factual evidence that is verifiable, but also that all cases of a certain type of "worth" would be judged in the same manner.

ii) Evaluation of Student Achievement

The evaluation of student achievement is an integral part of the teaching-learning process. Its function is to assess how well the student has learned the educational objectives of a unit, a year, or a level of study. Its results are used mainly for reporting to parents, for placement of students at the next level of instruction, for certification, and for improvement in learning. It is important, therefore, that this evaluation should be objective, that is, based on empirical evidence and verifiable.

The evaluative process is essentially a two-step process: the

measurement of what the student has learned, and the judgement about the quality of this learning. The first step, the measurement, serves to provide objective evidence of student learning, hence it requires collecting data which are indicative of learning. What is meant by data which are indicative of learning? They are data which represent accurate and objective quantifications of student performance in relation to certain tasks deemed to be the objectives of learning.

Clearly then some source of these data is the student performance, in other words, the tests. In fact, Ahman and Glock define tests as measuring instruments for obtaining "a quantitative representation of the degree to which a pupil reflects a trait."[1] However, if this quantitative representation is to be accurate and objective, then the measuring instruments used to arrive at this representation should be proven to be valid and reliable. That is, measuring instruments should yield precise measures of student achievement, and they should do so consistently.

Achievement tests, or tests which are purported to measure student learning can be classified into two groups: classroom tests, and standardized tests. The former are constructed by a classroom teacher for use in a particular class, and so they are designed to measure the performance of a particular group of students in relation to what a particular teacher has taught. The latter, by contrast, are constructed by measurement specialists and are intended for use in all classes of a certain type, such as a specific level of instruction. These tests are designed to measure student performance in relation to

pre-set standards of achievement, or more simply, according to educational objectives which are taken as standard learning requirements of a particular level of instruction.

However, the difference between the two types of tests can best be presented in the light of the two dimensions within which tests are normally evaluated: their validity,[2] and their reliability.[3] A test is said to be reliable if it produces the same results on different occasions with different examiners. In the case of standardized tests, the fact that they have been pretested gives some measure of assurance of their validity and their reliability, thus ensuring that their results represent a precise measure of student achievement, which is reliable and verifiable.

In the case of classroom tests, it may be argued that their validity might be ensured by the fact that testing items have been chosen by the teacher. In fact, he/she is in the best position to judge the appropriateness of tests for his/her pupils. However, although this may be true, it is also possible that this is not so in all cases. Therefore, the validity of classroom tests is, at least in some cases, questionable. Furthermore, as these tests are arbitrarily constructed and intended to be used in a particular class, it is questionable whether they can produce the same results on different occasions and/or with different examiners, hence they cannot be said to be reliable.

A further distinction to be drawn between the two types of tests is one with respect to educational objectives,. In standardized tests these objectives are defined as the required learning of a particular level of instruction, and so they represent pre-set standards. This

ensures that test results are always interpreted in relation to the same educational objectives. Hence, test scores are comparable not only in so far as they indicate a certain quantity of achievement, but also, and more validly so, they are comparable because they refer to the same learning objectives.[4]

On the other hand the learning objectives of classroom tests are set by each teacher and, for this reason, tests used in different classes may in fact present different selections of learning objectives. Test scores, in this case, are not comparable, since the method of scoring and the test items vary with different teachers. Therefore, the assessment of student achievement should consist of data collected through the use of standardized tests, and more specifically, as stated in the previous section, criterion-referenced tests.

The second step of the evaluation process is concerned with a qualitative judgement of student achievement. It involves the assignation of symbols such as words, "excellent", "good", "satisfactory", "poor" - or letters such as A, B, C, D - to quantitative assessments of student achievement. This assignation of symbols represents a value-judgement about what has been learned, and it serves to define "how well" the student has mastered the educational objectives of a given period of study.

This evaluation, then, assumes that learning is a continuum of knowledge acquisition with ranges from no proficiency at all to perfect performance, and that student performance indicates a degree of knowledge acquisition which falls within this continuum. Hence, the symbols

"good", "poor", etc., are used to represent degrees of achievement within this continuum.

The difficulty with evaluation arises when these symbols are used improperly, that is the criteria for these symbols are arbitrarily chosen, because in such cases, it is possible that different evaluators would assign the same symbol to different quantitative assessments of student achievement. This would mean that evaluation of student achievement, rather than representing an objective and verifiable measure of the quality of knowledge acquisition, would simply reflect a subjective and unverifiable opinion. It is therefore necessary that the evaluative symbols should be used in a uniform manner.[5] That is, it is necessary that what is intended by "good", "poor" or A, B should be properly defined, and these definitions should be the norms or standards for using them.

The use of standards makes value-judgements objective because they purportedly yield an accurate and consistent interpretation of what the case is. In fact, a standard is a description which enables the evaluator to compare student performance to a symbol representing a degree of proficiency. For instance, if the expression "A=80%" were taken as a standard, it would mean that the evaluator could assign an A grade only to a student performance assessed as 80 percent. Thus standards also ensure that value-judgements be verifiable.

However, an evaluation of what has been learned is not only concerned with assignation of grades. It should also define a point within the continuum of knowledge acquisition which represents an acceptable minimum of proficiency, which the student must obtain in

order to be allowed to the next level of instruction. In other words, evaluation of student achievement must necessarily be concerned with ensuring that the students are learning at a level which will facilitate and promote subsequent learning. Hence, evaluation must also set the standards for judging student performance in terms of failure and success: failure to indicate lack of minimum attainment; success to indicate attainment of a stipulated minimum of achievement. It follows then that those students whose performance has indicated a lack of attainment should be retained; those whose performance has indicated a satisfactory achievement, should be allowed to the next level of instruction.

It may be argued that this state of affairs would mean a return to the lock-step system of the 1950's, which, in the last two decades, has been judged undesirable, especially by those who defend the "non-failure" policy. While the validity of their claims has been discussed in the previous section, we shall defend the desirability of judging student performance in terms of success/failure on the grounds that the concept of evaluation does not only imply a process, but also a purpose.

This purpose is two-fold: to prove what has been learned, and to improve the learning. Furthermore this purpose is intrinsic to the concept of evaluation. Therefore it makes no sense to talk about "formative" evaluation, and "summative" evaluation as if they were two different kinds. The terms "formative" and "summative" define the purpose of evaluation whether it occurs during or at the end of the year. Therefore to judge student performance in terms of failure (and

success) is justified whenever retention is required in order to improve student learning.

Therefore the educational system, because it is publicly supported, has a moral obligation to its taxpayers to adopt a policy which is based on standard criteria of evaluation. This system, in fact, justifies its control over education not ony by claiming that it imparts knowledge, but also, and more strongly, by claiming that its policies facilitate the acquisition of that knowledge. Hence it has the obligation to provide for an honest and clearly defined way of evaluating student achievement in terms of failure and success.

CONCLUSIONS

The aim of the present project was to critically investigate two alternative evaluation policies in elementary school in order to determine whether the practice of "continuous promotion" is a viable and tenable alternative to traditional promotion. For this reason I analyzed the issue in three main areas: a) the effect of the policy in the classroom and on the instructional organization of the school, b) the assumptions on which the policy is based, and c) the justification of its practices.

The first objective was to establish whether the practice of "continuous promotion" enforced by the "non-failure" policy offered a measure of benefit not found in the traditional evaluation policy. Hence, I looked at the effects that each policy had in the classroom and on the instructional organization of the school. I have tried to show that there are significant differences between ways in which the two policies affect the process of schooling and that the "non-failure" policy, contrary to what is purported, militates against the efficiency of the educational process.

In fact, the practice of "continuous promotion" creates difficult structural and instructional situations as it fosters classes with multilevels of abilities which require modifications of the curriculum. Since such classes, as discussed in the first chapter, have many negative effects on the learner as well as on the quality of education which it produces, we must conclude that the practice of "continuous

promotion" fails to fulfill the promise of improving student achievement. hence, it is not a feasible alternative to the evaluation practice which makes use of failure to indicate lack of satisfactory progress.

The second objective of the project was to assess the validity of the assumptions on which each policy is based. It was found that the "non-failure" policy, which is grounded on psychological and social theories, is based on the assumptions that: a) "self-concept" (in the sense of self-worth) is a necessary condition to learning, and b) "failure", which undermines the learner's self-concept, has a negative effect on his achievement.

A review of self-concept research has failed to provide evidence which would support these claims. In fact, empirical findings in self-attribution research have yielded divergent and inconclusive results with respect to a causal relationship between a) self-concept and achievement and b) failure and self-concept. Hence, the assumptions on which the "non-failure" policy is based as assertions of causal relationships which are not justified by the evidence. Furthermore, the argument in support of the self-concept thesis is logically incorrect as it falls to the " *post hoc ergo propter hoc* form of causal fallacies which concludes that a negative self-concept is caused by failure because the former follows the latter.

The last objective was concerned with the theoretical justification of the practice of "continuous promotion". It is claimed that the "non-failure" policy is postulated by a specific theory of

education based on the assumption that learning is a continuous process. In fact, it is argued that since continuous learning implies the child's continuous progress, a traditional summative evaluation of the child's achievement would be inconsistent with the notion of continuous progress.

This conclusion, however, is not valid because the notion of summative evaluation is not incompatible with the notion of continuous progress. Evaluation is not merely a judgement about what has been learned. Furthermore the concept of learning as a continuous process does not imply "non-failure" in the achievement of a goal. In fact, the continuity of the learning process does not depend on the student being "promoted" to subsequent levels of instruction, but rather on the student's continuous acquisition of knowledge. To say that a student has failed to acquire a given knowledge within a given period of time does not imply that his/her learning process has been or should be interrupted; it is a recognition that the student must improve the quality and/or quantity of his knowledge in order to be able to continue to learn.

A further argument in defense of the "non-failure" policy is that such policy employs evaluation only in the formative sense, hence its function is only prescriptive of further learning experience. But even this is not an adequate justification. In fact, evaluation, even in the formative sense requires criteria by which to judge different performances, and when the judgement is made the implication is that we should act accordingly. The "non-failure" policy, however, fails to recognize that the decision of "non-failure" cannot be justified if it

applies to all kinds of achievements, regardless whether this achievement represents attainment or lack of attainment of the educational goal.

Hence, we must conclude that the "non-failure" policy is neither a viable nor a tenable alternative to the evaluation policy wich makes use of failure, and that the pass/fail tradition is a more viable practice which enhances the effectiveness of education.

SOME SUGGESTIONS FOR TEACHERS

It is hoped that this project will give teachers a better understanding of the problems associated with current evaluation policies and practices, so that they may perform more efficiently in their role of evaluators. Issues such as standards, criteria, and mode of evaluation require that decisions should be made only after investigation of all available alternatives and that these decisions should be supported by sound reasons.

In addition, teachers should be aware of the dangerous consequences of implementing "reforms" without questioning the validity of the claims and assumptions on which they are based. Decisions made at a higher level are sometimes the result of political expediency and at the expense of the learner's best interests.

NOTES

NOTES TO INTRODUCTION

1. Harold G. Shane, The Academic Score Decline: Are Facts the Enemy of Truth?", Phi Delta Kappan, October 1977, p. 83. "Today Americans have data to imperil their illusions and intensify their fears about pupil progress. Achievement test scores have been falling for 14 consecutive years. One might almost say that kids are on the skids, since there has been a consistent decline in aptitude and basic skills as measured by standardized tests."

   Carl Bognar, "Back to the Basics - An Introductory Survey", Interchange, Vol. 7, No. 4, 1976-77, p.2. "The Canadian Chambers of Commerce conducted a survey of employers. Of the respondents, half criticized the basic skills of secondary school graduates,..."

   Maureen Douglas, "Minimum Competency Testing", Compact Contents, Winter 1978, p.4. "According to the Natioal Assessment of Educational Progress, one in eight high school graduates can't read well enough to understand a simple traffic sign. Other estimates of functionally illiterate graduates range from 10 to 20 percent nationally."

   David F. Labaree, "Setting the Standard: Alternative Policies for Student Promotion". Harvard Educational Review, Vol. 54, No. 1, February 1984, p. 67. "Educators, parents and citizens in general have become concerned about the large number of students who are not mastering grade-level basic skills."

   Oswald Hall, Richard Carlton, Basic Skills at School and Work, Toronto: Ontario Economic Council, 1977, p. 1. "One of the underlying concerns, that current training in 'basic skills' might be inadequate to the needs of an industrialized society, provided the focus for the research reported here. This uneasiness regarding levels of achievement in basics was clearly shared by other segments of our society."

2. National Commission on Excellence in Education, "A Nation at Risk: The Imperative for Educational Reform", Elementary School Journal, November 1983, p. 115. "Some 23 million American adults are functionally illiterate by the simplest tests of everyday reading, writing, and comprehension."

Ibid., p. 120. "Secondary school curricula have been homogenized, diluted, and diffused to the point that they no longer have a central purpose. In effect, we have a cafeteria-style curriculum in which the appetizers and desserts can easily be mistaken for the main courses."

David F. Labaree, op. cit., p. 73. "The most frequently voiced criticism is that current promotional policies represent an abandonment by public schools of their once dominant concern with student achievement The much publicized decline in student scores on standardized achievement tests in recent years has led many people to question whether the schools are doing their job."

Wayne Reilly, "Competency Testing", Compact Contents, Winter '78, p. 7. "the movement, in some of its varied forms, has its good points. It places a much needed spotlight on basic education which in some places has become lost in a sea of expanded curricula and social problems.

Oswald Hall/Richard Carlton, op. cit., p. 23. "Variability or lack of basic skills on entry is attributed to policies of innovation, experimentation, and social promotion in the primary setting."

3. R.E. Slavin, "Realities and Remedies", Elementary School Journal, Nov. 1983, p. 136. "The commission report focuses on reforming the high school, yet it is the elementary school that should be the first focus of concern if we are to reverse the apparent decline in basic skills."

Carl Bognar, op. cit., p. 1. "The fall of 1976 seems to have marked the birth of a renewed "back to the basics" movement in Canadian education,... Enthusiasm for core curricula, standardized tests, and clear academic objectives can be discerned at some level of the education hierarchy in virtually every province."

Maureen Douglas, "Minimum Competency Testing", Compact Contents, Winter '78, p. 4. "The 'back to the basics' movement in California is supported by Assembly-man Leroy Greene, chairman of the education committee, who feels higher standards in education are needed because 'school curricula have expanded to the extent of diluting the basics...' he says."

Daniel P. Resnick, Lauren B. Resnick, "Improving Educational Standards in American Schools", Phi Delta Kappan, Nov. 1983, p.180. "We maintain a generally undemanding common curriculum in the middle schools, in the hope of reaping social benefits from a school system that does not sort children according to academic performance."

Thomas L. Wells, "Adjusting the Pendulum", Interchange, Vol. 7, No. 4, 1976-77, p. 4. "Much of the present activity and discussion concerning education today centres on the term 'the basics'."

4. Walt Haney, George Madaus, "Making Sense of the Competency Testing Movement", Harvard Educational Review, Nov. 1978, p. 474. "In part the minimum competency movement is one aspect of the back-to-basics movement, part of a backlash against the "funsie-wunsie open education" philosophy of the 1960's (Kilpatrick, 1977)."

W. James Popham, Stuart C. Ranking, "Minimum Competency Tests Spur Instructional Improvement", Phi Delta Kappan, May 1981, p. 637. "In 1977 Michigan had no statewide minimum competency testing requirements, ...yet in August of that year Jefferson issued a memorandum to the Detroit Board of Education calling for a creation of a high school competency testing in Detroit. ..Rather, the major mission of the proposed program was to improve the basic skills instruction in Detroit. Jefferson believed that the testing program could motivate instructional improvement in the basic skills and that those improvements would spread in time to other curricular areas."

David F. Labaree, op. cit., p. 81. "Educators, parents, and general public are frightened by the widely publicized declines in standardized test scores in the recent years and by the growth in the number of high school graduates who have failed to master basic skills. A policy of merit promotion offers a way out of this dilemma by promising to increase the academic demands which schools place on students and to motivate students to meet these demands."

5. Goldwin J. Emerson, Maryann Ayim, "Dewey and Peirce on Curriculum and the Three R's", The Journal of Educational Thought, April 1980, p. 23. "In the 1960's Canada found itself in a period of growth and expansion. ...In the 1970's economic restraint has forced educators to think in terms of compulsory curriculum and getting back to the basics."

Carl Bognar, op. cit., p. 1. "...Thomas Wells, Ontario's Minister of Education, who announced the institution of a core curriculum..." and "On Nov. 1, 1976, British Columbia Minister of Education Pat McGreer announced that B.C. schools will return to the basics by the fall of 1977. A core curriculum has been established...and goals are prescribed for each level of instruction."

6. Thomas L. Wells, op. cit., p. 6. "Teachers and parents clearly want more objective information about how well pupils are achieving and that is what we intend to provide."

7. Ibid., p. 4. "In Ontario, the initial seeds of curriculum refinements were sown early in 1975 when we introduced the new curriculum policy for elementary schools, "The Formative Years. ...With the new structure of the unified Curriculum Branch established, we followed in October and in November 1976 with two policy changes related to the curriculum beyond grade 6."

   It should be noted that since 1975 the Ontario Ministry has issued a series of curriculum guidelines not as separate documents but as extensions of the main circular, The Formative Years.

8. Thomas L. Wells, op. cit., p. 1.

9. Ibid., p. 6. "As a result, the area of evaluation and testing was the subject of considerable study within the Ministry for many months during 1976."

10. Ontario Schools: Intermediate and Senior Division, Ontario Ministry of Education, 1984, p. 85.

    In must be noted that, at present, the Ontario Ministry of Education has issued the above document to introduce structural changes into the intermediate and senior levels of schooling. However, although it contains a section for the elementary level (grades 7 and 8) it is mainly intended as a guideline for secondary schools. Furthermore as I have pointed out in my discussion, there are little changes regarding evaluation policies, hence this document does not address problems which exist in elementary schools.

11. Ibid., p. 35. "Procedures for evaluating student progress should be sufficiently varied to meet the requirements of different individuals and groups of students, different courses, the three levels of difficulty, and a variety of learning environments. For most purposes, it is recognized that the most effective form of evaluation is the application of the teacher's professional judgement to a wide range of information gathered through observation and assessment."

12. Ibid., p.2.

13. The Formative Years, Ontario Ministry of Education, 1975, p. 4.

14. Ontario Schools: Intermediate and Senior Divisions, Ontario Ministry of Education, 1984, p. 35.

NOTES TO CHAPTER I

1. Living and Learning, Ontario Department of Education, 1968.

2. Thomas L. Wells, "Adjusting the Pendulum", Interchange, Vol 7, No. 4, 1976-77, p. 4. "However, in some respects the momentum of the early and mid-1960's carried certain aspects of change a little too far, and today in Ontario (as elsewhere) the pendulum is being eased back somewhat."

3. R.H. Anderson, "The Nongraded School", The National Elementary Principal, Nov. 1967, p. 7. "Without exception, the emphasis is upon individualized instruction and upon developing each individual up to his full potential for physical, social, intellectual and civic accomplishment. Without exception too, there is reference to the fact that provision should be made for both differentiated rates of pupil progress and variations in the kind of programmes offered to this child and that."

   O. Hall, R. Carlton, Basic Skills at School and Work, Toronto: Ontario Economic Council, 1977, p. 53. "If there was a keystone to reform structure which gradually took shape, it was caught in the cliche "child-centred",...ideals of individualized and continuous progress were translated into ability level group work and social promotion,.... Curricula were enriched and expanded to provide variety as well as a stronger sense of immediate relevance to the pupil."

4. H. Furth, H. Wacks, Thinking Goes to School: Piaget's Theory in Practice. New York: Oxford University Press, 1974, p. 45. "Each child must be left alone to work within the structure at his own rate and in his personal style."

5. O. Hall, R. Carlton, op. cit., pp. 53-54. "With the recognition that self-concept had been an important factor intervening between ability and achievement, the new school environment was designed to maximize the student's positive self-image, both through the encouragement of his personal expression and the deliberate suppression of any negative feedback on his performance.

6. D. Kuhn, "The Application of Piaget's Theory of Cognitive Development to Education", Harvard Educational Review, Aug. 1979, p. 352. "Piaget (1971) specifies "equilibration",...as the critical process whereby the individual's system of mental actions, or operations, is reorganized into a new, more advanced structure.... (The) optimal mismatch theory, is implicit in some of his early writing, and was favoured by some early interpreters of Piaget, such as Hunt (1961) and Bruner (1960).... Optimal mismatch theory has become widely accepted in educational circles."

7. W.B. Brookover, A. Paterson, S. Thomas, "Self-Concept of Ability and School Achievement", Report of Project No. 845, East Lansing: College of Education, Michigan State University, 1962, p. 3. "It is hypothesized that the child learns what he perceives he is able to learn. It is further hypothesized that his self-perception is acquired during interaction with significant others who hold expectations of the student as learner."

   W.B. Brookover, E.L. Erickson, Society, Schools and Learning, Boston: Allyn and Bacon Inc., 1969, p. 106. "...a significant number of students are being needlessly hindered by low self-conceptions of academic ability...strategies must be developed for enhancing the self-conceptions of ability for a larger portion of our students."

8. D. Kuhn, op. cit., p. 352. "Optimal mismatch theory has become widely accepted in educational circles. It provides an elegant theoretical rationale for the popular view that curricula must be tailored to each child's particular level of competence."

9. W.B. Brookover, A. Paterson, S. Thomas, op. cit., p. 76. "If self-concept is subject to modification, as theoretically postulated and if modification in the images and expectations which others hold for the student takes place, then significant enhancement of achievement may be possible."

10. H.M. Glick, M. Schubert, "Mainstreaming: an Unmandated Challenge", Educational Leadership, Jan. 1981, p. 326. "The term mainstreaming is familiar to educators in connection with effort to implement PL 94-142., the Education for All Handicapped Children Act.... But it can also mean that, depending on the school district's services and the student needs, students may spend most or all of the school day in a regular classroom with special help being provided to the regular classroom teacher."

    In Ontario the same policy has been adopted by the Ministry of Education to be effective in 1985, this is known as Bill 82.

11. See J. Piaget's account of cognitive development in The Psychology of Intelligence, Totowa, N.J., Littlefield Adams, 1972, p. 25.

12. O. Hall, R. Carlton, op. cit., p. 50. "...factors responsible for weakness in pre-school skills and motivation are generally still at work throughout the whole elementary career of the pupil."

Robert L. Ebel, "Failure of Schools Without Failure", Phi Delta Kappan, Feb. 1980, p. 387. "One consequence of automatic promotion, and of the educational attitudes and values associated with it, has been progressively accumulating deficits of learning for many pupils. Evidence of those deficiencies is too substantial and too extensive to be denied or explained away, as some have tried to do."

13. Ibid., p. 387. "Highly individualized instruciton is necessarily quite inefficient. The teacher's efforts are diffused. Preparing individualized instructional programs takes a great deal of time."

14. O. Hall, R. Carlton, op. cit., p. 71. "Teacher workload is substantially increased; the strain of preparing for and instructing several skill levels is compounded by insecurity and dissatisfaction in achievements."

15. Ibid., p. 69. "The most serious, however, are the difficulties which have arisen in coping with the range of skills brought into each classroom by the social promotion policies."

16. W. Glasser, Schools Without Failure, New York: Harper & Row, 1969, p. 26. "Very few children come to school failures, none come labelled failures; it is school and school alone which pins the label of failure on children. Most of them have a success identity, regardless of their homes or environments. In school they expect to achieve recognition and with the faith of the young, they hope also to gain the love and respect of their teachers and classmates. The shattering of this optimistic outlook is the most serious poblem of the elementary schools."

17. O. Hall, R. Carlton, op. cit., pp. 68-69. "Pupils can't read the encyclopedia or do research while the others can: Their self-concept goes down. The student knows he is reading things with fewer words, smaller words...faces this every day...a subtle failure: perhaps worse than saying: Repeat this grade."

18. Ibid., p. 70. "We just don't have time for the very low group. I've been told: concentrate on the middle group - the low group will always be poor. I had one kid in grade three reading a pre-primer."

19. Ibid., "Many students, we were told, worked toward the minimum levels of competency required by the formal grading and promotion system, levels which had been adjusted downward to accommodate the very weakest pupils in a broadened skill spectrum. For the teachers, this underachievement posed serious motivational problems related to the work attitudes described above.", p. 76.

"Finally, there is a perplexing overall discontinuity between avowed recent aims (greater student interest, enjoyment, involvement and responsibility) and admitted outcomes (greater apathy, detachment, unwillingness to work, and responsibility).", p. 86.

"The results of remedial courses in the basic skills are not encouraging. The deficiencies in these skills are frequently mired in a morass of poor work habits, ineffective study patterns, careless attitudes toward academic matters and antipathy toward teachers. More than a remedial course is needed to offset this constellation of difficulties. Moreover, the optimal time for such remedial effort is long past in the career of the student.", p. 248.

"For the whole system, one could say that concern for high standards of performance tended to crumble as the philosophy of living superseded that of learning.", p. 252.

20. Social promotion, here, is intended as promotion according to age, regardless of the learner's achievement: for example, a seven year old is promoted to grade two, not because of what he has learned, but because he has already spent a year in grade one.

21. O. Hall, R. Carlton, op. cit., p. 70. "The weaknesses of the students promoted beyond their true skill levels are amplified by the interdependence of basic skill areas and special subject fields. Individualized and programmed language study material...may permit a grade five student, for example, to work in reading or composition at a grade three level. Nevertheless, the science, health or other subject materials encountered by the same student will generally demand a grade five level of vocabulary and comprehension and may present a barrier to his optimum work within these subjects as well."

22. Sindey P. Rollins, Developing Non Graded Schools, Itasca, Illinois: F.E. Publishers, Inc., 1968, p. 13.

NOTES TO CHAPTER II

1. Arnold J. Lien, Measurement and Evaluation of Learning, Dubuque, Iowa: W.C. Brown Co., 1967, p 18. "Father of achievement testing is a name generally given to Joseph Rice. Through his surveys of pupil learning in spelling, arithmetic and language he was able to survey the status of pupil learning through objective types of testing. At the same time, Alfred Binet should be credited with the real beginning of intelligence measurement."

2. B.S. Bloom, et al., Taxonomy of Educational Objectives, Handbook 2: Affective Domain. New York: D. McKay, 1964.

3. G. Moskowitz, Caring and Sharing in the Foreign Language Class. Rowley, Mass.: Newbury House Pub. Inc., p. 18. "...education should deal with both dimensions of humans - the cognitive or intellectual and affective or emotional...".

   R.A. Magoon, Education and Psychology - Past, Present and Future. Columbus, Ohio: Charles E. Merrill Pub. Co., 1973, p. 112. "...the process of education is vitally concerned with the development of values and attitudes..."

4. Phi Delta Kappa National Study Committee on Evaluation, Educational Evaluation Decision Making, Itasca, Illinois: F.E. Peacock Publishers, Inc., 1971, p. 12. "Second, evaluation no longer focused solely on the student. Instead it could also be used to provide insights about the curriculum and about educational procedures."

5. Ibid., p. 121. "Emphasis on the management of programs has shifted to center on the interrelationship between goals and objectives and resource allocation at the various levels of the educational enterprise."

6. Ralph W. Tyler, Basic Principles of Curriculum and Instruction, Syllabus for Education 360, Chicago: University of Chicago Press, 1950, p. 69. "The process of evaluation is essentially the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction. However, since educational objectives are essentially changes in human beings, that is, the objectives aimed at are to produce certain desirable changes in human behaviour patterns of students, then evaluation is the process for determining the degree to which these changes are actually taking place."

7. Phi Delta Kappa National Study Committee on Evaluation, op. cit., p.32. "Scriven and Stake have espoused a distinction

between "formative" evaluation (evaluation concerned with program improvement) and "summative" evaluation (evaluation concerned with determining overall effectiveness)."

8. J.S. Ahmann, M.D. Glock, Evaluating Pupil Growth. Boston: Allyn and Bacon Inc., 1975, p. 13.

9. D.G. Lewis, Assessment in Education. New York: J. Wiley and Sons, 1975, p. 51.

10. R. Glasser, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions." American Psychologist, 1963, Vol. 18, 519-521, p. 519.

11. R.L. Ebel, "Some Limitations of Criterion-Referenced Measurement." Testing in Turmoil: A Conference of Problems and Issues in Educational Measurement. Greenwich, Conn.: Educational Record Bureau, 1970, p. 35.

12. W.M. Gray, "A Comparison of Piagetian Theory and Criterion-Referenced Measurement." Review of Educational Research, Spring, 1978, Vol. 48, No. 2, p. 223-249, p. 227.

13. The abbreviation CRM will be used in this paper to signify Criterion-Referenced Measurement.

14. The abbreviation NRM will be used to signify Norm-Referenced Measurement.

15. V.R. Martuza, Applying Norm-Referenced and Criterion-Referenced Measurement in Education. Boston: Allyn and Bacon, Inc., 1977, p. 7. "While special aptitudes, personality variables, etc., are defined normatively and hence, are meaningful only within a norm-referenced framework, the measurement of achievement can frequently be conducted within either a norm-referenced or criterion-referenced framework."

16. A.L. Costa, "Affective Education: The State of the Art." Educational Leadership. Jan. '77, 260-263, p. 260. "Self-concept, self-esteem and attitudes are probably the only facets of the affect for which standardized measures have been developed with any degree of reliability and validity."

17. Ibid., p. 263.

18. M. Priestly, Performance Assessment in Education and Training: Alternative Techniques. Englewood Cliffs, New Jersey: Educational Technology Pub. 1982, p. 149.

19. Ibid., p. 149.

20. H.G. Macintosh, D.E. Hale, Assessment and the Secondary School Teacher. London: Routledge and Kegan Paul, 1976, p. 108.

21. A.L. Costa, op. cit., p. 261.

22. M. Priestly, op. cit., p. 149.

23. The traditional evaluation policy is also known as the merit policy described in Chapter I.

24. "To the Root of the Issue", NSPRA, 1972, p. 7. "In the decade of the 1960's, however, new doubts arose over the effect of academic competition and the striving for high grades on the well-being of children, causing much debate and widespread change in methods of grading and reporting."

25. W.B. Brookover, A. Peterson, S. Thomas. "Self-Concept of Ability and School Achievement." Report of Project No. 845. East Lansing, Michigan: Michigan State University, 1962, p. 3.

26. Ibid., p. 76.

27. Robert L. Ebel, "The Failure of Schools Without Failure", Phi Delta Kappan, Feb. 1980, p. 386. "Humanistic educators often argue that the threat of failure, or the experience of it, are ineffective motivators. They refer to "many studies" that show that pupil efforts to learn are motivated by success, not by failure."

28. W. Glasser, M.D., Schools Without Failure. New York: Harper & Row, 1969, Introduction XIII.

29. Ibid., p. 16.

30. Ibid., p. 25. "The task was difficult, perhaps impossible, because the current philosophy of education, which emphasizes failure prevents the students from developing a feeling of self-worth."

31. Ibid., p. 37. "...the certainty principle...also dominates the rules of the school relating to the children's behaviour."

"In addition to the certainty principle, education, like much of our society, is dominated by the measurement principle.", p. 38.

"The certainty principle, with its total inability to provide students with emotional satisfaction commensurate with their effort is an important cause of educational failure.", p. 40.

32. Ibid., p. 26. "...it is school and school alone which pins the label of failure on children."

33. Robert L. Ebel, op. cit., p. 386. "For all these reasons, failure and retention lost favor. They were replaced by automatic or "social promotion". "Keep children with their age-mates" was the slogan. "Do what is best for the individual child."

34. Ibid., p. 12.

35. Ibid., p.4. "Congregating in the central section of any major city are increasing numbers of people whose common denominator is failure....; it is their definition of themselves."

36. Ibid., p. 12. "There appear to be many kinds of failure, of which school failure is usually considered only one. This appearance is misleading: there are not many kinds of failure."

37. People who are very successful in one aspect of their life, for instance in business, might be eccentric such as Howard Hughes and shy away from human contact, hence they appear unable to relate to others.

38. W. Glasser op. cit., p. 8. "The traditional psychiatric-sociologic approach is ineffective because it assumes that school problems are almost entirely a reflection of individual problems, poor home, environment, poverty and racial discrimination."

"Even a relatively warm and successful home will not counter-balance school failure, although children from such homes rarely fail.", p. 14.

"In wealthier neighbourhoods, where homes are successful and the environment strongly motivates towards success, deficient education does not so often lead to failure.", p.11.

39. Ibid., p. 25. "Schools fail to teach children to gain and to maintain a successful identity through the need pathways of social responsibility and self-worth."

40. Ibid., p. 27.

41. Ibid., p. 4. "If school failure does not exist other handicaps can be more easily overcome."

"Failure, which should be prevented throughout school...", p. 27.

"There are not failures at any level;...", p. 93.

"To keep a child working in school, we must let him know...it is not possible to fail.", p. 96.

"Because grades emphasize failure much more than success and because failure is the basis of almost all school problems, I recommend a system of reporting a student's progress that totally eliminates failure.", p. 95.

42. Ibid., p. 59.

43. W. Glasser, M.D., The Effect of School Failure on the Life of a Child. Washington: N.E.A.. 1971, p. 18.

44. Ibid., p. 19.

45. W. Glasser, Schools Without Failure, p. 95. "I recommend a system of reporting a student's progress that totally eliminates failure."

46. Ibid., p. 95. "...I suggest that no student ever at any time be labeled a failure or led to believe he is a failure through the use of the grading system."

47. Ibid., p. 77. "the required self-confidence is usually not developed unless the person has experienced success in school. Without confidence in themselves, failing children stall at making a decision to avoid the failure they believe will result no matter what course they take."

W.B. Brookover & E.L. Erickson, Society, Schools and Learning, Boston: Allyn and Bacon Inc., 1969, p. 106. "...a significant number of students are being needlessly hindered by low self-conception of academic ability...strategies must be developed for enhancing the self-conception of ability for a larger portion of our students."

48. Morris Rosenberg, Conceiving the Self, New York: Basic Books Inc., 1979, p. 9.

49. Ibid., p. 21.

50. W.B. Brookover & E.L. Erickson, Sociology of Education, Homewood, Illinois: The Dorsey Press, 1975, p. 277. "Students conceptualize about themselves in many different areas, and it is important to specify which self-conceptions are being referred to and to determine whether these self-conceptions are relevant for the behaviour in question."

51. B. Borislov, "Self-Evaluation and Academic Achievement", Journal of Counselling Psychology, 1962, Vol. 9., p. 253.

52. W.B. Brookover & E.L. Erickson, op. cit., 1975, p. 277.

53. Living and Learning, Ontario Department of Education, 1968, p. 56. "Children need to feel that they are accepted and that their efforts are appreciated. Failure in our society too often takes the form of a public stigma and unfortunately the loser in the early years of school acquires an image of himself as a failure, which becomes deeply ingrained in his psyche."

Ontario Schools: Intermediate and Senior Divisions, Ministry of Education, 1984, p. 11. "How well individuals function depends largely on their feelings of self-worth."

54. R.A. Carlton, "Self-Concept Research: The Gap Between Social Theory and Educational Practice", Canadian Journal of Education, 1981, p. 83. "Researchers have generally been very cautious about reporting their findings in causal terms. The great bulk of empirical enquiry has so far yielded only descriptive evidence of modest correlations, which in turn lend themselves to quite divergent causal meanings."

55. Ibid., p. 80. "In part, however, self-concept enhancement was sought as a means to higher levels of motivation and achievement."

56. W.R. Marx and P.H. Winne, "Self-Concept and Achievement: Implications for Educational Programmes", Integrated Education, 1975, p. 30. "Some researchers imply that self-concept is a fundamental determinant of academic performance. Others hold the opposing view, that achievement is a prominent determiner of self-concept. The recent upsurge of humanistic philosophies of education has shaded these interpretations of causality, favouring one directed from self-concept to academic achievement."

57. K.L. Wilson & A. Portes, "The Educational Attainment Process: Results from a National Sample", American Journal of Sociology, 1975, pp. 81-82.

58. R.A. Carlton, op. cit., pp. 81-82. "Most of the researchers writing about self-concept have warned of the dangers of reviving a construct which is hypothetical and purely utilitarian. To the working scientist a self-referent construct is an inferred postulate:..."

59. H.W. Marsh, L. Cairns, J. Relich, J. Barnes, R.L. Debus, "The Relationship between Dimensions of Self-attribution and Dimensions of Self-concept", Journal of Educational Psychology,

1984, p. 4.

60. Ibid., p. 3.

61. Ibid., p. 3.

62. Ibid., p. 4.

63. Ibid., p. 6.

64. Ibid., p. 6.

65. Ibid., p. 6.

66. Ibid., p. 6.

67. Ibid., p. 9.

68. Ibid., p. 11.

69. Ibid., p. 29. "The self-serving bias is a descriptive label for the tendency to self-attribution of responsibility to be more internal for success and external for failure."

70. Ibid., p. 29.

71. Ibid., p. 16.

NOTES TO CHAPTER III

1. J. Stanley Ahmann, Marvin D. Glock, Evaluating Pupil Growth, Boston: Allyn and Bacon, Inc., 1975.

2. Ibid., p. 220-221. "In educational measurement, validity is often defined as the degree to which a measuring instrument actually serves the purposes for which it is intended."

3. J. Kleining, Philosophical Issues in Education, New York: St. Martin's Press, 1982, p. 188. "What about their reliability? A given mark or grade is said to be reliable if it is consistent with the mark or grade which another competent examiner would give, or which the same examiner would give on another occasion."

4. Victor R. Martuza, Applying Norm-referenced and Criterion-referenced Measurement in Education, Boston: Allyn and Bacon Inc., 1977, p. 335. "Fourth, the prior problem is compounded when we recognize that the meaning of identical grades on a test purporting to cover the same content can vary widely across teachers because of such factors as different objectives, item selection and scoring procedures."

5. Phi Delta Kappa National Study Committee on Evaluation, Educational Evaluation Decision Making, Bloomington, Indiana: Phi Delta Kappa, 1971, p. 19. "Finally, when such multiple values are applied, will it not almost inevitably be the case that the same data, when interpreted in terms of different value standards, will give rise to antithetical evalutions?"

BIBLIOGRAPHY

Ahman, J.S., Glock, M.D. Evaluating Pupil Growth. Boston: Allyn and Bacon, Inc., 1975.

Anderson, R.H. "The Nongraded School", The National Elementary Principal. Nov. 1967.

Bloom, B.S., et al. Taxonomy of Educational Objectives, Handbook 2: Affective Domain. New York: D. McKay, 1964.

Bognar, C. "Back to the Basics - An Introductory Survey", Interchange. Vol. 7, No. 4, 1976-77.

Borisolv, B. "Self-Evaluation and Academic Achievement", Journal of Counselling Psychology. 1962, Vol. 9.

Brookover, W.B., E.L. Erickson. Society, Schools and Learning. Boston: Allyn and Bacon Inc., 1969.

Brookover, W.B., A. Paterson, S. Thomas. "Self-Concept of Ability and School Achievement", Report of Project No. 845. East Lansing: College of Education, Michigan State University, 1962.

Carlton, R.A. "Self-Concept Research: The Gap Between Social Theory and Educational Practice", Canadian Journal of Education. 1981

Costa, A.L. "Affective Education: "The State of the Art", Educational Leadership. Jan. 1977.

Douglas, M. "Minimum Competency Testing", Compact Contents. Winter 1978.

Ebel, R.L. "The Failure of Schools Without Failure", Phi Delta Kappan. Feb. 1980.

Emerson, G.J., M. Ayim. "Dewey and Peirce on Curriculum and the Three R's", The Journal of Educational Thought. April 1980.

Furth, H., H. Wacks. Thinking Goes to School: Piaget's Theory in Practice. New York: Oxford University Press, 1974.

Glasser, R. "Instructional Technology and the Measurement of Learning Outcomes: Some Questions", American Psychologist. Vol. 18, 1963.

Glasser, W., M.D. Schools Without Failure. New York: Harper & Row, 1969.

Glasser, W., M.D. The Effect of School Failure on the Life of a Child. Washington: N.E.A., 1971.

Glick, H.M., M. Schubert. "Mainstreaming: An Unmandated Challenge", Educational Leadership. Jan. 1981.

Gray, W.M. "A Comparison of Piagetian Theory and Criterion-Referenced Measurement", Review of Educational Research. Spring 1978.

Hall, O., R. Carlton. Basic Skills at School and Work. Toronto: Ontario Economic Council, 1977.

Haney, W., G. Madaus. "Making Sense of the Competency Testing Movement", Harvard Educational Review. Nov. 1978.

Kleinig, J. Philosophical Issues in Education. New York: St. Martin's Press, 1982.

Kuhn, D. "The Application of Piaget's Theory of Cognitive Development to Education", Harvard Educational Review. August 1979.

Labaree, D.F. "Setting the Standard: Alternative Policies for Student Promotion", Harvard Educational Review. Feb. 1984.

Lewis, D.G. Assessment in Education. New York: J. Wiley and Sons, 1975.

Lien, A.J. Measurement and Evaluation of Learning. Dubuque, Iowa: C. Brown Co., 1967.

Living and Learning. Ontario Department of Educaton, 1968.

Macintosh, H.G., D.E. Hale. Assessment and the Secondary School Teacher. London: Routledge and Kegan Paul, 1976.

Martuza, V.R. Applying Norm-Referenced and Criterion-Referenced Measurement in Education. Boston: Allyn and Bacon Inc., 1977.

Marsh, H.W., L. Cairns, J. Relich, J. Barnes, R.L. Debus. "The Relationship Between Dimensions of Self-Concept", Journal of Educational Psychology. 1984.

Marx, W.R., P.H. Winne. "Self-Concept and Achievement: Implications for Educational Programmes", Integrated Education. 1975.

Magoon, P.A. Education and Psychology Past, Present and Future. Columbus, Ohio: Charles E. Merrill Pub. Co., 1976.

Moskowitz, G. Caring and Sharing in the Foreign Language Class. Rowley, Mass.: Newbury House Pub. Inc.

National Commission on Excellence in Education "A Nation at Risk: The Imperative for Educational Reform", Elementary School Journal. Nov. 1983.

Ontario Schools: Intermediate and Senior Divisions. Ontario Ministry of Education, 1984.

Phi Delta Kappa National Study Committee on Evaluation, Educational Evaluation Decision Making. Itasca, Illinois: F.E. Peacock Publishers Inc. 1971.

Piaget, J. The Psychology of Intelligence. Totawa, N.J.: Littlefield Adams, 1972.

Popham, W.J., S.C. Rankin. "Minimum Competency Tests Spur Instructional Improvement", Phi Delta Kappan, May 1981.

Priestly, M. Performance Assessment in Education and Training: Alternative Techniques. Englewood Cliffs, New Jersey: Educational Technology Pub., 1982.

Reilly, W. "Competency Testing", Compact Contents. Winter 1978.

Resnick, D.P., L.B. Resnick. "Improving Educational Standards in American Schools", Phi Delta Kappan. Nov. 1983.

Rollins, S.P. Developing Non Graded Schools. Itasca, Illinois: F.E. Publishers Inc., 1968.

Shane, H.G. "The Academic Score Decline: Are Facts the Enemy of Truth?", Phi Delta Kappan. Oct. 1977.

Slavin, R.E. "Realities and Remedies", Elementary School Journal. Nov. 1983.

The Formative Years, Ontario Ministry of Education, 1975.

"To the Root of the Issue", NSPRA, 1972.

Tyler, R.W. Basic Principles of Curriculum and Instruction, Syllabus for Education 360. Chicago: University of Chicago Press, 1950.

Wells, T.L. "Adjusting the Pendulum", Interchange. Vol. 7, No. 4, 1976-77.

Wilson, K.L., A. Portes. "The Educational Attainment Process: Results from a National Sample", American Journal of Sociology. 1975.