CHAPTER 14•
EVALUATION AND ASSESSMENT: EXAMINATIONS AND TESTS

Generally speaking, evaluating, assessing and administering examinations and tests are the most common and everyday activities that a teacher has. They are also the most complex and difficult of all of his professional responsibilities. As far as pupils are concerned, evaluation is certainly the most important part of the teaching situation because the quality of their input and their progress in teaching and learning situations give sense and meaning to their existence in the school.

1. WHY DO WE EVALUATE IN EDUCATION?

All teachers know that they are continually involved in conscious and sometimes unconscious evaluation. Everyone who has worked with a class or a number of classes for a school term has grouped the pupils into different categories. For example, in every grade, there can be a small group of brilliant achievers, a large group of average ability and a number of individuals who will probably fail. In addition to these categories, there is usually a small group of pupils in every class who are troublesome, a few who never really do their homework, a number of pupils who are regularly absent, certain individuals who continually try to attract the teacher's attention, etc. All of these statements that reflect part of the daily teaching situation result from evaluations. Although the evaluations are not formalized, this kind of subjective assessment leads to classifications, groupings and to important strategies of testing in the classroom. This kind of assessment is subjective in the sense that no objectively developed criteria were used to conclude that pupils C, R, S and Y are always a nuisance in the class. In these circumstances it could not be stated that the pupils had been tested, examined, assessed of formally evaluated.

• This chapter did not appear in Inleiding tot die Didaktiese Pedagogiek but was added by G.D.Y., slightly edited, from the authors' English translation/rewrite published as Chapter 10 in Van der Stoep, F. and Louw, W. J.: Didactics. Pretoria: Academica, 1984.

However, when matters such as determining a child's intelligence quotient are involved, or when aptitude, learning achievements and the promotion of pupils to the next grade are considered, subjective assessment is unacceptable. The teacher simply does not sit down with a list of names in his hand to mark those who pass or fail, although he knows very well who is bound to pass and who probably will not. These decisions must be based on objective information that has been collected objectively. Therefore, a variety of tests is introduced to insure that the evaluation is accountable in terms of objective criteria. Tests do not supersede the teacher's judgment: they actually supplement it. In this sense, tests are an independent means of assessing the nature and level of the child's achievements.

Consumers of education are most interested in the teacher's procedures of evaluation. The outcome of these procedures indicates the nature and scope of the progress parents expect of their children. In this context some teachers have learned through bitter experience that if examination and test marks are kept as high as possible parents will be satisfied, irrespective of whether their child has, in fact, reached the level suggested by the mark or percentage. These teachers reason that a few added points cost nothing, that parents will not ask awkward questions and that the child's self-image will not be endangered. In this case the teacher uses a strategy of evaluation that avoids conflict, thereby hoping to establish an equable learning and teaching environment. It is obvious that this strategy cannot be justified pedagogically, didactically or in any other way. Nowadays this kind of subjective evaluation is difficult to use because the marks achieved by pupils in all classes are continually determined by standardized tests that have been stored in secure data banks. Standardized tests are used regularly by schools to verify the quality of the pupils' learning achievements.

Other consumers of education are also interested in evaluation, especially employers. Schooling always leads to the achievement of a certificate. Certificates are statements of ability or of a level of education (schooling) that has been reached. This is important for the pupil as well as for the employer. For the pupil, the certificate

declares his competence to proceed with his formal education (e.g., by enrolling in a tertiary educational institution) or to start training outside the formal school system, in which case his certificate serves as a qualification for entrance. On the other hand, his certificate provides him with direct access to employment. The fact that employers place high value on these pronouncements of the educational sector and their comments regarding the capabilities of their employees (e.g., their ability to use language correctly) are important for educational institutions because they serve as a barometer to control the quality of schooling.

Government and the "powers that be" are another sector of society most interested in evaluation practices. The "trustworthiness"• (reliability and validity) of the procedures of evaluation in schools under their control not only determines the status, prestige or functional value of the education that they largely finance; this "trustworthiness" offers every citizen the reassurance that his education will provide him with the opportunity to lead a worthy vocational life and to enjoy the status associated with his vocation. Authorities, therefore, expect schools to contribute to forming independent, dependable, loyal and economically productive citizens. Trained manpower is the greatest asset any country can have. Therefore, the quality of the training must be carefully and continually monitored and controlled by means of standards especially designed for every form of education. Nothing could ever damage the image of a university more than to suggest that its standards are suspect.

Strategies of evaluation are of great importance in society as a whole. Education and the possession of certificates provide access to job opportunities for learners. Work generates income, i.e., economic independence. Generally, the better the qualification, the higher the income, and level of income generally determines the level of status in society. This is especially important in developing

• In this context I (G.D.Y.) have translated "reliability" as "trustworthiness". Clearly the authors do not use "reliability" only in the technical, psychometric sense of "consistency" or "stability"; by this term, they simultaneously imply that one can rely on a test result really meaning what it is being interpreted to mean (i.e., it's validity). For me, the non-technical term "trustworthy" deals with this ambiguity, while preserving it, because it simultaneously embraces the psychometric concepts of reliability and validity.

countries because it provides the mechanism for social mobility and means that one can aspire to higher social status on the grounds of good education and training. The stratification of society by classes, levels or ranks is always criticized but it cannot be ignored or wished away. It seems that it has become a part of every society. The certificates of educational institutions are an important key for unlocking these doors for the consumer of education.

This brief overview hopefully explains why "trustworthy" evaluation procedures and standards in education are of general importance in every country and community.

Naturally there is widespread and intense criticism of tests and examinations. Criticism is aimed at all levels and kinds of testing and examining, from measuring the intelligence of a young child to doctoral examinations. Reasons for the criticism are well known: it encourages the overemphasis of examination results, children do not learn to think, they only memorize; it prejudices the teaching of the gifted; it is discriminatory in education and in society as a whole, etc. But nobody has as yet suggested an acceptable strategy to replace tests and examinations. In fact the contribution modern technology has made to insure the "trustworthiness" of test and examination procedures has increased their prestige.

The aim of this chapter is not to provide the prospective teacher with comprehensive training in evaluation procedures. This is in itself a study or training course that is included in most teacher training curricula. The aims of this chapter are:

- to make student teachers aware of the importance and problems of responsible test and examination procedures; and
- to explain the basics of evaluation practices as clearly as possible.

The selected bibliography at the end of this book contains many excellent works on the subject, including works of South African authors. These works contain the necessary descriptions of testing techniques, statistical procedures and interpretations of a wide variety of test results as well as their statistical meaning.

To summarize: awarding marks is one of the activities in which the teacher is always involved. The circumstances in which he carries out his teaching contribute importantly to the strategies he will use. He must organize his test and examination procedures to grade pupils effectively according to their achievements. This is a very sensitive and important task. The child's self-image, his image of the future (direction of study, vocation), his relationship to his parents, his prestige in the class all depend on [really assume] the "trustworthiness" of the teacher's evaluation procedures. A large part of this is subjective. The subjective factor in assessment can never be completely eliminated; therefore, every teacher must use evaluation strategies as correctly as possible to guarantee the "trustworthiness" of his evaluations.

The marks awarded to a pupil, that to a large extent eventually reflect his ability, aptitude, acceptability, usefulness, etc., are usually used in the following ways:

- as a profile of his progress in school;
- as a basis for his promotion to the next grade;
- as encouragement to make greater learning effort;
- to serve as a guideline for choosing a direction of study and for vocational guidance;
- to draw a profile of his personality and character;
- to serve as motivation to study for scholarships, awards, etc.;
- to serve as criteria to assess his worth when he applies for a job;
- to provide data for research in a variety of circumstances, but especially for educational purposes.

For these reasons test and examination results may not be shrouded in mystery or considered to be an aspect of the teacher's responsibility that is not open to public scrutiny. Pupils must know at all times what the teacher's procedures are as well as the criteria he will use in his assessment.

2. WHAT AIMS ARE VALID FOR EVALUATION?

From what has been said above, the following aims of evaluation are important.

2.1 Certification of pupils' achievements

Although many teachers doubt the desirability of grading pupils on the basis of examinations (A, B+, B, C or 80%, 65%, 50%), it cannot be denied that well constructed examinations (tests) do provide a "trustworthy" criterion to determine the level of achievement and especially to decide whether the child can be promoted to the next grade. Reasons for measuring achievements were mentioned above. Surely it is inconceivable that a teacher could decide intuitively to fail a pupil! In addition to determining scholastic achievement, there is a whole series of equally valid aims when testing and examining in order to make responsible educational decisions such as the kind of school the child should attend, or to provide a basis for advising the parents. Without the substance of evaluative data the teacher cannot function effectively as an educationist.

The following are conditions where assessing and testing must be as objective as possible in order to make responsible and accountable educational decisions.

2.1.1 Grouping pupils

This is a common practice in teaching. Grouping pupils in grades, subjects, remedial classes, etc. are examples. When grouping pupils, the teacher seeks either homogeneity or individual differences in order to provide more effective teaching; for example, in selecting highly gifted pupils. In certain classes, pupils can also be grouped according to learning tempo (quick, average and slow). In practice, pupils are never grouped only according to tests. The teacher's observations are at least as important as the test results. Grouping also is never used in an attempt to eliminate individual differences.

2.1.2 The organization of specific learning programs

"Trustworthy" test results can provide valuable information in organizing learning programs; for example, the development of abilities, specific talents (art) or the solution of learning problems. They also provide an opportunity to construct a curriculum or design teaching material in accordance with pupils' needs.

2.1.3 To determine aptitude

This is a normal practice in school as well as vocational guidance. Tests provide indispensable information for choosing the right subjects, in determining the kind of teaching that must be provided (e.g., vocational education) and in deciding on the most applicable course of study for a specific pupil (e.g., the humanities or the natural sciences).

2.1.4 Changing schools

It is often very difficult for teachers to determine a pupil's level of achievement when the child moves to a new area and school. It would be unwise to attempt to establish a child's level while he is trying to adjust himself to a new school and circumstances. The use of standardized scholastic tests is a guide and reasonably "trustworthy" way to determine the newcomer's level of achievement and give him the necessary attention in class.

2.1.5 Research

The importance of well-designed and carefully executed evaluation procedures for educational research is self-evident. Various aspects have already been referred to: giftedness, learning problems, data banks to provide standardized examinations, determining realistic levels of achievement for different age groups, the development of "trustworthy" predictors of academic or vocational success for all categories of learners, and last but not least, school readiness. These are examples of the positive contribution "trustworthy" evaluative procedures can make to education.

2.2 Evaluating the effect of teaching

Good testing can be a valuable support for the teacher in assessing the effect of his teaching. Although standardized tests and examinations are available and are increasingly used in education, it is the teacher who actually decides on its contents. Whatever appears on the examination reflects what the teacher considers to be important and how he wants the pupils to deal with it. In the

explanation of the development of the curriculum in Chapter 11, the importance of aims was stressed. What exactly must be achieved in teaching specific aspects of contents? It follows that the teacher evaluates (selects contents) in terms of aims and objectives identified in the curriculum.

If the teacher wants the pupils to know (remember) certain facts or formulas then he must test accordingly. The same applies to aims such as creative thinking, learning methods for solving problems, transferring insights to new situations, etc. This imperative is true even for the smallest teaching unit. If a teacher teaches without first determining exactly what his aim is, he does not know what he is doing. If he then tests without knowing what he is testing, the same applies. This means the teacher expects specific achievements from his pupils that must be reflected in his test, that he has carefully identified these achievements and competencies in his marking scheme and that he knows exactly the criteria for which credit will be given and the weight they will have for each answer. In effect, he is, in part, assessing the influence of his own teaching through the quality of his pupils' achievements. When grading papers, it is often disconcerting to a teacher to see the quality of his teaching reflected in the answers his pupils give.

2.3 Encouragement

Tests and examinations can also encourage pupils, especially because they are so intent on succeeding. There are few motivations in life that are as encouraging as success. Good achievements by pupils encourage the teacher to attempt to improve his teaching even more. The same applies to the pupil. For the achiever, the examination should never be a crisis but an opportunity to show his steel. Badly devised tests frustrate this aim. The preparation of a good test requires much time and hard work but it has its own rewards.

2.4 Programmed and computer-assisted teaching

As explained in Chapter 10, immediate evaluation is an integral part of programmed and computer-assisted teaching. Each answer of the pupil is immediately assessed and marked right or wrong. If

teachers develop these programs for their own teaching, they must include immediate assessment in the program. The details of programmed and computer-assisted teaching were discussed extensively in Chapter 10 and will not be considered any further.

3. THE TEACHER AS EVALUATOR

Nowhere in our society are tests and examinations conducted more regularly and more extensively than in our schools. The teacher is the center of this educational activity; in fact, in the most literal sense of the word, he is the hub of the evaluation wheel. Therefore, it goes without saying that the teacher must be a knowledgeable and "trustworthy" evaluator, examiner and judge of his pupil's achievements. The contribution he makes to his pupils' future can never be overestimated. For this reason, accidental or casual factors have no place in his evaluative responsibilities. He must know the theoretical basis of evaluation and must be able to apply the related principles in his classroom. This not only applies to the child's learning achievements in a specific subject, it applies equally to the evaluation of him as a person including his aptitudes and potentialities as a human being.

Many of the tests the teacher uses are highly specialized. Education authorities appoint teacher specialists to administer specialized tests (e.g., intelligence tests). Generally speaking, schools make use of tests (i.e., all forms of examination and evaluation) as instruments to maintain and promote the quality of teaching. By means of these tests the teacher constructs a profile of every individual child's achievement. The aim is to eventually help each child attain the highest level of achievement by individualizing the teaching and learning situation. This is the most important reason for introducing computer-assisted teaching in schools all over the world. In their striving to increase positive teaching effects in this context, education planners and curriculum compilers are continually looking for ways to stabilize the teacher's evaluation. Homogeneous grouping of pupils is one example of what is meant. Programs to prepare school beginners are another. The teacher is examiner by virtue of the fact that he teaches people.

Most dictionaries agree that the concept "examination" can be described in the following terms: an examination is verbal or written proof established by means of questions and answers and its aim is:

- to assess the knowledge, capabilities or advancement of a pupil, student or candidate in a subject or series of subjects in a specific science or area of knowledge; and
- to examine a person's ability to accept certain responsibilities.

The teacher can approach the problem in different ways in the classroom. We are all aware of them. He can instruct his pupils to write an essay or paragraph on Napoleon's foreign policy or on the symbolism of T. S. Eliot's poetry. This would qualify as an essay examination. However, the teacher can also make use of objective tests of which there are many variations. These tests are considered to be objective because they are not influenced by the teacher's opinions. The questions (items) and answers have been previously standardized. When the essay on Napoleon's foreign policy is marked, the assessment is subjective because there is no standardized answer. It hardly ever occurs that two examiners award the same marks to an essay because their own interpretations of the answer will differ. Examples of objective examination items are given later.

4. NORM DIRECTED EVALUATION

One of the most important objectives in evaluation is to establish criteria in terms of which the teacher can describe the child's achievement. It is important for a teacher to remember that a mark (score), irrespective of its nature, is relatively useless unless he knows what the basis or background is to which the mark refers. For example, in administering a scholastic test, a mark of 10 can indicate exceptional giftedness for a school beginner but serious retardation for an older or secondary school student. The mark can be interpreted only if one knows the *norm* to which it can be related.

There are mainly two kinds of comparative groups that can be used to give evaluative meaning to a child's achievement on a test. One

can compare the child's achievement with that of other children of his age, grade level or within his class. The comparability of his achievement is the important consideration. The norms used to interpret the individual child's achievements are generally classified as follows.

4.1 Age norms

These norms compare the individual child's achievements to those of children of the same age. A clear example of this is found in studies of babies. At a certain age a child should weigh so much, be able to sit without help, be able to crawl, etc. However, as children grow older, it becomes more difficult to use age as a norm because its "trustworthiness" becomes suspect. Differences in learning achievement cannot be described only in terms of comparative ages.

4.2 Class or grade norms

Grade level norms are established by calculating the average score or mark of pupils in the same grade. A test is given to a representative sample of pupils who, although they come from different areas, socio-economic backgrounds, and although their ages might differ, are all at the same grade level. In using this method, comparisons are possible because all of the children [presumably] have enjoyed the same range and level of education. In this way, a norm is established for a specific grade level in school with respect to particular aspects of the curriculum, e.g., fractions in arithmetic. When such a standardized test is given to a pupil, his achievement can be compared and judged with respect to the average achievement of pupils in this specific grade.

4.3 Percentile norms

When using percentile norms, the percentages of pupils are determined with which the individual pupils' achievement can be compared: if 75% of the pupils with whom pupil A's achievements are compared scored lower marks than he did, this means that pupil A has a percentile rank of 75. Pupil A therefore achieved a higher score than 75% of the pupils who took the test, but he achieved a lower score than 25% of the group.

The particulars of norm directed procedures of evaluation imply the statistical analysis of test results. As was pointed out, the aim of this chapter is not to go into the details of such statistical analyses. Nobody doubts the value or accuracy of results analyzed in this way; nobody doubts the "trustworthiness" of the conclusions that can be drawn from these analyses. The question of statistical accountability will, therefore, not be gone into any further. There are, however, a large number of different everyday considerations of a more general didactic nature that should be explained.

5. CRITERION DIRECTED EVALUATION

Norm directed evaluation is aimed at devising procedures to compare a pupil with other pupils in different ways. However, many evaluation experts are of the opinion that this is not the best way to assess a pupil's achievements. Their aim is to identify a set and fixed standard of evaluation in terms of which the individual pupil should be evaluated. In this case, the individual pupil is not assessed in terms of the achievements of another pupil or group but according to criteria derived from the contents, as such. The level of achievement of all of the other pupils regarding the criterion is irrelevant when the individual is assessed because the criterion does not reflect the achievement of a group of other pupils. The criterion is found in the contents themselves. A criterion directed test will, therefore, indicate the degree to which an individual pupil has grasped a specific theme or problem, for example the ability to correctly spell certain categories of words. It is quite irrelevant that 80% of all pupils of his age and grade level can spell these words correctly; only the individual pupil's achievement is important. The test will indicate the child's difficulties or problems, irrespective of the difficulties of the pupils in his age group or any other group.

Criterion directed evaluation leans heavily on the importance of aims and objectives and insists that the child's achievement is assessed in terms of the specific teaching aims and objectives and the contents that were chosen for these purposes. Data banks and question banks are primarily associated with this type of evaluation because the questions and answers have been extensively tested and

verified; the discrimination value has been established empirically and the criteria (answers) are absolutely constant.

6. EXAMINATIONS AND TESTS

The importance of the authentic formulation of teaching and learning aims in curriculum development has been stressed in Chapters 7 and 11. The contents that are selected must insure the realization of these aims in a general and specific sense. At the same time, the ordering of these contents must place the learning task within the child's grasp, and it must insure a realistic level of access for the child while revealing the specific nature (elementals, essences) of the contents. Curriculum evaluation cannot occur outside of the framework of these aims. Also, the curriculum is not evaluated only after it has been designed: this evaluation is a continuous, summative procedure that occurs even while the curriculum is being constructed. When a practicing teacher is testing pilot curricula as part of a research team (i.e., when it is made operational in didactic situations), assessing the relationship between aims and contents is the teacher's fundamental concern. If the learning aims are not fully covered by the contents, i.e., if they cannot realize the desired learning effect, the pilot curriculum must be reviewed.

The same principle holds for all examination and test procedures used by a teacher. He always tests and examines in terms of the teaching and learning aims formulated in the curriculum. He determines his own specific teaching objectives in terms of curriculum aims and incorporates them in his lesson designs (plans). This is an inexorable and inescapable demand for every aspect of his evaluative practice. This principle never loses its validity in education, whatever the circumstances might be. Therefore, it is self-evident that a teacher cannot design a lesson without carefully and explicitly formulating his objectives. Any other procedure will put him in the position of a dog chasing its own tail.

The same principle applies to examinations. Examinations are concerned only with the formulated teaching and learning aims. Any other approach is unprofessional, unethical and

unquestionably dishonest. Therefore, a teacher must not decide what the questions will be for this afternoon's test or examination while riding a bus to school in the morning. Learning objectives, test questions and marking or scoring rubrics are a unified trilogy. A number of acts and particulars are of importance in this context when a teacher considers his test procedures.

6.1 Informal testing techniques

These techniques can be of great importance in drawing a profile of a pupil's achievements over a long period. Norm- and criterion-directed evaluations formalize the evaluation situation in a specific sense. They lead the teacher to objective measurement procedures that have a certain merit in their own right: they are unprejudiced, relatively free from subjectivity and they should provide "trustworthy" test results under almost all circumstances. However, we all know that the teaching situation is not objective by nature or character. In practice, there are no standard pupils or schools and this means that a complete perception of an individual pupil's achievements cannot really be obtained by making exclusive use of objective tests. Important aspects of a pupil's personality and character can be determined only by careful *observation* by the teacher. The teacher's observation of a pupil in a series of situations, often over a long period of time, can be of decisive importance in explaining the pupil's profile of achievement that is established by means of objective tests, and to place them in perspective. A good example of this kind of evaluation is found in pre-primary and primary education where standardized test have only limited value (except in the testing for school readiness). Many important decisions regarding children are based on the teacher's observations. The key to this kind of evaluation is that it must be systematic. The teacher must decide beforehand what areas of the child's achievements should be observed. These particulars must be written down carefully and regularly.

Education offers many opportunities for this kind of evaluation. In addition to the circumstances already referred to, the teacher can make unobtrusive notes on the child's self-confidence, bravado, apparent indifference, venturing attitude, aggression, the quality of his concentration, his ability to think logically, his behavior in a

group (e.g., his constant efforts to draw attention to himself), his preference for certain learning situations (projects or tasks), over-dependence on help from others, the quality of his language and articulateness, exceptional gifts (singing, drawing, dramatization), learning tempo, reading skills, perfectionism, etc. All of these factors have an important influence on a child's stability and his participation in teaching situations.

Teachers often overlook the importance of stability. When pupils are conspicuous in one way or another (rowdiness, copying other children's work, etc.), one is inclined to concentrate on that particular aspect of behavior instead of considering all of the factors that determine the stability of the child's learning input. Where inconspicuous pupils are concerned, teachers seldom or ever go to the trouble of consciously observing the ways their psychic life is manifested in the learning situation. These observations by the teacher are of the most fundamental importance because they provide the background against which formal test results can be assessed and interpreted. It has to be emphasized once again that these observations need to be planned and that a record must be kept of each pupil's particulars in order to make a comparison with formal test results possible.

6.2 Class tests and examinations

Generally speaking, teachers seldom make use of standardized tests and examinations. This is partly because they are not always available and partly because teachers formulate their own teaching and learning objectives that they prefer to evaluate by constructing their own tests and evaluations. However, with the emergence of the computer, standardized items are more readily available (via question and data banks) and, therefore, more attention should be paid to the advantages these items offer for evaluation. The advantage of the objectivity of these tests has already been discussed. The fact is, they are more reliable (psychometrically) than tests constructed on an *ad hoc* basis by the teacher; this is largely because they have been administered to a large number of pupils over a long period and because statistical analyses have established the (psychometric) validity of the item (question). Validity, in this sense, means that it has discriminative value. The

item discriminates (differentiates) between those pupils who know and those who do not know. During the formulation of the item and the establishment of its validity, attention is given to factors such as the possibility that the correct answer can be guessed, that it can be inferred from the way the question has been asked or that the answer can be discovered from the information provided in the phrasing of the question. Although the general teaching situation does not provide for the extensive use of these controls, the demands for reliability concerning the results of the teacher designed test are equally binding. The items on the teacher's test also must be valid and have discriminative value. If a specific question is answered correctly by 90% of the pupils, it has very limited discriminative value. These requirements will be discussed below from time to time. The fact remains that the teacher has an important responsibility to keep (psychometric) validity and reliability in mind as criteria when designing a test. In addition, it is an unavoidable requirement that the teacher's teaching and learning objectives must be reflected in the contents of his test for him to be able to properly evaluate the effect of his teaching.

When compiling a test or examination, the teacher is, in fact, informing his pupils of what he considers to be important in the syllabus. Therefore, one can understand why pupils are constantly analyzing copies of old examinations in their effort to discover what the teacher is likely to ask. As far as the teacher is concerned, his pupils' answers are an important source of information regarding the quality of his teaching as well as the progress of his pupils as a group and as individuals. He can determine from the answers which aspects of the syllabus have not been mastered adequately and which pupils need extra help because their progress is unsatisfactory.

One fact has become very clear: good tests do not drop from the sky; they are carefully planned and constructed with great consideration. Nor must a teacher ever assess pupils' answers without a scoring rubric; that is, he must know exactly what he expects the answers to be, the number of points to allocate to each question and for what he will penalize the pupil. He must also accurately judge how much time a pupil will need to answer a question. At the time of developing the test, the teacher also

composes the scoring rubric. It also is important to discuss the questions with the pupils when returning their test papers, especially those answers deserving high marks and those awarded only low marks. General mistakes and factors that influenced pupils' achievements (e.g., the amount of time spent on each question or mistakes in interpreting questions) must be discussed. This is a very important aspect of the teacher's teaching responsibility and is, in fact, an important aim when organizing tests and examinations.

6.3 Kinds of aims

To teach without an aim is like going on a journey without a destination. In the literature dealing with teaching aims, one encounters many variations. Aims may vary from subject to subject but, generally speaking, there are three main categories, i.e., three kinds or types of aims that are intended to actualize learning by means of teaching.

6.3.1 Cognitive aims

Cognitive aims are closely associated with knowledge. Knowledge, as such, covers a wide spectrum of abilities, e.g., subject detail, subject terminology, subject methods, generalizations and abstractions, etc. These aims are also aimed at understanding contents, the application of knowledge, the analysis of data and the synthesis and evaluation of information. As is well known, these aims always appear in the form of an instruction (verb), for example:

Aim	*Instruction* (verbs)
Knowledge	Define, prove, differentiate between, explain, identify, name.
Understanding	Change, write in your own words, illustrate, translate, interpret, determine the value, arrange, estimate, generalize, make conclusions, infer.

Analysis	Apply to, calculate, classify, choose, develop, use, transfer to, rearrange.
Synthesis	Combine, make deductions, develop, formulate, change, plan, design, make a suggestion, tell, write.
Evaluation	Reason, judge, consider, decide, value.

It is once more important to note that cognitive aims imply a question that contains an instruction that is introduced by a specific verb that tells the pupil exactly what he must *do* when answering the question.

Pupils often misread instructions; they decide what they think the question is and simply write down everything they know about the topic. Therefore, teachers must pay attention to examination skills during their teaching. The pupil must learn how to write a test or examination; how to read the questions; how to interpret the question and how to organize his answer. This does not happen by itself and pupils do not simply know these aspects of examinations without receiving careful instruction from the teacher. The fact that a pupil must be able to recognize and differentiate between these aims is just as important as the teacher's knowledge of aims. For this reason, the teacher should never write ambiguous questions. If he genuinely doubts the correctness of any two answers, the question must be a bad one, because the results of the pupil's answers will be misleading and the teacher will be unable to make valid deductions concerning the quality of his pupils' learning or his own teaching.

6.3.2 Affective aims

Affective aims are directed at the quality of the pupils' input. Because these aspects of achievement cannot always be observed directly, they pose specific problems for the examiner in effectively ordering and in designing authentic criteria for evaluation. Aims such as optimizing the child's learning intention, attitude, appreciation, principles, etc., are difficult to evaluate. One cannot simply award marks for qualities of existence although they are

certainly intrinsic aspects of all teaching and learning situations. One often finds that affective learning aims are coupled to specific contents like poetry, music and art where the accountability of a child's appreciation has a great deal to do with the way he has identified with the contents. His answers are often an unfolding of his experiences of the world that he attempts to articulate. His receptivity, the amount of attention he gives to the contents and the ways he makes them his own are, therefore, important aims within the field of teaching.

6.3.3 Psychomotor aims

The concept "psychomotor aims" means that the teacher will assess the directly observable skills of his pupils. During the primary phase of schooling, psychomotor aims are to learn to write and read correctly. In the secondary school, the teacher seeks proficiency in the pupils' handling of a musical instrument or a machine such as a lathe or typewriter used in vocational preparation. Therefore, the items in the test are formulated in terms of observable skills. Words like draw, construct and indicate initiate the tasks and instructions normally contained within these items.

7. STEPS IN CONSTRUCTING A TEST

In terms of what has been said above, the following is a framework within which the teacher can order and evaluate a test.

i) Make sure what contents are relevant. A subject teacher responsible for different classes may easily become confused and ask for applications of knowledge that have not been dealt with or that have been dealt with in one class but not the other.
ii) Write down the aims of the test precisely, i.e., say exactly what the test is intended to evaluate.
iii) Plan the type and scope of the test (e.g., essay, multiple choice).
iv) Make a provisional draft of the items (questions).
v) Plan the length of time needed to answer these items.
vi) Reconsider the items. Delete those that are unsatisfactory. Determine whether the question is clears and unambiguous as it stands.

vii) Order the items, e.g., from easy to difficult. Make sure that even the weakest pupil can pass if he has really prepared for the test. Make sure that the difficult items are not beyond the ability of the best pupils.
viii) Make sure the test instructions are clear.
ix) Draw up a marking (scoring) rubric. At this stage the teacher must be able to determine whether the test is too long, too difficult or too easy. Review the test accordingly.

The following are valuable guidelines in assessing a test.

i) Are the items *relevant,* i.e., are they in accord with the teaching and learning aims and the aims written down for the test?
ii) Do the items reflect that the aims are in *harmony* with one another, i.e., are the different kinds of aims (cognitive, affective, psychomotor) of the test in harmony with one another?
iii) Is the *time* allocated for the test *realistic?*
iv) Will the items have *discriminative value?*
v) Is the *level of difficulty* in accord with the teaching that was offered?
vi) Is the *formulation* of all of the items *fair?*
vii) Will the answers provide a *"trustworthy"* reflection of the *learning effect* of the class as a whole?

8. KINDS (TYPES) OF TESTS

8.1 The essay test

This is the oldest and one of the most popular forms of testing, but it is also the most unreliable. Many research projects have been launched to examine this point and it has been shown beyond doubt that it is an unreliable form of examination. It has also been shown that the subjective aspect of the assessment mars its reliability. Percentages awarded for the same language essay ranged from 50 to 98% and from 30 to 100% for the same history essay. Even in the case of a trigonometry test, the scores ranged from 28 to 95% for the same answer. Factors such as neat handwriting and neat presentation were often of decisive importance for the mark awarded to the essay. When, for instance, the essay test was

modified to include writing short paragraphs on various topics, its reliability increased. For this reason the construction of a good marking rubric is of great importance when using an essay examination. The filling-in test format is an important and useful variation of the essay format and will be discussed more fully later.

When drawing up an essay test, the teacher must keep the following in mind.

i) What is the aim of the question? The aim can be to determine whether the pupils can apply certain facts or whether they can make an evaluation (e.g., of a poem). Therefore, it is only logical that the teacher will refer to his aim in the question.

ii) Formulate the question carefully and coherently. Do not use complex sentences. Formulate the question in simple sentences to insure that every pupil in the class will understand.

iii) Give direct instructions by choosing the verbs with care, e.g., compare, name, give an example. Be careful not to begin the essay question with words such as why, who or which. They are not *verbs* and the danger is that the instruction will be vague (the aim of the question is therefore vague). Avoid instructions such as: "Write what you know...." "What is your opinion....", etc. This gives the child the opportunity to write down irrelevant details in the hope that the teacher will award points although the facts he mentions are not included in the grading rubric.

iv) Determine the area of knowledge to be covered by the question as accurately as possible. A general question, for example, dealing with the causes of the First World War, can be clearly formulated if the question refers to categories of causes such as political, economic, etc. When answering generally formulated questions, pupils are inclined to follow devious paths to write long-winded introductions and generally to "pad" their answers with many words. These problems can be avoided by identifying the area to be covered in the answer and by carefully defining the limits in the formulation of the question.

v) Be sure that the question is formulated in such a way that it does not discriminate to the advantage of a particular group, e.g., to the advantage if the weaker group in the class. On the other hand, if 20% of the pupils do not understand the question, then it must be unfair.

vi) Be careful of too wide a choice in the essay question. A teacher can arrive at a much better understanding of the pupils' command of the contents if they all answer the same questions because the frame of reference is the same for all.

Marking or scoring the essay question is just as thorny as formulating it. The following suggestions can possibly help the young teacher solve this problem.

- Take considerable trouble in composing a model answer or marking rubric. By doing so important criteria can be developed to assess the answers effectively and to be consistent in allocating marks.
- It at all possible, do not attempt to identify the writer of the answer. The pupil's personality and his behavior in class can influence the teacher's allocation of marks. Sometimes the pupil receives the benefit; sometimes not.
- Mark question 1 for the whole class, then question 2, then question 3, etc. This enables the teacher to evaluate all of the answers to a particular question which is important for maintaining the same standard and therefore for the reliability if the markings.
- Write comments, but not only negative remarks. Indicate why a pupil has been penalized but give him some encouragement as well. Comments are often of greater didactic importance than the mark allocated to the answer.
- If possible, ask a colleague to control or verify your assessment. This gives the teacher an excellent opportunity for self-assessment and for the sharing of information, hints, etc. The beginner needs all the help he can get to establish the correct marking standard and to account for the "trustworthiness" of his markings, especially of essay questions.

The advantages and disadvantages of essay questions have been referred to but it is nevertheless important to summarize briefly what has been said.

The main *advantages* are the following.

- It offers the pupil the opportunity to formulate his thinking, to order the contents, to reason, to make syntheses and evaluations, to motivate judgments, etc.
- It forces pupils to answer in their own words and not quote from the textbook. This contributes to language development. In this sense, it has positive teaching potential.
- It encourages neatness, care (e.g., spelling) and concentration on the essences of the contents.

The most important *disadvantages* are as follows.

- The time allocated to tests always creates a problem. Most pupils do not complete the examination.
- The essay cannot really be objectively marked.

8.2 Objective tests

Objective tests were introduced into evaluation procedures to counter the disadvantages of the essay examination and also because they are intrinsic aspects of certain systems such as computer-assisted teaching.

Tests considered to be objective qualify as such because the method of marking is determined at the time the item is written. The correct answer to the question is determined before the test is given and, therefore the evaluator's task is to determine whether the pupils' answers correspond to the correct answers. The claim of objectivity rests exclusively on the way the answer is assessed and marked. The content of the item and the way it reflects the contents of the syllabus can be just as subjective as the essay examination. The teacher can also make a mistake regarding the correctness of an answer to a test item. Therefore, this aspect of objectivity is not always above suspicion. The fact is that the correct answer is determined prior to the test and thus assessment is not a problem during the evaluation of answers after the test.

Every objective test or examination has a set structure in that all the candidates answer the same questions. This means the questions should be interpreted in the same way by all pupils. As we know, this is one of the greatest problems of the essay question. In most cases, in objective testing, the question (item) is presented in such a way that the candidate recognizes the answer instead of trying to remember it. This is always the case in the presentation of multiple-choice tests. The explanation of the answer, the ordering, formulation, handwriting, etc. have no influence on the assessment. Various kinds of objective tests include the following.

8.2.1 True-false tests

Although true-false tests are often used in assessing young children, there can be no doubt that this is the least satisfactory kind of objective test. Its popularity is probably due to the teacher's belief that it is a very simple test to construct. Nothing is further from the truth. In the majority of true-false tests, even good ones, pupils can average 50% by merely guessing. Besides it is extraordinarily difficult to identify items that are absolutely true or false. For example:
Automation in industry causes unemployment
True False

Young teachers experience difficulties in maintaining class disciplined
True False

Both statements are only partially true, i.e., they are true with certain reservations. Automation is not the only factor responsible for unemployment. Teachers should attempt to list a hundred true-false items and then determine, through their pupils, whether certain items are not biased; whether some items suggest the answer and whether some are partially true as well as partially false. Generally speaking, this kind of test should be avoided completely.

8.2.2 Completion tests

These tests are designed in such a way that the pupil will have to fill in a blank space with a word, phrase, number, formula or other particulars to complete the item.

a) Jan van Riebeeck landed at the Cape in _____
b) A soccer ball is _____ and a rugby ball is _____
c) At what age does one qualify for a drivers license? _____
d) What is the current percentage of the General Sales Tax? _____

The first two examples are genuine completion questions. The other two are really short-answer questions. But in both kinds of example the pupil is expected to provide the answers; i.e., there is nothing in the question that enables him to recognize the answer—as in the case of other variations of objective test items. Consequently, it is an effective format for testing factual knowledge (names, dates, formulas, etc.). This quality is also its limitation in that it is not suited to assess insight, formulations, ordering, application, etc.

When this form of evaluation is used, the following hints can be valuable.

i) Use clear, everyday language.
ii) Insure that the realization of the aim is reflected by the item.
iii) Avoid items that can be correctly answered on the basis of good general knowledge.
iv) Do not suggest the length of the answer by the length of the blank space.
v) Avoid confusion by not having too many blank spaces. Too many blank spaces can make the question unintelligible and thus unanswerable.
vi) Attempt to formulate the question so that the missing information is at the end of the sentence. This insures that the pupil will read the question before he answers it.
vii) Formulate the question as directly as possible. Do not forget that factual knowledge is being tested.

8.2.3 Pairing

In pairing, the pupil is expected to pair a fact in one column with a fact in a second column.
Example:

a) Waterloo	i) Napoleon Bonapart ii) Metternich
b) Epistles	i) Matthew ii) Paul
c) The story of an African farm	i) Langenhoven ii) Schreiner iii) Brink iv) Small

It is obvious that this format also tests factual knowledge, i.e., names, dates, definitions, etc. Within limits, the greater the number of facts from which one must be chosen, the more "trustworthy" the question's discriminative value becomes. Teachers often think it is effective to arrange the different subsections of a subject haphazardly in a pairing test. This leads to confusion rather that to making the test more difficult. Try to order (organize) the questions in a structural sequence; for example, it would be wise in history to start with dates, then follow with names, then battles and then peace treaties. This helps the pupil order (organize) his input. Teachers often deliberately confuse pupils by including irrelevant inferences (distracters or incorrect choices).
Example:

Renaissance	a) Charlemagne b) Pope Pius VII c) Michelangelo d) Bellini e) 1453

This is generally a bad strategy, especially for average achievers, and the test results are often "untrustworthy".

8.2.4 Multiple-choice questions

Multiple-choice questions are generally accepted as the most effective objective test format. The item basically consists of two parts:

- the *root* [*stem* in American measurement vocabulary] in which the problem is formulated; and
- a list of choices (distracters) among which one is the is the correct (key) answer.

There are many ways in which the root can be formulated. The most popular form is an uncompleted statement that is completed by means of the answer that is keyed as correct. Many inexperienced teachers prefer to formulate the root as a question. In a multiple-choice question (with the key response and distracters), the pupil indicates the correct answers.
Example:

Question:
What is the most popular South African wood for making furniture?
a) Birch
b) Poplar
c) Stinkwood (key)
d) Pine

Uncompleted sentence:
A traditional South African wood for making furniture is _____
a) Birch
b) Poplar
c) Stinkwood (key)
d) Pine

Multiple-choice questions can be used for testing any kind of aim in any school subject. Everything depends on the skill of the examiner to assess aims such as facts, vocabulary, cause and effect, insight, understanding, analysis, application, interpretation and appreciation. However, it is not an easy way of evaluating the pupils' achievements. In fact, it often takes twice as long to construct an effective multiple-choice examination as it takes to mark an essay examination. The effectiveness of this format depends largely on the choice of effective distracters that will force the pupil to marshal all of his knowledge, insights and learning

inputs to choose the correct (keyed) answer. It is for this reason that item banks devote so much time to multiple-choice items and especially to the empirical research necessary to determine the discriminative value of the items.

The following advice can be considered when constructing a multiple-choice examination.

i) The root should be in the form of a carefully considered problem and not merely the introduction to the keyed answer and distracters. Distracters must not be a series of facts that have no relationship to each other; for example, a poorly formulated root would be:

Natal is our garden province because
a) there are many parks
b) the National Park Board is located in Natal
c) sugar-cane is cultivated there
d) a great many vegetables are grown there
e) the landscape is generally green

A better way to formulate the same item is, for example:

Natal is considered to be our garden province because of –
a) its rolling landscape
b) its strict nature conservation
c) its luxuriant and varied vegetation (key)
d) its popularity as a holiday resort
e) its variety of parks

ii) The distracters ((i.e., the incorrect possibilities) must be acceptable alternatives for the key (correct answer) to the question or problem that is formulated in the root. If this strategy is not followed, any intelligent pupil can eliminate distracters as possible answers and arrive at the correct choice by careful reasoning.

iii) Do not suggest the correct (keyed) choice by formulating the distracters differently, e.g., by making them shorter or longer. If one of the choices is formulated differently and it is also the keyed choice, then the validity of the item becomes suspect. Teachers sometimes use this strategy to trap pupils

who guess, i.e., by making one of the distracters conspicuous in the expectation that those pupils who do not know the answer will choose it as the key.

iv) Do not make the root of the question excessively long in the belief that it will raise the level of complexity or in the belief that the discriminative value is increased. The following example illustrates this error:

"In the history of the western world, the university, next to the Church, is one of the oldest organized social institutions. Since the establishment of the precursors of the modern western university during the 12th century, the university has had a long history of development. Notwithstanding the fact that universities were generally aloof of the great streams of thought (e.g., during the Renaissance, the period of Humanism and the period of the Reformation they generally remained outside the stream of renewal) they did undergo change during this period of time. The modern university is the product of a process of gradual change and development as a result of a variety of pressures such as philosophical directions, scientific discoveries, wars, and political and religious perceptions."

The important task of the university is

a) to train high level manpower (key)
b) to provide community service
c) to provide pupils with opportunities to study further
d) to raise and consolidate the status of teachers.[•]

In this example, the problem raised in the root is not explained in the detail given of universities and, therefore, could have been left out completely. The quotation serves only to confuse the pupils and waste their time.

v) Insure that items are formulated correctly and always maintain the standards of good language.

vi) Vary the placing of the key among the distracters. If the correct answer is always a) or d), pupils will soon discover the pattern. In the same way, avoid a set sequence in placing the

[•] Louw, J. B. Z.: Owerheidsbeleid en –administrasie van universiteite in Suid-Afrika, p. (xi)

key for different questions, e.g., first a), then c), then b), then a) again, following the rest pf the pattern.
vii) Avoid distracters that indicate that the answer is not contained in any one of the possible choices. To follow the example above (iv): if one of the distracters states: e) None of the above, the candidate is fully entitled to reason that research is the most important task of the university while the examiner meant a) to be the key. The same applies to a distracter that considers a) and b) to be correct in the same example, an answer that in fact is correct but that is not clearly an answer to the problem stated in the root.
viii) State the problem positively. One is inclined sometimes to ask what a certain matter is *not*, thereby, in fact, concealing the problem (aim) to a certain extent; for example:

A voter cannot be identified by means of his
a) driver's license
b) credit card
c) check book
d) passport
e) identity document

It is obvious that the pupil will experience many difficulties in arriving at e). If the problem had been stated positively, the aim would have been much clearer, e.g.:

A voter can only identify himself at the voting booth by producing his
a) driver's license
b) credit card
c) check book
d) passport
e) identity document (key)

ix) Develop items in accordance with your teaching aims. The items must therefore test those matters you have in mind, be it knowledge, insight, application, interpretation, etc. For this reason the examiner must be critical of his own attempts. He must test the validity of his items in in the classroom by determining their discriminative value. Ask the following

questions. What is the degree of difficulty of the item? Is the item directed at teaching and learning aims? How many pupils gave the correct answer? Which pupils gave the correct answer? Build up an item bank for yourself on cards on which you indicate the discriminative value of each item. After a few years of experience the teacher collects considerable information in this way and develops a great number of valid test items. This achievement in itself imparts a feeling of professionalism.

x) Practice writing items. It is a time-consuming process that cannot be left to the day before a multiple-choice test is to be used.

xi) Make very sure that the answers you have chosen as correct are in fact correct. Nothing causes greater embarrassment than controversy about the accuracy of an answer when discussing the results of the test with the class.

It is generally accepted that very few student teachers were ever confronted with well-constructed and valid multiple-choice test or examinations during their school, university or college careers. Nobody advocates this format is the only one that should be used exclusively. However, it is a format for testing the realization of teaching aims in specific circumstances quickly and effectively—if the teacher has access to valid items that he has developed himself or that are available by means of an item bank.

8.2.5 The fill-in test or examination

In essence, this strategy is an ingenious variation of the traditional forms of evaluation, especially the essay question. It also combines certain objective examination methods. It is often used effectively in the natural sciences but it is equally useful in testing subjects in the humanities. The basic rationale is that the pupil fills in his answer in the space provided by the examiner. The length of this space is determined by the examiner after he has developed the marking rubric. He allows for an additional quarter or third of the space for the pupil to use for his answer. The aim is to force the pupil to answer the question instead of writing down everything he knows about the topic. The teacher varies the question (items) according to his aims. The greatest advantage of this method is that

it forces the pupil to consider carefully, to sift and order the facts related to the question before attempting his answer. Short question items generally coincide with what was described regarding the objective tests.

When constructing a fill-in test, the following suggestions can be considered.

i) The level of difficulty of the questions must not exceed the level of achievement of the pupils. In this kind of test the examiner is inclined to ask questions or parts of questions that do not necessarily indicate what exactly is asked.

ii) Teaching and learning aims are especially important because the length of the answer is limited. The examiner must determine what the length of the answer must be. Remember, some pupils have large handwriting; others have smaller. The exact determination of what must be tested is just as important as it is with multiple-choice items. Therefore, pupils should be taught how to write this kind of examination.

iii) Include strategies in your teaching that facilitate this kind of examination; fo example, models, sketches, tables, schemes, comparisons, etc.

iv) Fill-in questions focus easily on explanations, motivations, causes and effects, etc. Formulate the short questions to the point and apply the criteria for fill-in questions.

v) Be very accurate when awarding marks. Because the answer is expected to be short and to the point, the inclination is to award too many or too few marks for the answer. The level of difficulty is especially important. If the question implies ten facts in the answer spaces (e.g., dates) and the pupil can earn only five marks, or on the other hand, twenty marks, the teacher is held to account.

vi) Avoid the temptation to mark negatively, i.e., to subtract marks from a small total for mistakes made in a large number of facts. If a pupil earns only five marks for twenty facts and after five mistakes and earns nothing for the question, he becomes dissatisfied and frustrated.

vii) Give clear and unambiguous instructions (directions).

viii) Do not overestimate the pupil's skills in answering questions briefly and to the point. Be careful not to include

a question that you yourself had difficulty answering in a set number of words. If you do, you overestimate the pupil's ability. For this reason class exercise tests are very necessary.

In conclusion, the following observations can be helpful. Remember that all examination procedures are a part of your teaching strategy. If examinations and test do not contribute to improving the child's learning achievements and your own teaching, then, educatively speaking, they are being used in vain. For this reason it is *always* imperative, whatever the circumstances, to discuss the examination and the pupil's answers by means of the marking or scoring rubric as soon as possible after the test or examination. This is generally the most fruitful learning situation one can organize because good and poor achievements are dealt with in a singularly receptive atmosphere. It is also the occasion to deal with general class as well as individual pupil shortcomings and mistakes. A pupil is given the opportunity to ask questions about his answers and to receive a clear reply. He is even given the opportunity to raise objections about his marks. The teacher who is reluctant to place himself at the disposal of his pupils for these purposes is generally unsure of his testing procedures as well as the validity of his testing methods. It creates trust on the part of pupils if their test or examination is openly discussed with them and they are given the opportunity to compare their answers to the norm that is given in the marking or scoring rubric.