THE DEVELOPMENT OF A TEST OF SCIENCE SKILLS AND PROCESSES

Submitted in partial requirement for the degree of
Master of Education at the University of Waikato,
Hamilton, New Zealand, by

Dennis Michael McGrath

February, 1973

ACKNOWLEDGEMENTS

The author wishes sincerely to thank the following for assistance in this research project:

the New Zealand Educational Institute for a grant-in-aid towards the development and testing of the Science Skills and Processes Test;

the Principals and staffs of Fairfield College and Fairfield Intermediate who cooperated in the preliminary testing; the Principals and staffs of Melville Intermediate, Melville High, Cambridge Intermediate and Cambridge High Schools for their cooperation in the final testing programme;

Professor P.S. Freyberg for his thorough and most helpful criticisms and for his supervision of this study;

Mrs. E.D. Hartstone who most ably typed this final copy;

and to my wife, Judy, who with patience, understanding and careful reading of draft copies made the writing of this dissertation much easier than it might have been.

CONTENTS

ABSTRACT

This study reports the development of a test of science skills and processes for use in Forms 2, 3, 4. The test was constructed to meet two of the objectives of the Form 1-5 syllabus.

The test characteristics, mean 50.60 per cent, standard deviation 15.43 per cent and standard error of measurement 3.09 were calculated from a sample of 341 pupils in Forms 2, 3 and 4. The test-retest reliability was 0.75 and Kuder-Richardson internal consistency 0.77. Two subtests, one related to skills, the other to processes had means of 48.31 per cent and 50.50 per cent respectively. The standard deviations were 17.48 per cent (skills) and 17.00 per cent (processes) and standard error of measurement 1.79 (skills) and 2.10 (processes).

It is suggested that the test could be used to assess achievement and diagnose weakness in the attainment of the science skills and processes objectives of the Form 1-5 syllabus. Tables for converting raw scores to T-scores are provided in the Appendices.

CHAPTER 1
INTRODUCTION

The purpose of this dissertation was to develop a test of science skills and processes for the New Zealand Forms 1-5 science syllabus. The middle levels (i.e., Forms 2, 3, 4) were selected for testing as they were considered of particular importance, being around the transition from intermediate to secondary school. The complete list of seven objectives for the syllabus is discussed in Chapter 2, but the objectives relevant to this study include the development of skills appropriate to science and the application of scientific method. To give some rationale for considering skills and processes of such importance to warrant the development of a separate test, it appears necessary to consider first what is meant by science.

Science is concerned with man's knowledge of his physical environment and with the methods he has developed to understand and control it. In Schwab's (1964) terminology, science has both a substantive and a syntactical structure. The former structure is concerned with the content of the discipline and the latter with the processes and methods of validation. Thus science is not just a set of facts to be learned or even a list of problems to be solved. Science can be regarded as an attitude of learning using inquiry skills. These inquiry (or science process) skills are the concern of the Science Skills and Processes test (SSAP), the development of which is reported in this dissertation. Skills such as measuring, using numbers, hypothesising, experimenting, inferring, predicting, classifying and comparing are subsumed under the rubric of science skills and processes. Items designed to measure such skills as those listed above form the basis of the 40 item, multiple-choice, SSAP test.

Objectives concerned with skills and processes are more difficult to evaluate than objectives concerned with knowledge and its application, hence the present study was not only concerned with the development of the SSAP test but attempted to focus the attention of teachers on the skills and process objectives. This focussing attempt was, in some measure, realised, for when the original blueprint (Appendix A) was returned from the science educators chosen to assist in validating the student behaviours, comments from teachers indicated that the blueprint was useful in the classroom for evaluating student progress.

There has been an increasing awareness in New Zealand of the importance of science skills and processes (see for example Flynn and Munro, 1970, or Phillipps, 1971, or even the 1971 Draft Syllabus Science: Infants to Standard 4), but there remains a lack of published experimental work in New Zealand in methods of evaluating such objectives. The SSAP test, which was an attempt to fill this apparent lack, was developed to be used from Form 2 to Form 4 and consequently the content of the questions in the test was not derived from the syllabus but from the discipline of science at large. "Content-free" thus refers to the freeing of the subject matter of questions from syllabus restrictions.

Incidental to the development of the SSAP test were attempts to ascertain whether sex differences existed in science achievement at these class levels and to compare the science concept development of students in New Zealand with King's (1963) English study.

The number of students involved (N = 339) in this study would suggest that the norms reported (Table 10 and Appendix E) could best be regarded as tentative. However, the blueprint used in constructing the SSAP test (see Appendix B) may well serve as a model for further development in evaluating such student behaviours as "the development of an understanding of the process of science" (Science for Forms 1 and 2, 1967, p.3).

CHAPTER 2

LITERATURE REVIEW

The literature review falls into four major fields: science skills and processes, science process tests, science concepts and sex differences on science tests. Each of these fields will be discussed separately to lend some coherence to the diversity of research cited. The final section in this chapter deals specifically with New Zealand research related to science skills and processes and the Forms 1-5 science syllabus.

SCIENCE SKILLS AND PROCESSES

Any review of literature related to science education and curriculum development in the last ten to fifteen years soon unearths objectives concerned with "scientific method", "science skills", "methods of science", "scientific thinking", "problem-solving", "critical thinking" and many other similarly expressed notions. When Dressel (1960, p.45) suggests that these objectives really mean much the same and may be used interchangeably "without undue confusion" it is a source of concern that some attempt has not been made to at least communicate these objectives more succinctly to teachers.

These objectives are not peculiar to recent science curriculum projects. Dewey, for example, had suggested that "scientific method is not just a method which has been found profitable to pursue in this or that abstruse subject for purely technical reasons. It represents the only method of thinking that has proved fruitful in any subject..." (Dewey, 1910, quoted in Smith, 1963, p.210).

The analysis of scientific method by Dewey undoubtedly lead to the statements of the logical steps that scientists supposedly follow. These were perpetuated in school science text books for many years. These steps were usually paraphrased as:

1. Identification and statement of a problem
2. Suggesting hypotheses
3. Obtaining evidence to test hypotheses
4. Assessing the validity of the hypotheses
5. Revising the hypotheses (if necessary) in the light of evidence obtained.
6. Applying conclusions to new and similar problems.

Unhappily, statements such as these were the only emphasis on scientific method in science courses, the major emphasis until comparatively recent times being on the acquisition of knowledge. This lack of emphasis on scientific process is considered unfortunate because, as Dressel (1960) suggests, methods must be taught and not picked up incidentally. One would presume that the student being "taught" must be actively involved in the processes.

Lewis (1965) in a review of the objectives in science education points up the changing emphases of science curricula. He surveyed the objectives in science teaching from the 1930's until the early 1960's and reported a number of trends. These included an increasing awareness of science as part of our culture (one would suspect an even greater appreciation of this aspect today with an increasing understanding and awareness of such concepts as environment and pollution); the recognitiion that there were different levels of development in children and students, and thus objectives should differ at different levels of the schooling process; and, lastly, an increasing emphasis on skills, and on the application and evaluation of information.

This latter emphasis, he suggested, indicates that higher order skills are becoming of more concern and these may well be tested by using objective tests.

The skills and processes mentioned in the review of Lewis are still stated at a general level. One has to look at some of the science curriculum projects from the United States in the early 1960's to appreciate the concern for a redefining of the six-step model of scientific method proposed by Dewey. The lock-step process does not recognise the role of creativity in scientific endeavour and most recent writers (e.g. Nay, 1971 or Gagné, 1966) suggest that there are more adequate descriptions of what scientists do. To illustrate this point Gagné (1966) refers to a meaning of process that centres upon the idea that what is taught to children should resemble what scientists do, that is, the processes that they carry out in their own scientific activities. He lists such activities as observing, classifying, measuring, infering, making hypotheses and experimenting. The linkage of these processes to the Dewey scientific method should be apparent. Gagné's theory of the psychology of learning has influenced the course, Science - A Process Approach which was developed initially for grades K-6 in the United States. He writes of the process approach that

> "...it rejects the 'content approach' idea of learning highly specialised facts or principles of any particular science or set of sciences. It substitutes the notion of having children learn generalisable process skills which are behaviourally specific, but which carry the promise of broad transferability across many subject matters ... the point of view is that if transferable intellectual

processes are to be developed in the child for application to continued learning in sciences, these intellectual skills must be separately identified, and learned, and otherwise nurtured in a highly systematic manner." (Gagné, 1965, p.4)

In the development of the test of science skills and processes (SSAP), representative skills have been separately identified and behaviourally stated in an attempt to more effectively evaluate such learnings.

Schwab (1963) identified what he called two sides to the ability to use the skills involved in science. One side he referred to as the constructive mode, which seems to refer to a student's ability to contribute to scientific activities. Perhaps this could also be referred to as a creative aspect. The second side or analytical mode referred to a student's ability to understand and judge the soundness of science processes when they are reported. In an objective test such as the SSAP test there is no real opportunity for the individual to use the constructive mode. Instead situations are presented from which the student must select an alternative explanation or answer. It would seem, then, that the SSAP is dealing with Schwab's analytical mode and that only by putting a student in a practical situation might one test what appears to be a more open-ended constructive mode.

Brown (1968) also commented on the number of apparently similar ways science skills were stated. He gathered together a committee with the express purpose of defining what was meant by science processes, preparing instructional materials based on this definition and developing methods to evaluate pupil growth in science processes. Brown's article essentially records a hierarchy in a somewhat analogous manner to Bloom's (1956) taxonomy of educational objectives. Five major levels were identified, based mostly on data collection and use. The levels were :

1. Application of generalisations to new situations
2. The collection of data
3. The analysis of data
4. The synthesis of data
5. The evaluation of data

Many behaviours, specifically stated, were given as representative of these levels. For example, the identification of a problem and the ability to delimit the problem are subsumed under data collection; hypotheses formation and skill in testing hypotheses are listed under the analysis of data. Although this comprehensive list is most useful in designating types of student behaviour, no information on the evaluation methods developed was included in the article. One would assume that, having stated specific behaviours for teachers, evaluative procedures might include checklists of how well .these behavioural outcomes were attained.

Amos (1970) is another writer who sought to clarify objectives. He pointed up the difficulty of assessing teachers' understandings of what objectives meant in a survey of English 0-level biology teachers. Amos composed a list of some 19 objectives for G.E.C. 0-level biology courses and asked teachers to rank these. The objectives were categorised according to Bloom's taxonomy and ranged from knowledge of specific biological facts (Bloom category 1.12) to higher order skills such as ability to interpret experimental data and draw reasonable generalisations there from (Bloom 2.30, 5.30), and ability to formulate a hypothesis from evidence available (Bloom 5.30). He discovered that teachers found it difficult to separate objectives concerned with scientific method from each other but grouped them as "scientific thinking". Although teachers stated a need to include these scientific method objectives in their teaching, it was evident that examinations influenced the teachers' priorities with regard to objectives. As a result of examinations teachers tended to rate knowledge and understanding objectives more importantly than the skills objectives.

A new Australian Science Education Project (A.S.E.P.,1970) aims to provide experiences to develop skills and attitudes important for scientific investigation. Such skills aimed at enabling children "to inquire efficiently and solve relevant problems" include observing and ordering observations, determining patterns and relationships, formulating problems, obtaining information relevant to a problem (both from library and experimental sources) and interpreting findings critically. Once again the implication seems clear, that behaviours must be stated in terms that are observable in the students undergoing the instructional programme.

One final report, that of Nay and associates (1971), is included in this survey for it seems to hold particular promise for science educators concerned with the development of science processes and skills. The Nay project was concerned with the development of a Junior High School process-approach science curriculum. Nay identified an inventory of processes used in scientific inquiry. He excluded the complex mental operations such as communication and critical mindedness and concentrated on processes as such. Four phases in an investigation were identified. These were:

1. Initiation (which includes identification of a problem, predicting, hypothesising...)

2. Collection of data (the processes of observing, identifying limitations...)
3. Data processing (the organising and representing of data...)
4. Conceptualisation of data (the interpretation of data...)

A fifth phase that of "Openendedness" was suggested as emerging at each stage of an investigation and should emphasise the fact that the steps involved in an investigation are not completely independent of each other. Some fourteen separate processes and a number of subprocesses are represented in these four major phases. Some of these processes have been indicated above.

A major contribution of the Nay research lies in the firmly held idea that process and product in science are not separate entities but are interrelated and should be taught as such. There has been a growing concern in the literature that "process" is being emphasised to the detriment of "content" (see, for example, Atkin, 1966, Ausubel, 1963, or Roth, 1971) and that process and content should not be polar entities in any reasonable science curriculum. A balance between the two is necessary and is called for.

This balance places an emphasis on the teacher's interpretation of what is meant by science processes, for as Hurd (1969, p.16) writes "...it is undoubtedly true, a teacher's concept of what science is influences not only what he teaches but how he teaches." If the teacher's understanding of science processes are not congruent with current interpretations then instructional outcomes will not be representative of science.

Carey and Strauss (1971) report an analysis of teachers' understanding of the nature of science. They cite studies which indicate that there is little difference in the understanding of the nature of science between science teachers and non-science graduates or high school students. These disturbing findings should bring serious concern to those involved in science education if Hurd's suggestion that a teacher's understanding of what science is about influences his teaching of the subject is upheld. One would hope that teachers would have a better understanding of the nature of science than students and that if this understanding appears to be as lacking as Carey and Strauss indicate then efforts must be made to give teachers an adequate concept of what science is about and how inquiry is conducted scientifically.

In summary, science processes and skills have been described in many terms by different writers. Objectives relating to science processes and skills have been regarded as important for a great many years although only in recent years has there been a concerted attempt to involve students in these processes. This involvement has been facilitated by suggestions for teachers of writers such as Gagné, Brown, Schwab and Nay, who have attempted, in often analogous ways, to break up the complex intellectual set of behaviours that comprise the "scientific method" and present them in simpler, observable activities appropriate to the psychological development of the student. Piaget (1971), in writing of the development of such capabilities as combining hypotheses and experimentally verifying hypotheses states strongly that it is a duty of the schools to emphasise the importance of research and discovery methods in the teaching of the physical sciences rather than relying on mere repetition. The justification for inclusion of objectives relating to science processes has been upheld and recognised. Is it appropriate for this Form 2 - 4 age level and how have such processes been evaluated? These questions will be dealt with in the following section.

SCIENCE PROCESS TESTS

> "...at about 11 - 12 years of age, there begins a fourth and final period of which the plateau of equilibrium coincides with adolescence. This period is characterised in general by the conquest of a new mode of reasoning, one that is no longer limited exclusively with objects or directly representable realities, but also employs 'hypotheses', in other words, propositions from which it is possible to draw logical conclusions without it being necessary to make decisions about their truth or falsity before examining the result of their implications" (Piaget, 1971, p.33.)

To Piaget, then, there is at the particular age level that the SSAP test is aimed at, the beginnings of a psychological readiness amongst the students. Between the ages of 11 - 12 and 14 - 15 students "spontaneously acquire" (Piaget, 1971, p.51) all the intellectual equipment needed for experimentation. If this is so, then instruments to measure the attainment of these processes should be available to enable teachers to evaluate science process objectively. Welch and Pella suggested that, to the contrary, most of the evaluation, up to the present time, has been concerned with the measurement of content achievement while "little has been accomplished with regard to the objective of understanding science processes." (Welch and Pella, 1968, p.64.)

One of the earlier reports of an attempt to test scientific thinking was by Burmester (1953). She developed a test which did not assume mastery of science content or scientific terminology and which she used to test high school seniors. She reported a correlation of 0.64 with grades the students obtained in natural science courses.

The Secondary Schools Examinations Council in Great Britain was also concerned, in the early 1960's with ways of measuring student attainment other than by conventional written papers. Examination Bulletins were published and circulated. These often presented very open-ended types of questions which aimed, for example, at testing 'scientific thinking'. For example, a problem situation was set, alternative explanations suggested and the student was asked to discuss each possibility (or hypothesis) and suggest the most likely explanation; alternatively, the design of an experiment to test a hypothesis was called for. One is reminded of Schwab's (1963) two modes of using science skills by these types of questions: in the first example it appears an analytic mode is called for and for the second example more creativity is required and hence a constructive mode may be employed. These types of items were used by Flynn and Munro (1970) in New Zealand schools and this work is described below. The Secondary Schools Examinations Council also generated the collection of papers by Eggleston and Kerr (1969). Of special interest is an attempt by Fox (1969) to assess scientific ability through special studies or investigations carried out by individuals or groups of students. Abilities such as analysis, deduction and inference were found to be more readily assessed in a special study than by other forms of test. The age group of the students appears to be rather higher than that the SSAP test has been developed for.

Welch and Pella (1968) report a study concerned with the development of an instrument for making an inventory of knowledge about the processes of science. They listed a number of elements of scientific processes from a variety of philosophical viewpoints on science, validated these by presenting the elements to fourteen research scientists and composed a 150 item instrument - the Science Process Inventory (SPI). Students are presented with a variety of statements and simply asked to express agreement or disagreement with each statement. The student's response is assumed to indicate his knowledge of the idea contained in the statement. The instrument was designed for secondary school pupils but it may be simply measuring how well students had been taught correct responses to statements and not how much practice the student had been given in a practical, problem-solving, experimental approach to science teaching. Welch and Pella report correlations (product-moment coefficients) of 0.61 - 0.64 with a test of mental ability, i.e. within the usual range of attainment tests of various kinds.

A very good example of a test of science skills and processes was reported by Tannenbaum (1971). This example is one of the very few documented for the age level grades 7, 8 and 9. Indeed, Butzow and Sewell (1972), in describing the use of the Test of Science Processes (TSP), refer to the TSP as the only reported development for this age range. The TSP identified eight science processes: observing, classifying, comparing, quantifying, measuring, experimenting, inferring and predicting. These processes are the same as those developed in the Science - A Process Approach (S-APA) curriculum project but Tannenbaum suggests that his test is unique because it deals with an age group beyond the S-APA project. The use of science processes is, of course, not confined to any one test or project. It is the student behaviours stated in the blueprint which gives a test its unique characteristics.

Tannenbaum validated a list of science processes against the opinions of science educators and produced firstly a blueprint of 53 behaviours and then a 98 multiple-choice test with 35 mm colour slides and photographs for many of the items. The test gave a Kuder-Richardson formula 20 reliability overall of 0.91 and reliabilities on the eight subscores from 0.30 to 0.80. Reliabilities are not reported for a test-retest correlation method. An interesting feature was an attempt to establish some evidence of criterion - related validity by asking a teacher to rate a small group (N = 35) of pupils on a 0 - 9 scale for each of the eight processes in the test. These scores were correlated with the students' scores on the TSP. The evidence, although equivocal, did indicate some degree of criterion-related validity.

Butzow and Sewell (1972) have used the TSP to measure pre- and post-test scores on the subcategories in an introductory physical science programme. They suggest a possible hierarchy from simple to complex:

1. observing, classifying, predicting (most simple)
2. comparing, inferring, experimenting,
3. measuring,
4. quantifying. (most complex)

The latter levels of this hierarchy are included in the skills objective in the SSAP test under investigation in this study, while the earlier levels are included under the processes objective.

There is an increasing use reported in the literature of the S-APA material. A typical example is that of Raun (1967), who measured the changes in cognitive and affective behaviour brought about by students (grades 5,6) using some of the strategies of science. Using some of the strategies outlined in S-APA, after five months instruction, he found no consistent pattern of behavioural change among the grades and, in fact, grades 5 and 6 showed regressive tendencies. This, to Raun, supported the argument that there is rather slow development of science process beyond grade 5. One wonders if this is not a plea for earlier intervention in science education!

Morgan (1971) raises a number of questions and suggests a different way of resolving the problem of evaluating science skills. He cogently argues that the process approach is concerned with a set of "unobservable activities that take place in the mind (processor) of the student". (p.77.) If one attempts to evaluate observing, measuring (input processes), there must be things to observe and measure. Linguistic descriptions, he argues, give the game away. If the input relies on the printed word then an unsatisfactory output may be simply a reflection of lack of communication and not inadequate internal processing. The solution is to use tests based on students' interaction with materials and equipment. This often presents difficulties through the amount of time required, so Morgan's resolution involved the use of 8 mm film loops. These are shown to the pupils who record responses for checking. This is an interesting, albeit more expensive than pencil-and-paper, innovation in testing science processes.

In summary, it is clear that although science processes are capable of being developed at the Form 2 - 4 age levels and that these processes have formed objectives of many science programmes there has, until very recently, been little effort to consciously evaluate these processes and skills. There are standardised tests (e.g. Sequential Tests of Educational Progress - Science) but often, as in the STEP tests, these instruments incorporate subject matter as well as science skills. This inclusion of content makes the test unusable for a New Zealand science syllabus. Tests such as those of Tannenbaum (1971) and the SSAP test serve important functions of providing blueprints of behavours that may assist teachers in selecting appropriate objectives for their classes and provide both diagnostic and attainment evaluative functions.

SCIENCE CONCEPTS

Ramsey and Howe (1969 b) in an analysis of research on instruction in secondary school science in the United States, report that papers directed towards knowledge of content in science by far outnumbered those which attempted to measure higher order skills. Most of the research, however, did not deal with students' knowledge of science concepts as such but rather with evaluations of science curriculum projects such as the Biological Sciences Curriculum Study (BSCS), the Physical Science Study Committee (PSSC) and the Chemical Education Material Study (CHEM Study). The same authors, in another report, suggest that concept studies were more concerned at finding relationships "between the child's level of maturity and the understanding of a particular concept". (Ramsey and Howe, 1969 a, p.32.) Such concepts as the particle nature of matter, relativity, force, acceleration seem most frequently used as examples.

In fact, there have been few examples of research related to the development of a range of scientific concepts in children. King's (1961) exhaustive study of the responses of some 1,200 children to seventy questions dealing with basic scientific concepts such as length, time,direction, volume, weight and shadows remains as a landmark because of the range of concepts tested. King's study concentrated on an age range of children from 5 to 12 years which is below the age for which the SSAP test was developed so this work will not be discussed further, except to comment that it was from the data obtained in 1961 that King's second study, reported in 1963, developed.

King (1963) described the responses of some 800 children from age 5 to 17 years to twenty science questions. The questions will be described in some detail for the same test was used in the research being reported in this dissertation. The questions were reproduced on a standard form and student responses were objectively checked. The questions ranged over concepts of :

Time (estimate a 15 second interval)

Volume (if a stone is dropped into water will the level be higher or lower than before?)

Shadows (predict where a shadow will fall from a light source)

Rotation of cog wheels

Physical (are all things that move alive?)

Shapes (what is shape formed when a cone is cut vertically?)

This test was administered together with a verbal test (Simplex Junior Intelligence Scale) and a nonverbal test (Raven Progressive Matrices, Revised order, 1956).

There were significant differences (0.01 level) in mean scores for both Simplex, Matrix and to a lesser degree the Science Schedule (0.05 level) between boys and girls. There were no significant sex differences for the intercorrelations of the four variables (King included age as a variable). These intercorrelations are reported in Table 1.

TABLE 1

	Simplex	Matrices	Science Concept Schedule
Age	0.682	0.570	0.504
Simplex		0.767	0.677
Matrices			0.573

(From King, 1963, p.243)

The Science Schedule is reported as having a reliability of 0.604.

An analysis for each of the science questions was also carried out. This analysis was by age, verbal and non-verbal score. Some of the questions, particularly numbers 17, 18, 19, 20 (see Appendix D for the test) appeared to present considerable difficulty, since less than 60% of the children, at whatever age, gave correct replies to these questions.

King concluded that there appeared to be a high correlation between scores on a verbal intelligence test and scores on the Science Schedule. The rate of growth of correct science replies compared with achievement on the verbal and non-verbal tests was almost identical for boys and girls so that any differences in science attainments of boys and girls observed at later ages may be due to "non-academic influences" such as teaching, interest and self motivation. (King, 1963, p.247.)

Rowland (1965) and Klein (1971) reported investigations of differences in science achievement between children from different socio-economic groupings. In the former study pupils of high socio-economic status achieved better than lower groups when science background and intelligence were controlled. Klein reported significant differences (0.01 level) between the mean scores from three different socio-economic samples on a test of science concepts and an intelligence test. The children used in these studies were somewhat younger than those for whom the SSAP test was developed but a consideration of socio-economic status of parents was included and is reported in Chapter 4 of this study.

SEX DIFFERENCES IN SCIENCE ATTAINMENT

It has been mentioned above that King (1963) identified a sex difference between boys and girls scores on the science concept test (0.05 significance). A brief survey of recent literature suggests possible reasons for differences in science achievement. Rowell (1971) for example, used a group of 234 children in South Australian schools and gave them a test of 10 questions which required constructed answers. Some of the questions bear close resemblance to those problems used by Inhelder and Piaget (1958). Significant differences $0.001 < p < 0.01$ were detected between the mean score of boys over the girls' mean score. Although there was this significant difference overall, some classes of children did not show significant differences. Rowell therefore devised a rating scale to measure teacher expectations of science achievement. Although the number of teachers involved was small (N = 12) Rowell tentatively suggested that teacher expectation may be playing a part in the differential science achievement of boys and girls and this may account for some classes not showing the significant differences between girls and boys that other classes did.

Maccoby (1966) rather suggests a difference in intellectual functioning as a result of sex-typed personality traits, such as aggression - independence in boys and conformity passivity in girls. A dependent conforming person is passive, waiting to be acted on by the environment. If we stress a discovery approach and insist on pupil involvement in science processes we may well be penalising girls in a science programme. Certainly there may be other factors. McMurray (private correspondence) in a survey of some 1200 Form 1 teachers for the New Zealand Educational Institute reported that some 46.83% of teachers found the syllabus content of great interest to less than half of the girls compared with only 28.41% of boys in the same category. Similarly, teachers checked a category "more than 50% of group actively participated" less frequently (65.0%) for girls than for boys (80.5%). In New Zealand schools girls do not appear as interested or as involved as boys in science at Form 1 level.

THE NEW ZEALAND SITUATION

As long ago as 1944 the Thomas Report on secondary education in New Zealand in a discussion of the objectives of general science for secondary schools suggested that one of the ends to be sought should be "...to illustrate by demonstration, experiment and discussion the use of scientific method as a tool of accurate thinking" (The Post Primary School Curriculum, 1944, p.30). In 1963, a Forms 1-4 Science Revision Committee was set up to prepare a new syllabus. This syllabus, in keeping with the recommendations of the Currie Commission, was to exhibit continuity between Forms 1 - 2 and Forms 3 - 4. Until this time the official syllabus (1948) for primary schools was based on a descriptive form of Nature Study and scientific method, which had been evident in the 1929 syllabus, had been de-emphasised. However, Watson (1964) reported that in 1958 nearly half of the intermediate schools had established some physical science programme with, presumably, some concomitant attempt at developing scientific method.

The early 1960's saw an influx into New Zealand of overseas, particularly United States, science curricula. PSSC physics was introduced on a trial basis in 1961 and similarly CHEM study in 1963. Both of these projects laid emphasis on scientific processes and pupil involvement in science and when the guidelines for the Form 1 - 4 science revision were published in 1963 and a draft syllabus published in 1965 it was not surprising to see included as aims

> "the fostering of a spirit of inquiry and a sense of wonder, the development of an understanding of the process of science together with an appreciation of the developing

nature of science..."

Science for Forms 1 and 2, 1967, p.3.

The syllabus for Forms 3 and 4 was published in 1968 and spelled out even more clearly some seven objectives which it suggested, were classified along lines used by Bloom. These objectives were :

1. Knowledge of basic facts, principles and theories
2. Development of basic concepts
3. Application of scientific method
4. Development of skills appropriate to science
5. Development of scientific attitudes
6. Recognition of the significance of science in society
7. Development of a continuing interest in science

(The SSAP test focusses on objectives 3 and 4 and these will be discussed more fully in Chapter 3.)

How do teachers view these objectives? In a small (N = 15) unreported study conducted by the author, Form 1 - 2 teachers in Hamilton were asked to rank these objectives. Scientific method and skills were ranked third and fourth behind concepts and attitudes yet ahead of facts and principles and the other objectives listed above. Duncan et al (1971) reported a national survey which focussed on what the objectives of Form 6 Chemistry were deemed to be and how much emphasis there was placed on different objectives. They listed seven objectives (pp.46 - 47), the objectives of concern for this study being "Skills : ability to communicate chemical knowledge to others, orally and in writing, to set up and use chemical apparatus; to plot and interpret graphs" and "Scientific Method : ability to propose and evaluate methods for solving chemical problems new to the pupil." When the respondents were asked to rank the seven objectives in order of priority, scientific method was listed second to understanding and little weight was attached to skills acquisition.

Teachers both at Form 1-2 and at Form 6 thus appeared to rank an objective relating to scientific method (or processes) rather highly. The problem remains as to how these process objectives are to be evaluated. The Report on 1969 School Certificate Science Examination went some way to indicate possible procedures teachers might use in evaluating science skills and processes. The grid or blueprint set up still used broad objectives, however, and it is not possible to determine exactly the criteria expected to be used for devising questions under such process goals as "ability to use scientific method as defined in the Form V syllabus".

A rather different approach was that of Flynn and Munro (1970). They attempted to evaluate Nuffield Science courses which were being taught in New Zealand Form 3 classes. Course objectives were identified and instruments designed to measure the following abilities were constructed:

Ability to :

> "analyse given data (verbal, tabular, graphical and pictorial), suggest testable hypotheses consistent with given data, devise experiments to test hypotheses, and to extrapolate from given data".

(Munro, undated, private correspondence 1972.)

Problem situations were set up in such a way that to "solve" the problem pupils had to display clear evidence of the mental abilities the tests purport to measure. The problems were along the lines of those suggested in the Secondary Schools Examination Council Examination Bulletins and afforded opportunities for what Schwab (1963) refers to as analytic and constructive modes. The tests were content - free, the only difficulty being establishing marker reliability. To meet this demand a matching-type marking schedule was developed and tests indicated that items in the schedule could be matched to cover 90 per cent of the responses surveyed.

Lastly, Phillipps (1971), in an article describing the "new ways of science", traces changes in the framing of objectives and after describing a content - process grid and methods of evaluating objectives, calls for an attempt "to use and revise techniques which will show the extent to which we are achieving our new goals" (p.25).

SSAP is in partial answer to that call.

CHAPTER 3

DEVELOPING THE SCIENCE SKILLS & PROCESSES TEST

Determining the Test Objectives

"In constructing a new test, the specification of student outcomes to be measured is by far the most difficult task. This is especially true if the desired outcomes are other than recall or application of subject matter content." (Cooley & Klopfer, 1963, p.73.)

In the development of a test of science processes and skills, with objectives other than recall or application, Cooley's and Klopfer's words still hold true today. It was decided that the test would concentrate exclusively on assessing student's ability to use the skills and scientific method outlined in the Form 1-5 science syllabus. These objectives are but two objectives of seven listed and their substance may be adjudged by a consideration of the objective statements.

"3. Application of Scientific Method

This implies the ability to identify a problem, to bring to bear earlier experience appropriate to the problem to formulate explanations and hypotheses, to test such explanations and hypotheses by experiment or other means, to accept, modify or reject, and to draw conclusions.

The achievement of this objective depends, not what is taught but rather on how it is taught.

4. Development of Skills Appropriate to Science

Skills such as the ability to manipulate scientific equipment correctly, to measure correctly, to construct and interpret tables, charts and graphs, to be able to find relevant information from available reference sources, are also included. The development of such skills naturally depends on constant practice in the classroom and laboratory."
(Science for Forms 3 and 4, 1968, p.2.)

It was further decided to go beyond the objective 3 statement and include objectives of observing, classifying and comparing. These latter objectives are commonly expressed in lists of science processes, (e.g. Walbesser, 1963, A.A.A.S., 1965, and Tannenbaum, 1971) and it is suggested these are underlying skills of the ability to identify problems and formulate hypotheses, or the ability to measure accurately and construct tables or charts.

The concentration of the test on science processes and to a lesser extent science skills, meant that the test itself could be independent of the content of the syllabus. The removal of content restrictions meant that the test could be administered to pupils at three different class levels (Forms 2, 3, 4) and that the examples used in the test items could be chosen from any branch of the sciences.

The first major step was the construction of a specification table or blueprint for the test. Essentially this consisted of taking the objectives 3, 4, mentioned above, and writing specific behaviours that might be expected from pupils at the Forms 2 - 4 class level.

An example will serve to illustrate the procedure described above. Objective 3 was broken into a series of partial objectives and desired student behaviours written for each partial objective. The first partial objective "to identify a problem ..." thus generated three student behaviours, namely: The student will show his ability to identify a problem by

(a) recognising that meaningful questions may be asked about a given problem;

(b) rephrasing the given problem in such a way that he can begin to formulate suggestions as to possible answers;

(c) recognising the need for specific techniques to solve the problem.

(The final list of partial objectives and student behaviours is included in Appendix B.)

The statement of science skills and processes was mailed to a number of people, throughout New Zealand, with especial interest in science education. This followed the established lines of determining curricular validity; e.g. Bloom, 1971, p.27, writes of the necessity to "create a committee or other type of consensus mechanism to develop a set of specifications" against which an evaluation instrument can be constructed and validated. The 32 science educators were asked to rank

(a) the importance of the objectives

(b) the relevance of the student behaviour to the stated objective

(c) the clarity of expression of the desired behaviour and

(d) the difficulty of achievement of the behaviour for a 12 - 15 year old student.

Appendix A contains the statement sent to the experts and a copy of the instrument for responding to the statement.

A total of 19 usable replies were received from three broad categories, namely, (a) teachers college lecturers and university science lecturers, (b) secondary school teachers, (c) primary school teachers, science advisers, science curriculum workers. The number of returns in each group was insufficient to warrant intra-group comparisons but the 19 returns did ensure a wide range of viewpoints pertaining to the desired student behaviours.

The responses from the science educators were scored on a simple 1, 2, 3 basis, according to category checked, and were totalled and averaged for the 19 replies (see summary in Table 1). In column 4 the summary records the proportion of respondents scoring the item as too difficult.

A number of arbitrary criteria were selected as guidelines for the inclusion of behaviours for the final test blueprint. These were:

(1) No new behaviours could be added to the final blueprint.

(2) No behaviour would be included if the mean score on adjudging the relevance of the student behaviour, Column 2, was greater than 2.0 (i.e., less than moderately relevant).

(3) No behaviour would be included if the mean score on the expression of clarity, Column 3, was greater than 2.0 (i.e. less than moderate in clarity).

(4) No behaviour would be included if more than 40% of respondents said it was too difficult (Column 4).

With minor rewriting in a few instances some 33 student behaviours subsequently formed the basis of the blueprint. The final statement of processes and skills (see Appendix B) thus consisted of some 33 behaviour outcomes under 12 partial objectives. Two of the partial objectives could not be tested in a pencil-and-paper situation but were included in the blueprint for completeness and logical continuity. On the basis of this blueprint some 80 test items were written for pre-testing.

An unusual feature of the blueprint is that it differs markedly from the more traditional two-dimensional classification. This latter two-dimensional chart lists a content classification along one axis and a behavioural classification along the other. Items are written to give weightings to all behaviours and content areas, though, of course, the extent of the weightings will vary with both the behaviours and the content. The Science Skills and Processes Test (SSAP) relies on no syllabus-bound content and can thus be regarded as one-dimensional. The behavioural classification forms one major dimension and "science", for this was essentially where the content was derived from, is the other, minor, dimension. It was the former dimension which had to be carefully designed and specified in the SSAP.

TABLE 2 : Mean scores of science educators responding to validation of behaviours instrument. See Appendix A for instrument.

Objectives		Col.1	Col.2	Col.3	Col.4
I	a		1.42	1.76	.106
	b	1.26	1.85	1.88	.636
	c		2.00	1.76	.371
II	a		1.79	1.94	.371
	b	1.52	1.42	1.83	.265
	c		1.96	1.94	.583
III	a		1.37	1.67	.333
	b	1.68	1.58	1.78	.280
	c		1.63	1.55	.318
IV	a		1.32	1.73	.212
	b		1.37	1.67	.265
	c		1.86	2.06	.636
	d	1.53	1.63	2.12	.636
	e		1.91	2.12	.557
	f		1.53	1.73	.333
V	a		2.21	2.24	.636
	b		1.53	2.00	.265
	c	1.47	1.63	2.18	.530
	d		1.47	1.65	.424

Objectives		Col.1	Col.2	Col.3	Col.4
VI	a		1.96	1.55	.265
	b		1.79	1.83	.424
	c	2.27	1.79	1.88	.371
	d		1.21	1.61	.053
VII	a		1.96	1.78	.212
	b		1.79	1.44	.053
	c		1.53	1.44	.053
	d	2.24	1.32	1.44	.159
	e		1.96	1.73	.477
	f		1.16	1.38	.000
VIII	a		1.37	1.44	.053
	b		1.32	1.47	.106
	c		1.37	1.44	.000
	d	1.79	1.42	1.61	.106
	e		1.32	1.61	.212
	f		1.58	1.61	.371
	g		1.85	2.00	.689
	h		1.26	1.28	.053
IX	a		1.96	2.06	.557
	b		2.00	2.24	.778
	c	2.21	1.32	1.44	.000
	d		2.06	2.45	.765

Objectives		Col.1	Col.2	Col.3	Col.4
X	a		1.44	1.55	.056
	b		1.10	1.38	.000
	c	1.26	1.42	1.61	.280
	d		1.55	2.18	.410
	e		1.74	1.83	.477
XI	a		1.10	1.28	.000
	b	1.88	1.32	1.61	.265
	c		1.44	1.94	.280
XII	a		1.16	1.11	.056
	b	1.83	1.42	1.64	.235
	c		1.37	1.64	.389

Construction and Administration of the Pre-Test

On the basis of the blueprint some eighty items were written, edited and assembled into two 40 minute test booklets. The items were of a multiple-choice nature with five choices for each and the usual procedures regarding construction (e.g. Brown, 1966, Gronlund, 1971) were followed. At least two items were written for each behaviour in the blueprint.

The pre-test was used for a number of purposes. Firstly to identify defective items, secondly by use of item analysis to choose items for the final test, thirdly to check the time allowed for students answering the questions and fourthly to check ambiguities in instructions.

The pre-test was administered to 178 Forms 2, 3 and 4 children in a Hamilton suburb. The Form 2 children were all at one intermediate school, which was the principal contributing school for the one secondary school the Forms 3 and 4 children attended. The location of the schools was chosen to represent a wide socio-economic grouping and in all cases (two classes at each of the three school levels) the classes were either not streamed or chosen to give a representative range of abilities and achievement.

The tests were administered by the same person for each class. Both tests were administered within a maximum period of three days during the same school week early in September, 1972.

The answer sheets from both tests were combined to give a total score (possible 80) and the items were then analysed. Three factors were considered :

(1) the number choosing the correct answer (hence the difficulty index could be obtained)

(2) the numbers choosing each of the distractors

(3) the discrimination index was determined by a consideration of the differences between the proportions of the top 27% and the bottom 27% of students passing each item (see Furst, 1958, or Anastasi, 1968).

As a result of this item analysis the final SSAP test was constructed using the best items available from the pre-test. For two behaviours none of the items from the pre-test were considered acceptable and items purporting to test these particular behaviours (see Appendix B) were thus omitted from the final test. The items chosen from the pre-test had a mean difficulty index of 0.49 and a discrimination index range from 0.20 to 0.69. The range was necessary to include items considered representative of the 29 student behaviours of the final blueprint (Appendix B).

CHAPTER 4

PROCEDURES USED TO INVESTIGATE TEST CHARACTERISTICS

Description of Sample

The students chosen for testing with the final version of SSAP were selected from Forms 2, 3, and 4 at four schools. The schools were selected to give representation of both a range of socio-economic status and urban - rural background. The two Hamilton City schools were a secondary school and its contributing intermediate school. The schools were in close proximity and were assumed to draw on similar populations for their school rolls. The particular area in which the schools were located was itself representative of wide socio-economic groupings. The two rural schools chosen were located in a small town (6,500 population). In the town there was but one intermediate school and one secondary school so again, the schools were assumed to be drawing on similar populations although there might well be a higher proportion of rural pupils at the secondary school.

The schools were asked to arrange for two classes of students at each of the three levels. These classes were to be representative of a full range of abilities and achievement. A total of 339 pupils in 12 classes sat the three tests used in the testing programme, 161 boys and 178 girls. This number gave at least 100 pupils at each of the three school levels.

Instruments

Three tests were administered to the students and the features of these tests will be described in detail below. Briefly the tests were

(a) the OTIS test of mental ability (NZCER, 1969)

(b) a science concept test developed by King (1961, 1963) and

(c) the final version of the Science Skills and Processes Test.

(a) THE OTIS TEST OF MENTAL ABILITY

The Higher Test Form A was used for the students tested. This test, although not standardised for New Zealand, does provide norms that are within 1 - 2 points of the revised Intermediate Test norms (NZCER, 1969, p.1) and has been designed to test mental abilities in the age range of the students used in this research.

The test was administered, marked, and scores converted to intelligence quotients according to the instructions in the manual of directions (NZCER, 1969). Two of the fourth form classes had sat the OTIS Higher Form A two years previously but the practice effect "would in most cases have disappeared over time" (NZCER, 1969, p.5).

It was also recognised that a practice effect may result from the OTIS to the SSAP test but the instruction manual is again reassuring: "...the practice effect of the OTIS test on other multiple-choice tests administered within one week has been shown to be negligible for the majority of pupils." (NZCER, 1969, p.5). The three tests were administered in all cases over a period of time not more than one week and it was assumed that by sitting the OTIS Test first there would be no practice effect on either the concept test or the SSAP test.

(b) THE SCIENCE CONCEPT TEST

King, in the earlier 1960's tested a large number of pupils in Britain with a number of questions, some multiple-choice, relating to children's understanding of basic science concepts. King's earlier work (1961) reported the growth of children's scientific knowledge with increasing chronological age. This growth was measured by responses to some seventy questions. King (1963) reported an extension of the original study and provided a copy of a twenty-question test dealing with the concepts of time, volume, space, physical and a miscellaneous category.

This test was used in the present study as an indicator of the students' scientific concept attainment. The test was reproduced exactly as reported by King (1963). The test was administered by the writer to eight of the twelve classes, and the administration instructions were clearly indicated to the class teacher in the other four classes so that no ambiguities could arise. The test and instructions may be found in Appendix D.

The total score in the test was used as an indicator of science concept attainment. A sample of the 13 year-old students' scores on the individual questions is included, along with a comparison with King's 1963 data for this age group in Appendix J.

(c) THE SCIENCE SKILLS AND PROCESSES TEST (SSAP)

This test comprised 40 multiple-choice (5 choice) items selected from the pre-test procedures mentioned above. The test was constructed to measure two major objectives science skills and science processes as outlined in the Form 1-4 science syllabus (op.cit.). These two major objectives were tested by 23 items (processes) and 17 items (skills) covering some twenty-nine behaviours (see Test blueprint, Appendix B).

The items selected from the pre-test were edited and in a few instances alternative distractors were provided. The correct choices were randomised and an answer sheet was constructed so that a student's choice alternated between A, B, C, D, E and F, G, H, I, J on alternate questions. This simple alternation provided a built-in check so that respondents could better ensure they were answering in the correct answer space. The test booklet was constructed using many of the guidelines suggested by Thorndike (1971b).

The instructions to the students were printed on the test booklet and the test administrator (which was the author in every case) read the instructions through with each class. A copy of the test and instructions may be found in Appendix C.

The test had a time limit of 40 minutes which, from observation both of the classes sitting the test and from the answer sheets, proved more than adequate for the majority of pupils (in fact over 90% of students finished the test). The 40 minutes time limit was convenient for testing in secondary schools which work on a 40 or 45 minute class period. It is suggested that the time limit had a negligible effect on student performance, because of the liberality of the time limit and because students were able, if finished before the 40 minutes, to return and check any questions of which they may have been unsure.

Data Collection

The four schools used in the testing programme were all visited in late October or early November 1972. The OTIS test was given to each class first, then either the concept test or the SSAP test. The usual procedure for the secondary schools involved testing on three separate occasions. The OTIS test was of 30 minutes duration, 15 minutes was allowed for the CONCEPT test and 40 minutes for the SSAP test. The intermediate schools, with a more flexible timetable arrangement, were visited twice. On the second occasion the CONCEPT test and the SSAP test were administered, in that order, with a break of some 5 - 10 minutes between tests.

In all the classes tested the writer talked with the students about their role in a research programme aimed at finding out how students think in the ways similar to scientists. This was a subjective attempt to allay any fear of the testing programme being used as a means of passing or failing the students.

It was assumed in selecting the classes to be tested, that representative (i.e. in terms of ability and achievement) groups would be formed at each of the class levels. Unfortunately this was not so, as evidenced in the results section below. The third form grouping, on the basis of the OTIS test, do seem to be a much more homogeneous sample than was desired.

Data Analysis

From the description of the instruments used it should be apparent that for each individual student three variables were obtained. The student's age was also obtained. The means for these variables, together with the standard deviations, are reported separately for boys and girls at each of the three class levels (see Table 3).

Correlations

It was decided to seek intercorrelations of variables at the three class levels. Product - moment correlations were determined for

(a) OTIS score and CONCEPT score

(b) OTIS score and SSAP total score

(c) CONCEPT score and SSAP total score

(d) SSAP subscores, i.e. the skills and processes tests.

These intercorrelations were calculated separately for boys and girls and are reported in Table 4. One of the minor focusses of this study was to seek evidence of a sex difference in science-based tests, hence the separate reporting of the sexes at each class level.

Intercorrelations were also calculated for the total group (N = 339) between

(a) age and CONCEPT score

(b) age and SSAP total score (Table 6)

S.S.A.P. Analysis

The major purpose of the study was the development of a test of the science skills and processes and to this end the major portion of the data analysis pertains to this objective.

Norms and quartile ranks were determined for each class level (Table 9) although, as stated above, there is a restricted sample in the Form 3 class level.

An extensive test and item analysis was carried out on the SSAP test. This was done by computer. Cards were punched for each student with a record of his responses for the 40 items. These cards were checked randomly with the individual student's answer sheet to ensure accuracy. The following information was obtained for the whole test :

1. Mean score and standard deviation.
2. Frequency distribution of percentage scores (Table 8).
3. A histogram expressing the number of students scoring in each of the interval classes (Figure 1).
4. A question analysis showing the number of students answering each choice and hence the efficacy of the distractors. This is recorded in Appendix C.
5. An index for discrimination for each item (Appendix G).
6. An index of difficulty for each item (Appendix G).

This procedure was repeated for the two subtests of skills and processes (see Appendices H and I).

It should be noted that no correction for guessing was used in marking the SSAP test. This decision was based on "practical convenience as much as psychometric conviction". (Thorndike, 1971a, p.61, footnote). There were clear instructions in the test booklet that no penalty would be accrued for guessing.

Reliability of the S.S.A.P. Test

The reliability of a test can be determined in a number of ways (see, for example, Stanley, 1971, p.370), but reliability of the SSAP test was determined by three procedures. Firstly, the same test was given to 70 Form 2 pupils. This test - retest procedure was carried out over a time difference of four weeks. This time lag was in an attempt to reduce any memory of previous responses although the number of items (40) in the test would also assist in reducing the memory of previous responses. There had also been no feedback from the first test session and the motivation of the students did not appear lower when they were informed that once again they were to be involved in an important science research programme. This latter point about motivation is recognised as a subjective evaluation of the tester (the writer) but, nevertheless, is regarded as significant by the writer.

A second measure of reliability was to use the split-half method, corrected by the Spearman-Brown formula $r_{tt} = \frac{2r_{ab}}{1 + r_{ab}}$ (Stanley, 1971, p.408).

The test was already divided into two subtests so the next step was to further subdivide the subtests. This subdivision ensured that there was a distribution of behaviours, as far as possible, to the two halves of each subtest. For example, if one of the partial objectives had two items then one was placed in each half of the subtest.

The third measure involved the use of the Kuder-Richardson Formula 20 (Stanley, 1971, p.413). This formula can be used without splitting the test and assumes, as does the Spearman-Brown formula above, that the large majority of candidates have considered all items. The time allocation has been mentioned above and over 90% of candidates completed the test so that it was considered appropriate to use both these reliability measures. The K-R 20 was calculated for each of the subtests as each subtest was measuring a different major objective, and for the test as a whole.

Validity of the S.S.A.P. Test

The validity of the SSAP test is of obvious importance and yet it involves measures which are difficult to obtain.

The content validity, or the extent to which the curriculum content and the student behavioural outcomes are reflected in the test, is difficult to assess in the traditional sense of matching the above two variables in a blueprint. It has been suggested that one of the unique features of the SSAP test is its essentially one-dimensional nature so in one sense the notion of the content-validity does not appear to apply to the SSAP test. However, because the behavioural outcomes were "validated by experts" it is suggested that there is a high degree of content validity in this test.

Criterion-related validity is also difficult to assess because there are very few tests which could serve as criteria for the SSAP to be validated against. This difficulty may well be overcome by the development and acceptance of tests such as the SSAP test.

CHAPTER 5
RESULTS

This chapter deals with three major sets of findings; firstly the results from the testing programme which involved the three tests, secondly the characteristics of the SSAP test as a whole will be described and thirdly characteristics of the two subtests, skills and processes will be set out.

Results from the testing programme

The means and standard deviations of the four variables, for boys and girls separately at the three class levels are given in Table 3.

The average age of the Form 4 boys exceeded that of the girls by 2.58 months and this difference was significant of the 0.05 level. The difference in means of boys' and girls' scores for Form 3 on both the CONCEPT test and the SSAP test were significant at 0.02 and 0.01 levels respectively. The mean score of the Form 4 boys on the CONCEPT test was also significantly higher (0.02 level) than that for girls. There were no other significant intra-class differences.

The standard deviations for the OTIS scores (the range was 11.1 to 13.8) are lower than the OTIS manual suggests would be a typical range, that is 13-15 points. This may indicate a restricted range of students. The Form 3 sample, does seem a higher ability group, as measured by the OTIS test, than either of the other two classes. The smaller standard deviations for the Form 3 girls, for age, OTIS and CONCEPT test, again indicates a less representative sample of students had been tested than at either of the other class levels.

TABLE 3 : Means and Standard Deviations of Four Variables for Three Class Levels

Class	Sex	Means				Standard Deviations			
		Age	Otis	Concept	SSAP	Age	Otis	Concept	SSAP
Form 2	Boys (N = 60)	157.9	100.97	13.38	16.33	6.85	13.29	2.52	5.26
	Girls (N = 63)	157.3	104.33	12.62	16.05	5.72	12.87	3.18	4.87
Form 3	Boys (N = 53)	169.2	114.9	16.22	24.22	5.55	13.10	3.43	4.96
				++	+++				
	Girls (N = 58	168.78	113.5	14.76	21.60	4.39	11.34	2.60	4.68
Form 4	Boys (N = 48)	182.14	103.85	16.00	23.10	5.22	13.80	3.01	5.24
		+		++					
	Girls (N = 57)	179.56	105.33	14.53	21.18	6.20	11.09	2.77	5.50

+ significant at 0.05 level; ++ significant at 0.02 level; +++ significant at 0.01 level

Product - moment intercorrelations between the OTIS, the CONCEPT test and the SSAP test were calculated separately for boys and girls (Table 4). All correlation coefficients are significant at the 0.01 level.

TABLE 4 : Correlations between three variables - boys and girls separately.

FORM 4

Boys (N = 48)

	Otis	Concept	SSAP
Otis	-	0.66	0.68
Concept		-	0.54

Girls (N = 57)

	Otis	Concept	SSAP
Otis	-	0.54	0.68
Concept		-	0.46

FORM 3

Boys (N = 53)

	Otis	Concept	SSAP
Otis	-	0.45	0.65
Concept		-	0.43

Girls (N = 58)

	Otis	Concept	SSAP
Otis	-	0.43	0.67
Concept		-	0.38

FORM 2

Boys (N = 60)

	Otis	Concept	SSAP
Otis	-	0.33	0.67
Concept		-	0.40

Girls (N = 63)

	Otis	Concept	SSAP
Otis	-	0.63	0.68
Concept		-	0.53

The correlations were tested for significance of differences between the sexes using Fisher's Z - function transformation (King, 1969, p.87). All except one pair of correlations were not significantly different. There was a significant difference at the 0.05 level between the boys' score and the girls' score on the OTIS-CONCEPT test correlation at the Form 2 level.

The correlation coefficients for boys and girls combined are recorded in Table 5.

TABLE 5 : Correlation coefficients between three variables.

FORM 4 (N = 105)

	Otis	Concept	SSAP
Otis	-	0.56	0.68
Concept		-	0.49

FORM 3 (N = 111)

	Otis	Concept	SSAP
Otis	-	0.44	0.66
Concept		-	0.40

FORM 2 (N = 123)

	Otis	Concept	SSAP
Otis	-	.51	0.67
Concept		-	0.47

Once again, all these correlations are significant at the 0.01 level. The OTIS score appears the better predictor of total SSAP score. It is significantly better at the 0.05 level for all three class levels.

Finally correlations were determined between age and CONCEPT test scores and age and SSAP test scores. These are reported in

Table 6.

TABLE 6 : Correlation between Age and Concept Test and Age and SSAP Test.

Combined Girls and Boys (N = 339)

	Concept	SSAP
Age	0.24^{++}	0.29^{+++}

++ significant at 0.02 level

+++ significant at 0.01 level

There are small yet significant correlations between age and attainment in both the CONCEPT test and the SSAP test.

The SSAP Test characteristics

Table 7 summarises the results for the SSAP test as a whole. The number of students (341) included two students whose results were not incorporated in the determination of correlations outlined in the previous section because results of one of their tests were not available.

TABLE 7 : SSAP Test ; Summary of results

Possible Score	40
Mean Score	20.24
Mean Percentage Score	50.60%
Standard Deviation	6.18 (15.43%)
Standard Error	3.09
Number of Students	341

The frequency distribution of percentage scores for the SSAP test is given in Table 8 and a histogram of the computer printout of the SSAP test analysis accompanies this table (Figure 1). The histogram reflects a distribution of scores, over the total sample of students, which approaches a normal curve.

The item analysis indicated a mean difficulty index of 50% and discrimination indices ranging from 0.14 to 0.59 (mean value of 0.38). The discrimination indices and difficulty indices may be found in Appendix G. A graphical interpretation is included in Figure 2 which shows the distribution of item statistics to be arranged about a discrimination index of 0.4 and a difficulty index of 50%.

TABLE 8 : Frequency Distribution of Percentage Scores for SSAP Test. (N = 341)

Interval Number	Percentage Range	Number of Students
1	0 - 4.9	0
2	5 - 9.9	0
3	10 - 14.9	1
4	15 - 19.9	5
5	20 - 24.9	6
6	25 - 29.9	16
7	30 - 34.9	24
8	35 - 39.9	32
9	40 - 44.9	32
10	45 - 49.9	32
11	50 - 54.9	42
12	55 - 59.9	42
13	60 - 64.9	37
14	65 - 69.9	28
15	70 - 74.9	20
16	75 - 79.9	15
17	80 - 84.9	7
18	85 - 89.9	2
19	90 - 94.9	0
20	95 - 100	0

The separate class mean scores and standard deviations, for both boys and girls, have been referred to above (see Table 3). Percentile ranks were derived from a smoothed ogive of raw scores.

SCIENCE PROCESS TEST

QUESTION ANALYSIS SHOWING FOR EACH QUESTION THE NUMBER OF STUDENTS ANSWERING EACH CHOICE.
NON RESPONSES ARE EXCLUDED FROM THIS ANALYSIS.

HISTOGRAM 3

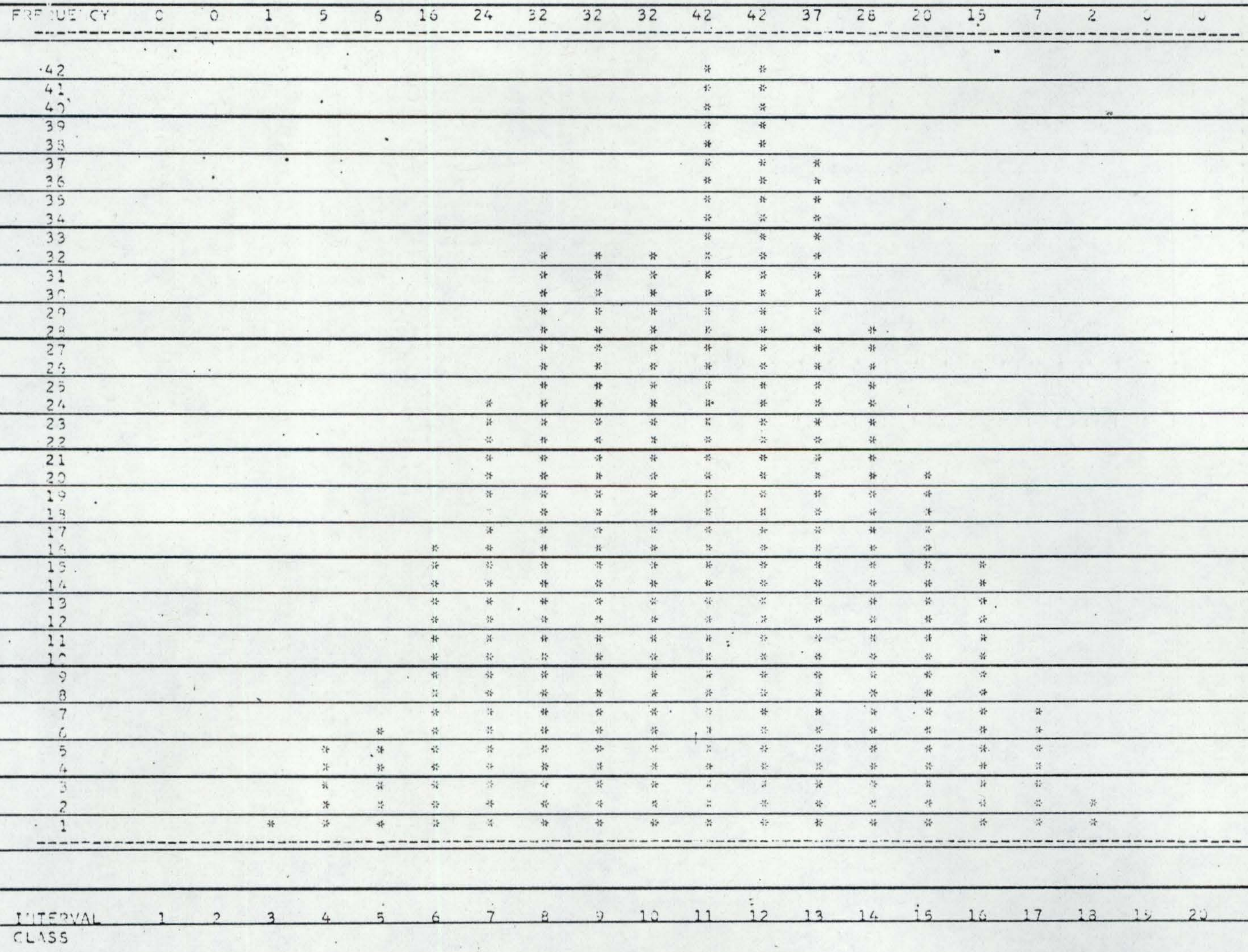
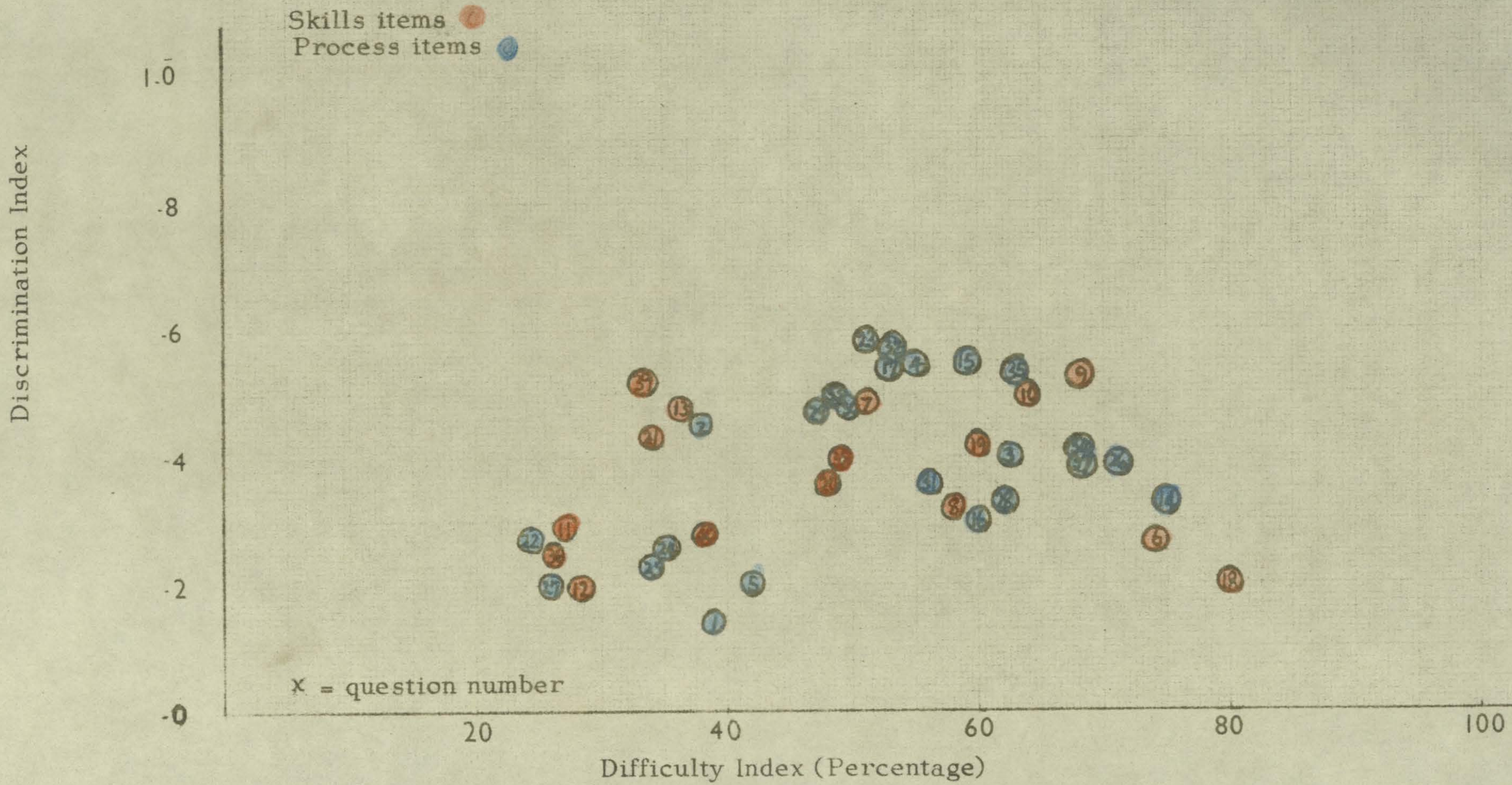


Figure 1 : Histogram of results for SSAP test

Figure 2. Difficulty and Discrimination Indices for the SSAP test.



These percentile ranks are included in Appendix E. Quartile ranks are given for the three class levels in Table 9. The semi-interquartile range ($\frac{Q3 - Q1}{2}$) indicates

TABLE 9 : Quartiles, range and semi-interquartile range for the three class levels, SSAP Test.

	Form 2	Form 3	Form 4	All Forms
Q1	12.60	19.40	19.67	15.61
Q2	15.70	23.0	22.45	20.04
Q3	19.56	26.35	26.61	24.78
$\frac{Q3 - Q1}{2}$	3.48	3.48	3.47	4.58
Number of Students	123	111	105	339
Range of Scores	23	24	26	30

that the dispersion of marks between the three groups is similar.

The raw scores on the SSAP test were used to formulate ogives for T-score determinations. Separate conversion scales are provided for the three class levels (Table 10). It should be pointed out that, as mentioned above, the Form 3 group were probably a "brighter-than-average" group and the conversion table can best be regarded as tentative for this class level.

TABLE 10 : The conversion of raw-scores to T-scores.

	T - SCORES			
Raw Score	Form 2	Form 3	Form 4	All Forms
5	22	20	27	22.5
6	27	21	28	25
7	30	22	29	28
8	33	23	30	29.5
9	35.5	24	31	31.5
10	38	25	32	34.5
11	40	26.5	33	36
12	42.5	28	34	37
13	44.5	30	35.5	39.5
14	46.5	31.5	36.5	42
15	48	34	38	43
16	50	36	39	44.5
17	52	37	40.5	45.5
18	53.5	39	42	47
19	55	41	43.5	48
20	56.5	43	45	50
21	58	45	47	51
22	60	47	48.5	53
23	61.5	49	50	54.5
24	63	51.5	52	56
25	65	53	53.5	58
26	67	55	55	59.5
27	69	57	57	61
28	71	59	59	63
29	74	61	61	65
30	76.5	63	63	68
31		66	66	70
32		68	68	73
33		71	71	75
34		74	74	77.5
35		77	78	78
36				
37				
38				
39				
40				

The ogives from which these conversion tables were constructed are included in Appendix F.

The reliability of the SSAP test was determined in three ways. Firstly 70 pupils at Form 2 were given the same test after a period of four weeks and a product - moment correlation coefficient for test-retest yielded a value of 0.75. A second method, the split-half with correction for length using the Spearman-Brown formula, was carried out on the two subtests, i.e. the skills test and the process test. This was necessary because the skills and processes items were regarded as having different major objectives. The split-half method yielded values of 0.59 for the skills test and 0.69 for the process test. Lastly, the Kuder-Richardson method, again for the two subtests, gave values of 0.61 for the skills test and 0.67 for the process test. A Kuder-Richardson determination for the whole test gave a value of 0.77. Item 34 was omitted from the process test reliability determination and Item 37 from the skills test. Both these items duplicated items which purported to test the same objectives and were omitted to enable the split-half correlation to be calculated. The items were not included in the Kuder-Richardson calculations for the subtest.

The sub-test characteristics

The skills test comprised 16 items and the processes test 22 items. Table 11 summarises the results for both these tests.

TABLE 11 : Skills and Processes Tests; Summary of Results.

	Skills Test	Processes Test
Mean Score	7.73	11.11
Mean Percentage	48.31%	50.50%
Standard Deviation	2.80 (17.48%)	3.75 (17.00%)
Standard Error	1.79	2.10
Number of Students	341	341
Possible Score	16	22

These results, for the tests across all three levels, indicate that the skills test was slightly but not significantly more difficult than the processes test. Separate mean scores for boys and girls at the three levels may be found in Table 12. Significant differences between boys' scores and girls' scores on both tests, exist at the Form 3 level only. In both cases the boys' score was higher.

TABLE 12 : Means and S.D.s for Skills and Processes Tests at three class levels.

Class	Sex	Mean Scores		Standard Deviations	
		Skills	Processes	Skills	Processes
Form 2	Boys (N = 60)	6.64	8.95	2.67	3.15
	Girls (N = 63)	6.05	8.50	2.05	3.23
Form 3	Boys (N = 53)	9.21 +	13.59 ++	2.63	3.04
	Girls (N = 58)	8.21	11.98	2.36	3.06
Form 4	Boys (N = 48)	8.33	12.87	2.88	3.37
	Girls (N = 57)	8.21	11.63	2.72	3.56

+ significant at 0.05 level

++ significant at 0.01 level

Correlations based on the Kuder-Richardson method have been described above for both the subtests. The intercorrelation of the subtest scores was 0.59.

An item analysis of the two subtests was also carried out and details are included in Appendices H and I.

CHAPTER 6

DISCUSSION OF THE RESULTS FROM THE SSAP TEST

General test characteristics

As mentioned above (Chapter 5, p.1) the rather smaller standard deviation in age for Form 3 girls (4.39 months) compared with the remainder of the groups (a range of 5.2 to 6.8 months) suggests a more restricted sample was tested at this class level. Moreover, the total Form 3 sample does seem a higher ability group, as measured by the OTIS, than either of the other two class levels. These figures indicate a need for caution in using results from the Form 3 sample to derive T-scores and Percentile Ranks. Both the Form 2 and Form 4 OTIS scores and age ranges suggest that Table 10 could be used with some measure of confidence to convert raw scores for those classes to T-scores.

There is an expected trend in the scores on the CONCEPT test from Form 2 to Forms 3 and 4. This trend is reflected in the correlation (0.24) between age and concept score. This is a much smaller correlation than King's (1963) value of 0.504. King, however, used a much wider range of ages in his testing programme (the range was from 5 - 17 years of age) and on the basis of this the scores on a 20 item test would give a greater spread of marks and, predictably, a higher correlation with age than was evident in the present study.

Sex differences

The situation regarding sex differences, which are frequently noted in achievement tests in science (see, for example, Rowell, 1971, or King, 1963) is not at all clear from the results of the present pupil sample. The difference between the age of the Form 4 boys and girls (0.05 level) may well explain some of the difference between the boys' and girls' scores on the CONCEPT test (0.02 level) for it has been noted above that a small positive correlation exists between age and CONCEPT test score (0.24).

This age difference does not explain the differences at the Form 3 level however. The girls scored less than the boys on both the CONCEPT test and the SSAP test. King's (1963) study revealed that a significant difference (0.05) level did exist between girls' and boys' scores for his total sample of children. The observed difference in this study is in the same direction. King suggests that "boys' superiority on science questions does not come about until after the children have left the primary school" (King, 1963, p.247). This suggestion is also reflected in the Form 3 scores reported in this study.

These results may well be an example of what Rowell (1971) refers to as a teacher-expectation effect. At the secondary school level children are coming into contact with subject specialists for probably the first time. In the four Form 2 classes used in this study, science was taught by the classroom teacher, not a science specialist. Rowell suggests that in many classes teachers have preconceived notions about girls' lack of success in science and these expectations become reflected in the differential attainment of girls. The small number of teachers involved in this study (there were only three teachers - all male - at each of the Form 3 and Form 4 levels as one teacher at each level taught two classes) would preclude any investigation of this teacher-expectation effect, but nevertheless it may well be playing a part in the differential attainment of boys observed in the Form 3 classes. There appears to be no logical explanation as to why these significant differences observed at Form 3 level all but disappear at Form 4. Perhaps these differences are a reflection of the atypical cross-section tested at the Form 3 level.

The report by McMurray (private correspondence, 1970) cited above, suggested that teachers perceived girls at Form 1 as not as interested in science as boys. When the scores of the Form 2 girls and boys are compared both on the CONCEPT test and the SSAP test, there are no significant differences between them, so it would appear that in this sample, at least, lack of interest does not impair performance on either a concept test or tests of science processes and science skills.

Relationship with other cognitive measures

In the development of achievement tests it is common practice to provide measures of the relationships with other cognitive measures such as intelligence tests (Welch and Pella, 1968, King, 1963, or Burmester, 1953). The SSAP yielded a correlation range of 0.67 to 0.68 with the OTIS (Table 5). This highly consistent result is in agreement with the Welch and Pella (1968) reported range of 0.61 - 0.64. Tannenbaum (1971), one of the very few studies to report the development of a test of science processes, does not give any correlations between his test and an intelligence test, but one would presume that Pearson coefficients of about these values would be in the usual range for attainment tests. The Burmester (1953) study also reported a correlation of 0.64 with her test of scientific thinking and a natural science achievement test.

Elley and Livingstone (1972, p.108) report a longitudinal study of the relationship between OTIS scores and science results in New Zealand School Certificate examinations for pupils at a large boys' high school. The reported correlation of 0.47, while significant, is difficult to interpret because of the changing and often unstated objectives of School Certificate Science examinations. One wonders to what extent objectives related to skills and processes are measured if an examination blueprint is not provided. This blueprint provision has become more prevalent in recent years (see for example the Report on 1969 School Certificate Science Examination).

The Pearson coefficient range for the OTIS-CONCEPT test correlation (0.44 to 0.56) is smaller than King's (1963) correlation between the CONCEPT test and the Simplex Junior Intelligence Scale (0.677). Once again one is reminded of the much larger age range King used in his study and there are obvious difficulties in extrapolating both from a smaller age range and from an OTIS measure to the Simplex so that a direct comparison of the two sets of figures may not be very profitable.

One of the objectives of the present study was to develop a test that would require students to use the higher cognitive levels of Bloom's (1956) taxonomy. The CONCEPT test is mainly recall with perhaps some application. The range of Pearson coefficients (0.40 - 0.49) for correlation between the CONCEPT test and the SSAP is therefore within an expected range, that is, a positive but only moderate correlation between the two tests.

The two Piagetian stages of cognitive development of concern in this study were the concrete and formal operational stages. The formal operational stage begins to develop about 11 - 12 years of age and, once attained, enables hypothesis-making to be used as a strategy in solving science problems. One would predict higher scores on the SSAP with increasing age of students if the test is measuring these higher order mental skills. The Pearson coefficient of 0.29 (Age - SSAP test) while small is nevertheless significant at the 0.01 level and does support the above prediction. A recent survey of Australian school children in similar age ranges to this study noted a significant difference ($0.05 < p > 0.025$) between boys and girls when an experiment requiring hypothesis formation was used. (Dale, 1970.) Dale also reported a gradual, almost linear increase in problem-solving ability for that particular problem and stated that children in Australia were reaching the formal operational stage later than the children reported by Inhelder and Piaget (1958). An interesting extension of this present study would be to determine the correlation between the SSAP test and the conceptual levels of students as determined by Piagetian tasks.

Reliability of SSAP test

The test-retest method yielded a correlation of 0.75. This coefficient is lower than is desirable but still explains some 56% of the variance between the scores on the two test occasions. The time interval between tests (4 weeks) was considered sufficiently long for individuals to forget responses made the first time. A check indicated that scores on the retest had a mean of 17.3 marks compared with 16.2 marks on the first test occasion.

The split-half method, with correction by the Spearman-Brown formula, indicated that the skills subtest had a lower internal consistency than the process subtest. Two factors appear important in considering the reliabilities of the subtests. The first was the small number of items in the subtests. There were 16 skill items and 22 process items and this meant that, for the split-half method, there were 8 and 11 items respectively in each subtest. The second factor was based on the assumption that the same objectives were being tested in each of the half tests. This would appear to be an invalid assumption, for although the overall objectives were concerned with either skills or processes, the partial objectives were often concerned with different cognitive abilities. There would thus appear to be a difficulty in splitting such a test.

The values of K-R consistency coefficients reported by

Tannenbaum (1971) ranged from 0.30 to 0.80 on eight subtests and an overall value of 0.91 was recorded. The K-R coefficients for the Skills subtest (0.61) and the Process subtest (0.67) were within this reported range.

Validity of the SSAP Test

The difficulty in validating the SSAP test was mentioned above in Chapter 4. It bears repeating at this point that, on the basis of the collective opinions of 19 persons associated with science education from primary level to university level, the behaviours listed in the blueprint (Appendix B) seem representative of the behavioural domain to be measured by the SSAP test. In this sense the test can be adjudged as having content validity. The number of items for each behaviour are also included in Appendix B so that the relative importance of each can be determined. It has been repeatedly stressed that one of the dominant features of the SSAP test is that the "content" of the questions is not syllabus-bound. This latter characteristic means that only the behaviours and not the content can be assessed for content validity.

The question of criterion-related validity for the SSAP test remains largely unanswered. A desirable next step would be to initiate follow-up studies which could well be along lines suggested by Tannenbaum (1971). He used a rating scale for teachers to indicate whether pupils were exhibiting, in the classroom, the behaviours set out in the blueprint. The use of observation instruments would seem to be a worthwhile extension in assessing a student's ability to use science processes and skills, but there are also difficulties associated with the subjective use of rating scales. There are at present no published tests suitable for New Zealand schools which attempt to measure achievement of the same objectives as the SSAP purports to measure.

Item analysis of the SSAP test and subtests

Items for a test should not be selected on statistical properties alone. To measure changes in student behaviours toward the stated behavioural objectives (see Appendix B) it is suggested that items should be chosen for content-validity rather than difficulty value (D.V.) or discrimination index (D.I.). Two examples may serve to illustrate this point. Firstly, in a mastery test, if D.V. were a major criterion, items could be rejected as being too easy and secondly, if a particular skill has not been adequately taught, then a low D.V. may indicate the item as being too difficult and thus lead to rejection of the item. In both cases the items may well be important in terms of content-validity. This is not to decry the use of item analysis data for it can yield important information about items. The statistical data must, however, be used along with content-validity information.

The analysis data from the SSAP test has been reported graphically in Chapter 5. The average difficulty index of 50 per cent is in close agreement with the mean score for the SSAP test of 50.60 per cent. The range of difficulty indices was from 24 per cent to 80 per cent. It has been suggested that many kinds of test items have low intercorrelations and that a distribution of item difficulties clustered around the 50 per cent level would "often approximate the distribution required to obtain maximum discrimination throughout the range of scores." (Henrysson, 1971, p.152.) For multiple choice items the ideal mean D.V. is probably higher than this 50 per cent level and the 50 per cent level obtained in the SSAP test. Henrysson mentions a value of around 60 per cent for a 5 choice item. Thus the mean difficulty index of 50 per cent would suggest the SSAP test is probably a little too difficult for the range of classes tested. One solution to this problem would be to not recommend the test for Form 2 classes but more acceptable solutions might be either to direct teacher attention, by providing lists of validated student behaviours, to skills and processes objectives to ensure a more adequate coverage of these objectives, or, alternatively, to revise some of the items with the specific intention of raising the difficulty indices.

The discrimination index range of 0.14 to 0.59 (mean 0.38) is rather lower than is desirable. However, only one item (Question 1) was below a D.I. of 0.20. There are items (see Figure 2) which could well be revised, for example, Question 1 (D.I. = 0.14, D.V. = 34%), but as many of these items measure student behaviours that have been validated by 'experts', they should not be deleted from the test on statistical grounds alone.

The skills subtest shows a wide range of D.V. from 26 per cent to 81 per cent (mean 48 per cent). This mean is reflected in the mean score for the skills subtest of 48.31 per cent (Table 11). The D.I. range of 0.24 to 0.63 (mean 0.43) shows a greater discriminability of this test than the whole SSAP test. The skills test does show better discrimination between upper and lower groups than the SSAP test. (Compare discrimination indices of the relevant items from Appendices H and I.)

The processes subtest yields figures in close agreement with the skills subtest. The range of D.V. from 23 per cent to 73 per cent (mean 50 per cent) and D.I. range of 0.23 to 0.63 (mean 0.44) again indicates good discriminability.

Generally, the item analysis data are indicative of a test that discriminates more than adequately between the upper and lower groups. However, further refinement of some of the items would doubtless improve the SSAP test.

CHAPTER 7

CONCLUSIONS

This dissertation has reported the construction of a test of science skills and processes for the Form 1-5 Science syllabus. The test is a multiple-choice, 40 item instrument (17 items related to skills and 23 to processes) and was designed to be free from syllabus restrictions in terms of content. It has a time limit of 40 minutes.

On a local sample the test had a reliability coefficient of 0.75 on a test-retest method and an internal consistency of 0.77 using Kuder-Richardson formula 20. The internal consistency of the skills subtest was 0.59 and 0.69 for the processes subtest. Both these coefficients were determined by the split-half method with correction by the Spearman-Brown formula (the K-R formula 20 figures were 0.61 and 0.67 respectively). The rather low test-retest reliability suggests the need for further analysis. It has been conjectured (Chapter 6 above) that because of the diverse range of behaviours grouped under a general heading of either skills or processes, this may well lead to a low reliability because the different items may be testing the attainment of these different behaviours which collectively contribute to skills or processes. The Tannenbaum (1971) test reported above, used a much larger number of items (98) and eight subtests, each of which purported to measure <u>separate</u> skills. With an increased number of items and wider range of subtests it would doubtless be possible to increase the internal consistency of the SSAP test. A factor analysis of inter-item correlations might assist in isolating groupings of items, which, when considered together, might give higher sub-test reliabilities.

Any test claiming to measure student attainment of particular objectives will have results affected both by the teachers' perception of the objectives and by the teaching methods used. The sample of teachers used in this study was too small for such an analysis but it is suggested that a worthwhile direction for further study could be an attempt to correlate student scores on the SSAP test with their teachers' ranking of their achievement of the objectives of the syllabus.

Teaching methods are often polarised into reception learning, in which the teacher provides descriptions and explanations and the student does not have an active role and discovery methods in which students work at problem-solving guided by teacher questions and prompting. Rowell (1972, private communication) has recently attempted to measure pupils' perception of their teachers' teaching method by using a continuum with reception learning at one pole and discovery learning at the other. Students mark a position on the continuum representing their perception of the teaching method. Clearly the discovering-learning method would enable students to more effectively develop science skills and processes. The teaching method employed could also prove important in an investigation of the factors that contribute to success on the SSAP test. Yet another factor that could be investigated is that of cognitive style. Maccoby's (1966) point about different cognitive styles affecting learning, especially in science, has been mentioned above (Chapter 2).

The results from the blueprint validation were somewhat disappointing in that only 19 science educators replied to the 32 sent out. There was a tendency, again unsupported by a statistical analysis because of the small groups of respondents, for educators furtherest removed from the classroom (for example, curriculum workers and teachers college lecturers) to be more likely to regard the behaviours set out in the original blueprint (Appendix A) as appropriate for students in the age range. Teachers, apparently a more pragmatic group, had lower expectations and rejected many of the behaviours listed as unsuitable for the age group.

This validation of the blueprint by experts does bring into focus one of the problems of the SSAP test construction. It must be assumed that the authority of the group used to validate the blueprint was acceptable. There was no validation check of the actual items used in the SSAP test. To do this would require lists of items to be forwarded to science educators and this would place a demand on the time of the experts but would perhaps have established a greater content validity for the SSAP test.

It was suggested in the introductory chapter that one of the minor objectives of this study was to identify sex differences, if present, in science achievement for the particular age groups. Contrary to expectation sex differences were not marked throughout the three levels. Indeed, the Form 3 sample was the only group to yield statistically significant differences when the age difference of the Form 4 sample was considered. Possible reasons for this sex difference have been discussed in Chapter 6. Another minor objective was to compare the science concept achievement to similar age groupings to King (1963). The graphical comparison (Appendix J) of 13 year old students' answers to the 20 questions revealed a consistency in attainment between the British and New Zealand groups.

It would seem important in using such skills as measuring, quantifying and predicting that a knowledge of science concepts similar to those in the CONCEPT test would be desirable. The range of results on the CONCEPT test would suggest that remedial work in establishing these basic concepts might well be in order before serious attempts to develop facility in science skills are made. The CONCEPT tests developed by King were largely influenced by many of the early Piagetian experiments related to children's thinking. An attempt to establish individual students' levels of cognitive development (as defined by Piagetian tests) might therefore contribute to a better understanding of the SSAP test.

The paucity of research related to science skills and processes evaluation has been stressed (Chapter 2). The educational significance of tests such as the SSAP test lies in the continuing emphasis on the syntactic structures of disciplines and not just on the substantive structure. With this emphasis it is important for teachers to understand not only what is meant by the syntactic structure of science but what kinds of student behaviours might reasonably be expected from students in the classroom. The SSAP test goes some way to meeting the demands of new ways of measuring the outcomes of instruction by providing an instrument to yield both attainment and remedial information related to science skills. Science tests must no longer focus on the lower levels of Bloom's (1956) taxonomy and the SSAP test seeks to evaluate higher levels of the taxonomic hierarchy. It is also desirable not to base evaluations on one kind of instrument but on the accumulation of a variety of test results which supply profiles of an individual's attainment. In this respect, also, an improved version of the SSAP test would appear to offer a set of validated student behaviours which could be used in conjunction with other science attainment tests to further our knowledge of students' science attainment.

BIBLIOGRAPHY

American Association for the Advancement of Science (1965) The Psychological Bases of Science - A Process Approach.

Amos, R. (1970), "Teachers' opinions about the importance of scientific method in English Ordinary-Level biology courses", J. Research in Science Teaching, 7, pp.303-313.

Anastasi, A. (1968), Psychological Testing, Third Edition, New York: Macmillan.

Atkin, J.M. (1966), "Comments on two approaches", Science, 151, p.1033.

Australian Science Education Project (1970), Newsletter Number 2, November 1970.

Ausubel, D.P. (1963), "Some psychological considerations in the objectives and design of an elementary school science program", Science Education, 47, pp.278-284.

Bloom, B.S. (ed.) (1956), Taxonomy of Educational Objectives Handbook I : Cognitive Domain. New York: David McKay.

Bloom, B.S. (1969), "Some theoretical issues relating to educational evaluation", in Tyler, R. (ed.), 1969, op.cit., pp.26-50.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. (1971), Handbook on Formative and Summative Evaluation of Student Learning, New York: McGraw-Hill.

Brown, J. (1966), Objective Tests. Their Construction and Analysis. London: Longmans, Green.

Brown, W.R. (1968), "Defining the processes of science : a statement by educators in India", The Science Teacher, 35, pp.26-28.

Burmester, M.A. (1953), "The construction and validation of a test to measure some of the inductive aspects of scientific thinking", Science Education, 37, pp.131-140.

Butzow, J.W. and Sewell, L.E. (1972), "An investigation of introductory physical science using the test of science process", J. Research in Science Teaching, 9, pp.267-270.

Carey, R.L. and Stauss, N.G. (1970), "An analysis of experienced science teachers' understanding of the nature of science", School Science and Mathematics, 70, pp.366-376.

Cooley, W.W. and Klopfer, L.E. (1963), "The evaluation of specific educational innovations", J. Research in Science Teaching, 1, pp.73-80.

Dale, L.G. (1970), "The growth of systematic thinking : replication and analysis of Piaget's first chemical experiment", Aust. J. Psychology, 22, pp.277-286.

Department of Education (1967), Science for Forms 1 and 2, Primary School Syllabus, Wellington.

Department of Education (1968), Science for Forms 3 and 4, Secondary School Syllabus, Wellington.

Department of Education (1971), Suggestion for a Draft Syllabus Science - Infants to Standard 4, Curriculum Development Unit.

Dressel, P.L. (1960), "How the individual learns science", in Henry, N.B. (ed.), Rethinking Science Education, N.S.S.E. Yearbook, pp.39-62.

Duncan, J.F. et al (1971), "The objectives of lower sixth-form chemistry teaching", N.Z.J. of Educational Studies, 6, pp.45-61.

Eggleston, J.F. and Kerr, J.F. (1969), Studies in Assessment, Liverpool: The English Universities Press Ltd.

Elley, W.B. and Livingstone, I.D. (1972), External Examinations and Internal Assessments, Wellington: N.Z.C.E.R.

Flynn, H.E. and Munro, R.G. (1970), "Evaluation of a curriculum", British J. of Educational Psychology, 40, pp.271-282.

Fox, D.H. (1969),"Assessing scientific ability through special studies", in Eggleston, J.F. and Kerr, J.F. (1969), op.cit., pp.153-168.

Furst, E.J. (1964), Constructing Evaluation Instruments, New York: D. McKay.

Gagné, R.M. (1965), "Psychological issues in Science - A Process Approach", in A.A.A.S., op.cit., pp.1-8.

Gagné, R.M. (1966), "Elementary science: a new scheme of instruction", Science, 151, pp.49-53.

Gronlund, N.E. (1971), Measurement and Evaluation in Teaching, Second edition, New York: Macmillan.

Guilford, J.P. (1965), Fundamental Statistics in Psychology and Education, Fourth edition, New York: McGraw-Hill.

Henrysson, S. (1971), "Gathering, analysing and using data on test items", in Thorndike, R.L. (ed.) op.cit., pp.130-159.

Hurd, P.D. (1969), New Directions in Teaching Secondary School Science, Chicago: Rand McNally.

Inhelder, B. and Piaget, J. (1958), The Growth of Logical Thinking from Childhood to Adolescence, London: Routledge & Kegan Paul.

King, W.H. (1961), "The development of scientific concepts in children", British J. of Educational Psychology, 31, pp.1-20.

King, W.H. (1963), "The development of scientific concepts in children (2)", British J. of Educational Psychology, 33, pp.240-252.

King, W.H. (1969), Statistics in Education, London: Macmillan.

Klein, C.A. (1971), "Differences in science concepts held by children from three social-economic levels", School Science and Mathematics, 71, pp.550-558.

Klopfer, L.E. (1971), "Evaluation of learning in science", in Bloom, B.S., Hastings, J.T. and Madaus, G.F., op.cit., pp.559-641.

Lewis, D.G. (1965), "Objectives in the teaching of science", Educational Research, 7, pp.186-199.

Maccoby, E.E. (ed.) (1966), The Development of Sex Differences, California: Stanford University Press.

McMurray, B. (1970), A Report prepared for the New Zealand Educational Institute on the teaching of science at Form 1 and 2. Private correspondence.

Morgan, D.A. (1971), "STEPS - a science test for the evaluation of process skills", The Science Teacher, 38, pp.77-79.

Munro, R.G. (1972), Tests of the Science Test Development Project mimeographed, undated, private correspondence.

Nay, M.A. and associates (1971), "A process approach to teaching science", Science Education, 55, pp.197-207.

New Zealand Consultative Committee on the Post-Primary Curriculum (1944), The Post-Primary Curriculum, Wellington: Government Printer (The "Thomas Report".)

New Zealand Council for Educational Research (1969), OTIS Tests of Mental Ability, Manual of Directions, Wellington: N.Z.C.E.R.

Piaget, J. (1971), Science of Education and the Psychology of the Child, London: Longman.

Phillipps, R.W., (1971), "New ways in science", Education, 20, pp.18-25.

Ramsey, G.A. and Howe, R.W. (1969a), "An analysis of research related to instructional procedures in elementary school science", Science and Children, 5, pp.25-36.

Ramsey, G.A. and Howe, R.W. (1969b), "An analysis of research on instructional procedures in secondary school science, Part I - outcomes of research", The Science Teacher, 36, pp.62-70.

Raun, C.E. (1967), "The interaction between curriculum variables and selected classroom student characteristics", cited in

Ramsey, G.A. and Howe, R.W. (1969a), op.cit., p.31.

Report on 1969 School Certificate Science Examination, Department of Education, 1970.

Roscoe, J.T. (1969), Fundamental Research Statistics for the Behavioural Sciences, New York: Holt, Rinehart & Winston.

Roth, R.A. (1971), "Princeton project and the process approach", School Science and Mathematics, 71, pp.44-48.

Rowell, J.A. (1971), "Sex differences in achievement in science and the expectations of teachers", Aust. Journal of Education, 15, pp.16-29.

Rowland, G.W. (1965), "A study of the relationship between socio-economic status and elementary school science achievement", reported in Ramsey, G.A. and Howe, R.W. (1969a), op.cit.

Schwab, J.J. (1963), Biology Teachers' Handbook, New York: John Wiley & Sons.

Schwab, J.J. (1964), "Structure of the disciplines: meanings and significances", in Ford, G.W., and Pugno, L. (eds.), The Structure of Knowledge and the Curriculum, Chicago: Rand McNally, pp.1-30.

Secondary Schools Examination Council (undated), Information series : Examinations Number 2. "Account of an exploratory examination in science", mimeographed.

Sequential Tests of Educational Progress: Science. Cooperative Test Division, 1956-1959, Grades 7-9.

Smith, H.A. (1963), "Educational research related to science instruction for the elementary and junior high school", J. Research in Science Teaching, 1, pp.199-255.

Stanley, J.C. (1971), "Reliability" in Thorndike, R.L. (ed.), 1971a, pp.356-442.

Tannenbaum, R.S. (1971), "The development of the Test of Science Processes", J. Research in Science Teaching, 8, pp.123-136.

Thorndike, R.L. (ed.) (1971a), Educational Measurement, Second edition, Washington: American Council on Education.

Thorndike, R.L. (ed.) (1971b), "Reproducing the test" in Thorndike, R.L. (ed.) op.cit., pp.160-187.

Tyler, R.W., (ed.) (1969), Educational Evaluation : New Roles, New Means, N.S.S.E. Yearbook, Chicago: National Society for the Study of Education.

Walbesser, M.H. (1963), "Curriculum evaluation by means of behavioural objectives", J. Research in Science Teaching, 1, pp.296-301.

Watson, J.E. (1964), Intermediate Schooling in New Zealand, Wellington: N.Z.C.E.R.

Welch, W.W. and Pella, M.O. (1968), "The development of an instrument for inventorying knowledge of the processes of science", J. Research in Science Teaching, 5, pp.64-68.

APPENDICES

Appendix A	Original test blueprint and instructions to validators.
Appendix B	Final SSAP test blueprint.
Appendix C	Final test and student instructions.
Appendix D	Concept test and instructions.
Appendix E	Table for converting raw scores to percentile ranks.
Appendix F	Ogives used to formulate T-scores.
Appendix G	Item analysis data - SSAP test.
Appendix H	Item analysis data - Skills subtest.
Appendix I	Item analysis data - Processes subtest.
Appendix J	Comparison of science concept attainment of 13 year old students.

APPENDIX A

The original blueprint and instructions sent to science educators. The results from this survey are shown in Table 2.

INSTRUCTIONS TO SCIENCE EDUCATORS

APPENDIX A

Dear

I am presently engaged in an attempt to construct a test of the science processes and skills mentioned in the Form 1 - 5 Science syllabus. I have taken the objectives from the syllabus, allocated them to a taxonomy, and proposed student behaviours which would indicate that a student had attained the particular objective.

I am asking for your support and cooperation in this venture because of your known interest in the field of science education. I would be grateful if you could assist in the following way:

1. You will note that the accompanying sheets have lists of the objectives, student behaviours and ranking columns.

2. Would you please rank (a) the objectives and (b) the student behaviours in the appropriate column. For example, in considering Objective I;

   (i) if you consider the objective is an essential one ring number 1 in column 1.

   (ii) now consider the student behaviours and rank them according to relevance to the syllabus objectives (column 2); clarity of expression of the behaviour (column 3); and difficulty of the behaviour with respect to the age level i.e. 12 - 15 years of age (column 4).

3. You will note that objectives X, XI and XII are not specifically mentioned in the syllabus but rankings for these would also be appreciated.

4. Any comments you may offer, e.g. you may consider other objectives should be included, would be appreciated.

5. A copy of the syllabus objectives is appended for your ready reference.

6. Would you please return the sheets in the enclosed envelope by ______________________________ .

Thanking you in anticipation.

Yours sincerely,

Dennis M. McGrath
Lecturer in Science

R A N K I N G S

Objective or partial objective	Col.1 How important is the objective?	Student behaviours related to the objectives.	Col.2 How relevant is the student behaviour to the syllabus objective?	Col.3 How clearly is the student behaviour expressed?	Col.4 How difficult would the achievement of this behaviour be for 12-15 year old student?
	1 = Essential objective 2 = Very important objective 3 = Desirable objective 4 = Valuable objective if time available		1 = Highly relevant 2 = Moderately relevant 3 = Low relevance	1 = High clarity (clear) 2 = Moderate 3 = Low clarity (unclear)	1 = Appropriate 2 = Too hard 3 = Too easy
I "To identify a problem...	1 2 3 4	The student will show his ability to identify a problem by			
		(a) recognising that meaningful questions may be asked about a given problem	1 2 3	1 2 3	1 2 3
		(b) rephrasing the given problem in such a way that he can formulate suggestions as to possible answers.	1 2 3	1 2 3	1 2 3

		(c) recognising the need for specific techniques to solve the problem.	1 2 3	1 2 3	1 2 3
II ...to bring to bear earlier experiences appropriate to the problem...	1 2 3 4	The student will show his ability to bring to bear appropriate experiences relevant to the problem by			
		(a) analysing the problem in such a way as to suggest a method of solution.	1 2 3	1 2 3	1 2 3
		(b) applying previously aquired scientific principles to the problem in hand	1 2 3	1 2 3	1 2 3
		(c) isolating relevant from irrelevant data	1 2 3	1 2 3	1 2 3
III ...to formulate explanations and hypotheses	1 2 3 4	The student will show his ability to formulate hypotheses and explanations by suggesting			
		(a) hypotheses related to the problem at hand	1 2 3	1 2 3	1 2 3
		(b) hypotheses consistent with known data	1 2 3	1 2 3	1 2 3
		(c) hypotheses that can be experimentally tested	1 2 3	1 2 3	1 2 3

IV ... to test such explanations and hypotheses by experiment or other means...	1 2 3 4	The student will show his ability to test such explanations and hypotheses by			
		(a) deciding what data are necessary to solve the problem	1 2 3	1 2 3	1 2 3
		(b) selecting the main variables to be experimentally varied or controlled	1 2 3	1 2 3	1 2 3
		(c) distinguishing between dependent and independent variables	1 2 3	1 2 3	1 2 3
		(c) deciding how to control the independent variables	1 2 3	1 2 3	1 2 3
		(d) deciding how to control the independent variables	1 2 3	1 2 3	1 2 3
		(e) ranking hypotheses on usefulness so that an early experiment may well eliminate other possibilities	1 2 3	1 2 3	1 2 3
		(f) describing cause and effect relationship	1 2 3	1 2 3	1 2 3
V ...to accept, modify or reject and to draw conclusions	1 2 3 4	The student will show his ability to draw conclusions or to accept, modify or reject them by			
		(a) distinguishing between assumptions, hypotheses, theories and principles	1 2 3	1 2 3	1 2 3

		(b) recognising that sufficient evidence must be collected to form generalisations (i.e. to withhold judgement where necessary)	1 2 3	1 2 3	1 2 3
		(c) evaluating evidence for (i) reliability (ii) validity	1 2 3	1 2 3	1 2 3
		(d) detecting trends in data that can be used to predict or formulate conclusions.	1 2 3	1 2 3	1 2 3
VI manipulate scientific equipment correctly...	1 2 3 4	The student will show his ability to manipulate scientific equipment correctly by			
		(a) selecting or designing appropriate apparatus to assist in data collection	1 2 3	1 2 3	1 2 3
		(b) selecting apparatus capable of providing appropriate physical measurements	1 2 3	1 2 3	1 2 3
		(c) recognising the measuring limits of certain pieces of scientific equipment	1 2 3	1 2 3	1 2 3
		(d) demonstrating his skill in laboratory situations	1 2 3	1 2 3	1 2 3

VII ...to measure accurately...	1 2 3 4	The student will show his ability to measure accurately by			
		(a) Demonstrating a knowledge of the cardinal and ordinal aspects of numbers (both positive and negative) in relation to common measurement scales	1 2 3	1 2 3	1 2 3
		(b) demonstrating a knowledge of fractions, percentages and decimals in relation to common measurement scales	1 2 3	1 2 3	1 2 3
		(c) measuring using standard units	1 2 3	1 2 3	1 2 3
		(d) selecting appropriate units for particular measurements	1 2 3	1 2 3	1 2 3
		(e) estimating probable results in terms of these units	1 2 3	1 2 3	1 2 3
		(f) recording measurements appropriately and accurately.	1 2 3	1 2 3	1 2 3
VIII ...to construct and interpret tables, charts and graphs	1 2 3 4	The student will show his ability to construct and interpret tables, charts and graphs by			
		(a) correctly labelling axes on graphs	1 2 3	1 2 3	1 2 3

		(b) choosing appropriate scales for graphs	1 2 3	1 2 3	1 2 3
		(c) stating units and headings for a given graph	1 2 3	1 2 3	1 2 3
		(d) selecting a line of best fit for a given set of points	1 2 3	1 2 3	1 2 3
		(e) detecting trends in data	1 2 3	1 2 3	1 2 3
		(f) using trends to extrapolate and interpolate	1 2 3	1 2 3	1 2 3
		(g) deriving an unique original generalisation from various data	1 2 3	1 2 3	1 2 3
		(h) constructing from a set of data appropriate graphs or charts.	1 2 3	1 2 3	1 2 3
IX ...to find relevant information from available reference sources	1 2 3 4	The student will show his ability to find relevant information from available reference sources by			
		(a) recognising the appropriate authorities	1 2 3	1 2 3	1 2 3
		(b) recognising differences in the reliabilities of various source material	1 2 3	1 2 3	1 2 3
		(c) selecting relevant data from a given table of data	1 2 3	1 2 3	1 2 3
		(d) recognising the limitations in existing data.	1 2 3	1 2 3	1 2 3

X Observing	1 2 3 4	The student will show his ability to observe by			
		(a) identifying the objects that interact within a system	1 2 3	1 2 3	1 2 3
		(b) identifying differences and likenesses in a wide variety of objects	1 2 3	1 2 3	1 2 3
		(c) identifying and describing the results of interactions in terms of initial and final states	1 2 3	1 2 3	1 2 3
		(d) discriminating data relevant to an experience	1 2 3	1 2 3	1 2 3
		(e) distinguishing between observations and inferences.	1 2 3	1 2 3	1 2 3
XI Classifying	1 2 3 4	The student will show his ability to classify by			
		(a) dividing a total set of objects into two or more subsets differing in some characteristics	1 2 3	1 2 3	1 2 3
		(b) correctly subdividing a set of objects into several categories of increasing or decreasing inclusiveness	1 2 3	1 2 3	1 2 3
		(c) classifying likenesses and differences among a variety of different events (e.g. rates of change).	1 2 3	1 2 3	1 2 3

XII Comparing	1 2 3 4	The student will show his ability to compare and contrast by describing in terms of physical properties the similarities or differences between two or more			
		(a) objects	1 2 3	1 2 3	1 2 3
		(b) interacting groups of objects	1 2 3	1 2 3	1 2 3
		(c) processes (e.g. chemical, geological).	1 2 3	1 2 3	1 2 3

APPENDIX B

FINAL SSAP TEST BLUEPRINT

APPENDIX B

FINAL BLUEPRINT FOR SSAP TEST

The letter designation (e.g. (a)) in Student Behaviour column refers to original blueprint (see Appendix A).

Objective or Partial Objective	Student Behaviour	Question
I "to identify a problem..."	The student will show his ability to identify a problem by (a) recognising that meaningful questions may be asked about a problem.	1
II "...to bring to bear earlier experience appropriate to the problem..."	The student will show his ability to bring to bear appropriate experiences relevant to the problem by (b) applying previously acquired scientific principles to the problem at hand.	2, 3
III "...to formulate explanations and hypotheses..."	The student will show his ability to formulate hypotheses and explanations by suggesting	
	(a) hypotheses related to the problem at hand	5, 14
	(b) hypotheses consistent with known data	4
	(c) hypotheses that can be experimentally tested.	15, 16
IV "...to test such explanations and hypotheses by experiment or other means..."	The student will show his ability to test such explanations and hypotheses by	
	(a) deciding what data are necessary to solve the problem	No suitable item available from pre-test data
	(b) selecting the main variables to be experimentally varied or controlled	17
	(f) describing cause and effect relationships.	22

Objective	Behaviour	Question
V "...to accept, modify or reject and to draw conclusions."	The student will show his ability to draw conclusions or to accept, modify or reject them by (b) recognising that sufficient evidence must be collected to form generalisations (i.e. to withhold judgement where necessary).	23, 24
VI "...manipulate scientific equipment correctly..."	The student will show his ability to manipulate scientific equipment correctly by (a) selecting or designing appropriate apparatus to assist in data collection	11
	(c) recognising the measuring limits of certain pieces of scientific equipment	12
	(d) demonstrating his skill in laboratory situations.	Not able to be tested in written test.
VII "...to measure accurately..."	The student will show his ability to measure accurately by (a) demonstrating a knowledge of the cardinal and ordinal aspects of numbers (both positive and negative) in relation to common measurement scales	6, 10
	(b) demonstrating application of knowledge of fractions, percentages and decimals in relation to common measurement scales	8, 9
	(c) measuring using standard units	Not able to be tested in written test
	(d) selecting appropriate units for particular measurements.	7
	(f) recording measurements appropriately and accurately.	13

Objectives	Behaviour	Question
VIII ...to construct and interpret tables, charts and graphs	The student will show his ability to construct and interpret tables, charts and graphs by (a) correctly labelling axes on graphs	20
	(b) choosing appropriate scales for graphs	39
	(c) stating units and headings for a given graph	21, 37
	(d) selecting a line of best fit for a given set of points	No suitable item available from pre-test
	(e) detecting trends in data	19, 40
	(f) using trends to extrapolate and interpolate	38
	(h) constructing from a set of data appropriate graphs or charts.	36
IX ...to find relevant information from available reference sources	The student will show his ability to find relevant information from available reference sources by (c) selecting relevant data from a given table of data.	18
X Observing	The student will show his ability to observe by (a) identifying the objects that interact within a system	25
	(b) identifying differences and likenesses in a wide variety of objects	26
	(c) identifying and describing the results of interactions in terms of initial and final states.	28
XI Classifying	The student will show his ability to classify by (a) dividing a total set of objects into two or more subsets differing in some characteristic	29, 30

Objective	Behaviour	Question
	(b) correctly sub-dividing a set of objects into several categories of increasing or decreasing inclusiveness	31
	(c) classifying likenesses and differences among a variety of different events (e.g. rates of change).	27
XII Comparing	The student will show his ability to compare and contrast by describing in terms of physical properties the similarities or differences between two or more	
	(a) objects	32, 34
	(b) interacting groups of objects	33, 35

APPENDIX C

THE SSAP TEST, INSTRUCTIONS AND STUDENTS' ANSWER SHEET

The figure in front of each of the item distractors is the percentage choice of that distractor.

SCIENCE SKILLS AND PROCESSES TEST

Instructions:

This is a test of how well you can do some of the things that scientists have to do. It is not a test of how well you can remember facts about science. Many of the questions may involve things you have not discussed at school. Do not worry about this but read the question and choose the best answer from the alternatives given Do not spend too much time on any one question. If you cannot think of an answer either guess or leave it blank.

DO NOT BEGIN UNTIL YOU ARE TOLD TO DO SO

PLEASE DO NOT WRITE ON THE QUESTION PAPER

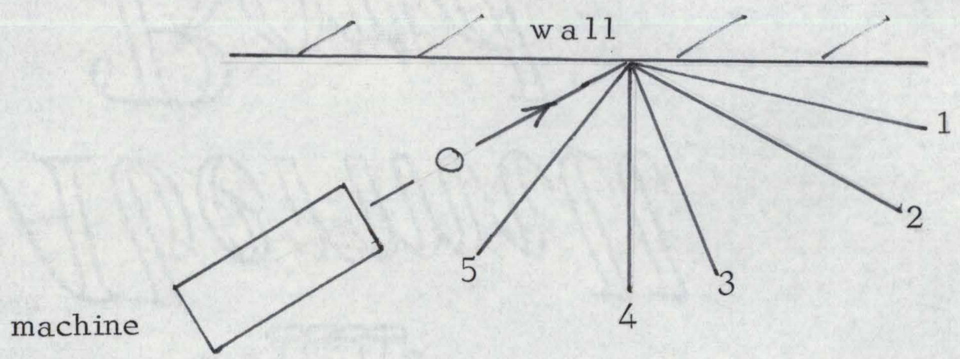
1. A girl throws a basket ball against a very high concrete wall. The ball bounces on the ground. Which of the following is <u>least</u> important to someone studying bouncing?

9.9	A.	What the ball is made of
14.9	B.	What the ground surface is made of
38.6	C.	How high the wall is
15.9	D.	Gravity
19.4	E.	What the wall is made of

2. You can make salt solutions with the same amount of salt in them by adding salt until a fresh egg floats in the solution. This assumes that all fresh eggs have almost the same

11.9	F.	volume
42.1	G.	weight
8.1	H.	shape
33.1	I.	density
3.8	J.	size

Questions 3 and 4 refer to this diagram.



3. A machine throws rubber balls at a wall so that they bounce off. The direction the ball will bound off is best shown by

3.5	A.	1
64.1	B.	2
15.7	C.	3
10.2	D.	4
5.2	E.	5

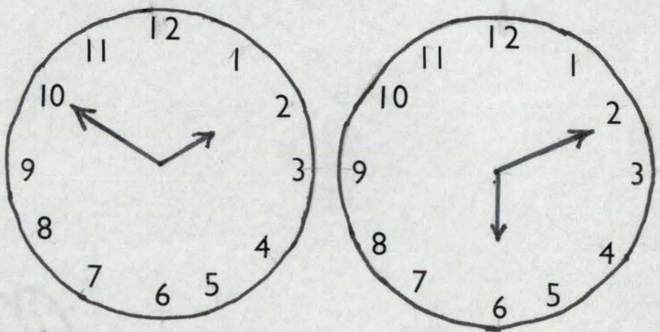
4. The best explanation of the direction the ball makes leaving the wall is

0.6	F.	The balls can be thrown softly
13.6	G.	The path will change whether the balls are fired hard or soft
6.1	H.	The machine can be shifted
22.3	I.	The angles which the ball leaves the wall can vary
55.7	J.	The angle the ball makes with the wall is the same going and leaving.

5. John noted that when the weather was below freezing point his hand was more likely to stick to metal than wooden materials. Which of the following would best explain this?

47.3	A.	Metal conducts heat better than wood
21.2	B.	Metal radiates heat better than wood
3.2	C.	Metal is more shiny than wood
25.8	D.	Metal feels colder than wood
1.2	E.	Metal and wood are both solids.

6. Here is a drawing of a clock at different times.



The time elapsed is

9.8	F.	6 hours 20 mins
4.6	G.	5 hours 20 mins
70.7	H.	4 hours 20 mins
7.3	I.	7 hours 20 mins
3.2	J.	17 hours 20 mins

7. Which of the following units are used tomeasure area?

5.5	A.	metre
28.7	B.	cubic inch
49.0	C.	square metre
10.7	D.	yard
4.3	E.	centimetre

8. A pupil reports a distance measurement as $\frac{16}{100}$ of a metre. What would be the expression as a decimal?

55.1	F.	0.16 m
5.8	G.	1.6 m
6.4	H.	16 m
25.5	I.	0.016 m
5.2	J.	0.0016 m

9. A weather chart showed rain on 7 days of the 28 days in February. What percentage is this?

4.1	A.	75%
0.9	B.	50%
68.7	C.	25%
11.3	D.	20%
11.0	E.	7.5%

10. Which temperature is 25 degrees lower than 15°C?

2.3	F.	15°C
6.1	G.	10°C
71.3	H.	-10°C
9.0	I.	0°C
8.4	J.	-35°C

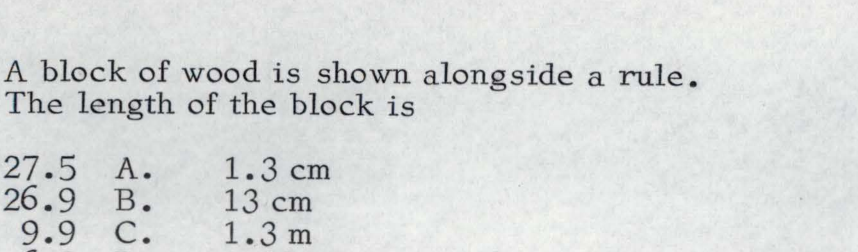
11. You are given (I) stopwatch
(II) weight and string
(III) clock
(IV) tape measure

You are to study the relationship between the length of a pendulum and how long it takes to make one to-and-fro motion. Which things would be best to use?

29.9	A.	I and IV
22.3	B.	I, II, IV
35.4	C.	I, II
2.9	D.	III, IV
7.3	E.	I, II, III, IV

12. Which of the following could <u>not</u> be measured using a mercury thermometer?

17.4	F.	body temperature
18.6	G.	the melting point of ice
24.4	H.	the temperature of a candle flame
24.1	I.	the boiling point of alcohol
13.9	J.	the temperature at which butter melts.

13. 

A block of wood is shown alongside a rule.
The length of the block is

27.5	A.	1.3 cm
26.9	B.	13 cm
9.9	C.	1.3 m
6.9	D.	1.30 cm
26.7	E.	1.3 mm

Questions 14 - 15.
Dennis decided to find out something about glues for a science project.

Dennis got several materials and tried to stick them together with some glue. He tried

1. paper and wood
2. paper and paper
3. wood and wood
4. wood and metal
5. metal and paper

He found that only the first three experiments worked. This is because

14.
2.9 F. metals are solids
2.0 G. the glue was too thin
12.5 H. paper can absorb glue better than metal
4.4 I. paper is a soft material
77.1 J. you need two materials that absorb glue to make a good join.

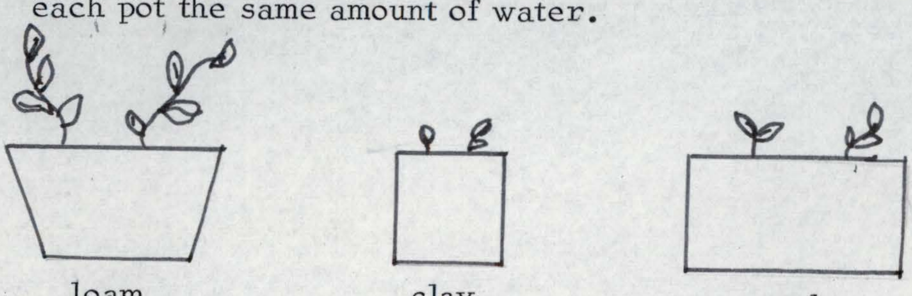
15. If Dennis wanted to find out whether the results he got in Question 14 were true for all glues he should

8.1 A. clamp the metal and wood tightly
17.1 B. try different materials
6.1 C. scratch the metal surfaces first
57.4 D. try a variety of glues
10.2 E. make sure the glue is properly applied.

16. John wondered if sound is able to travel through water. To prove that sound can travel through water he should

2.6 F. Ask his teacher
3.7 G. Hit 2 stones together above water and listen to the sound
58.9 H. Put his head under the water and hit 2 stones together in the water
17.4 I. Put his ear next to the water and hit 2 stones together above the water
16.2 J. Try to read about sound in a science book.

17. Tom wanted to find out which of 3 types of soil would be best for growing beans. He found 3 flowerpots, put a different type of soil in each and planted the same number of beans in each. He placed them side by side on the windowsill and gave each pot the same amount of water.



Why is Tom's experiment not a good one and does not prove prove loam is best?

15.1	A.	the plants in one pot got more sunlight than plants in the other pots.
57.7	B.	the amount of soil in each pot was not the same
9.0	C.	one pot should have been placed in the dark
4.6	D.	Tom should have used 3 kinds of seed
11.9	E.	Tom should know what soil is best without doing experiments.

18. The following table comes from a chemistry book.

Element	Volume(units)	Boiling point	Atomic number
Carbon	5.4	over 4000°C	6
Silicon	11.6	3310°C	14
Germanium	13.3	2840°C	32
tin, white	16.4	2690°C	50
tin, grey	20.6	2690°C	50

Which element has the highest atomic number?

10.2	F.	Carbon
2.3	G.	Silicon
2.0	H.	Germanium
4.4	I.	White tin
79.7	J.	Both white tin and grey tin.

19. Here are some results Rangi obtained when he measured height and time in an experiment.

Height (cm)	12	7	4	3
Time (secs)	1	2	3	4

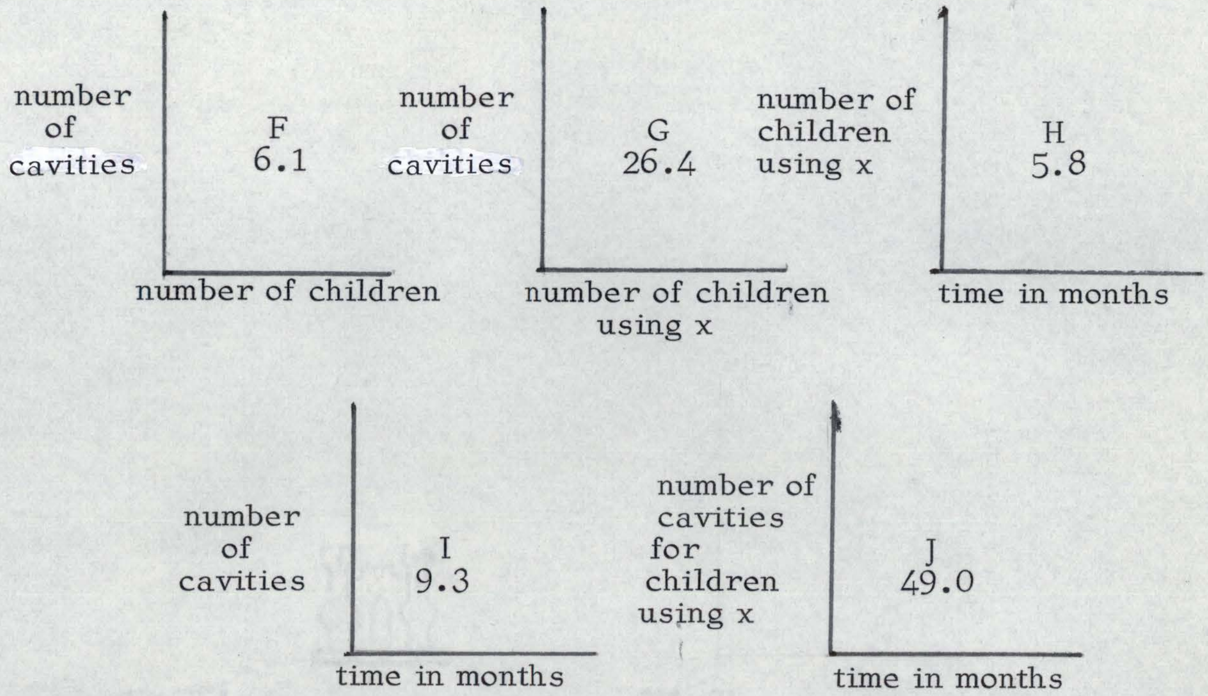
What is the best statement you could make about these results?

3.2	A.	The height increases with time increase
64.4	B.	The time increases with height decrease
3.2	C.	Time and height remain the same
8.1	D.	As the time decreases the height decreases
20.0	E.	There is no relationship between height and time.

Questions 20 - 21

A toothpaste manufacturer decides to test his new toothpaste "x". He supplies his toothpaste to half of the children in a small town, ensures that all children eat similar food and asks the town dentists to keep a note of the number of cavities in children's teeth for 12 months. The results are pleasing.

20. Which graph would be best to indicate the results of the experiment?



21. What would be a suitable heading for the graph?

13.1 A. "Toothpaste x really works wonders"
37.4 B. No. of cavities for children who used toopaste x
3.5 C. No. of children who used toothpaste x
3.8 D. No. of cavities for children
41.2 E. Toothpaste comparison graph.

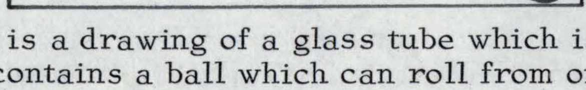
22. A student, on observing a bean seed after a six day interval notices that it has a root and a shoot. Previously no root or shoot were visible. What is the most likely cause of the change in the seed?

32.5 F. Light
6.1 G. Fertilizer
13.9 H. Heat
22.0 I. Water
24.4 J. Plant Food

23. A scientist is open-minded about his work if he

29.0 A. Discusses most of his ideas with others
50.7 B. Considers ideas which go against his own
7.8 C. Thinks up many new ideas for experiments
8.1 D. Agrees with the ideas of other scientists
2.9 E. Reads widely.

24.



This is a drawing of a glass tube which is sealed at both ends and contains a ball which can roll from one end to the other.

If you want to find the average time it takes for the ball to roll down the tube when it is tipped vertically (upright), about how many times should you time it.

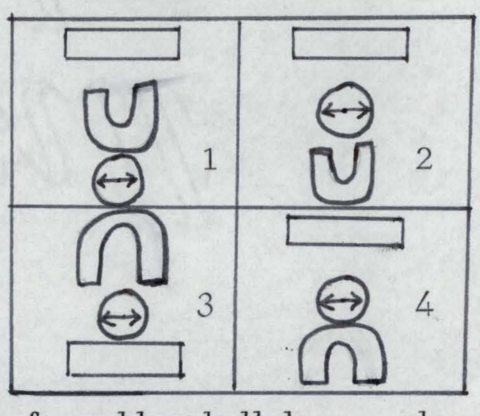
9.3	F.	1
41.5	G.	2
35.7	H.	15
6.1	I.	250
5.5	J.	5,000

25. You have a closed box with a small hole in one corner. You can smell an odour. 1 hour later the same smell persists. What is the best explanation?

7.3	A.	It is either a liquid or a gas but not a solid
3.7	B.	It is either a liquid or something that rapidly becomes a liquid
15.4	C.	It is either a solid or a liquid but not a gas
36.3	D.	It is either a gas or something that rapidly gives off a gas
35.4	E.	It is too difficult to say what is in the box.

26. This is a set of 4 drawings. Each drawing shows a compass, bar magnet, and horseshoe magnet. In which two drawings are the three things arranged in the same way?

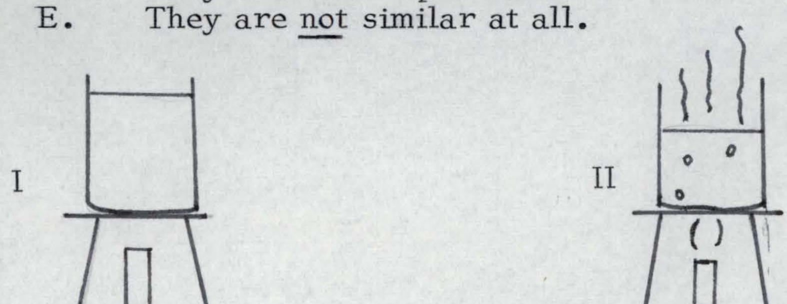
1.7	F.	1 and 3
15.7	G.	2 and 4
71.3	H.	2 and 3
2.3	I.	1 and 4
7.3	J.	1 and 2



27. In what way is the motion of a rubber ball down a slope and a plant growing similar?

2.6	A.	They are both living materials
25.5	B.	They are both constantly changing
16.5	C.	They can both be stopped
4.6	D.	They are both experiments to do with temperature
48.4	E.	They are not similar at all.

28.

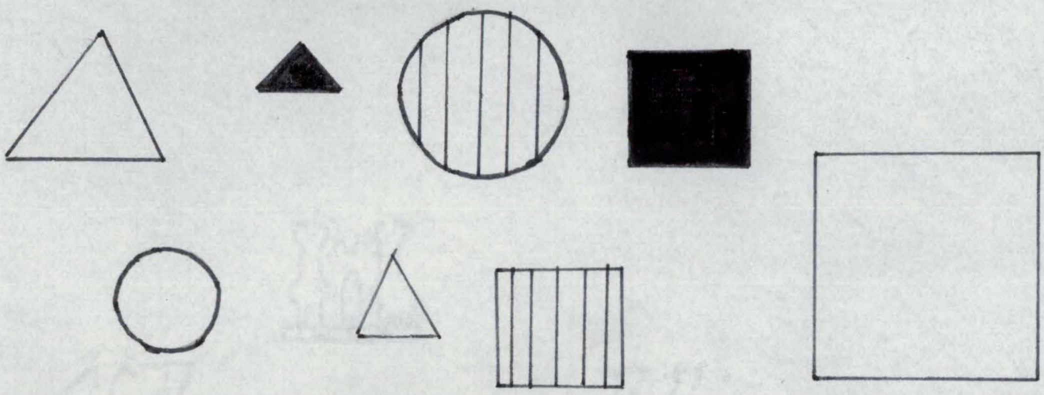


The drawings are of a beaker of water on a tripod. Drawing

II was made 10 minutes after drawing I. Which choice is the best way of telling that there has been a change?

10.2	F.	The water is boiling in drawing II
5.2	G.	The gas is on in drawing II
22.0	H.	The water gets hot when the gas is on
1.7	I.	The water is not boiling in drawing I
58.9	J.	The water is boiling in drawing II but not in drawing I.

Questions 29 - 30 refer to these figures



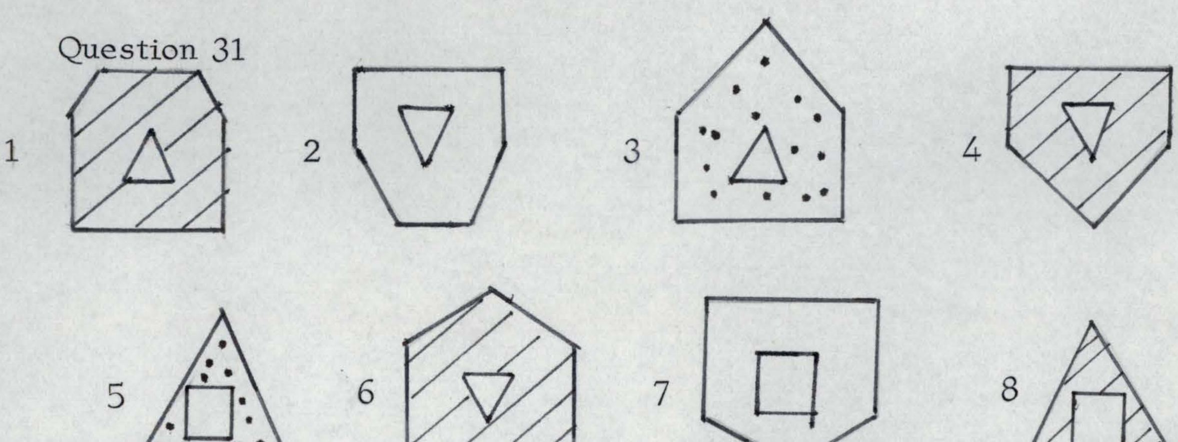
29. If you are to group the 8 figures by the amount of shading what is the smallest number of groups you would form?

6.1	A.	1
37.4	B.	2
39.4	C.	3
11.9	D.	4
3.2	E.	5

30. If the 8 figures were grouped by shape what is the smallest number of groups you would form?

7.8	F.	1
38.0	G.	2
45.2	H.	3
4.9	I.	4
1.5	J.	5

Question 31

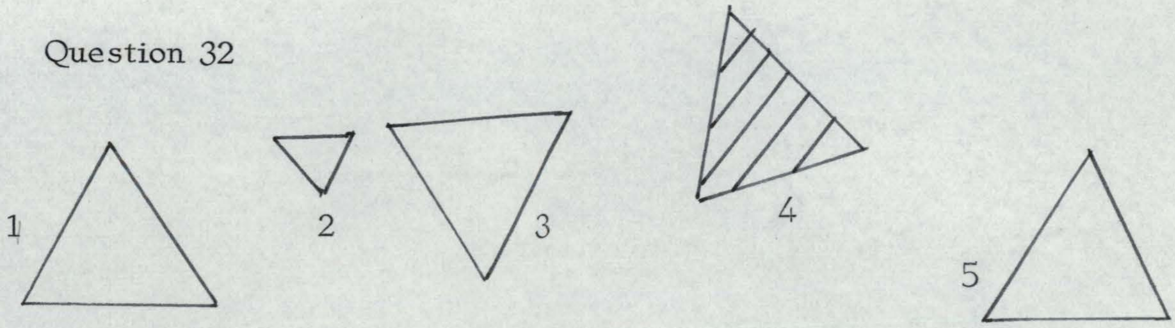


31. Which choice includes only the striped figures with a

triangular hole?

56.5 A. 1, 4, 6
5.8 B. 1, 2, 3, 4, 6
1.5 C. 5, 8
30.7 D. 1, 4, 6, 8
3.2 E. 4, 6

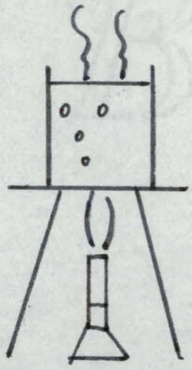
Question 32



32. Which of the following tells you <u>exactly</u> how these triangles are different?

11.3 F. 4 is a different shading
5.5 G. 2 is smaller
49.9 H. 2 is smaller and 4 is a different shading
22.0 I. 1, 3, 4, 5 are the same size
9.0 J. 4 and 2 are different from each other.

Question 33



beaker of water

candle

33. In which way are these two events the same?

51.0 A. Something is burning in both and heating something else
4.9 B. There is a solid burning in both
9.6 C. There is a liquid burning in both
13.1 D. There is no way in which the two are the same
18.0 E. They are both experiments.

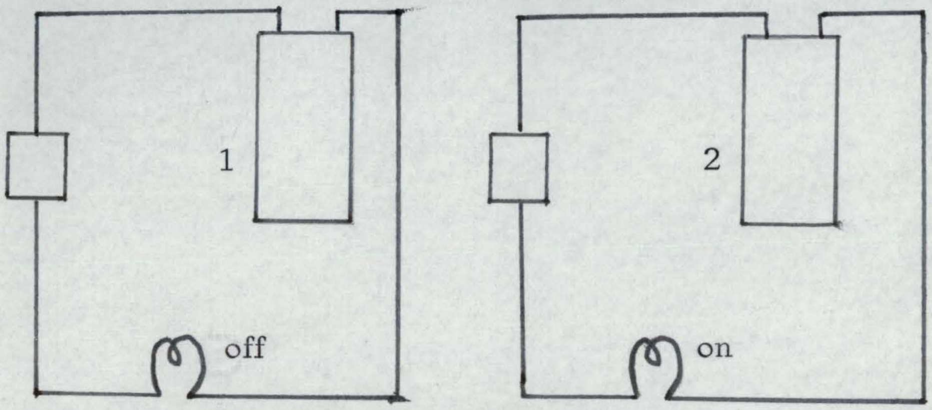
34.

Animal I	Animal II	Animal III	Animal IV	Animal V
hair	feathers	hair	scales	shell
backbone	backbone	backbone	backbone	muscular foot
claws	claws	no claws	claws	no claws

The animal that could be put in a group by itself is

4.9	F.	I
6.9	G.	II
5.5	H.	III
9.0	I.	IV
69.9	J.	V

35. These are two drawings of a battery, a bulb and a switch and some wires. Which is the only thing you can be sure is different between drawing 1 and drawing 2.



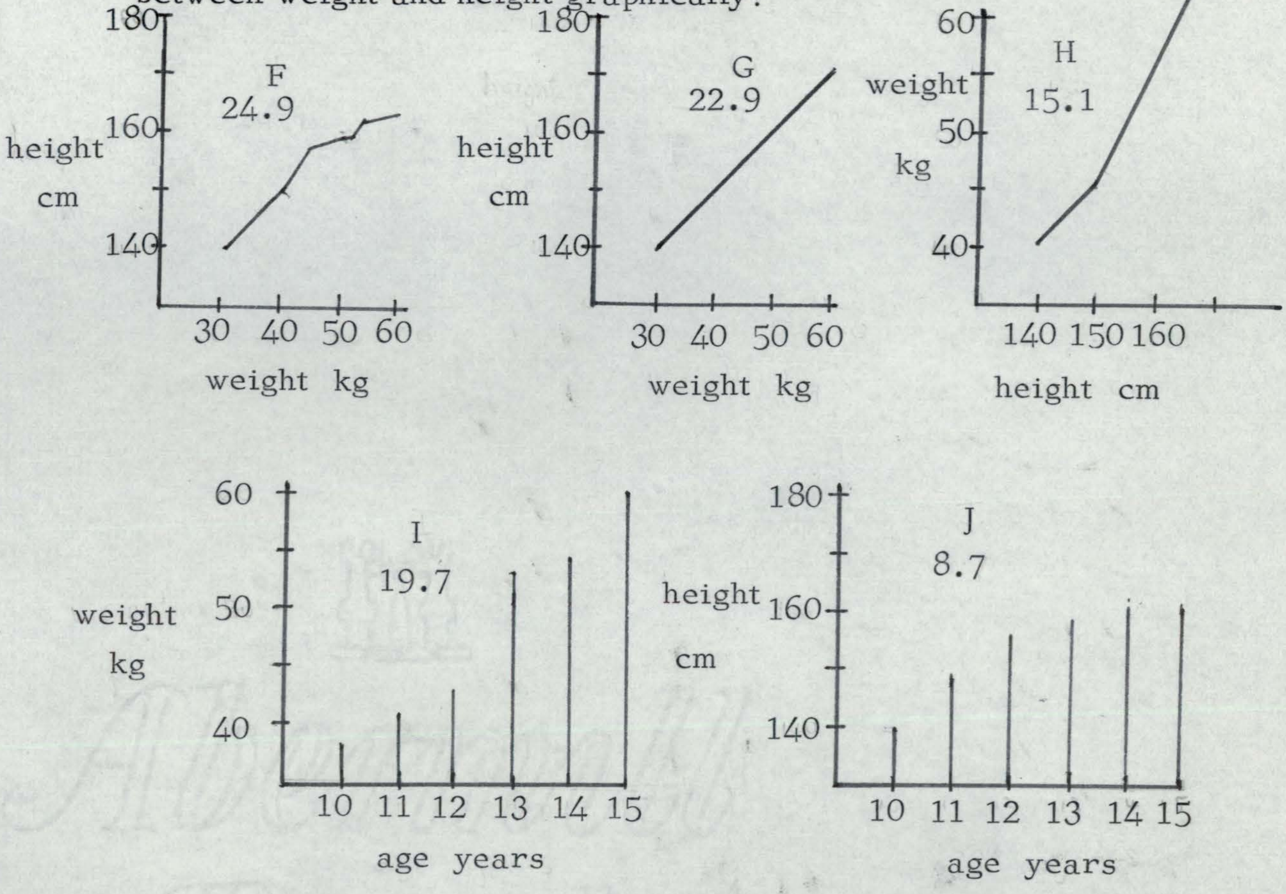
3.8	A.	The bulb was replaced
11.6	B.	The wires were tightened
4.4	C.	The bulb was screwed in
7.8	D.	The battery was recharged
66.4	E.	Electricity is flowing through the bulb.

Questions 36 - 39

The following table gives the ages, average weights and average heights of girls in New Zealand in 1970.

Age	Height (cm)	Weight (kg)
10	140	32
11	148	41
12	155	45
13	158	53
14	161	55
15	162	60

36. Which would be a suitable way of showing the relationship between weight and height graphically?



37. What would be a suitable heading for the graph?

1.5	A.	New Zealand girls' height
3.5	B.	New Zealand girls' weight
68.7	C.	The relationship between weight and height of girls
11.6	D.	The age of New Zealand girls and their height
8.1	E.	The taller you are the more you weigh.

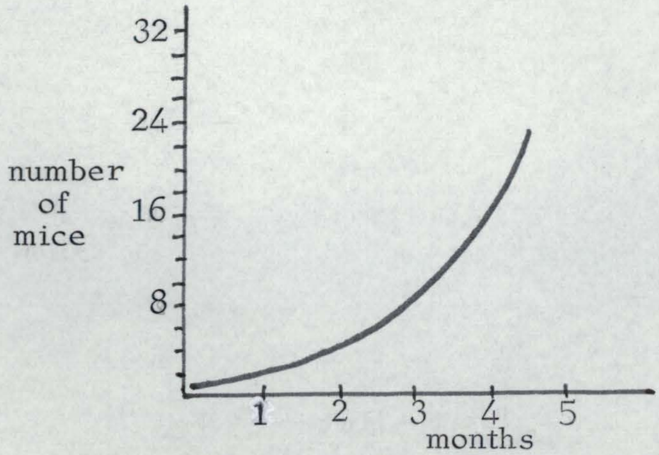
38. What would a girl 150 cm tall probably weigh?

12.5	F.	41 kg
6.7	G.	35 kg
50.8	H.	42 kg
11.9	I.	45 kg
9.3	J.	53 kg

39. If you wished to draw another graph showing the relationship between age and height what would you choose for the scales?

30.7	A.	Age in years and height in cm
19.4	B.	Age and height
15.1	C.	Age in years and weight in kilograms
13.3	D.	Weight and height
12.5	E.	Weight in kilograms and height in centimetres.

Question 40 is based on this graph



40. Every month the population of mice

38.3	F.	doubles
27.8	G.	increases by 50%
6.1	H.	grows to the limit of the food supply
5.5	I.	grows more slowly than the month before
13.3	J.	increases by 75%

ANSWER SHEET

NAME: .... **SCHOOL:** .... **CLASS:** ....

Instructions: Put a circle around the letter indicating which you think is the best answer to each question. (The number alongside the letter is simply to assist in scoring.)

Example: [101] A_1 B_2 C_3 (D_4) E_5

Do not mark more than ONE letter. If you wish to change your answer, show the error clearly with a cross before marking your new choice.

Example: [102] F_1 ~~G_2~~ H_3 I_4 (J_5)

-----

(1)	A_1	B_2	C_3	D_4	E_5	(21)	A_1	B_2	C_3	D_4	E_5
(2)	F_1	G_2	H_3	I_4	J_5	(22)	F_1	G_2	H_3	I_4	J_5
(3)	A_1	B_2	C_3	D_4	E_5	(23)	A_1	B_2	C_3	D_4	E_5
(4)	F_1	G_2	H_3	I_4	J_5	(24)	F_1	G_2	H_3	I_4	J_5
(5)	A_1	B_2	C_3	D_4	E_5	(25)	A_1	B_2	C_3	D_4	E_5
(6)	F_1	G_2	H_3	I_4	J_5	(26)	F_1	G_2	H_3	I_4	J_5
(7)	A_1	B_2	C_3	D_4	E_5	(27)	A_1	B_2	C_3	D_4	E_5
(8)	F_1	G_2	H_3	I_4	J_5	(28)	F_1	G_2	H_3	I_4	J_5
(9)	A_1	B_2	C_3	D_4	E_5	(29)	A_1	B_2	C_3	D_4	E_5
(10)	F_1	G_2	H_3	I_4	J_5	(30)	F_1	G_2	H_3	I_4	J_5
(11)	A_1	B_2	C_3	D_4	E_5	(31)	A_1	B_2	C_3	D_4	E_5
(12)	F_1	G_2	H_3	I_4	J_5	(32)	F_1	G_2	H_3	I_4	J_5
(13)	A_1	B_2	C_3	D_4	E_5	(33)	A_1	B_2	C_3	D_4	E_5
(14)	F_1	G_2	H_3	I_4	J_5	(34)	F_1	G_2	H_3	I_4	J_5
(15)	A_1	B_2	C_3	D_4	E_5	(35)	A_1	B_2	C_3	D_4	E_5
(16)	F_1	G_2	H_3	I_4	J_5	(36)	F_1	G_2	H_3	I_4	J_5
(17)	A_1	B_2	C_3	D_4	E_5	(37)	A_1	B_2	C_3	D_4	E_5
(18)	F_1	G_2	H_3	I_4	J_5	(38)	F_1	G_2	H_3	I_4	J_5
(19)	A_1	B_2	C_3	D_4	E_5	(39)	A_1	B_2	C_3	D_4	E_5
(20)	F_1	G_2	H_3	I_4	J_5	(40)	F_1	G_2	H_3	I_4	J_5

APPENDIX D

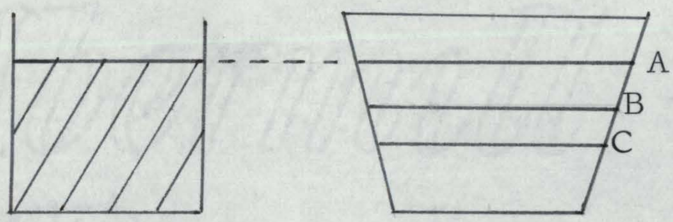
SCIENCE CONCEPT TEST

(after King, 1963)

SCHOOL . . . .

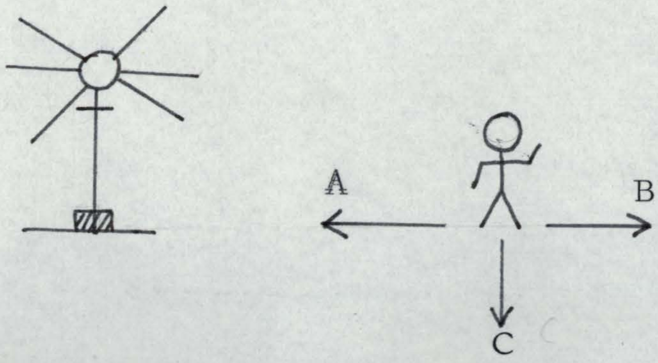
NAME . . . . FORM . . . .

AGE . . . . .

		Qu.	Answer
A1.	Estimate the time in seconds between two taps.	1	______
B1.	A jar is partly filled with water. A stone is gently dropped into the water. Is the level of the water now higher, lower or the same as it was before?	2	______
B3.	 Look at the diagrams. The first represents a glass of water. All the water is poured into the second glass. If the bottoms of both glasses are the same size, will the new level of water be at A, B or C?	3	______
B7.	Imagine two pieces of plasticine which have the same size and weight. One lump is rolled into the shape of a pencil. Will the two pieces now have the		
	(a) same weight? (yes, no?)	4	______
	(b) same volume? (yes, no?)	5	______
C1.	Ice turns into water, sugar dissolves in water.		
	Does steam turn into water? (yes, no?)	6	______
	Does salt turn into water? (yes, no?)	7	______
	Does hail turn into water? (yes, no?)	8	______
	Do clouds turn into water? (yes, no?)	9	______

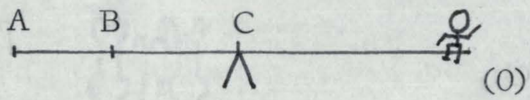
Qu.	Answer

C2.



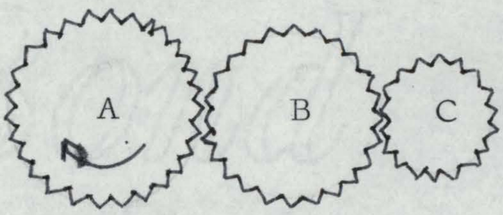
The drawing shows a man standing in a road in the light of a lamppost. Will the man's shadow fall along the lines A, B or C? 10 ____________

C3



Here is a diagram of a seesaw with a girl sitting at one end (O). If a boy is twice as heavy as the girl where would he sit to balance the seesaw - at A, B or C? 11 ____________

C6



Here are three cog wheels A, B, C. Which cog wheel turns in the same direction as A. Is it B, C or neither? 12 ____________

Which cog wheel turns the fastest? Is it A, B, or C? 13 ____________

D1. Is a flower alive? (yes, no, don't know) 14 ____________

D3. Are all things that move living? (yes, no, don't know) 15 ____________

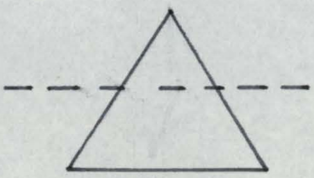
E2. A needle is stuck upright in a cork. The cork floats on water. Draw two lines showing position of cork and needle. 16 ____________

E4. A circle cut from cardboard is held edge on to an electric light. The shadow of this cardboard falls on a screen. Draw the shadow you would see on the screen. 17 ____________

	Qu.	Answer
E5. A rubber ball is cut through the middle. Draw the shape made by the cut.	18.	
E5f. The top of a cone is cut off as in the diagram. Draw the shape made by the cut.	19	
E5g. The cone is cut in half as shown in the diagram. Draw the shape made by the cut.	20	

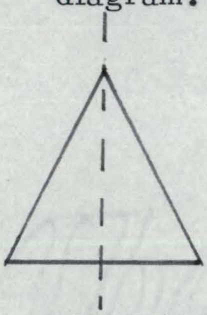
E5. A rubber ball is cut through the middle. Draw the shape made by the cut.

E5f. The top of a cone is cut off as in the diagram.



Draw the shape made by the cut.

E5g. The cone is cut in half as shown in the diagram.



Draw the shape made by the cut.

<u>Give reasons for answers to any of the questions that interest you.</u>

<u>In each case, give the number of the question.</u>

APPENDIX E

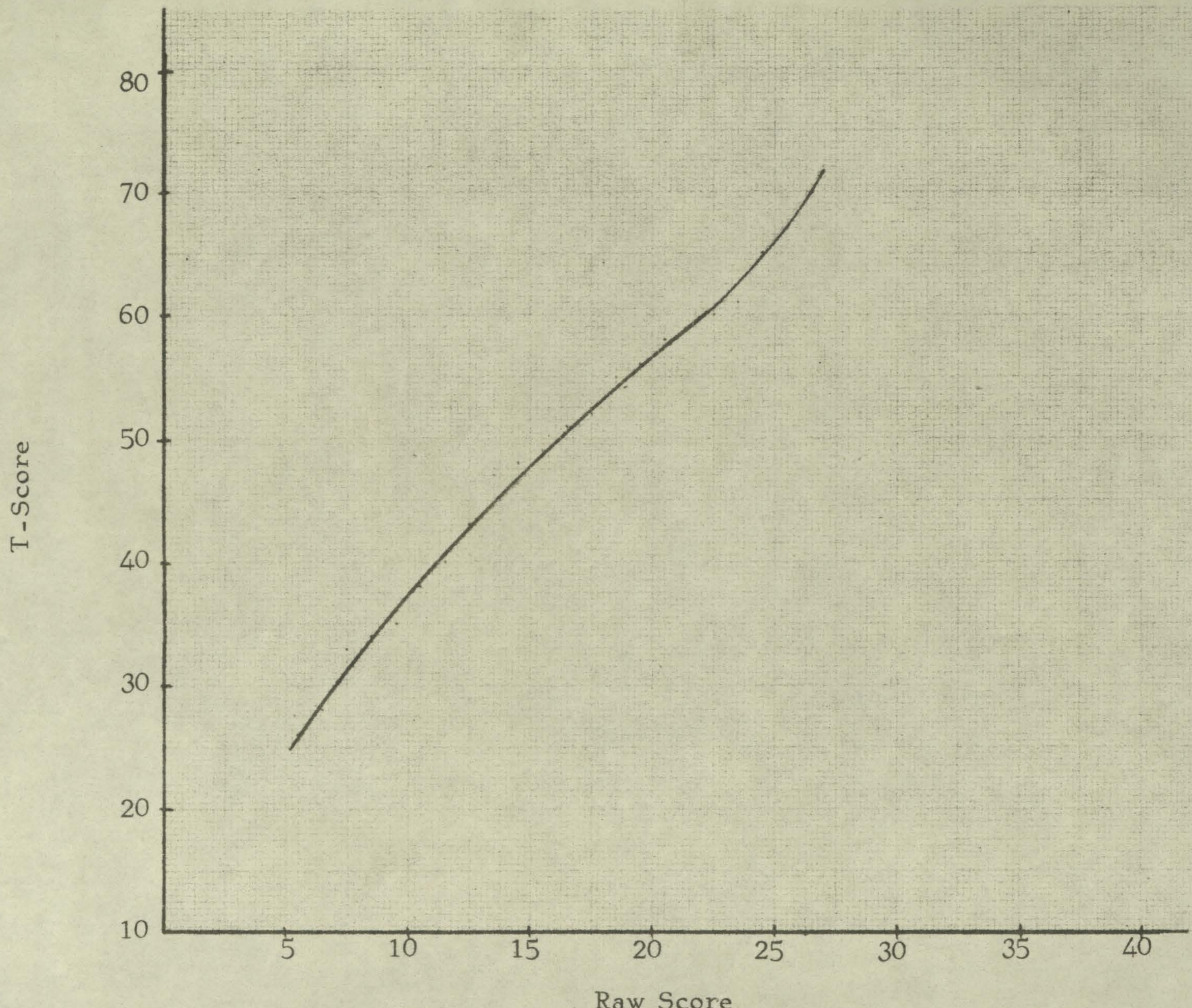
PERCENTILE RANKS FOR RAW SCORES ON THE SSAP TEST

	Form 2	Form 3	Form 4	All Forms
Raw Score	Percentile Rank	Percentile Rank	Percentile Rank	Percentile Rank
5	1			1
6	2			2
7	3			3
8	6			4
9	8			5
10	10	1	2	6
11	12	2	4	9
12	21	2	8	11
13	25	5	9	16
14	38	8	10	22
15	43	11	14	26
16	52	15	17	32
17	58	20	21	37
18	66	25	23	44
19	73	30	26	50
20	78	36	32	54
21	81	42	38	59
22	86	48	48	65
23	90	55	58	70
24	93	63	67	74
25	95	68	70	78
26	96	75	76	83
27	97	80	81	87
28	98	85	85	91
29		90	88	94
30		93	93	95
31		94	94	96
32		95	95	97
33		97	97	98
34		99	98	99
35			99	100
36				
37				
38				
39				
40				

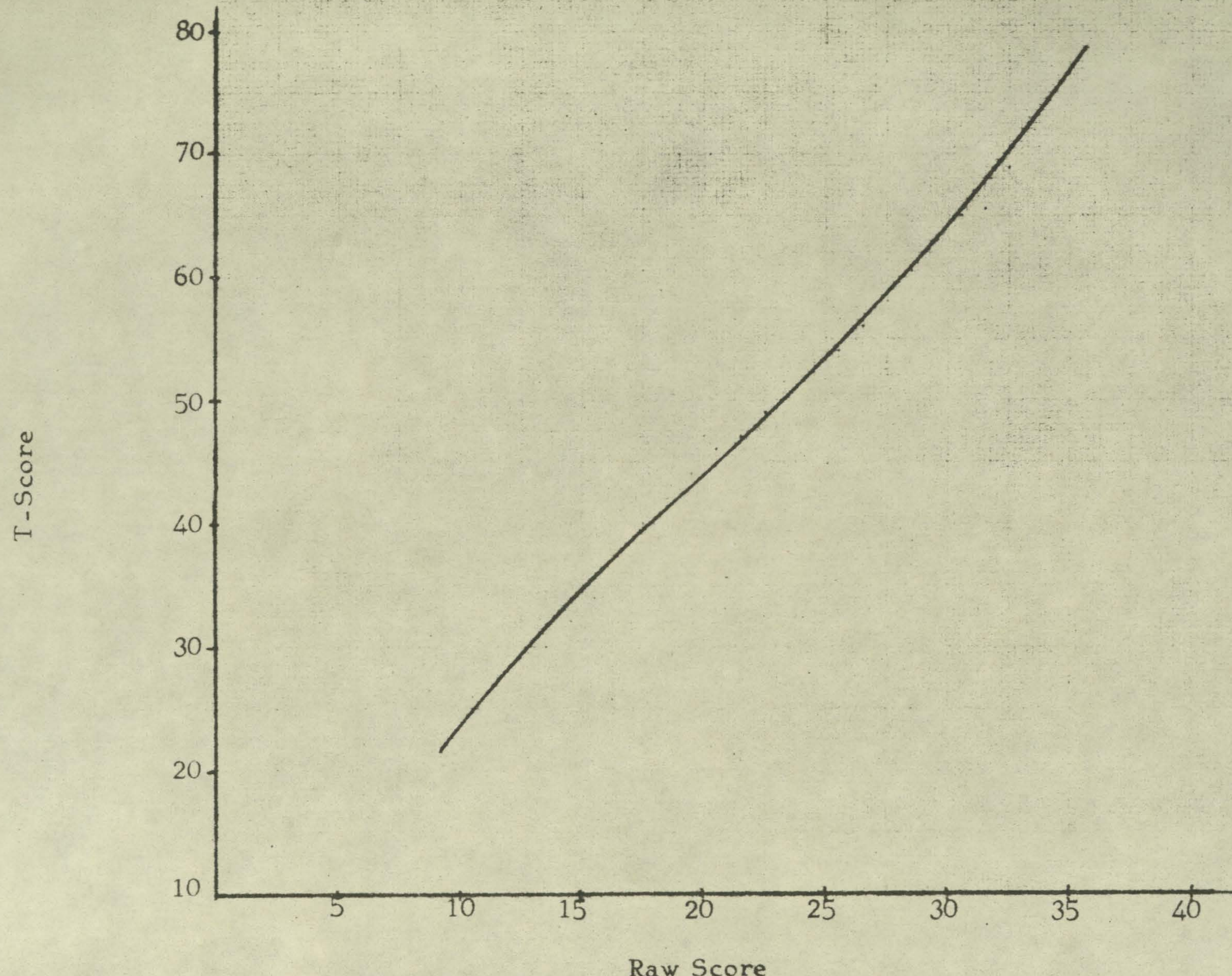
APPENDIX F

Ogives for separate classes, Forms 2 - 4, and for all classes combined which were used to construct the conversion table (Table 10) whereby raw scores may be converted to T-scores.

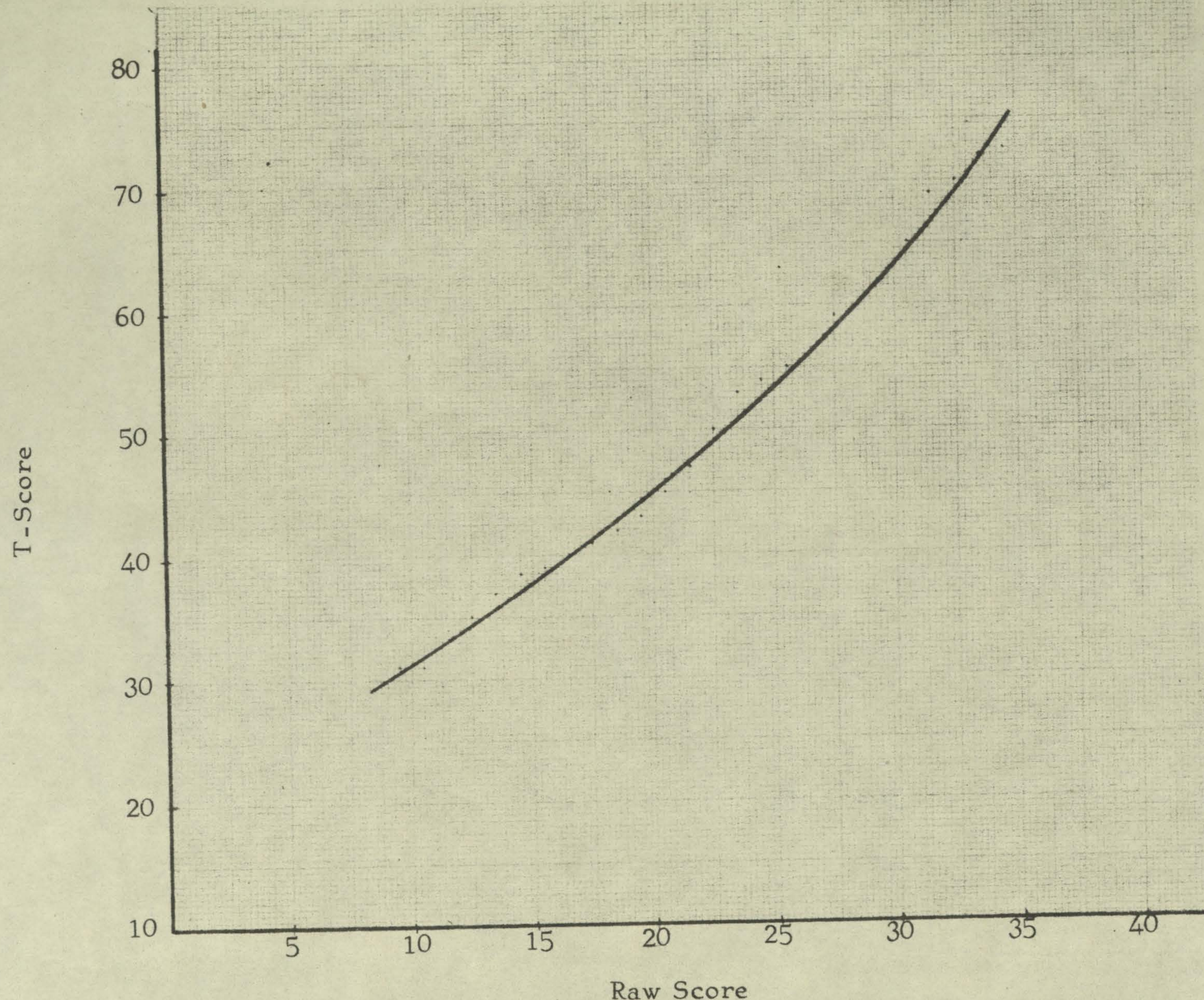
OGIVE for SSAP Test - Form 2



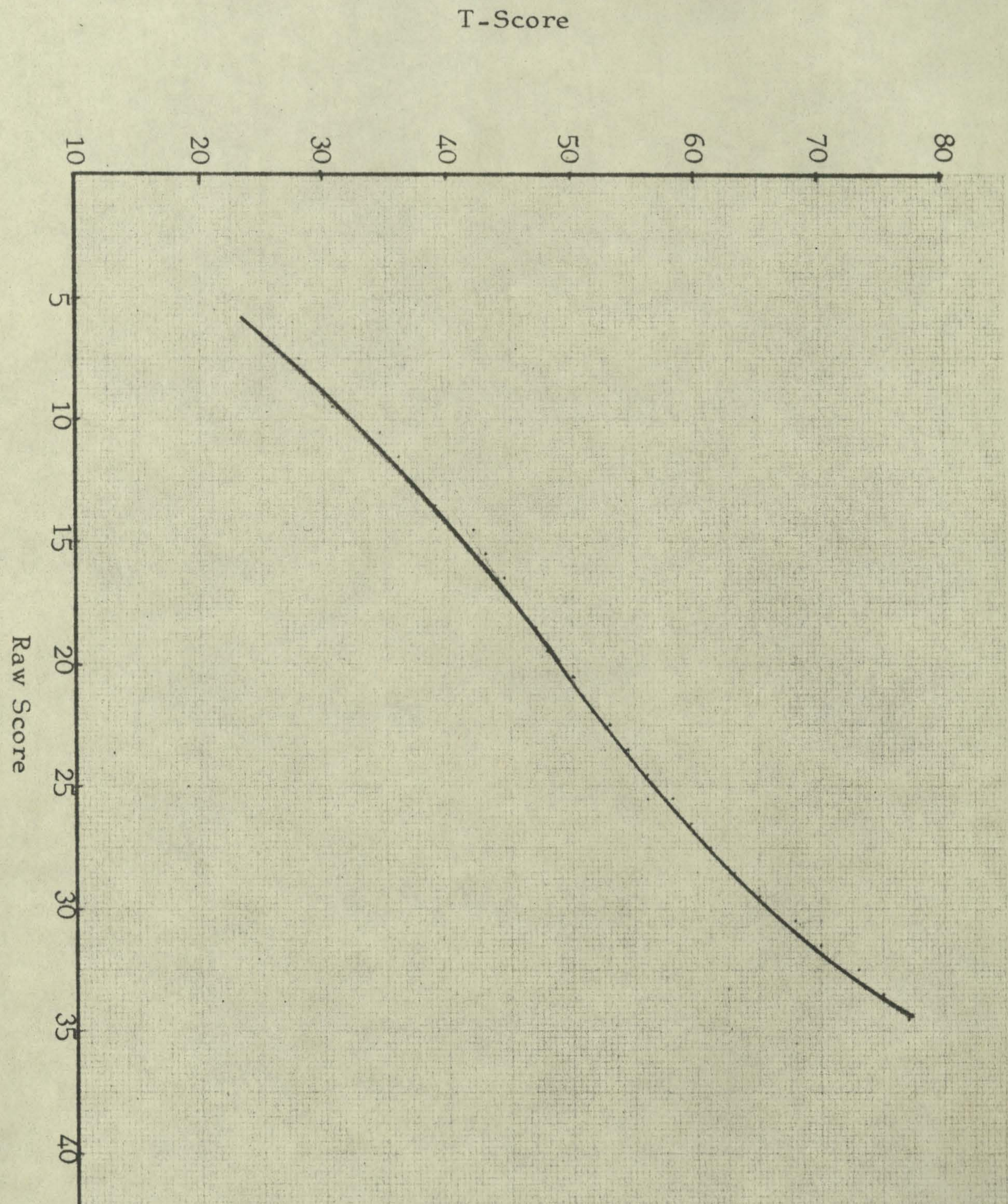
OGIVE for SSAP Test - Form 3



OGIVE for SSAP Test - Form 4



OGIVE for SSAP Test - All Classes



APPENDIX G

DIFFICULTY AND DISCRIMINATION INDICES FOR THE SSAP TEST

Item Number	Difficulty Index Percentage	Discrimination Index
1	39	0.14
2	38	0.45
3	63	0.40
4	55	0.55
5	42	0.20
6	74	0.27
7	51	0.49
8	58	0.32
9	68	0.53
10	64	0.50
11	27	0.29
12	28	0.20
13	36	0.48
14	75	0.33
15	59	0.55
16	60	0.30
17	53	0.54
18	80	0.20
19	60	0.42
20	48	0.36
21	34	0.43
22	24	0.27
23	51	0.59
24	35	0.26
25	34	0.23
26	71	0.39
27	26	0.20
28	62	0.33
29	47	0.47
30	50	0.48
31	56	0.36
32	53	0.57
33	49	0.48
34	68	0.41
35	63	0.53
36	26	0.25
37	68	0.40
38	49	0.40
39	33	0.52
40	38	0.28

APPENDIX H

DIFFICULTY AND DISCRIMINATION INDICES FOR THE SSAP SKILLS SUBTEST

Item Number	Difficulty Index Percentage	Discrimination Index
6	71	0.35
7	47	0.57
8	55	0.33
9	64	0.63
10	69	0.42
11	29	0.32
12	26	0.24
13	35	0.35
18	81	0.32
19	63	0.42
20	49	0.43
21	39	0.42
36	27	0.33
38	49	0.57
39	36	0.61
40	36	0.42

APPENDIX I

DIFFICULTY AND DISCRIMINATION INDICES FOR THE SSAP PROCESSES SUBTEST

Item Number	Difficulty Index Percentage	Discrimination Index
1	39	0.32
2	35	0.37
3	64	0.49
4	55	0.61
5	44	0.32
14	73	0.36
15	60	0.55
16	61	0.35
17	55	0.57
22	23	0.23
23	50	0.63
24	36	0.26
25	33	0.29
26	72	0.40
27	26	0.23
28	58	0.32
29	48	0.46
30	51	0.52
31	57	0.39
32	51	0.53
33	47	0.47
35	61	0.60

APPENDIX J

A comparison of the science concept attainment of 13 year-old students

The responses of the 62 13 year-old (± 3 months) Form 2 children to each of the twenty questions in the Science Concept Test were collated and a percentage correct response (p.c.r.) was determined for each question. These p.c.r. are plotted on the accompanying graph. Data from King's (1963) research (82 13 year-old pupils) was added for a comparison.

It is clear that there are only minor differences between the two groups - an exception appears to be question 18 which asks the children to draw the shape made when a ball is cut.

A comparison of science concept attainment of 13 year-olds