A PRELIMINARY CONSTRUCTION AND VALIDATION OF A TEST ON THE NATURE OF SCIENCE AND SCIENTIFIC THINKING

STEPHEN PATRICK NORRIS

[illegible]

A PRELIMINARY CONSTRUCTION AND VALIDATION OF A TEST ON THE NATURE OF SCIENCE AND SCIENTIFIC THINKING

A Thesis

Presented to

the Faculty of Education

Memorial University

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

by

Stephen Patrick Norris

April, 1975

ACKNOWLEDGEMENTS

The author wishes to thank Professor Robert K. Crocker for his dedicated and informed assistance given freely throughout the entire writing of this thesis. However, special thanks are in order for the many hours he struggled with contrary computer programs during the analysis period.

This thesis is dedicated to Betty, my wife-to-be, who while typing the thesis became as committed to its completion as I. Without her patience, encouragement, ability to spell and punctuate, the thesis would not have progressed so far, so soon.

ABSTRACT

This study concerned itself with the construction and validation of a test on the nature of science and scientific thinking, to be used with grade X and grade XI high school students and with grades I to XI science teachers. The instrument was grounded on the theoretical framework specified in Bloom's Taxonomy of Educational Objectives: Cognitive Domain and upon a model of the nature of science gleaned from the writings of several philosophers of science and compiled by the author. This theoretical grounding gave support to the instrument's content validity.

As an additional check on validity, a preliminary form of the test, containing ninety-one four alternative multiple-choice items, was given to a panel of ten validators, including practicing scientists, science educators, and an epistemologist. Using the results of their analysis, sixteen of the items were excluded from the test, several were modified, and an answer key was devised based upon 70 percent agreement among the panel members. The form of the test which resulted was administered to approximately two hundred fifty grade X and grade XI students and forty-five grades I to XI science teachers.

Both item and factor analyses of the student's responses indicated that the most plausible interpretation of the manner in which they answered the test was that their responses were essentially random. Such a result was suspected from the beginning, since the students had never received instruction in the content examined by this

instrument. However, the item analysis results were still helpful in the rewriting of several test items. Also, the item analysis of the teacher responses indicated that they performed significantly better than the students at the 0.001 level of significance.

While the factor analysis results reflected the random response given by the students, an attempt was made to discover to what degree the hypothesized factor structures for the instrument matched the factors which were extracted from the students responses. It was believed that more interpretable results would have been received in this regard had the students been instructed in the content examined by the test.

TABLE OF CONTENTS

LIST OF TABLES

Chapter I

THE PROBLEM

Introduction

Mathematics, which deals in part with the manipulation of numbers, and English Grammar, which concerns itself with the conventional usage of words, have very seldom been indicted on the grounds of irrelevency to the school curriculum. This fact would seem to stem from the observation that the graduated high school student, who may never open another book to study for his remaining years, must still possess some facility with numbers and with words, if he is to function effectively in present day society.

Science, on the other hand, has fared differently. While some knowledge of the content of science most certainly increases a person's copeability in this highly "techno-scientific" world, it can be argued that enough "common sense science" can be learned through personal experience for one to go through life unimpeded. In addition, many educators have felt that the applications of science are what should be taught, as these can be of practical use to students, while the theoretical framework should be left to the scientists.

To counteract these ideas, many scientists have claimed that science should be taught for science's sake. Their contention is that the study of science is an inherently rewarding activity, irrespective of its applications. However, these scientists unfortunately misconceive the general public's attitudes towards science as being congruent with their own.

Realizing that the logic of proposing to teach science for its own, intrinsic value was not satisfying to the layman, many scientists and science educators analyzed the structure and nature of science looking for components of more service to the average student than quickly forgotten and transitory facts, theories, and laws. Elements were identified which pedagogically seemed to be of benefit to the non-scientist, since theoretically they were on the one hand more transferable to other areas of knowledge, and on the other hand more retainable facets of the scientific endeavour. Examples of things identified as being transferable are certain intellectual traits essential to the perpetuation of the scientific enterprise, such as the habit of making careful and accurate observations, the habit of holding views tentatively, the habit of basing judgment on fact, the habit of intellectual honesty instead of exaggeration and rationalism, the habit of criticalness as well as of self-criticism, the belief that the world can be understood in rational terms, and the belief that the acceptance or rejection of an idea cannot be based on the personal qualities of its protagonist.

Cited as being more retainable are certain understandings of the nature of science and scientific processes, such as the ability to distinguish fact from theory, an awareness that the laws and theories of science are approximations to the truth and are hence subject to change, a knowledge of the arbitrariness of definitions and schemes of classification found in science, and a capacity to distinguish between the aims and products of science and the aims and products of technology.

That these aspects of science are recognized as significant is manifested in an extensive study conducted by Pella, O'Hearn, and Gale

(1966). In this research one hundred documents, found in three science journals and three science education journals, published during the period 1950 to 1964 inclusive, were found to contain referents to scientific literacy. TABLE I indicates the referents identified and their frequency of appearance in these one hundred articles (Pella, et. al., 1966, p. 200). Of the six frames of reference mentioned, the study being conducted by this investigator is concerned with three: the "Ethics of science", with regards to the aims, purposes, and methods of science; the "Nature of science", referring to science as an evolving, idea generating activity rather than as simply a body of knowledge; and "Science and technology", referring to the distinctions and coordination between these two fields.

TABLE I

Referents for Scientific Literacy Found in 100 Documents

Referent	Frequency
Science and society	67
Ethics of science	59
Nature of science	51
Conceptual knowledge	26
Science and technology	21
Science and humanities	21

Many steps have to be taken before a curriculum is available which can teach these facets of the nature of science. Textual materials, teaching strategies, and evaluation techniques are in need of development. Some progress has been made in these areas, but a much stronger thrust is required. The aspect of this problem undertaken by the present study is described in the next section.

Statement of the Problem

This study concerned itself with the construction of a valid, reliable, and usable test for assessing certain aspects of the nature of science and scientific thinking. The instrument was constructed for use with grades X and XI high school students and grades I to XI science teachers. The items were based on a comprehensive model of the nature of science derived from writings in the philosophy of science and from the models used to develop previous instruments in this area. Also, items were written to sample as wide a range as possible of the categories contained in the Taxonomy of Educational Objectives, Handbook I: Cognitive Domain (Bloom, et. al., 1956). This technique allowed for the evaluation of the examinees at levels requiring a mere recall of specific facts through to levels necessitating an analysis and evaluation of situations based upon this knowledge.

Besides the main problem of constructing the instrument, two secondary problems also arise as a consequence of the first. These are related to designing a model of the nature of science and to classifying items into Bloom's cognitive levels. Both these problems, alluded to here, will be addressed in detail in later sections.

Justification for the Study

The construction of this instrument is defensible when it is observed that a trend has started toward process oriented science teaching in primary and elementary school and toward high school science courses which stress facets of the nature of science rather than just conceptual knowledge. If teachers are to operate effectively in such curricula, they must possess an adequate knowledge of the courses' underlying philosophy. An application of this instrument to a sample of science teachers may indicate deficiencies in their comprehension of the categories tested. These results would have implications for science teacher education programs.

Also, the model upon which the test was constructed can be used as a blueprint for developing a science curriculum which specifically directs itself to teaching the components of science inventoried by the test. The instrument could then be used as a model for designing other tests to examine students' achievement in these courses.

Used in a survey, the instrument could also be used for diagnostic purposes to indicate students' weaknesses and strengths in this area of knowledge. Such information would be of much import when making decisions upon teaching strategies and course selection.

Nor is this instrument a redundance of existing tests. Firstly, it does not evaluate certain aspects of the nature of science that some other instruments do, for example, the relationship between science and society or the role played by scientific societies and journals. Secondly, the dimensions spanned are a synthesis of those discovered in several instruments. Thirdly, whereas prevailing tests contain few items requiring intellectual abilities and skills on the part of the examinees, but concentrate primarily on knowledge level or opinionative questions, the instrument developed in this study requires mental operations at all levels of Bloom's cognitive taxonomy.

Chapter II

REVIEW OF RESEARCH ASSOCIATED WITH THE PROBLEM

Since it was believed that the instrument developed in this study could be used in conjunction with science teacher training programs and with diagnosing students' comprehension of the nature of science and scientific thinking, it seemed reasonable to include in this literature review some indication of the results of research conducted to establish the level of students' and science teachers' understanding of the nature of science. However, since the amount of credence given to these findings should increase with the validity of the instruments used, it was thought appropriate to begin this review with a brief résumé of existing instruments used to measure scientific understandings.

AVAILABLE INSTRUMENTS FOR MEASURING UNDERSTANDINGS OF SCIENCE

An extensive search of the literature in science education from the 1930's onwards revealed a surprisingly small number of instruments for measuring understandings of the nature of science and scientific thinking. None of the discovered tests adequately examined the range of categories desired by this investigator, even though most were of some utility for suggesting facets of the nature of science which were desirable in the instrument developed in this study. Included in this review are those tests of greatest applicability to the development of this researcher's instrument.

Science Support Scale

The "Science Support Scale - Tri-S", developed by Schwirian (1968), solicits people's attitudes regarding science and scientists and the role science should be playing in the world. The theoretical basis for the scale is founded on Barber's (1962) list of five cultural values he deems necessary for science to flourish as an enterprise within society; rationality, utilitarianism, universalism, individualism, and progress and meliorism. In this inventory, the respondents are requested to indicate their degree of agreement or disagreement with a statement by using a numerical rating scale on which "5" corresponds to "strong agreement" and "1" corresponds to "strong disagreement". Utilizing this form of rating scale, Schwirian has found that the best forty items of the sixty item scale yield a reliability coefficient of 0.837, based on a sample of 513 university undergraduates.

Of interest to the study reported here is that Simpson, Shrum, and Rentz (1972) have demonstrated that this instrument is probably inappropriate for use with high school students since the reliability ranged from only 0.338 to 0.558, and a factor analysis of the items indicated that the divisions of items into the five Barber Scales was not interpreted with similar meaning by the students.

However, the test was of limited use in the present study since some of Schwirian's items suggested categories of the nature of science that were of interest and also provided cues for writing distractors to certain questions. For example, with respect to the functions or aims of science, the following items from the "Science Support Scale" were pertinent.

Item I. "One important function of science is to demonstrate the wonder and orderliness of God's universe."

Item 10. "An important function of the scientist today is to question what man says he believes."

Item 21. "One important function of science is to teach people to be critical thinkers, not believing everything they are told."

Test on Understanding Science

One of the most referred to instruments in the literature is Cooley's and Klopfer's "Test on Understanding Science, (TOUS)" (1963). This test is an attempt to evaluate comprehensions of the scientific enterprise, scientists, and the methods and aims of science. Each of these three areas is subdivided into a number of "themes", as is outlined below. (Cooley and Klopfer, 1961)

Area I - The Scientific Enterprise

Theme 1. Human element in science

2. Communication among scientists

3. Scientific societies

4. Instruments

5. Money

6. International character of science

7. Interaction of science and society

Area II - The Scientist

Theme 1. Generalizations about scientists as people

2. Institutional pressures on scientists

3. Abilities needed by scientists

Area III - Methods and Aims of Science

Theme 1. Generalities about scientific methods

2. Tactics and strategy of sciencing

3. Theories and models
4. Aims of science
5. Accumulation and falsification
6. Controversies in science
7. Science and technology
8. Unity and interdependence of the sciences

Of the eighteen themes listed, only those included under Area III are categories examined in the instrument developed in this study. The items of TOUS that sample knowledge of the themes in this Area were of much help when developing items for the test constructed in the present research. However, the instrument constructed here includes categories other than those catalogued in Area III. (see Chapter III, p. 50)

The initial preliminary form of TOUS, Form Z, consisted of 120 four-alternative multiple choice items, which were administered to approximately 900 high school students. On the basis of these results and a complete item analysis, the remaining items were rewritten and the second preliminary form of TOUS, Form Y, was prepared.

The final form, Form W, consists of sixty items, extensively validated with the assistance of scientists and science educators. The overall reliability of this form was found to be 0.76 with a standard error measurement of 3.49. The product moment correlation of TOUS with the "Otis Mental Ability Test", Form Am, was calculated from the data of a sample of 2980 students in grades nine through twelve to be approximately 0.65. This means that about 58% of the variance in TOUS scores can be accounted for by characteristics other than those measured by the Otis test.

Scientific Attitude Inventory

Moore and Sutman (1970) developed and validated the "Scientific Attitude Inventory", designed to test attitudes towards science. However, the test seems to measure more than scientific attitudes, since it is divided into three intellectual categories as well as three emotional categories. Each division is characterized by a positive and a negative position statement on the basis of which thirty positive and thirty negative statements about science were constructed. Students' understanding is believed to increase in relation to their agreement with the positive statements and their disagreement with the negative statements.

In a pretest - posttest experimental study conducted by Moore and Sutman (1970), the test - retest reliability coefficient obtained, based on pre and posttest scores of the control group, was 0.934. The validity of the instrument was evaluated by instructing the experimental group specifically in the traits measured by the test. The higher posttest means obtained by this group were used to infer construct validity.

There are three position statements which seem to be measuring more substantive knowledge than attitudes toward science, and hence were applicable to the instrument developed in this study. These statements are listed in TABLE II (Moore and Sutman, p. 230).

TABLE II

Applicable Position Statements Found in the Scientific Attitude Inventory

Category Number	Position Statement
1 - A	The laws and/or theories of science are approximations of truth and are subject to change.
2 - A	Observation of natural phenomena is the basis of scientific explanation. Science is limited in that it can only answer questions about natural phenomena and sometimes it is not able to do that.
4 - A	Science is an idea-generating activity. It is devoted to providing explanations of natural phenomena. Its value lies in its theoretical aspects.

Nature of Science Scale

Kimball (1967) developed the "Nature of Science Scale, (NOSS)" using position statements in much the same manner as Moore and Sutman. The function of the scale is to elicit opinions about important characteristics of science. The twenty-nine position statements appearing in the final version are based on eight assertions about the nature of science, which are consistent with the views of Conant and Bronowski (Kimball, 1966, p. 111). These assertions are typified by the following and are ones which also appear in the model of the nature of science used to construct the instrument in the present study.

1. "The fundamental driving force in science is curiosity concerning the physical universe. It has no connection with outcomes, applications, or uses, aside from the generation of new knowledge."

2. "There is no one "scientific method" as often described in school science textbooks. Rather, there are as many methods of science as there are practitioners."

3. "Tentativeness and uncertainty mark all of science. Nothing is ever completely proven in science, and recognition of this fact is a guiding consideration of the discipline."

The content validity of the test was checked by a panel of science teachers, science supervisors, and science professors. Also, construct validity was supported by deleting items which failed to discriminate in favour of science graduates over college graduates from other fields. The split-half reliability based on data gathered from ninety-seven subjects, about half of whom had majored in science as undergraduates, was 0.72.

Test of Science Comprehension

Nelson and Mason (1963) designed a "Test of Science Comprehension" intended to measure the degree of understanding or comprehension of particular situations from a scientific standpoint. These comprehensions are not specifically outlined by the authors of the test but they could be delineated as: choosing viable hypotheses, analyzing data, distinguishing facts from hypotheses, selecting justifiable conclusions, picking mathematical rules to fit data, choosing between opposing theories based on evidence, and deciding upon the relevancy of evidence.

Nelson's and Mason's test is similar to the one constructed in this study in that it frequently requires the examinees to operate at levels of Bloom's Taxonomy higher than the knowledge category. Many items could be classified into the Analysis, Synthesis, or Evaluation domains. However, the instrument fails to cover a wide range of the nature of science, leaving out important aspects such as the tentativeness of scientific explanations, the driving force of science, the aims of science, and the distinctions between science and technology. Also, there was no indication of any attempt to validate the test.

Wisconsin Inventory of Science Processes

Aikenhead (1973) describes another instrument designed to inventory knowledge of the scientific enterprise, "The Wisconsin Inventory of Science Processes, (WISP)". The test consists of ninety-three statements concerned with the assumptions, activities, objectives, and products of science, which the examinee judges as being accurate, inaccurate, or not understood. However, since both the response "not understood" and the response "inaccurate" are taken as opposites to "accurate" when scoring the test, the examinee is misled into thinking there are three distinct answers.

The usual statistical information and norms for the test have not as yet been completed, except a reported reliability coefficient of 0.82, of unidentified origin, based on a sample of grade twelve students (Aikenhead, 1973).

Science Process Inventory

Welch (1966) developed an instrument, the "Science Process Inventory, (SPI)", designed to evaluate the level of knowledge of scientific processes possessed by secondary school pupils. Elements of the scientific process were derived from the writings of Beveridge, Conant, Kemeny, Lachman, Nash, and Wilson. To be included in the inventory an element had to appear in three or more of these six references (Welch and Pella, 1967).

The elements were divided into four categories and presented to fourteen research scientists for validity judgment and were revised using their suggestions. The resulting design, presented in TABLE III, became the guideline for construction of the inventory (Welch, 1966).

Besides the expert judgment of research scientists and science educators as to the content validity of the items, the validity of the test was supported by an item analysis of the responses obtained from a sample of 380 high school students, by the fact that scientists score higher than science teachers, who in turn score higher than students, and by an acceptable reliability coefficient of 0.79.

TABLE III

Categories of Scientific Processes Used in the Science Process Inventory

Main Category	Elements Found in Each Category
I. Assumptions	A. Reality B. Intelligibility C. Consistency D. Causality
II. Activities	A. Observation B. Measurement C. Classification D. Experimentation E. Communication F. Mental Processes
III. Nature of the Products	A. Probability B. Tentativeness C. Theories D. Models E. Laws
IV. Ethics and Goals	A. Goals and Motivation B. Objectivity C. Anti-authoritarianism, Skepticism D. Amorality E. Repeatability F. Parsimony

Elements of interest in the study reported here are those contained in Category III and some of those contained in Category IV. The others were not considered in the development of the instrument described in this report.

Summary

Although all of the reviewed tests have some relevance to the present study, none covers completely the range of the nature of science of interest. Also, it was desired in the present research to construct an instrument covering a wider scope of cognitive levels than those examined in existing instruments. Only the "Test of Science Comprehension" seemed to make a deliberate attempt to develop items designed to require the examinees to function at the higher levels of Bloom's cognitive taxonomy.

Several of the tests which required the examinees to indicate agreement or disagreement with particular statements seem to be investigating opinions rather than substantive knowledge about the nature of science. It was thought by this investigator that such a test structure would create a more casual atmosphere for the person taking the test, who might, as a result, tend not to think about the answers as much as might be desired. Also, while in this agree/disagree type of test format it would be possible to examine, for example, whether or not the examinee has knowledge of the idea that generally in science a simpler theory is more desirable than a complex theory, the test developer would be in a quandary if he tried to evaluate, in this test format, whether or not a person can choose between competing theories on the grounds of one being simpler than the other. In other words, this type of instrument lends itself to the testing of knowledge level items but is not suitable for the measurement of higher level cognitions, such as analysis or evaluation.

UNDERSTANDINGS OF THE NATURE OF SCIENCE

Although this study addresses itself, for the most part, to the construction and validation of a testing instrument and less specifically to the uses to which the instrument can be put, a test can only make sense in the context in which it will be used. As stated previously in this chapter, since it is expected that the test developed in this research could be used for the in-service training of science teachers and for identifying areas of misunderstanding of the nature of science possessed by students, then it seemed relevant that the results obtained by application of similar instruments be briefly summarized.

Examples of uses these instruments have served include: determining the level of students', science teachers', and prospective science teachers' understanding of the nature of science; how the attainment of scientific literacy is affected by type of instruction; and how understandings of science correlate with other variables.

Students' Understanding of the Nature of Science

In a study of 672 ninth grade students attending two junior high schools in Philadelphia, Moore (1971), using his "Scientific Attitude Inventory", found neither strong acceptance of the test's positive position statements nor strong rejection of the negative statements. The students exhibited a lack of understanding of the nature of scientific theories and laws and of the difference between science and technology.

Isserstedt (1971), using an unpublished instrument, found that even among above average high school students an unfavourable and inaccurate view of science and scientists was the rule rather than the exception. He also discovered that after an eight week exposure to advanced science courses and research experience with scientists their understanding did not significantly improve.

Related to the study by Isserstedt, Jones (1969) has shown that students doing a general science course designed to teach an understanding of science fared significantly better on the "Test on Understanding Science" than the students doing professionally oriented physics and chemistry courses.

Science Teachers' and Prospective Science Teachers' Understanding of the Nature of Science

A study by Wood (1972), utilizing "The Wisconsin Inventory of Science Processes", revealed that future science teachers demonstrate no lack of understanding in any particular areas of the nature and processes of science. The areas best understood were scientific observations, experimentation, and communication of knowledge. However, it was found that the number of university science credits was negatively correlated with scores on WISP.

Olstad (1969) found that for future elementary science teachers knowledge of science subject matter, as measured by the "Advanced General Science Test", is significantly independent from science understanding, as measured by the "Test on Understanding Science", to warrant separate instruction for both facets of knowledge.

Kimball's (1967) results with the "Nature of Science Scale" showed that science teachers' and scientists' understanding of science was not significantly different. However, he discovered that philosophers possessed the best understanding of science and, indeed, a significantly better understanding than that of scientists.

Correlates of Scientific Understanding with Other Variables

Schwirian and Thomson (1972), using the "Science Support Scale", discovered no significant differences in scores obtained on the Scale by undergraduate university students with respect to the following variables: age, sex, religion, father's education, mother's education, major subject, and type of high school attended. Significant differences were found with reference to the father's occupation only - the higher the occupational status of the fathers, the more positive the students' scientific attitudes.

Baumel and Berger (1965) found for grade nine students very little correlation between scores on a test designed to measure students' understanding of the nature and processes of science and end of the year science marks.

Type of Instruction and Its Effect upon Understanding the Nature of Science

Two studies were found contrasting students' understanding of science, as measured by the "Test on Understanding Science", in traditional and PSSC physics courses. The two reports, one by Trent (1965) and the other by Crumb (1965), although in disagreement when the statistical significance of the results are compared, are quite in agreement if educational significance is the focus of concern. In his experiment, Crumb found that students taking the PSSC Physics course increased their scores on TOUS by a significant amount while those in the traditional course did not. Trent concluded from his results that neither group showed significant changes in TOUS scores.

A look at the actual data from both studies in TABLE IV indicates that it is indeed questionable whether or not either course effectively teaches for an understanding of the nature of science. Educationally significant results seem to be lacking from both studies, from the point of view of choosing one course over the other on the basis of the results.

TABLE IV

TOUS Scores Obtained in Two Independent Experiments - Highest Possible Score is 60

Physics Course	Investigator	Pretest	Posttest
PSSC	Crumb	37	40*
	Trent	37	39
Traditional	Crumb	35	38
	Trent	34	37

*Significant at the .01 level

Other studies have been conducted by individuals who have developed their own course to teach scientific literacy. Klopfer and Cooley (1963) compiled a "History of Science Cases, (HOSC)" and found that students doing a course based on HOSC achieved significantly higher on TOUS and equally as well on tests of science content as students doing more traditional science courses.

Lowery (1967) found that the attitudes of grade five students towards science can be changed in a positive way when experiencing a science curriculum that guides them to experiment individually.

Kahn (1962) reported that students who learned about scientists by reading current events stories and who afterwards analyzed the stories for revelations concerning the nature of science achieved a significantly better understanding of the characteristics of scientists than students who merely read the stories.

Summary

It can be seen that the research reviewed in the previous section is of direct relevance to the purposes of this present study. Outlined in point-for-point format, these studies say:

1. Students in our high schools possess somewhat less than desirable understandings of science, scientists, and the scientific enterprise;

2. if it is to be a goal of science education that students increase their competence in these areas, then teaching strategies have to be designed with these objectives in mind;

3. although science teachers' and prospective science teachers' knowledge in this field is more comforting, there remains room for improvement in this area of teacher education.

It is thought by this investigator that the test developed in this study could be instrumental in identifying deficient areas of student and teacher knowledge and in assisting curriculum development in this area. It is also believed that the instrument might be used along with other instruments as a tool for course evaluation. Courses could be modified in an attempt to overcome certain knowledge deficiencies which might be hypothesized by the manner in which students or teachers respond to the items on the tests.

CHAPTER III

GUIDELINES FOLLOWED DURING THE TEST DEVELOPMENT

Any test construction must take into account certain considerations in order for the instrument to be valid, reliable, and usable. The guidelines for the test constructed in this study can be categorized into three areas:

(i) theoretical considerations - dealing with the model of the nature of science on which the test was based, a model of cognitive operations used to write items requiring different levels of thinking ability, the validity of the model of the nature of science with regards to its covering a representative sample of the content area and with regards to its consistency with widely accepted, though not necessarily universally believed, ideas, and a factor analysis attempting to identify how the respondents perception of the instrument corresponds with the theoretical model;

(ii) statistical considerations - relating to the reliability of the test, an item analysis of students responses to the test, and an analysis of the correlations between scores obtained on the test constructed in this study and scores on other tests, such as reading level and intelligence inventories, which theoretically should be measuring different constructs;

(iii) practical considerations - encompassing decisions made with regard to ease of administration, time required for administration, facility of scoring, and ease of interpreting and applying scores.

THEORETICAL CONSIDERATIONS

Being of a theoretical nature, much of the content discussed in the following sections is not supported by empirical evidence. Much of the material has been derived from a concensus of opinions based on logical analyses of the nature of science, the nature of thinking, and the nature of the respondents' view of how the test is structured. However, this is not to say that the procedure followed in this section is not valid, since very often in education the methods adopted are not always ones that have been tried and proven but rather ones that seem pedagogically the most sensible.

A Model of the Nature of Science

If an attempt was made to develop an outline of the nature of science including all the characteristics of science and its various components, an almost insurmountable task would be met. This is so firstly, because, if each and every detail of the substance of science was to be included in the model, volumes would be needed and secondly, because there would be no way of deciding whether or not all of the characteristics of science had been covered.

This problem could be remedied by including only those facets of the nature of science considered to be important by the scholars in the field. However, this technique presents yet another problem. "Importance" only has meaning in the context in which the material is to be utilized. It was the opinion of this investigator that the above considerations provide some flexibility in choosing categories of the nature of science to be studied. Therefore, it was decided to incorporate in the instrument constructed in this study categories included under Cooley's and Klopfer's heading "Methods and Aims of Science" (1961, p. 4) and Pella's, O'Hearn's, and Gale's headings "Ethics of Science" (1966, p. 200). Sections dealing with such aspects of science as science and society, the scientist as a person, and science and the humanities were not embodied into the model. It was also believed that the arbitrariness of these decisions is justified on the grounds described in the preceding paragraph.

The rest of this section will describe each category that has been included in the model used to construct the test for this study.

Scientific Facts. Facts in science are supplied by observations which, under the same conditions, yield repeatedly the same results. These observations can be made directly by the senses, for example, the sense of vision tells us that spruce trees remain green all year; or by instruments, which are merely extensions of the senses, for example, when the sun is observed with the aid of a telescope sunspots are seen. In science, both of these phenomena would be considered facts.

Conant (1951, p. 35) says that facts are discovered in "experiments which have been often repeated with the same results (within certain limits of error); we shall assume that under the same set of conditions, the phenomena are in all details reproducible." He continues by observing that while it is a fact that a suction pump will draw water up to a height of about thirty-four feet, the statement that the earth is surrounded by an ocean of air whose weight causes pressure is a conceptual scheme or a hypothesis. (see Hypotheses, p. 28)

Hanson (1961, p. 31) cautions that we have to be careful when trying to decide upon what should and should not be considered fact. "What appears as one thing for one person may appear different for another, depending upon the individual interpretation which is based on past experience." The age-old argument about whether it is the sun that rises or the earth that falls illustrates this point.

Besides facts deduced from reproducible observations, science might also consider as factual an event that has been verified by a large number of people. For instance, although a lunar eclipse is not able to be repeated at the wish of the observer, its verity is assumed by the large number of people who have seen it.

However, it should be indicated that although science does claim knowledge of certain facts, many scientists would hesitate at saying they believe these to be absolutely true. These scientists would claim that even if an experiment has repeatedly given consistent results, there is a remote possibility the results could change at some later date.

Scientific Theories. The ultimate goal of science is the development of tested theories. Their usefulness rests on their ability to bring together what was previously an unrelated set of phenomena into a unified whole, where the explanation for the occurrence of each event can be explained by the fact that the other events exist. The significance of scientific theories, then, lies in their ability to explain and unify. Hanson (1961, p. 90) says that "Physical theories provide patterns within which data appear intelligible....Theories put phenomena into systems." Hempel (1966, p. 70) states: "Theories seek to explain regularities, and generally, to afford a deeper and more accurate understanding of the phenomena in question."

Besides explanatory power, a scientific theory possesses certain other important characterisitcs. Probably the most important of these is its ability to make predictions and retrodictions. "....a good theory will also broaden our knowledge and understanding by predicting and explaining phenomena that were not known when the theory was formulated" (Hempel, 1966, p. 76). This feature of scientific theories is also presented by Copi who says: "A scientific explanation for a given fact will have directly testable propositions deducible from it, other than the one asserting the fact to be explained (1968, p. 381). So, then, as well as serving the function of harmonizing already existing knowledge, scientific theories can be seen as catalyzing the discovery of new knowledge.

Another trait of all scientific theories is that they are tentative. However, in spite of the fact that theories in science are never presented as being the ultimate position on a particular area, once a theory has been accepted for use it dons a certain robustness that tends to save it from being overthrown, even in the face of arguments against it. Conant says, "A conceptual scheme is never discarded merely because of a few stubborn facts with which it cannot be reconciled; a conceptual scheme is either modified or replaced by a better one, never abandoned with nothing left to take its place" (1951, p. 173).

The whole problem of deciding upon the correctness of a theory in science also forms much of what scientists do. The situation, however, is confounded by another characteristic of theories - their inability to be proven true. This trait is an obvious consequence of the fact that theories are formulated on the basis of a limited number of observations and serve as generalizations from these data. A scientist could not possibly observe every specific instance of a phenomenon, so, while his theory is founded upon substantial evidence, he must admit to the possibility of disconfirmatory evidence which has not yet been observed. Juxtaposed with the characteristic that theories can never be proven true is the trait that before a theory can even be considered it must, at least in principle, be able to be proven wrong. Since science is a self-correcting enterprise, a theory formulated in such a way as to defy testing cannot be admitted.

Very often in science two different theories are available as explanations for the same phenomena. The choice as to which theory to accept is sometimes very obvious if one theory can explain more than the other. "One of the grounds for accepting one theory over another is if the one can explain all the other explains, and then some" (Toulmin, 1953, p. 113). The theory which is accepted by science "....will presumably continue to explain all the experimental laws that the earlier theory could explain, in addition to explaining experimental laws for which the earlier theory could not account" (Nagel, 1961, p. 87). In science, then, each successive theory subsumes its predecessors, which are then treated as special or limiting cases.

If it happens that both theories are equivalent when accounting for empirical facts, then very often one of the theories can be chosen on the grounds of its simplicity with respect to the other. Sheffler (1967, p. 9) states: "Any two theories of the same domain of phenomena may be compared to see if either is superior in accounting for the relevant empirical facts or, if equivalent on this score, if either surpasses the other in simplicity or convenience, etc."

To summarize, scientific theories are not facts or ideas which science holds unequivocally as being true. They are products of creative imagination, tentatively proposed as interpretations of natural phenomena. "To be accepted as a paradigm, a theory must seem better than its competitors, but it need not, and in fact never does, explain all the facts with which it can be confronted" (Kuhn, 1970, p. 17).

Scientific Hypotheses. The distinction between a hypothesis and a theory in science is very often considered one of degree rather than one of kind. In the first instance, whereas both theories and hypotheses are considered as being tentative explanations proposed for some natural phenomena, the theory is assumed to have accumulated much confirmatory evidence and is hence strongly accepted, while the hypothesis is still embryonic and considered doubtful by its proponents. "I will call such a proposition a hypothesis while it is still highly in doubt, and a theory when we have accepted it" (Kemeny, 1959, p. 93). Secondly, a theory is often considered to be a grand-scale hypothesis explaining a large number of observed facts, while a hypothesis is limited to interpreting a small number of isolated phenomena. Accordingly, a single theory could be composed of a large number of hypotheses.

The similarities between hypotheses and theories extend to the manner in which they are proposed and their method of being falsified. "...scientific hypotheses and theories are not derived from observed facts, but are invented in order to account for them" (Hempel, 1966, p. 15). "...the fact that a test implication inferred from a hypothesis is found to be true does not prove the hypothesis to be true" (Hempel, 1966, p. 8).

The resemblance of hypotheses to theories is also indicated by Copi's five criteria which are used to judge the worth or acceptability of hypotheses. "They may be listed as (1) relevance, (2) testability, (3) compatability with previously well-established hypotheses, (4) predictive or explanatory power, and (5) simplicity" (1968, p. 382). Copi continues: "A really fruitful hypothesis will not only explain the facts which originally inspired it, but will explain many others in addition" (1968, p. 392).

Scientific Laws. In their simplest form scientific laws are summarizations of large accumulations of observational data. They state rules of the form that when one event in nature occurs a certain other event always occurs also. A law in science "....asserts a uniform connection between different empirical phenomena...." (Hempel, 1966, p. 54). An example of such a statement is the one describing the condition for equilibrium of a lever: when the product of the force and moment arm on one side of the fulcrum equals the product of the force and moment arm on the other side, the lever will balance. Another example is Boyle's Law, which states that the volume of an ideal gas is inversely proportional to its pressure if the temperature remains constant.

Although scientific laws are usually stated in a definitive manner, seemingly disqualifying any exceptions to the rules they affirm, scientists, in general, would admit to the possibility that a law is not universally true. This is the only logically sound procedure, since a law is meant to be generalized to all cases of a particular phenomena and all instances will never be observed. "....we must needs conclude that all natural laws may be considered statistical laws. True, even statistical laws can lead to statements with so high a degree of probability that they are almost certain, but there can always be exceptions in principle" (Heisenberg, 1958, p. 35).

Like a theory, a law also possesses a certain sturdiness in the face of opposing evidence. A law which has received considerable support from observation is usually not discarded because of a non-conforming case but rather modified so that its scope does not include this particular instance any more. Hanson (1961, p. 115) claims: "When a genuine disconfirmatory instance does appear, rather than that their universality should be qualified, laws are usually saved. The law's universality is retained, but it is made inapplicable to the recalcitrant instance."

Once they have been formulated and given empirical support, laws can serve more than purely descriptors of what happens - they become prescriptors of what will happen. Where Boyle's Law initially described the relation between pressure and volume of gases, it can now be used to prescribe what the pressure of a gas will be when its volume and temperature are known. Toulmin (1953, p. 61) says that a new law "....allows us to extend the inferring techniques....".

Nagel (1961, pp. 74-77) describes four types of laws found in science. These four kinds of laws and an example for each are: (i) causal laws - when a sufficient negative charge is built up on a cloud, electricity will flow from the cloud, in the form of lightning, to a region of lower negative potential; (ii) laws asserting an invariable sequential order - the stages of growth of a butterfly consist of the egg, the larva, the pupa, and the adult; (iii) laws expressing invariable statistical relations - the Boltzmann Distribution Function giving the percentage of molecules of a gas which are travelling at a particular speed; and (iv) laws describing a functional dependence - Boyle's Law which says the volume of an ideal gas is inversely proportional to its pressure at constant temperature.

Aims of Science. While many scientists may consider it part of their responsibility as members of the human race to assist the world in the beneficial use of scientific knowledge, it may be said that science can be more accurately described as a way of searching for knowledge than as a means of applying this information. For instance, the task of explaining how electricity behaves belongs to the field of science, while the design of a new electrical appliance does not.

This view of the goal of science has been widely emphasized by scientists and science philosophers. Heisenberg (1958, p. 19) says that "....science subordinates every detailed question to the great task of understanding nature as a whole....". This aim of science is also indicated by Conant (1951, p. 24) who says, "....science is a way of explaining the universe in which we live" and by Copi (1968, p. 375) who claims that "The scientist seeks not merely to know what the facts are, but to explain them....". Nagel (1961, p. 4) asserts, "....it is the organization and classification of knowledge on the basis of explanatory principles that is the distinctive goal of the sciences."

Controversies in Science. Progress in science has usually been marked by the replacement of an old theory with a new or modified theory. Very often these transition periods have been permeated with energetic quarrels among the protagonists of each theory. These controversies have arisen over various matters, including religious or philosophical beliefs and have, with minimal exceptions, stifled scientific progress. As Kuhn (1970, p. 4) points out: "....the early developmental stages of most sciences have been characterized by continual competition between a number of distinct views of nature, each partially derived from, and all roughly compatible with, the dictates of scientific observation and method." Frequently, at least one of these competing theories has been based on some prejudice or self-evident truth, such as the early Greeks conviction that the Earth is at the centre of the universe because that is where it should be. These controversies have recurringly taken place because people could not understand how a scientific theory could co-exist with their philosophical or religious ideas.

Science and Technology. Although very often scientists will make extensive use of technology in their research, typically in the design of a piece of experimental apparatus, his ultimate goal is the utilization of the equipment to help carry out his work. The aim of the technologists, on the other hand, can be exemplified solely by the production of the instrument. The previous statement illustrates that science and technology cannot be completely divorced from each other even though their prime purposes are incongruous. It can be said quite positively that science and technology are not mutually exclusive but, instead, without advances in science progress in technology will be halted and without technological headway scientific progress will be brought to a standstill. "....technology has always been both the starting point and consequence of natural science" (Heisenberg, 1958, p. 16). Conant (1951, p. 326) says: "The applied scientist utilizes time and time again the new findings of those investigators working solely to advance science......if advance in pure science ceases, the applied scientist will run out of his most precious fuel - new ideas and new experimental results." Copi (1968, p. 373) also stresses the connection between science and technology by observing: "The practical value of science lies in the easier and more abundant life made possible by technological advances based on scientific knowledge."

There is a generalized statement, that for the most part appears to be true, which can be made concerning the differences between science and technology: While science is ultimately concerned with explaining why nature behaves as it does, technology tries to use these explanations for more practical objectives.

Interdependence of the Sciences. During the years when scientific knowledge was rudimentary with respect to today's standards, the distinctions between various branches of science were quite clear. The physicist, astronomer, alchemist, geologist, and biologist each worked with very little overlap into the other areas. At the present time, however, the different divisions of science are becoming less and less independent, as is illustrated by the emerging fields of biochemistry, biophysics, geophysics, physical chemistry, and chemical physics. Advances in cell theory depend upon progress in chemical kinetics which in turn are contingent upon the knowledge gained in atomic physics. The more modern view of science is that the traditional boundaries between the disciplines are artificial and serve more as a hindrance than an aid to progress.

The Driving Force of Science. From the writings of the early Greeks and before and from knowledge of present-day man, it can be shown that man is innately desirous of knowledge about why nature behaves as it does. Some people become so enthralled with this longing to know that they devote their lives to attempting to find the answers. These people are scientists. It appears, then, that the main impetus behind science is man's curiosity of how nature functions. Copi (1968, pp. 373-74) supports this view: "Science is knowledge and thus an end in itself......This intrinsic value is the satisfaction of curiosity, the fulfillment of the desire to know......Scientific knowledge......is itself a direct satisfaction of a particular desire, the desire to know." Also concerning the driving force of science, Nagel (1961, p. 4) comments: "It is the desire for explanations which are at once systematic and controllable by factual evidence that generates science....".

The Dynamic Nature of Science. While very many people would have science described statically as an organized body of knowledge, this description is in many ways too restrictive to give an accurate view of the nature of science. Since scientific knowledge is for the most part, if not entirely, tentative, any statement which implies permanence of scientific knowledge is a priori misleading. Besides being a body of evolving information, science is also the activity through which this knowledge is gained. This dynamic quality of science is referred to by Conant (1947, p. 24): "....we may say that science emerges from the other progressive activities of man to the extent that new concepts arise from experiments and observations, and the new concepts in turn lead to further experiments and observations." If Conant's implication of tentativeness is not clear from the previous statement, he is quite explicit in the following: "The dynamic view in contrast to the static regards science as an activity; thus, the present state of knowledge is of importance chiefly as a basis for further operations" (1951, p. 25).

Serendipity and Science. The people who describe the process of arriving at new scientific knowledge as consisting of five or six sequential steps would have you believe that new information in science is discovered in no other way. The history of science testifies that this is not so. Frequently, important advances in science have begun with a chance of serendipitous observation. The discovery of radioactivity is a usual example of such a case. The National Science Teachers' Association attests to this method of discovery in science by stating: "Major advances in science often result from the occurrence of serendipitous events during the "piecemeal" approach to scientific exploration" (Theory into Action, 1964, p. 30).

However, all chance occurrences in nature do not lead to scientific progress. As prerequisite to such an advance, the observer of the event must be capable of recognizing the observation as significant. Unless he can do this, the event will pass unnoticed. "In general such occurrences have no impact on knowledge unless the observer or experimenter has a "prepared mind" capable of recognizing the occurrence as important" (Theory into Action, 1964, p. 30). Toulmin (1953, p. 44) also points out the necessity of the experimenter being prepared to notice the event's meaning. ".....such discoveries are not made by accident, even though they may be made as a result of an accident......it is trained skill quite as much as imagination which guides him in the exploration once it is begun."

Conant (1947, p. 66) cautions that before a true discovery can be made from an accidental observation or the real significance of the observation known, systematic science must also take place: "....an accidental discovery may lead by a series of experiments (which must be well planned) to a new technique or a new concept or both...."

Simplicity of Explanation. Many people misconceive science as an activity which strives to achieve more and more complex explanations of nature. However, scientists find theories whose complexity causes inconvenience to be very annoying and, therefore, tend to seek the simplest explanations possible. In fact, science is guided by the principle that nature is parsimonious and can thus be understood in relatively simple terms.

This demand for simplicity is sometimes exhibited by the manner in which a choice is made between two competing theories or hypotheses explaining the same phenomena. If both have equal consistency with the observed facts but one explanation is obviously simpler than the other, then science assumes the simpler one to be more correct. "....if two hypotheses accord with the same data and do not differ in other respects relevant to their confirmation, the simpler one will count as more acceptable" (Hempel, 1966, p. 41).

Other philosophers have also expressed similar views to Hempel's:

> "A hypothesis that does not itself clash with experience may yet be given up in favor of an alternative hypothesis that explains more facts or is simpler, or easier to handle" (Sheffler, 1967, p. 9).
>
> "....the simplest theory which fits all the available facts is the one we tend to accept" (Copi, 1968, p. 386).
>
> "....a law or theory cannot be regarded as satisfactory if it's mathematical form is so complex that it cannot be used conveniently for purposes of calculation and prediction...." (Nagel, 1961, p. 321).

It is incorrect to think, however, that the choice between alternative explanations based on the principle of parsimony is a straight forward decision. Very often simplicity is not an easily identifiable characteristic, or it may be the case that one explanation is simpler in one respect while the other is simpler in another manner. In such cases the choice becomes more subjective and the usual procedure would be to wait for more relevant evidence to be discovered. Hempel (1966, p. 41) states: "....it is not easy to state clear criteria for simplicity in the relevant sense and to justify the preference given to simpler hypotheses and theories."

<u>Science and the Scientific Method</u>. There still appear some modern science books which describe to the reader a scientific method characterized by a list of ordered steps. These guidelines for experimentation are typified by the following: (i) identifying the problem, (ii) forming a hypothesis, (iii) collecting the data, (iv) analyzing the data, and (v) drawing conclusions. Yet, if the methods through which discoveries in science have been made are examined, it is found that while the above five steps describe adequately some methods in science, they are atypical of many others. Quite often data were available before a problem was seen, or the formulation of a hypothesis identified a problem rather than an existing problem stimulating the formation of a hypothesis.

Nagel (1961, p. 12) says: "There are no rules of discovery and invention in science, any more than there are such rules in the arts", and Conant (1951, p. 45) agrees: "....a careful examination of these subjects fails to reveal any one method by means of which the masters in these fields broke new ground." A more accurate description of the scientific method is given by Sheffler (1967, p. 68): "The scientific game imposes the constraints of descriptive accuracy, theoretical coherence, and logical discussion; it imposes no general limitations on passion, imagination, or flair."

Besides believing that there is a stepwise, closely adherred to scientific method, many people also think that if this method is followed satisfactory results will be guaranteed. This falsity is pointed out by Nagel (1961, p. 12): "...nor should the formula be read as claiming that the practice of scientific method effectively eliminates every form of personal bias or source of error which might otherwise impair the outcome of the inquiry....".

Definitions and Classification Schemes. Many people are under the impression that the definitions and schemes of classification found in science are inherent in the material studied. They believe that the electron's negative charge or the grouping of elements into the periodic table are intrinsic qualities of nature that have been discovered by scientists. Their views, however, are mistaken. That scientists say the electron has a negative charge is an arbitary definition which has evolved over the centuries. Science's conception of electricity would be just as complete if the electronic charge had been chosen to be positive. In addition, the manner in which the elements have been categorized in the periodic table, as well as the way biologists have grouped living things, have both been based on the observed similarities and differences of these objects which were of most utility to the scientists involved. Other classification systems were also possible. Yet, of all the possible classificatory permutations, only a limited number can be suitable for the particular paradigms under which the science is operating.

Concerning definitions and classification schemes, the following quotations are found in the literature of the philosophy of science.

> "A set of empirical "facts" can be analyzed and classified in many different ways most of which will be unilluminating for the purposes of a given inquiry" (Hempel, 1966, p. 13).

> "....we must realize that the adoption of this or that alternate classification scheme is not anything which can be true or false......The scheme of classification adopted depends upon the purpose or interest of the classifier" (Copi, 1968, p. 408).

> "One classification scheme is better than another......to the extent that it is more fruitful in suggesting scientific laws and more helpful in the formulation of explanatory hypotheses" (Copi, 1968, p. 409).

> "The chemist's definition is intended to attach to the word, as meaning, that property which in the context of his theory is most useful for understanding and predicting the behavior of those substances which the word denotes" (Copi, 1968, p. 93).

Experiments in Science. Experiments are performed in science so that the number of variables examined at any one time can be reduced to two, or at least if more than two are involved, just two are assumed to have significant effects. From such situations cause and effect relationships can be inferred.

Usually, experimentation is carried out to test a prediction that has been made from a hypothesis or a theory. Scientists rarely work in a haphazard manner just looking to see what will happen. "No competent scientist does pointless or unplanned experiments" (Toulmin, 1953, P. 66). Toulmin (1953, p. 67) continues by saying: "....the physicist does not enter his laboratory until he has some completely specific question to answer" and "....only when a regularity has already been recognized or suspected can the planning of an experiment begin" (1953, p. 111). Conant (1951, p. 54) also expresses this position: "The scientific experimenter......wants to test a deduction from a conceptual scheme (a theory)....".

Sometimes, an experiment can be used for purposes other than testing predictions. If the area being examined is devoid of sufficient theoretical structures, then, while the scientist may have some idea or hunch of what his experimental results will be, he possesses no sound basis for believing this. In such a case the scientist can proceed on a "try it and see" basis, hoping to make a new discovery. "Experimentation......is used in science not only as a method of test, but also as a method of discovery" (Hempel, 1966, p. 21).

The Limits and Scope of Science. Science addresses itself to the observation and explanation of the entire physical universe. Science is not characterized by the study of unobservables such as supernatural phenomena, even though individual scientists may still believe strongly in the supernatural.

While most scientists would agree that all natural phenomena are theoretically understandable, they are limited in their explanations of nature chiefly by their inability to make perfect measurements. At the present level of understanding in certain areas of science, probabilistic statements about the relationships between certain variables are the best that can be made. Nor does science believe that even with improvements in technology will the time come when measurements can be made completely free from error. In fact, one of the tenets of modern atomic theory, the Heisenberg Uncertainty Principle, states that on the atomic scale the very act of observation disturbs the system under investigation to the extent that both the mass and momentum of a particle cannot be known simultaneously. This inconvenience is not a fault of measuring instruments but an inherent quality of the act of making measurements and something science can theoretically never overcome.

The Falsification and Verification of Scientific Explanations. Since science can be described as an ever-evolving process in which old explanations are constantly being revised or replaced by newer theories, an understanding of the criteria used to falsify or verify scientific theories is crucial to an accurate understanding of the nature of science. A comprehension of this falsification/verification process is, however, unfortunately confounded by the fact that while theories in science can be proven false, they can never be proven true. Hempel (1966, p. 8) points out that: "Even if many implications of a hypothesis have been borne out by careful tests, the hypothesis may still be false", as well as the notion that "....scientific inquiry.....involves the acceptance of a hypothesis on the basis of data that afford not deductively conclusive evidence for it, but lend it only more or less strong "inductive support" or confirmation" (1966, p. 18). Sheffler (1967, p. 86) also conforms to this point of view by stating: ".... acceptance in science is never a matter of proof, and the tentative acceptance of a relatively unsupported hypothesis is compatible with acknowledgment of controlling tests to which future experience will subject it."

There is one tradition of science philosophy that maintains a rather simplistic view of the situation discussed in the previous paragraph. Nagel, et. al. (1969, p. 69) says that this school of thought claims: "....hypotheses can never be known to be true with certainty. But if even a small amount of contrary evidence does materialize, then the most celebrated of hypotheses is indeed known to be false." Nagel and his associates claim, however, that the falsification of a theory in science is never as uncomplicated as this position would have one believe. Conant (1947, p. 103) also disagrees with this view by asserting: "If a conceptual scheme is highly satisfactory to those who use it, neither a few old facts which cannot be reconciled nor a few new ones will cause the concept to be abandoned" as does Sheffler (1967, p. 78) who says: "....nor does contrary evidence alone force rejection of a paradigm; such rejection occurs only if a suitable alternative paradigm is available."

The truth of this last statement by Sheffler is not realized by many people. Scientists would rather operate under a theory whose verity they seriously questioned rather than with no theory at all. This is so because without any theory no direction can be given to future work. This position is supported strongly by Kuhn (1970, p. 77):

> "....the act of judgment that leads scientists to reject a previously accepted theory is always based upon more than a comparison of that theory with the world. The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgment leading to that decision involves the comparison of both paradigms with nature and with each other."

The statement by Kuhn begs the question: "What criteria are used for replacing an older theory with a newer one?" Nagel (1961, p. 87) gives a clear answer to this question: "The new theory will presumably continue to explain all the experimental laws that the earlier theory could explain, in addition to explaining experimental laws for which the earlier theory could not account."

Up to this point the discussion has included only aspects concerning the falsifiability of existing theories or their replacement with new or modified explanations. An important aspect yet to be covered is the procedure through which theories are initially verified.

Hempel (1966, pp. 33-38) gives a comprehensive summary of this process.

> "In the absence of unfavorable evidence, the confirmation of a hypothesis will normally be regarded as increasing with the number of favorable test findings" (p. 33).

> "....the increase in confirmation effected by one new favorable instance will generally become smaller as the number of previously established favorable instances grows" (p. 33).

> "....the confirmation of a hypothesis depends not only on the quantity of the favorable evidence available, but also on its variety...." (p. 34).

> "....it is highly desirable for a scientific hypothesis to be confirmed also by "new" evidence - by facts that were not known or taken into account when the hypothesis was formulated" (p. 37).

> "Support may also come......from more inclusive hypotheses or theories that imply the given one and have independent evidential support" (p. 38).

The implication derived from the above quotations is that the verifiability of scientific explanations is, at best, only a highly probabilistic activity, which is always in danger of making mistakes.

<u>Measurement in Science</u>. Many important advances in science have coincided with equally significant improvements in measuring techniques. Such consequential progressions as determining the speed of light, measuring the rate of drift of the continents, finding the speed of recession of distant stars, and estimating the distance to various bodies in the universe could not have been realized if the measuring techniques had not been mastered beforehand. Indeed it might be said that measurement is to discovery in science as counting is to adding in arithmetic.

Yet, no matter how accurate measuring methods may become, scientists work under the assumption that they will never be able to make perfect readings of nature's variables. Simultaneously, however, scientists are aware of the limits of precision dictated by the particular instruments they are using. This fact allows them to have confidence in a measurement if it is significantly different from that which could be reasonably expected by chance alone. "Important advances in science are based on quantitative measurements only if the measured quantity is large as compared with possible systematic and accidental errors" (Conant, 1947, p. 95).

It is also important to note that scientists do not go about measuring any variables that happen to be accessible. When a scientist measures a variable he is expecting to discover something and this expectation has usually been aroused as the result of an inference from a scientific theory. "Measurements by themselves do not yield new concepts; those scientists who have made great advances with the aid of new or improved measuring equipment have known what to measure because they were able to bring in a new concept or conceptual scheme at just the right moment in the drama" (Conant, 1951, p. 156).

<u>Models in Science</u>. Scientists use models in order to describe the behavior of something that is unfamiliar in terms of something whose behavior is familiar. Accordingly, in the Kinetic Theory of Gases, molecules are postulated as having the mechanical properties of billiard balls; in the Continental Drift Theory, the Earth's surface is pictured as consisting of a series of overlapping plates of crust. These various models are not meant to be mirrors of reality but, rather, are justified "....in the first place by the way in which they help us to explain, represent and predict the phenomena under question" (Toulmin, 1953, p. 37).

Much confusion has often arisen because of the way models are used in science. For example, at first consideration it may seem paradoxical that scientists describe light sometimes as a wave and other times as a particle. This, however, is not to say that sometimes light IS a wave and other times IS a particle. This would mean that nature is capricious, a characteristic that would bring science to a standstill. What scientists are trying to communicate is that sometimes light BEHAVES LIKE a wave and at other times BEHAVES LIKE a particle, without any reference to what light IS in reality. Toulmin (1953, p. 29) says concerning this point of confusion: "We do not find light atomized into individual rays: we represent it as consisting of such rays."

There is another type of model used in science other than the concrete or physical kind which has been the main focus of the preceding discussion. This type of model is abstract in that it has no familiar sensory cues. Such models are typified by the Schrödinger Equation for predicting the probable positions of electrons around the atomic nucleus. Such a model cannot be drawn or pictured in the ordinary sense, but has meaning only to those initiated into the mathematical concepts.

Imagination in Science. After being told how to proceed, any reasonably intelligent person could collect data from a scientific experiment by taking careful readings from a measuring instrument. However, given the data, it is not so obvious how to recognize a regularity in the numbers or to conceive of an explanation of why the variable under investigation has behaved as it has. This second step of seeing significance in the results of an experiment requires intellectual processes which cannot always be described as purely logical procedures. Discovery in science is not assured by the existence of the pertinent data. There are no rules to follow in the transition from data to theory. "The transition from data to theory requires creative imagination" (Hempel, 1966, p. 15). "....the discovery of important, fruitful theories in empirical science requires inventive ingenuity; it calls for imaginative, insightful guessing" (Hempel, 1966, p. 17).

The type of "imaginative guessing" that Hempel speaks of requires a particular kind of outlook. Just anyone cannot be expected to create a scientific explanation. This type of work needs the experience and insight of a trained mind. "....there are certain kinds of imagination which only a man with a particular training can exercise" (Toulmin, 1953, p. 44).

The Reproducibility of Results. While ancient philosophers based many of their explanations of nature on what they termed "self-evident truths" such as, "the planets travel in circular orbits since the circle is the only perfect geometrical figure" or "heavier objects fall faster than light objects because they are heavier", modern scientists tend to question anything reputed as being obviously true. Indeed, if modern scientists believe anything in nature is self-evident, it is that nature is not capricious. A list of five "Major Items in the Process of Science" compiled by the National Science Teachers' Association is headed by: "Science proceeds on the assumption based on centuries of experience, that the universe is not capricious" (Theory into Action, 1964, p. 21). Conant (1951, p. 33) says: "The assumption of a uniformity in nature - a belief in the reproducibility of phenomena - is basic to all science....", and furthermore "....we shall assume that under the same set of conditions the phenomena are in all details reproducible" (1951, p. 33).

It can be seen that since scientists require many independent checks of an observation before they are willing to concede its validity, then the assumption of repeatability is basic to all science. Science could not proceed without taking this for granted, even though it can never be proved. The National Science Teachers' Association takes the position that:

> "The conclusions which may emerge from a scientific inquiry require support by evidence obtained through carefully directed observation or experiment. Such evidence must either be replicable in principle, or must be capable of being independently confirmed through competent investigations" (Theory into Action, 1964, p. 30).

Scientific Evidence. No matter how obvious, scientists usually will not accept a statement or explanation, even tentatively, unless there is some logical reason for doing so. Having logical reasons implies, almost invariably, having empirical evidence. With regard to this point Copi (1968, p. 379) says:

> "Since every scientific explanation is regarded as a hypothesis, it is regarded as worthy of acceptance only to the extent that there is evidence for it."

> "The term evidence as used here refers ultimately to experience.....Science is empirical in holding that sense experience is the test of truth for all its pronouncements."

When trying to decide whether or not an explanation has enough evidential support to be accepted, there are no rules to follow. Whether or not two, three, or one hundred pieces of evidence are sufficient depends upon the particular case and must be decided by the scientists in the area. However, certain kinds of evidence tend to offer more support than others. For example, if a theory successfully predicts the occurrence of a phenomenon that was not even known when the theory was formulated, this evidence tends to be held in the highest esteem by scientists.

The matter is further complicated by the fact that often a piece of evidence can support two competing theories in the same area. What has occurred is that a significant point of discrepancy between the theories has not been addressed by this particular piece of evidence, and, therefore, more evidence will have to be found before a choice can be made between the theories.

The National Science Teachers' Association (Theory into Action, 1964, p. 30) indicates another characteristic of acceptable evidence:

> "...if the evidence is to have any merit at all, it must embody in some measure the idea of control - for example, the claim that a phenomenon occurs under certain conditions is worth little, if anything, unless the evidence provides reasons for believing that phenomenon does not occur in the absence of those conditions."

Summary. Admittedly, the Model of the Nature of Science presented does not describe all the important facets of science and at times takes a somewhat narrow and simplistic approach to the categories it does include. In particular, for instance, areas dealing with the scientific enterprise such as communication among scientists, scientific societies, instrumentation, funding of scientific work, the international character of science, and the interaction of science and society; and areas concerned with the scientist as a person such as institutional pressures on scientists, intellectual and other abilities required by scientists, have been omitted. Also, for example, while Popper's (1959) position on the falsifiability/verifiability of scientific explanations could have been a major contribution to the model, it was decided to take a more uncomplicated approach.

These restrictions have been imposed on the model purposely so that a more comprehensive sampling of a smaller area of the nature of science could be obtained and also so that the model would conform to the intellectual limitations of the high school students for whom the instrument is intended. However, it is believed that since the positions taken in the model are in agreement with those of several eminent philosophers of science, and since the categories themselves have been derived from those found in other instruments in this area whose validity has already been checked, the model presented is valid.

It is also believed that the scope of content covered is a matter of preference on the part of the test designer, with the possible restriction that the test not be so narrow that inferences about examinees' knowledge of the nature of science cannot be made from the test scores. It is believed by this investigator that a test based on the Model of the Nature of Science described will be broad enough in scope to make generalizations about the students' knowledge in this area.

It is difficult to decide what is the best way to classify the categories of the nature of science that have been described in the model. It is quite certain, however, that factor analysis results (see Factor Analysis, p. 56) will reveal that students do not view the test as having the twenty-two categories described. For this reason, a hypothesized factor structure, consisting of four factors, has been described in TABLE V. This schema yields four categories which are probably not independent of each other. In fact, this investigator doubts whether any of these categories can be expected to be mutually exclusive.

TABLE V

Classification of Nature of Science Categories into Four Main Areas

Area I What is Science?

1. Aims of science
2. The driving force of science
3. The limits and scope of science
4. Science and technology

Area II Forms of Scientific Knowledge

1. Facts
2. Theories
3. Hypotheses
4. Laws
5. Definitions
6. Classification schemes
7. Models
8. Evidence

Area III How Scientific Knowledge Emerges

1. Dynamic nature
2. Scientific method
3. Experiments
4. Measurement
5. Falsification and verification
6. Serendipity
7. Imagination
8. Controversies

Area IV Characteristics of Scientific Knowledge

1. Simplicity
2. Reproducibility
3. Interdependence of the sciences

A Model of Cognitive Operations

It has long been known that people possess different levels of thinking ability and that the knowledge of certain information does not mean, ipso facto, that this knowledge will be able to be used in an unfamiliar but related situation. It has also been widely assumed that the ability to use information in novel situations requires a level of cognition somewhat higher than merely the ability to recall this information. Similarly, it was the belief of this researcher that different cognitive levels are required to, for example, give a definition of a scientific theory, identify an example of a theory, be able to distinguish between a theory and a fact, decide whether or not a theory is consistent with given facts, indicate logical fallacies in a theory, and to choose between two theories on the grounds of explanatory power, precision, or simplicity.

Because it was the intention to devise test items which reflected both the different categories of the nature of science and the different levels of thinking ability, it was decided to use Bloom's Taxonomy of Educational Objectives: Cognitive Domain (1956) as a guideline for question development. This Taxonomy adequately covers the range of cognitive abilities which were considered important.

The structure of the Taxonomy is outlined in the following section and research into the validity of Bloom's assumptions are discussed in the second section following.

Taxonomy of Educational Objectives. The Taxonomy of Educational Objectives, Handbook I: Cognitive Domain (Bloom, 1956) is a scheme for classifying educational objectives according to the cognitive abilities needed to attain these goals. A similar handbook for the affective domain has also been published (Krathwohl, 1964) and a third book dealing with the psychomotor domain is presently being developed.

Handbook I is structured along six cognitive levels, the first being Knowledge; the others, called "Intellectual Abilities and Skills", are Comprehension, Application, Analysis, Synthesis, and Evaluation. Each of these levels is divided into various subcategories. A description of each level, adapted from the APPENDIX of Handbook I, is given in APPENDIX C.

Although the reliability of placing test items into the six major levels of Bloom's Taxonomy is not extremely high (Poole, 1971 and 1972), it was believed that the categorization presented in TABLE VI, as an example of the types of questions found at various Bloom levels, closely adheres to the Taxonomy's structure. However, it was realized that an analysis of the content only is not sufficient information to accurately place test items into the Taxonomy. This is so, since what might require a mere knowledge level process for one person may, in fact, require an analysis or even higher process for another. The assumptions made were that the examinees, in general, should be able to answer from memory the items classified under the knowledge category, while the questions listed under the other Taxonomic levels would require the intellectual abilities indicated by that level.

TABLE VI

Item Response Types Categorized into Bloom's Cognitive Levels

Response	Cognitive Level
1. Picks definition of a scientific theory.	1.11
2. Identifies an example of a theory.	4.10
3. Distinguishes between examples of facts and theories.	4.10
4. Decides whether a theory is consistent with given facts.	4.20
5. Tells whether a theory contains logical fallacies.	6.10
6. Chooses between two theories on the grounds of explanatory power, precision, or simplicity.	6.20

It was thought that structuring the test items into these cognitive strata might also have implications for the factor analysis results. It is possible that the examinees' responses could factor according to cognitive levels rather than nature of science categories. A more probable situation is one where the science categories are confounded with those of the Taxonomy. (See Factor Analysis, p. 56)

<u>Assumptions Underlying the Taxonomy</u>. The Taxonomy of Educational Objectives: Cognitive Domain seems to be based on three major assumptions:

(i) the process categories which it outlines are cognitive processes

(ii) these cognitive processes are learned or learnable behaviors

(iii) the arrangement of the categories is taxonomic, that is,
  (a) the ordering of the categories - Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation - is hierarchial according to complexity of cognitive process;
  (b) the hierarchy of the categories is cumulative, that is, each level subsumes all the processes below it in the hierarchy and, in addition, contains an extra element not found in any of the levels preceding it.

Validity of the Assumptions. Since the main purpose of using the Taxonomy in this study was to have a guide so that test questions covering various mental operations would be included, the problems concerned with the validity of the assumptions mentioned in the previous section are not overly significant. However, to base part of a study on material with seriously doubtful validity exposes the study to justifiable criticisms. For this reason, a brief summary of some of the research into the validity of Bloom's assumptions is presented.

Kropp, Stoker, Bashaw (1966) conducted an extensive study in order to determine the construct validity of Bloom's classification scheme. Three questions were considered in this study.

(1) Can the hierarchial structure of the Taxonomy be empirically supported?

(2) Can evidence be found to support the imputed generality of the cognitive processes concerning their transcendence of all content areas?

(3) Can each level of Bloom's structure be explained by more elemental cognitive abilities?

The method employed for data collection was to construct four tests, each consisting of six subtests corresponding to the main categories of the Taxonomy. The tests were administered to approximately

1600 students at each of grades nine through twelve.

The results indicated that the hierarchical structure of the Taxonomy is probably correct, since an inverse relationship between subtest score and taxonomic level was found. On the basis of these findings, "...the conclusion was drawn that there was a clear tendency for the empirical data to support the imputed hierarchical structure of the taxonomy" (Kropp, et.al., 1966, p. 168). With regards to the generality of the cognitive processes, there was a suggestion that scores were obtained by a complex interaction of process and content. The investigations into the aptitude content of the taxonomy were impaired by inadequate taxonomy-based tests and lack of sufficiently refined analytical techniques.

In two studies conducted by Smith (1968 and 1970), hierarchical syndrome analysis (McQuitty, 1960 and 1966) was used to test the assumptions of hierarchical and cumulative structure in Bloom's Taxonomy. In the analysis, reciprocal pairs of categories are sought. A pair of categories is reciprocal if category X is most like category Y and category Y is most like category X. "Evidence of a cumulative and hierarchical arrangement of cognitive processes is provided when the reciprocal pairs consist of adjacent Taxonomy subtests which are then joined in turn by the subtests which are theoretically closest to those forming the reciprocal pair" (Smith, 1970, p. 40).

The analysis of the two Smith studies generally supports the Taxonomy assumptions of a hierarchical and cumulative continuum of cognitive processes.

Although neither of the above studies supplies conclusive evidence for the validity of Bloom's assumptions, it is believed, that for the purposes of this study, enough support has been given to the Taxonomy to make it a credible guide for item construction.

Factor Analysis

By examining the Model of the Nature of Science and the Model of Cognitive Operations presented previously in this chapter, and by enumerating the categories, a two dimensional array of fifty separate categories can be discovered. However, it is unlikely that the examinees will view the test as having this many distinct subdivisions. It is more probable that their responses will indicate that they view the test as consisting of five or six different factors or variables.

Factor analysis is a systematic attempt to resolve test questions into a smaller number of categories. The principal object of the technique is to simplify the description of the data by reducing the number of relevant variables or categories which are needed to account for variance in test scores. In doing this, it becomes easier to identify what traits the test is actually measuring and, therefore, the construct validity of the instrument is strengthened.

The factor analytic technique can be used to subdivide the sources of variance contributing toward performance on any test. For example, suppose a test was found to load on three main factors with these correlations: factor 1 - .40, factor 2 - .50, factor 3 - .60. Also, the reliability coefficient was found to be .90. Then the variance of scores on this test can be accounted for as shown in TABLE VII. It can be seen that not all of the variance can be explained with the factors used. It might be possible, however, to either add another factor or change the three existing ones so that more of the variance can be explained.

TABLE VII

Percentage of Variance Accounted for in a Hypothetical Test

Source of Variance		Percentage Contribution to Total Variance
Factor 1		16%
Factor 2		25%
Factor 3		36%
Error		10%
Not explained		13%
	TOTAL	100%

The mathematical procedures of factor analysis are to a large degree iterative in nature. This fact all but precludes the use of factor analysis techniques unless a computer is available. However, for the sake of completeness, a simplified description of the mathematical computations involved in the analysis is given.

Whatever technique of factor analysis has been chosen, calculations will begin with the construction of a table of intercorrelations between various tests or between various items on the same test. The analysis will end with a factor matrix giving the correlation of each item with each of the factors. The extraction of the factors themselves presents the laborious task.

Initially, factors which seem logical are postulated. Examples of such factors might be the total score on the test, the sum of the scores of items which logically seem to be related, or the averages of the scores obtained for certain groups of items. These hypothesized factors can then be added to the correlation matrix and the correlations of each item with the factors calculated.

By representing the factors geometrically as frames of reference, the correlations of the items with the factors can be plotted on a graph. These axes (factors) can then be mathematically rotated so that all significant negative correlations are eliminated and each item has a significant loading on only one factor. The rotated factors can better represent the items in terms of positive correlations and in terms of each item being highly correlated with only one factor. It should be noted that simple factor structures are not always possible to find.

Validity of the Instrument

The validity of a testing instrument concerns what the test measures. In this regard, it makes little sense to ask the question "Is this test valid?" but, rather, it is sensible to ask "For what purposes is this test valid?". In attempting to answer this latter question, Cronbach (1960, p. 106) refers to four types of validity, each concerned with a different aspect of what a test examines. These are outlined in TABLE VIII. For the purposes of developing the instrument in this study, the content and construct validity were of most concern.

TABLE VIII

Types of Validation

Type of Validity	Question to Answer
Predictive	Does the test score accurately predict any future performance?
Concurrent	Does the test score accurately describe any present performance?
Content	Does the test give a representative measure of a certain domain of tasks?
Construct	Does the test measure certain theoretical mental traits possessed by the examinee?

The content validation of a test, very often referred to as logical or rational validity, requires a "systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured" (Anastasi, 1968, p. 100). This necessitates a complete description of the content area under consideration before test items are written. This content area should also be defined so as to include certain broad objectives, such as the application of principles and data interpretation (Anastasi, 1968, p. 100).

It was believed that the Models of the Nature of Science and Cognitive Operations described previously in this chapter possess content validity, since each covers an extensive range of its particular domain. That the content included in these models is significant is supported by the long list of scholars who attest to its importance.

To assure that the test items would be representative of the content encompassed by the two models, the outlines in TABLES IX and X were followed during item construction. The various weights given to the categories in the nature of science are representative of the stress placed upon them by writers in the field, in this researcher's opinion.

With respect to the distribution of questions throughout the cognitive categories, it was decided to attempt to have about equal numbers of questions examining levels 1.00 to 3.00 as levels 4.00 to 6.00. This procedure was adopted since Smith (1970, p. 40) has found that the "cognitive processes......tend to fall into two clusters or types based on the statistical distance between them. The two types are represented by (a) the knowledge of principle, interpretation, extrapolation, and application subtests and (b) analysis, synthesis, and evaluation subtests.....A logical analysis of the test items seems to indicate large difference in complexity between the application and analysis subtests." It was discovered, however, that restricting the test to multiple choice questions also limited the cognitive levels that could be adequately examined. For example, the Synthesis category, 5.00, is very hard to sample in a multiple choice format. But if Bloom's cumulative assumption is correct, placing a relatively large number of items in category 6.00, Evaluation, will also examine ability to perform at the Synthesis level. The comparatively large proportion of items in the 4.00, Analysis, level is justified similarly, since ability to operate at this level can be used to infer ability to operate at the two lower levels. However, it should be noted that inability to operate at the Evaluation or Analysis levels indicates nothing about the ability to operate at levels lower than these.

TABLE IX

Distribution of Items Throughout Bloom's Taxonomy

Cognitive Level	Proportional Number of Items
1.00 Knowledge	20%
2.00 Comprehension	10%
3.00 Application	10%
4.00 Analysis	30%
5.00 Synthesis	10%
6.00 Evaluation	20%

TABLE X

Distribution of Items Throughout Nature of Science Categories

Nature of Science Category	Proportional Number of Items
I. What is Science?	15%
II. Forms of Scientific Knowledge	50%
III. How Scientific Knowledge Emerges	25%
IV. Characteristics of Scientific Knowledge	10%

It was also realized that each of the categories in the nature of science model could not appropriately be examined, on a multiple choice test, at each of the cognitive levels. However, questions were constructed for each science category so that they would cover as wide a range of Bloom's Taxonomy as could be accomplished. The distribution of questions throughout both models is shown in TABLE XI. This matrix gives the distribution for the preliminary form of the test given to scientists for validation purposes (see Appendix A).

TABLE XI

Matrix of Item Percentages Falling under Both Science Model and Cognitive Model

Science Model \ Cognitive Model	1.00	2.00	3.00	4.00	5.00	6.00	Total
AREA I	7	0	1	5	0	2	15
AREA II	12	2	2	30	1	4	51
AREA III	9	2	1	5	2	10	29
AREA IV	2	0	1	1	0	1	5
Total	30	4	5	41	3	17	100

More inclusive than content validity is construct validity. To the extent that a test measures a certain theoretical trait or construct, such as verbal fluency, mechanical aptitude, etc., it may be said to have construct validity. Content validity is an obvious prerequisite to construct validity.

The establishment of the construct validity of a test is a long process consisting of an interplay between observation, reasoning, and imagination. No single piece of research or experiment can be used to indicate whether or not an instrument ~~has construct~~ validity. Two of the methods available for assessing a test's construct validity are factor analysis, discussed already on p. 56, and correlations with other tests, which will be discussed under the next section, STATISTICAL CONSIDERATIONS (see p. 68)

Summary

This section, THEORETICAL CONSIDERATIONS, has described models of the nature of science and of cognitive processes and presented some considerations which must be taken into account when test validity is being examined. The section has been described as theoretical because most of the material is by nature what educators hypothetically believe to be educationally the most sound. These beliefs are subject to change as new research evidence is accumulated.

The next section deals with entities that seem a bit more concrete and absolute, statistical data, but when it becomes time to interpret these results, hypothetical procedures again take over.

STATISTICAL CONSIDERATIONS

As was mentioned previously, the presence of statistical data does not automatically imply that meaningful or correct decisions can be made. The interpretation of statistical data is more important and, unfortunately, less unequivocal than the mere accumulation of this data. For example, it is one thing to compute a reliability coefficient and yet another to decide, once the reliability is known, whether or not it is suitable for the purposes at hand.

Reliability of the Test

Reliability refers to the consistency of the scores obtained by the same individuals on identical or equivalent forms of a test administered at different times. Once a reliability coefficient has been calculated, some information is available as to what extent differences in scores can be attributed to "true" differences in the characteristics under investigation and to what extent differences are attributable to other variables, such as chance error.

Since reliability is concerned with the degree of agreement between two independently obtained sets of scores, it can be expressed in terms of a correlation coefficient. The closer the coefficient is to one, the less susceptible is the test to random fluctuations due to changes in the examinee or the testing environment.

An obvious way for finding a reliability coefficient is to administer the same test to the same individuals on two separate occasions. The correlation between the two sets of scores would then be the test-retest reliability coefficient. Although straightforward, this type of reliability suffers drawbacks:

(i) The time interval between both test administrations cannot be specified definitively, since it has been found that the reliability coefficient calculated is dependent upon the amount of time lapsed between both applications of the test.

(ii) The practice obtained in writing the test initially can also have an effect on the subjects' responses to the second test. This is especially true of a test which requires the examinees to reason or use their ingenuity to think through a problem. The test-retest technique will yield a spuriously high reliability coefficient since, once having grasped the principle involved in the problem, the subject can reproduce the correct response at sometime in the future without having to repeat the intervening steps.

In order to avoid the practice effect which can sometimes seriously affect a test-retest reliability coefficient, it is more usual to use split-half reliability. From a single administration of one form of the instrument two scores are obtained for each individual by dividing the test into equivalent halves. The two sets of scores are then correlated in the usual way.

A serious problem with the split-half technique is how to divide the test so that the most comparable pairs of scores can be obtained. There are no firm rules available to help solve this, so in this study a decision was made to use the Kuder-Richardson 20 reliability. This technique is based upon dividing the test in half in every possible way. The split-half reliability for each of these possibilities is then computed. The KR-20 reliability coefficient is simply the average of all these split-half reliabilities.

Unless the homogeneity of the test items is very high, the KR-20 reliability coefficient will be lower than any single split-half coefficient. Thus, to a certain extent, the KR-20 reliability coefficient may be considered to be indicative of the test's heterogeneity or of interitem consistency.

Once the reliability coefficient has been calculated, a more useful statistic, the standard error of measurement, can be found. The standard error can best be explained using an example. Suppose there were available 100 scores one person obtained on the same test. Assuming no practice effects, these scores would still fluctuate within a certain range of scores and would theoretically form a normal distribution around a certain mean. This mean is taken as the person's "true" score, while the standard deviation of the distribution is called the standard error of measurement or standard error.

The standard error is useful because it allows predictions to be made concerning the probability of a true score being within a certain range once the actual score is known. For example, it can be said that 68% of the time a person's actual score will lie within one standard error, above or below, of his true score; 96% of the time the actual score will lie within two standard errors of the true score, etc. The standard error can be calculated from the reliability coefficient by the following formula:

$$\sigma_{meas.} = \sigma_1 \sqrt{1 - r_{11}}$$

where σ_1 is the standard deviation of the test scores and r_{11} is the reliability coefficient.

From the discussion it should be clear that a valid test must first be reliable. In the limiting case where the reliability coefficient is zero, then test scores occur completely at random. In such a case, the test can hardly be supplying the information sought, something it must do if it is to be valid. The logical question now is

"How high should the reliability be?". The answer to this question cannot be given with any degree of precision. However, one important consideration must be taken into account:

> the more important or serious the decisions made using the test score, the higher the reliability coefficient should be.

For example, if the test results are to be used to review the effectiveness of a certain curriculum, an instrument with low reliability, say .60, might be used successfully. If, on the other hand, the test scores are to be used to decide which students are to be placed in the special education class, reliabilities in the range of .90 or higher are needed.

Item Analysis

The logical procedures employed in item construction can be supplemented by a more quantitative procedure called item analysis. This technique gives answers to the following questions:

(1) How difficult is an item?

(2) Does an item discriminate between good students and poor students?

(3) How good are each of the distractors for an item?

The difficulty level of a test item is simply the ratio of the number of people getting the question correct to the total number of people who attempted it. Many educators claim that items having a 50% difficulty are the best, since they provide the best differentiation between students. Yet, easy and hard items are required if a test is to be truly representative of the content area in question. If an item is so difficult that only 5% of the students answer it correctly, this is no reason to discard the item. In fact, the item may provide valuable information for the teacher.

The discriminatory power of a test item gives the difference in the proportion of high achievers answering a question correctly to the proportion of low achievers answering the same question correctly. An item possessing a low or negative discriminatory index indicates that the items may contain ambiguities, such as a grammatical error, two correct answers, etc., which mislead the good students more than the poor students. Such items must be carefully examined in an attempt to identify the source of the problem.

For the purposes of this study items whose difficulty level was outside the range of .30 to .70 and/or whose discriminatory power was less than .20 were examined to see if they could be improved. These are arbitrarily set cutoffs but represent approximately average levels desirable.

The three incorrect choices on a four alternative multiple choice test must be designed to perform specific functions. First, of those that get an item wrong, approximately equal numbers should have chosen each of the three distractors. If no subjects chose one of the alternatives, it is obviously not serving as a distractor. Secondly, of those that choose an incorrect idstractor, a greater number should be from the group that did poorly on the entire test. If a greater number of the high achievers pick an incorrect alternative, then the question should be examined for ambiguities.

<u>Correlations with Other Tests</u>

Of greatest significance to the validity of any test is the answer to the question "What traits or characteristics is this test measuring?". The answers that have been given for tests in the past have esentially been inferences from indirect evidence. For example, if an intelligence test accurately predicts achievement in school, then the test's construct validity is increased, since achievement and intelligence are assumed to be positively correlated.

Another procedure for investigating the construct validity of a new test is to examine how the test correlates with other previously validated instruments. If the previous instrument is supposed to be measuring the same domain of characteristics, then the correlation between the two tests should be high, but not extremely high. If the new test correlates too well with the existent instrument without any enhancement of usability, then the new test is a needless replication.

If the previous instrument is purported to measure something different than the new test, then the correlations between the tests preferably should be low. For example, if a new mechanical aptitude test is found to correlate very highly with a reading ability test, then there is doubt as to whether the new test is measuring mechanical aptitude at all. Suppose the correlation between the two tests was .90. This means that 81% of the variance on the mechanical aptitude test can be accounted for by what the reading ability test measures. A highly undesirable situation, to say the least!

PRACTICAL CONSIDERATIONS

No matter how valid a test may be, if it, for any number of reasons, has low usability it will be of minimal service to education. Because of this, certain practical considerations related to the test's usability are essential to the development of a good instrument.

Ease of Administration

When a test is to be administered eventually by teachers or others with relatively little training in testing procedures, ease of administration is a particularly important quality for the test to possess. A test requiring handing out papers, explaining directions, and timing for various subtests is likely to create so much confusion as to make the results of an otherwise reliable and valid instrument questionable. Problems in administration can be lessened if concise, precise directions are supplied in the test manual.

Time Required for Administration

There are two main considerations to be taken into account when deciding how long a test should be. First, the time of administration should be chosen so that the test can conveniently be given in an ordinary school situation. This would eliminate, for all practicality, tests which are more than sixty minutes long. Second, the test has to be long enough so that reliable results can be obtained. The length of the test is important in this respect because reliability increases with increasing the length of the instrument.

Facility of Scoring

The use of separate answer sheets, entirely objective tests, and an easy to follow scoring key all contribute to less tedious and troublesome scoring. Once this has been accomplished, a test is more likely to be used and thus contribute to education.

The use of separate answer sheets and a scoring key also means that the same test booklet can be used over and over, and clerical staff can be used to mark tests and tabulate scores.

Ease of Interpreting and Applying Scores

Since, in the final analysis, the only value of any test is determined by the use to which it has been subjected, the test's manual must give unambiguous information concerning the interpretation and use of test results. Attention should be directed towards an easy conversion from raw to derived scores, and to the interpretations and misinterpretations that these scores can yield.

SUMMARY

This chapter has presented many of the considerations which were taken into account before the instrument development portion of this study was begun. Each aspect of this material is associated with the validity of the test to a certain degree. Any content validity which the test possesses is very heavily dependent upon the models of the nature of science and cognitive operations. To the extent that these models are content valid, and to the degree that the test examines a representative sample of each model, then the test itself will be content valid. The construct validity of the test is contingent upon the results of the factor analysis, item analysis, and correlational studies. These data are presented in the following chapter. The practical considerations are important in that, quite easily, extraneous variables can become major sources of variability in a test's scores. When this occurs, the reliability of the test suffers as well as the fact that the scores themselves are not indicative of what the test is actually designed to measure. Obviously, such perplexities expose the validity to serious doubts.

Chapter IV

PROCEDURES FOLLOWED DURING TEST DEVELOPMENT

The method utilized in the development of any test purporting to be valid is analogous to the method of successive approximations used in mathematics to solve sets of equations having more unknown variables than there are equations. At first, the test is just an unstructured idea with no precisely articulated objectives or definite structure. Gradually, after many trials and modifications, the instrument comes closer and closer to measuring what the examiner intended. However, just as with the successive approximations method, where an exact answer can never be expected, this method never produces a totally valid instrument. This is so, since one can never take into consideration all the possible variables affecting the test scores.

This chapter, then, will describe the various stages of development that were undertaken in constructing the test. It is believed that the many factors that have been considered in this development are critical in making the instrument valid.

CONSTRUCTION OF THE FIRST PRELIMINARY FORM

Before the writing of this form had begun, it was uncertain whether or not items could be constructed which seemed to satisfactorily examine the content intended. There was also no evidence that students at the grades ten and eleven levels could complete the type of test envisaged in a reasonable amount of time. This form, then, served as an informal trial for both the test constructor and the examinees. Some feeling for the feasibility of this type of test was sought.

Item Format

The first consideration was to decide into what form the test would be cast. Previous instruments in this area have used various item formats: essay, true-false, agree-disagree, agree-disagree-not sure, and multiple-choice. The first decision was to use one of the objective formats in preference to the essay style. This choice was made even though the essay test has the distinct advantages of easy construction, of testing more efficiently such mental processes as synthesis and comprehension, and of forcing the instrument to focus on the more general facets of a subject area. It was believed that the objective format possessed enough significant advantages over the essay format to justify this decision. The advantages and disadvantages of each style of test are given in TABLE XII, adapted from Gronlund (1965, p. 108).

TABLE XII

Comparative Advantages of Objective and Essay Tests

Criterion	Objective Test	Essay Test
Learning Outcomes Measured	Efficient for measuring knowledge of facts, understandings, thinking skills. Inefficient for measuring the ability to synthesize and to solve some types of problems.	Inefficient for measuring knowledge of facts. Efficient for measuring understandings, thinking skills, the ability to synthesize and to solve problems involving originality.
Preparation of Questions	A relatively large number of items are needed. Item construction is very time consuming and often difficult.	Only a few questions are needed. Item construction is easier than in objective tests.
Sampling of Content Area	Can extensively examine the subject matter because of the large number of questions.	The sampling of the subject area is limited because of the few questions that can be asked.

TABLE XII (continued)

Criterion	Objective Test	Essay Test
Control of Pupils' Response	The examinees' response is limited to the ones specified. It is difficult to bluff answers, but guessing can have a significant effect.	Because of free response, the writing skill of the examinee rather than his substantive knowledge may influence his score. Guessing, however, is minimized.
Scoring	The scoring is quick, easy, and reliable.	The scoring is slow, difficult, and unreliable because of the subjectivity involved.
Influence on Learning	Can encourage the development of factual knowledge, understandings, and higher thinking skills.	Encourages a concentration on the overall organization and structure of many ideas.

It can be seen that the choice to use the objective test format based on TABLE XII is consistent with the theoretical, statistical, and practical guidelines specified in Chapter III. First, this type of test is capable of comprehensively sampling the categories of both the Model of the Nature of Science and the Model of Cognitive Operations; second, the test is more likely to have a high reliability because of the objectivity in scoring and the restriction of the examinee's response; and third, the test is easier to score and yields more interpretable results than the essay test.

The next decision with regard to item format was to pick the multiple-choice format in favour of the other types of objective tests. The foundation for this decision can be described most accurately by the following excerpts.

> "The multiple-choice item is generally recognized as the most widely applicable and useful type of objective test item. It can more effectively measure many of the simple learning outcomes measured by the short-answer item, the alternative-response item, and the matching exercise. In addition, it can measure a variety of the more complex outcomes in the knowledge and understanding areas" (Gronlund, 1965, p. 140).

> "Multiple-choice test items are currently the most highly regarded and widely used form of objective test item. They are adaptable to the measurement of most important educational outcomes - knowledge, understanding, and judgement; the ability to solve problems, to recommend appropriate action, to make predictions. Almost any understanding or ability that can be tested by means of any other item form - short answer, completion, true-false, matching, or essay - can also be tested by means of multiple-choice test items" (Ebel, 1965, p. 149).

Another factor influencing the test format was the desire to categorize items into the various levels of Bloom's Taxonomy. Since the examinees possess a variety of learning experiences, it becomes difficult to classify questions into single Bloom categories with respect to all of them. What may require the ability to comprehend in one student, may necessitate the ability to analyze in another. At least two approaches for minimizing this effect have been made by other researchers. The first submitted items to a panel of judges to see if a concensus could be reached in categorizing the items according to Bloom (Klein, 1965, and Klinchman, 1964). However, it was found almost impossible to classify the questions without knowledge of the subjects' learning experiences. The second has included passages in the test which then become the learning experiences for each subject. These reading passages "....were selected on the basis of their probable interest value, probable ease of comprehension, and their unfamiliarity to students" (Kropp, Stoker, and Bashaw, 1966). It has been found that this latter method, in general, requires subjects to operate at the cognitive level intended by the test developer.

Since research seemed to be in favour of the second approach described above, it was decided to attempt to construct the First Preliminary Form along those lines. What was needed, then, was a content area which was probably unfamiliar to all students, so that the relevant material could be included in the test and the questions based on this material. The choice was made to deal with ancient Greek astronomy, since it was not included in any of the schools' curriculum and was, therefore, probably equally unfamiliar to all students and since it lended itself to many questions about the nature of science and scientific thinking.

Writing Multiple-Choice Items

While multiple-choice test items have certain distinct advantages over essay questions, they have the disadvantage of being difficult to construct. This problem was recognized in this study, and steps were followed which are recommended in the literature as aids in item writing. These directions, described below, are synthesized from the works of Gronlund (1965), Ebel (1965), and Nunnally (1959).

The stem of the item should be meaningful by itself, including as much of the item as possible. If the stem of a multiple-choice question is not meaningful or does not present a specific problem, the question can be considered a collection of true-false statements disguised as a multiple-choice question. This often results in the student having to decide the verity of three or four statements which are probably unrelated, rather than having to decide among three or four statements which are all possible solutions to the single problem presented in the stem. The following two versions of a test item are illustrative of these comments.

Poor: Scientists
- (a) generally have IQ scores over 150
- (b) work all day in a laboratory
- (c) usually do not enjoy going to movies
- (d) carry out experiments to test the truth of their hypotheses

Better: Scientists carry our experiments in order to
- (a) test whether hypotheses they have made are false
- (b) prove that the laws of nature are true
- (c) have a situation where measurements can be made free of error
- (d) all of the above

All the alternatives should be grammatically consistent with the stem. The main reason for taking care to have grammatical consistency between stem and alternative is not to bolster the use of correct English grammar. Rather, very often attention is given to the correct alternative making sure that it is grammatically correct while the distractors are neglected. As a result, alternatives which contain an error in grammar are automatically eliminated from being the correct answer. In this regard, particular attention was given to the tenses of the verbs and the use of the indefinite articles "a" and "an" in developing each form of the test.

Do not make the correct alternative consistently longer or shorter than the distractors. While this caution may seem trivial, it is well known that seasoned test-takers will seek and detect patterns which may give clues to the correct answer. This specific pattern, concerning the length of the correct answers, arises because the correct response usually needs to be qualified in order to be considered completely correct.

In this test, an effort was made to make all distractors approximately equal in length. Other times, two equally short and two equally long distractors were used. Often, when the correct answer could not be shortened to be comparable in length to the distractors, the distractors were lengthened by adding qualifying words. The following item is an example of this procedure.

Poor: A main goal of science is to
- (a) make practical discoveries
- (b) show the use of discoveries
- (c) provide logical explanations for events that occur in nature
- (d) improve human welfare

Better: A main goal of science is to
(a) make discoveries that have practical uses
(b) show the use of discoveries about nature
(c) provide explanations for events in nature
(d) improve human welfare as much as possible

All of the incorrect alternatives should be plausible. The purpose of incorrect alternatives is to distract the unknowing away from choosing the correct answer. In order to achieve this, these alternatives have to be as equally attractive as the correct answer.

In this test, care was taken to avoid having distractors which were obviously incorrect, and therefore could be eliminated. Whenever possible, distractors using misconceptions of the nature of science, which have been documented in other studies, were used.

The position of the correct answer among the alternatives should be chosen randomly. If the correct answer is placed among the alternatives in any other than a random manner, the chances of a detectable pattern emerging are increased. Test-takers are very apt to discover such patterns and use them to their advantage, thus invalidating test scores.

During the present test construction, the correct responses were placed randomly among the alternatives of each question using the table of random numbers found in Glass and Stanley (1970).

The use of certain alternatives such as "none of the above" or "all of the above" must be done carefully. These phrases can be added as the last alternative in a multiple-choice item in order to force the examinees to consider all the alternatives together.

The response "none of the above" can be used as the correct answer when no other correct answer is provided or as a distractor when a good answer is given. In best-answer type items, the use of "none of the above" is clearly inappropriate, since the examinee is asked to select the best of several alternatives of varying degrees of correctness. All of the alternatives possess some truth, if only minimal or questionable, and, therefore, "none of the above" cannot be the correct response. However, when using correct-answer type items, "none of the above" can be used effectively if all the other alternatives are clearly incorrect or as a distractor when, obviously, only one of the other alternatives is correct.

The use of "all of the above" is justified when attempting to evaluate whether or not the examinees realize that more than one answer can be correct. Also, it can be used as a distractor when clearly only one alternative is correct. Difficulties arise in such questions if the examinees read the first alternative, note that it is correct, and then do not bother to read further. In such a case, a correct answer gets marked wrong. This can be defended if the instructions indicate that it is the correct or the most correct answer which is sought. Another problem is that some examinees will know that two of the answers are correct and therefore correctly pick "all of the above" on the basis of partial knowledge.

It should be noted that to avoid patterns being detected, the alternatives "none of the above" and "all of the above" should appear as correct and incorrect choices approximately an equal number of times.

Another type of alternative was used in this test. The questions contain two regular alternatives and then two others, one "both of the above" and the other "neither of the above". As an example:

Scientific facts are discovered
(a) in experiments which have been often repeated with the same results
(b) in observations of nature which have been seen many, many times
(c) both a and b
(d) neither a nor b

No documentation on this type of item could be found in the literature, but it was considered by this investigator to be easier to properly construct than the "all of the above" type. The item had the flexibility of containing zero, one, or two correct responses from which to choose. Sometimes two documented misconceptions could be included opening the possibility for the unknowledgeable person to pick "both" as the correct answer.

Summary. Because of the inherent difficulties in using the types of items described in this section, special care was taken with these questions during the construction of each successive form of the test. However, during the construction of Form C, which will be described later in this chapter, certain distinct problems were identified with some of these questions. At this point some critical decisions were made which are also described in the later section.

Field Test of the First Preliminary Form

As was explained previously, the First Preliminary Form was constructed in order to get a general impression about whether or not this type of test was feasible with respect to the construction of test items, the difficulty experienced by students in reading the test, and the time required by students to write the test. The test consisted of fifty, four-alternative, multiple-choice questions based on case histories from ancient Greek astronomy and on general areas concerning the nature of science. Before this form was duplicated, slight changes were made on the basis of comments by three graduate science education students, two science education professors, and one philosophy/education professor. During this initial development, the various Bloom and nature of science categories were kept in mind and questions ranged throughout both these dimensions. However, no systematic attempt was made to insure that each level of both classifications was examined. It was believed that this procedure would not be of significant import to the stated purpose of this test form.

It should also be recorded that the quality of the duplicating process used with this form (alcohol machine) and the non-professional diagrams (drawn by the test developer) may also have had some effect on the ease of taking the test. Again, though, these are believed to be insignificant relative to the purposes outlined for this form.

The test was administered to approximately thirty grade X and thirty grade XI students in an all-boys high school. The instructions were written on the test as follows:

DIRECTIONS

> This is a multiple-choice test. You are given four answers from which to choose in each question. Put the letter of the answer you pick in the correct place on the answer sheet.
>
> If there are any words on this test you don't understand, please write them on the back of the answer sheet.
>
> If there are any sentences, or parts of sentences, you find confusing, please underline them.

Results of the First Preliminary Form

Both the grade X and grade XI students were given forty-five minutes to write the test. In this time period approximately 90 percent of both grades completed the fifty items on the test. Contrary to what was thought before the test was administered, no sentences were underlined to indicate that they were confusing. It is not certain whether this was an indication that the reading level was appropriate or whether the students forgot the instruction to underline confusing parts, when they became absorbed in the test. However, since very few questions related to reading comprehension were asked by the students and since, when questioned, students gave favourable remarks about their ability to read and understand the test, the reading level was considered reasonable. Also, since only one word, doughnut, was mentioned as being troublesome, the reading level appropriateness was given further support.

Before the test was administered, the developer also had doubts concerning whether or not the students could answer the questions with any greater accuracy than what could be expected by chance alone. However, the results of the First Preliminary Form, summarized in TABLE XIII, indicate that the students did answer the test somewhat differently than would be expected by chance alone. Many of the questions were answered far better than could be expected by guessing, while others were answered far worse.

TABLE XIII

Percentage of Grade X and Grade XI Students Choosing Each Alternative in the First Twenty Questions of the Fifty Question First Preliminary Form*

Question Number	Alternative			
	A	B	C	D
1	(10, 12)†	(83, 67)§	(3, 3)	(3, 3)
2	(20, 24)	(10, 15)	(23, 18)§	(47, 45)
3	(27, 0)§	(47, 33)	(10, 9)	(20, 58)
4	(33, 45)§	(17, 18)	(10, 3)	(43, 33)
5	(33, 24)	(30, 42)§	(30, 12)	(10, 18)
6	(33, 42)	(30, 24)§	(10, 0)	(30, 33)
7	(23, 9)	(13, 0)	(3, 12)	(63, 79)§
8	(13, 30)	(50, 39)	(27, 24)§	(13, 6)
9	(20, 24)	(0, 0)	(0, 9)	(83, 67)§
10	(7, 6)	(67, 45)§	(10, 21)	(20, 27)
11	(17, 6)	(23, 18)§	(10, 6)	(53, 67)
12	(13, 24)	(17, 50)	(27, 18)§	(47, 42)
13	(3, 3)	(33, 36)	(53, 39)§	(13, 21)
14	(47, 55)	(0, 0)	(53, 42)§	(0, 3)
15	(37, 36)	(30, 21)§	(13, 3)	(23, 42)
16	(7, 9)	(50, 58)	(30, 27)§	(13, 3)
17	(17, 18)	(40, 36)	(37, 36)	(10, 6)§
18	(7, 15)	(27, 36)§	(60, 39)	(10, 3)
19	(20, 18)	(17, 9)	(30, 39)	(30, 33)§
20	(37, 39)	(13, 21)	(37, 24)§	(20, 15)

*The patterns in the first twenty questions are similar to those found in the rest of the instrument.

†The first number of the ordered pair represents the percentage of grade X students.

§This is the correct alternative according to the Model of the Nature of Science.

From TABLE XIII, it can be seen that many of the questions were answered correctly more than 25 percent of the time. Large deviations from 25 percent are not to be expected by chance alone. Also, in several items the correct alternative was chosen by a much lower frequency than the 25 percent expected by chance, indicating that the students were purposely avoiding the correct response because of some misconception of the nature of science, or because of some characteristic in the way the alternative was written.

Summary

The construction and administration of the First Preliminary Form of the instrument was considered to be successful in meeting its objectives for the following reasons:

1. It gave the developer experience in writing these types of items and confidence that similar questions could be constructed centring around areas of science other than astronomy.
2. It indicated that grade X and grade XI students could complete the test in a reasonable amount of time.
3. It indicated that the reading level was at least approximately appropriate.
4. It showed that certain distractors were not functioning properly, since they were not being chosen by significant numbers of examinees.

CONSTRUCTION OF THE FORM FOR SCIENTISTS

Since the First Preliminary Form of the test gave some support to the feasibility of constructing the type of instrument envisaged, it was decided to attempt to expand the instrument to include other areas of science, and to examine all categories of the Model of the

Nature of Science and of Bloom's Taxonomy.

The Addition of Content Areas

Since the First Preliminary Form contained questions dealing with only one content area of science, astronomy, it was thought that the validity of the test might be questionable on these grounds. While the questions that were asked about astronomy could quite easily be asked about many other areas of science, it is conceivable that, because of some unknown factors, a person might be able to answer a question dealing with astronomy and not be able to answer the same question concerning atoms, say. For this reason, it was decided to extend the test to other areas of science while asking similar questions to those used in the astronomy section.

The criteria for deciding what content areas to choose were the same as those used in picking astronomy as an area: the subject matter had to be interesting, comprehens~~ible~~, and relatively unfamiliar to the students. In accord with these guidelines, it was decided to write questions based on both outdated and contempory ideas of continental drift and evolution.

Categorizing Items According to Bloom and the Model

In this form of the test, a systematic attempt was made to construct items ranging over all the Bloom and nature of science levels. While endeavouring to do this, certain limitations in the procedure were discovered. The first was that it was very difficult to construct items which could be classified as Synthesis. Furthermore, the two questions that have been categorized as Synthesis items are doubtful members of this class. However, this restriction becomes understandable when the criterion "Learning Outcomes Measured" in TABLE XII is examined. In general, objective tests are "inefficient for measuring the ability to synthesize and to solve some types of problems".

The second observation was that constructing test items at the various cognitive levels was applicable for only certain categories of the nature of science. For example, for the category "Controversies", only a Comprehension of Knowledge question could be devised, while for the category "Theories", questions covering a wide range of the cognitive domain were constructed, as indicated in TABLE VI, Chapter III.

Enumeration of the empty cells in TABLE XIV indicates that the instrument does not adhere to its ideal specifications of examining all areas in the Model of the Nature of Science and the Model of Cognitive Operations. This, to some degree, detracts from the test's content validity. But, for the instrument to remain manageable in length, such truncations were unavoidable. Also, it should be noted again, the very nature of objective tests dictates that certain cells be empty. This investigator suggests that one solution to this problem would be to devise four tests, one for each area in the nature of science classification. This would allow for the possibility of including questions in each Bloom category and still having tests of reasonable lengths.

TABLE XIV

Number of Questions for each
Cognitive and Nature of Science Category
in the Form for Scientists

Nature of Science Category	Bloom Category					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
AREA I						
1. Aims	2	-	1	1	-	1
2. Driving Force	2	1	-	1	-	-
3. Limits & Scope	1	-	-	1	-	-
4. Science & Technology	-	-	-	1	-	-
AREA II						
1. Facts	2	-	-	4	-	-
2. Theories	7	1	1	8	-	1
3. Hypothesis	1	-	-	4	1	-
4. Laws	1	-	1	2	-	-
5. Definitions	1	-	-	-	-	-
6. Classifications	-	-	-	1	-	-
7. Models	1	-	-	2	-	-
8. Evidence	1	-	-	7	-	2

TABLE XIV (continued)

Nature of Science Category	Bloom Category					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
AREA 111						
1. Dynamic Nature	-	-	-	1	-	-
2. Method	2	-	-	1	-	-
3. Experiments	-	-	-	1	-	-
4. Measurement	-	-	-	1	-	-
5. Falsification	4	1	-	-	1	8
6. Serendipity	-	-	1	-	-	-
7. Imagination	1	-	1	-	-	1
8. Controversies	-	1	-	-	-	-
AREA IV						
1. Simplicity	-	-	-	-	-	1
2. Reproducability	2	-	1	1	-	-
3. Interdependence	-	-	-	1	-	-

Diagramming and Duplicating

While the First Preliminary Form suffered from possible invalidities arising from the poor quality of the duplicating and diagramming, the form given to the scientists was much superior on these variables. The diagrams were drawn by trained and equipped artists and the duplication was done using an offset process rather than a spirit duplicator. It is believed that these improvements added greatly to the appearance and readability of the instrument, and therefore avoided the chance of the test-takers being unnecessarily distracted from the main purpose of the test.

Administering Test to Scientists

The administration of an intermediate draft of the test to a panel of scientists was considered a crucial step in validating the instrument. As indicated in the letter to scientists in APPENDIX A, the scientists were to answer the questions on the test and to pass any comments concerned with the context of the test items or with the structure of the items. These results were to be used to decide whether there was agreement on the answers to particular questions, and whether items needed to be rewritten to be made less ambiguous.

Choosing the Scientists

Initially, it was decided to limit the range of scientists to those who worked on the campus of Memorial University. Also, instead of approaching just any scientists, it was decided to ask the Assistant Dean of Science to suggest scientists who might cooperate. Of the approximately fifteen scientists he named, only six could be contacted because of summer vacations. Of the six approached, all were agreeable to taking the test.

In order to get a broader perspective of comments on the test, some science educators and a philosopher/educator, whose speciality was epistemology, were also asked to take the test.

The final validation panel, listed in TABLE XV, consisted of eleven people, ten of whom returned the test. It is believed that the outlook of this panel was general enough to insure that serious invalidities would not be overlooked. It is also believed that a key to the correct answers based on a concensus from this panel is also defensible.

Reading Level

The reading level of this and subsequent forms of the test was gauged to suit the high school students for whom the test was intended. An objective analysis of ten, one hundred word selections of the test was made using Fry's Readability Graph (Kennedy, 1974). This analysis showed that seven of the selections were below the grade nine level while the other three were between the grades ten and eleven levels. This reading level was assumed to be appropriate.

TABLE XV

Validation Committee Asked to Answer the Form for Scientists

Subject Speciality	Number of Professors
Physics	4*
Chemistry	2†
Biology	2§
Biochemistry	1
Psychology	1**
Philosophy	1

*Two of these people were science education professors.

†One of these people was a science education professor.

§One of these people was a science education professor

**This person did not return the test

Results

When results were received from all the validators, questions were examined individually and rewritten in accordance with the comments they had offered. Most of the changes involved changes in wording or phrasing; very few dealt with changes in the content of questions. Two questions, numbers 8 and 69, were dropped at this stage. The validators thought that the references to the Bible found in these two questions might arouse too much controversy to be worth including.

After the grammatical changes had been completed, a table giving each validator's answer to each question was constructed. This table was then examined, and questions for which seven validators gave the same answer were chosen to form part of the succeeding draft of the test. On this basis, sixty questions were chosen to remain in the next version.

When the overall response pattern was examined, it seemed as if there were two validators who frequently disagreed with the response chosen by the majority of validators. In order to investigate this observation more thoroughly, account was taken of the number of times each validators' answer was inconsistent with the majority response. These results appear in TABLE XVI. It can be seen that, indeed, two of the test validators, Validator D and Validator J, disagreed with the majority answer a significantly large number of times.

TABLE XVI

The Percentage of Times Each Validators' Response was Inconsistent with the Majority Response

Validator									
A	B	C	D	E	F	G	H	I	J
3%	14%	14%	30%	20%	20%	17%	14%	15%	50%

Because these two scientists' responses differed considerably from those of the other validators, it was decided to reexamine the items which had already been discarded for not receiving the same response from seven validators. This review led to the acceptance of eleven more items, all of which could be classed under one or the other of the following descriptions:

(i) the item obtained the same response from five validators, neither of which was Validator D or Validator J.

(ii) the item received the same response from six validators, only one of which was Validator D or Validator J.

The items selected on these bases were numbers 2, 12, 14, 18, 30, 45, 70, 74, 79, 90, and 91.

The remaining discarded items were then considered individually in an attempt to find reasons why they did not receive the same response from a majority of validators. During this process four more items were revised and accepted as valid questions. Each of these items is discussed separately below.

Item 33. This item received the following response format.

Alternative	Frequency chosen
a	5
b	2
c	3
d	0

By examining the alternatives to this item in APPENDIX A, it could be assumed that, if alternative b was not included, the people choosing this response would then pick alternative a. This assumption is based on the fact that the difference between hypotheses and theories is usually viewed as one of degree rather than one of kind. The item was thus accepted with the following modifications, and "a" as the correct alternative.

"This type of work in science can be described as
(a) hypothesizing
(b) experimenting
(c) neither a nor b
(d) both a and b"

Item 46. This item received the following response format.

Alternative	Frequency chosen
a	3
b	1
c	0
d	4

It was believed that much of the trouble with this item occurred because of the complexity of the alternatives. Each alternative required the examinee to consider two possible answers at once. However, by referring to APPENDIX A, it can be noticed that the people who chose either response a or response b, all agreed that the evidence offered in the stem of the question was "against the sun-centred explanation". With this in mind, the question was accepted as below, with "a" as the correct alternative.

"That this changing direction was not observed is evidence
(a) against the 'sun-centred' explanation
(b) for the 'sun-centred' explanation.
(c) against the 'earth-centred' explanation
(d) none of the above"

In this form, the item was accepted.

Item 83. This item received the following response format.

Alternative	Frequency chosen
a	0
b	0
c	4
d	6

It was believed that the problem with this item arose from trying to equate a type of work with particular people rather than with particular fields of endeavour. This assumption was supported by conversations held with three of the validators. Based on the comments they gave, the question was modified and accepted as written below, with "b" as the correct alternative.

"This type of work is characteristic of
(a) philosophy
(b) science
(c) technology
(d) any of the above"

Item 87. This item received the following response format.

Alternative	Frequency chosen
a	1
b	1
c	3
d	5

It was believed that problems with this item resulted from the validators equating the Greeks "self-evident ideas" with hypotheses or theories. This comparision was not a valid one, since the Greeks did not hold these ideas open to test. After some discussion with three validators, the item was modified and accepted as follows, with "d" as the correct answer.

"In science, these ideas would be called
(a) facts
(b) principles
(c) laws
(d) none of the above"

Summary

From the Form for Scientists, which initially comprised ninety-one items, a total of seventy-five items were selected. These items were valid to the extent that any ambiguities pointed out by the panel of validators were corrected, and that the chosen answer for each item was agreed upon by the majority of the panel.

These seventy-five items were further tested to study their usefulness with students and teachers. The procedures followed in this phase of the instrument development are described in the next section.

CONSTRUCTION AND ADMINISTRATION OF FORMS A AND B

At this point in the investigation, since a sizable pool of items had been collected and considered valid for the purposes of the instrument being developed, a decision was made to give a form of the test to samples from the populations for which the test was intended - grades X and XI students and grades I to XI teachers of science.

Constructing Form A and Form B

A pool of seventy-five items were already available from the analysis of the form given to scientists. These items were randomly ordered, except for certain items which had to appear sequentially. There were several groups of items which pertained to passages of information (see, for example, items 27 to 32 in APPENDIX A) and could not logically appear separated from these passages. These groups of items were placed randomly throughout the test. This collection and order of questions comprise Form A.

Since this test was intended to be given as a power test at this phase of development, it was hypothesized that items appearing near the end of the test would be affected by the examinees' degree of fatigue and other variables related to the time needed to complete the test. This effect would unnecessarily bias any analysis of these questions. To counteract this occurrence, Form B was constructed. Form B is essentially Form A reversed; that is item 75, Form A, became item 1, Form B, item 74, Form A, became item 2, Form B, etc. Again, exceptions were made in cases where questions were logically related to other questions and had to appear in a definite sequence. However, it was believed that this procedure guarded against biasing any item more than another with factors related to fatigue.

Choosing Samples to Take Forms A and B

Considering that the information gathered was intended to be used to make inferences about the instrument and not about the populations from which the samples were chosen, strictly random samples were not needed. It was argued that as long as the samples were approximately representative of the populations for whom the test was intended, then valid results could be obtained.

Student Sample. The student sample was obtained from one, large, coeducational high school. Four classes each of grade X and grade XI students were chosen. In both grades, students ranged from very high ability, through average ability, to very low ability. Approximately 150 grade X students and 110 grade XI students wrote the test.

Teacher Sample. Problems were encountered obtaining teachers to write the test which were not present when finding students. Whereas the students were more or less told to do the test by their principal, teachers had to be asked to write the test. This presented many difficulties with regard to the time needed for test administration.

In order to facilitate obtaining the teacher sample, four categories were considered: kindergarten to grade III teachers, grade IV to grade VI teachers, grade VII to grade IX teachers, and grade X and grade XI teachers. A sample of twenty-five teachers from each group was sought.

The desired number of teachers at the kindergarten to grade VI level were found in three schools. However, to obtain enough teachers at the other levels, requests had to be made to approximately six school districts. It is realized that this sample of teachers is biased to the degree that those teachers who agreed to write the test are different from those who would not agree. However, such limitations to the study were unavoidable.

Administering Forms A and B

To Students. The sample of students consisted of five grade X and four grade XI classes attending Beaconsfield Central High School, St. John's, Newfoundland. Copies of Form A and the corresponding answer sheet were distributed to every second student in each of the classes. The remaining students were given copies of Form B and the answer sheet. The directions were read and, when the students had no more questions, the test began. After one hour and forty-five minutes, approximately ninety-five percent of the students had completed the test. Since the students who had not finished had nearly half the test to complete, all forms were collected.

To Teachers. Approximately equal numbers of Form A and Form B were distributed to all the teachers in four schools: Holy Cross Primary, St. Joseph's Primary, St. Joseph's Elementary, and St. Joseph's Junior High School. All the schools were in St. John's, Newfoundland. About ninety-five percent of these teachers completed the test and returned the results. This sample totalled approximately forty teachers.

The attempt to get teachers at the high school level to do the test, by sending tests to various districts throughout the province, was all but futile. While approximately fifty tests were sent out, only about 25 percent of these were returned. The study had to be completed using this limited sample.

THE RESULTS DERIVED FROM ADMINISTERING FORMS A AND B TO STUDENTS

It was planned that, from the results of administering Forms A and B to grades X and XI students, one form of the test could be constructed which would be applicable to both grades. This rewriting would be based on an item analysis of the students' responses to the test.

Initially, four separate item analyses were run corresponding to each of the cells in TABLE XVII. These analyses yielded, for each of the four cells, each individual's score, the test mean and variance, a cumulative frequency distribution, a KR-20 reliability coefficient, and an analysis of each item giving a difficulty coefficient, the number of students choosing each distractor, and the number of students in the percentile ranges 0-20, 20-40, 40-60, 60-80, and 80-100 who chose each distractor.

TABLE XVII

Number of Students in Each Grade Who Wrote Either Form A or Form B

Grade	Form		Total
	A	B	
X	78	74	152
XI	58	55	113
Total	136	129	265

Analysis of Variance of Students' Results

If no significant differences existed between the results of Form A and Form B, or between the results of grade X and grade XI students, or between any combination of these, then the four sets of results could be combined to yield one item analysis. From this analysis a rewriting of test items could take place.

In order to check for significant differences a "two-factor, fixed-effects analysis of variance with unequal n's" was performed. The results of this analysis are described below. TABLE XVIII was derived from the item analysis results, since the values needed to compute the analysis were not explicitly produced in this program.

TABLE XVIII

The Mean Score for Each Group
Taking Either Form A or Form B

Grade	Forms		Row Means
	A	B	
X	$\overline{X}_{11}$ = 28.0	$\overline{X}_{12}$ = 28.7	28.4
XI	$\overline{X}_{21}$ = 25.9	$\overline{X}_{22}$ = 27.2	26.5
Column Means	27.1	28.1	Grand Mean 27.6

The ANOVA calculations were performed according to the description in Glass and Stanley (1970, p.436). The results of this analysis are summarized in TABLE XIX. A check of the F-ratio values shows that neither is significant at the .01 level. In other words, the hypotheses that there are no significant differences between the cell means and that there are no significant interactions between the two variables, grade and test form, cannot be rejected at the .01 level of significance.

Since the four groups of students were not shown to be statistically different groups by the ANOVA, the investigator in this study felt that the data for all the groups could be combined and used to give one set of item analysis and factor analysis results. The results of these two analyses are discussed in the following sections. In order to facilitate combining all test scores for the item analysis, the Form A items were rearranged to be in the same order as the Form B items.

TABLE XIX

Summary of the ANOVA Performed on the Results of Four Groups of Students

Source of Variation	df	MS	F	p
Students' Grade	1	214	3.91	>.05
Form of the Test	1	61.9	1.13	>.25
Interaction of Form and Grade	1	4.64	.084	>.25
Within Cells	261	14300		

Item Analysis of All Students' Responses

The item analysis procedures used in this study were performed by a computer program devised by the Division of Educational Research Services of the University of Alberta. The program, TESTØ1, scores a multiple choice test and provides most of the item statistics that are useful in improving the test. The following is the information outputed by the program:

1. the key used for scoring
2. each individuals ID number and test score
3. the test mean and variance
4. the KR-20 reliability for the test
5. a frequency distribution of test scores
6. normalized scores with mean of 13 and standard deviation of 4
7. a cumulative percentage distribution
8. cutting points for dividing the distribution into fifths
9. for each item the following are computed:
    (i) the number of subjects answering the test
    (ii) the number who omitted the item
    (iii) the proportion who answered correctly
    (iv) the keyed answer
    (v) the distribution of responses among the distractors
    (vi) the distribution of responses for each fifth of the total test score distribution
    (vii) test score and normalized test score means for the group of subjects choosing each distractor
    (viii) the biserial correlation between item response (right or wrong) and the total test score

Of all the information yielded by the item analysis, the ones considered to be most significant for improving the test were the test mean, standard deviation, KR-20 reliability, the item difficulty level, discriminatory ability, and the distribution of responses among the item distractors. The values of these statistics are given in TABLES XX and XXI.

TABLE XX

Test Statistics Results

Statistic	Value
Number of Subjects*	242
Test Mean	28.23
Test Variance	51.34
KR-20 Reliability	0.7001

*Only subjects who had completed every item were used in this analysis.

TABLE XXI

Item Statistics Results

Statistic / Item Number	Diff*	K†	Number Subjects Choosing Each Distractor				Disc§
			A	B	C	D	
1	0.488	C	26	69	118	29	.41
2	0.207	A	50	62	30	100	.10
3	0.558	B	31	135	47	29	.36
4	0.426	D	22	39	77	103	.13
5	0.298	D	52	56	62	72	.32
6	0.405	B	53	98	38	53	.33
7**	0.355	C	87	16	86	53	.21
8**	0.273	B	67	66	20	89	.06
9	0.256	B	34	62	101	45	-.14
10	0.260	A	63	64	53	61	.22
11	0.624	C	28	25	151	38	.57
12	0.409	C	36	73	99	34	.21
13	0.649	C	17	43	157	25	.55
14	0.562	A	136	31	40	34	.38
15	0.244	C	62	59	59	62	-.11
16**	0.831	C	20	16	201	5	.32

TABLE XXI (continued)

Item Statistics Results

Item Number \ Statistic	Diff*	K†	Number Subjects Choosing Each Distractor				Disc§
			A	B	C	D	
17**	0.202	A	49	22	9	162	-.03
18	0.339	A	82	33	61	66	.35
19	0.533	C	28	37	129	48	.75
20**	0.508	D	9	95	15	123	.54
21	0.269	C	69	36	65	72	.25
22	0.339	B	53	82	67	40	.12
23**	0.442	D	24	93	18	107	.54
24**	0.657	D	55	19	9	159	.52
25**	0.198	A	48	108	24	62	-.17
26**	0.632	C	18	11	153	59	.28
27**	0.620	A	150	40	38	13	.14
28**	0.434	D	85	16	36	105	.39
29	0.401	C	107	24	97	14	.29
30**	0.570	C	69	12	138	23	.52
31	0.169	A	41	131	37	33	.15
32	0.620	B	42	150	14	36	.43
33	0.252	B	75	61	46	60	.08
34	0.459	D	68	23	40	111	.41
35	0.421	D	67	30	43	102	.58
36	0.442	D	25	26	84	107	.40
37	0.211	D	61	85	45	51	.15
38	0.310	D	62	75	30	75	.30
39	0.318	A	77	59	54	52	.06
40	0.351	A	85	65	55	36	.41
41**	0.103	A	25	38	26	152	-.06
42	0.256	B	39	62	37	105	.22
43	0.521	A	126	21	48	47	[illegible]
44**	0.384	C	71	20	93	58	.28
45	0.360	D	96	29	28	87	.24
46**	0.194	A	47	51	121	23	-.28
47**	0.558	C	46	44	135	15	.68
48	0.310	A	75	61	42	62	.28
49	0.430	C	85	27	104	25	.43
50	0.649	A	157	30	37	17	.37
51**	0.165	B	47	40	136	17	-.02
52**	0.339	B	19	82	69	70	.16
53	0.298	A	72	65	69	35	.26

TABLE XXI (continued)

Item Statistics Results

Statistic / Item Number	Diff*	K†	Number Subjects Choosing Each Distractor				Disc§
			A	B	C	D	
54	0.347	B	42	84	24	90	.28
55**	0.314	D	115	36	14	76	.26
56**	0.219	C	20	21	53	147	-.08
57	0.244	D	36	70	76	59	.26
58**	0.174	B	21	42	100	78	.16
59	0.198	C	72	71	48	50	.20
60	0.248	D	69	50	62	60	.40
61**	0.062	A	15	28	64	135	-.09
62	0.227	D	45	97	45	55	.21
63**	0.463	A	112	25	16	89	.29
64	0.649	C	38	16	157	31	.32
65**	0.174	C	12	37	42	145	.00
66	0.132	A	32	57	104	49	-.07
67**	0.529	C	15	37	128	62	.41
68	0.438	D	28	71	37	106	.41
69	0.153	C	52	51	37	101	.06
70**	0.405	D	105	19	20	98	.28
71	0.583	B	15	141	71	15	.30
72	0.322	B	58	78	28	78	.02
73**	0.182	B	68	44	114	16	-.03
74**	0.376	C	20	90	91	40	.33
75	0.682	D	40	22	15	165	.54

*Item Difficulty Level; Proportion of Subjects Getting Item Correct
†The Correct Answer
§Index of Discrimination calculated by subtracting the proportion of students in the bottom 20% getting the item correct from the proportion in the top 20% getting the item correct
**This question contained at least one distractor which was chosen by less than 15% of the students answering incorrectly

An evaluation of the test based on the results presented in TABLES XX and XXI is presented in the following sections.

Test Mean. The obtained mean of 28.23 on the seventy-five item test indicates that for the sample of students selected, this test was very hard. However, such low results were expected, since the students had not received any systematic instruction in the type of material that this instrument examined. Such a result, however, does indicate that if the content examined by this test is considered an important objective of science education, then perhaps our science courses, as they now exist, suffer serious deficiencies.

Test Variance. The test variance of 51.34, converted to a standard deviation of approximately 7, indicates that the test possesses the ability to discriminate between subjects. If a normal distribution is assumed, a spread of six standard deviations would be needed to include all subjects. This would equal a range of scores of about forty to fifty on the seventy-five item test. Such a large range indicates a good discriminating ability on the part of the test.

KR-20 Reliability. The obtained reliability coefficient of 0.70 would necessitate great care when making any serious or far reaching decisions concerning students. However, it should also be noted that being a KR-20 reliability it is lower than one which would be calculated using an ordinary split-half technique.

Of more use than the reliability coefficient is the standard error of measurement (SEM) described previously in Chapter III, p. 66. Once calculated the SEM allows one to know how much confidence to place in a student's obtained score as being representative of his true score. The SEM for this test is calculated as follows:

$$\sigma_{meas} = \sqrt{51.34}\sqrt{1 - 0.70}$$
$$= 3.94$$

Since with a normal distribution a distance of $2.58\sigma_{meas}$ on either side of the mean includes 99 percent of all cases, it can be said that 99 times out of 100 a person's true score on this test will be within $2.58\sigma_{meas}$, or 10.17, of his obtained score.

Item Difficulty Level. In most books on psychological testing, it is usually recommended that test items be chosen so that they form a moderate spread of difficulty around a mean difficulty level of 0.50. Anastasi (1968, p. 164) reports on a study by Ebel which showed that a test whose items clustered around the 0.50 difficulty level yielded a wider differentiation in total scores and a higher reliability coefficient than other tests possessing a wider distribution of item difficulties.

However, since the instrument developed in this study could conceivably be used for diagnostic purposes, then the criteria of 0.50 difficulty may not be strictly applicable. For example, it is possible to imagine an item having a 0.10 difficulty level and still being a justifiable test question, since it points out a widespread deficiency in student knowledge. In this item analysis it was decided to follow the criteria specified in Chapter III, p. 68; that is, items whose difficulty level fell outside the range of 0.30 to 0.70 would be reassessed.

The frequency distribution of items for various difficulty levels for Forms A and B is given in TABLE XXII. It can be seen that forty-six of the total seventy-five items fall within the range of acceptability as defined in Chapter III. Of the remaining twenty-nine items, only one was easier than the 0.70 level. This seems reasonable when examined in the light of the fact that the sample of students had not received instruction in the areas tested by the instrument. These twenty-nine items were, however, reassessed with the understanding that the other forty-six items were not necessarily satisfactory just because they had an acceptable level of difficulty.

TABLE XXII

Number of Items at Each Difficulty Level
for Forms A and B

Difficulty Range	Number of Items
.05 - .10	2
.11 - .15	2
.16 - .20	10
.21 - .25	9
.26 - .30	5
.31 - .35	11
.36 - .40	7
.41 - .45	8
.46 - .50	3
.51 - .55	5
.56 - .60	3
.61 - .65	8
.66 - .70	1
.71 - .75	-
.76 - .80	-
.81 - .85	1

The items that were reexamined because they fell outside the acceptable range of difficulty were: items 2, 5, 8, 9, 10, 15, 17, 21, 25, 31, 33, 37, 41, 42, 46, 51, 53, 56, 57, 58, 59, 60, 61, 62, 65, 66, 69, and 73. These questions can be found in the copy of Form B contained in Appendix E.

Of these items, three - numbers 51, 61, and 65 - had originally received a poor concensus of answers from the scientists. They had been included in the test on the basis of receiving the same response from only five validators, neither of which was Validator D or Validator J. The justification for this procedure has been described previously in this Chapter under the section CONSTRUCTION OF THE FORM FOR SCIENTISTS. However, the item analysis revealed that the student responses were divided in the same way as the scientists', resulting in a high degree of difficulty. For this reason, these three items were eliminated at this point.

The remaining twenty-six items were examined in search of possible reasons why they were so difficult and, in the case of one item, so easy for the students. Relevant questions to be asked were: "Do these items tend to belong to only certain Bloom categories?" and "Do these questions belong to only certain Nature of Science categories?". TABLE XXIII, which has been extracted from Appendix D, gives the cognitive and science categorization for each of these questions.

TABLE XXIII

Cognitive and Science Classifications for the Items Whose Difficulty Level Was Outside the Range of 0.30 to 0.70

Item Number	Diff	Bloom Category	Nature of Science Category
2	.207	4.10	II-1
5	.298	4.10	II-1
8	.273	1.23	I-1
9	.256	4.20	II-2
10	.260	4.20	II-8
15	.244	1.11	II-1
16	.831	1.24	IV-2
17	.202	4.10	III-3
21	.269	4.10	II-3
25	.198	1.24	II-2
31	.169	1.24	IV-2
33	.252	1.24	I-2
37	.211	4.20	II-2
41	.103	3.00	II-2
42	.256	3.00	IV-2
46	.194	1.24	I-2
53	.298	4.10	II-4
56	.219	6.20	III-5
57	.244	4.10	III-2
58	.174	4.10	II-3
59	.198	4.10	II-3
60	.248	5.30	II-3
62	.227	1.11	II-4
66	.132	4.10	II-4
69	.153	4.30	I-4
73	.182	6.10	III-7

By inspecting TABLE XXIII, it can be seen that most of the questions which students found difficult fell into the Knowledge and Analysis categories, with eight into the former and thirteen into the latter. Based on the percentage of questions in the entire test which were classed as Knowledge or Analysis, one would expect to have seven Knowledge level and eleven Analysis level questions in these twenty-six items. A chi-square analysis showed that there was only a 25% chance that the observed and expected frequencies of these item types come from different populations.

It can also be seen that fifteen of the questions fall into category II of the science dimension, the Forms of Scientific Knowledge. From the proportion of category II items on the entire test, this is precisely the number of questions that would be expected to be answered incorrectly.

Several of the difficult items were classified in the 1.24 level of the cognitive domain, Knowledge of Criteria. Bloom states that these criteria "....are likely to appear complex and abstract to students and to acquire meaning only as they are related to concrete situations and problems" (1-56, p.70). It is not surprising, then, that students found these questions difficult.

Overall, the results indicate that most of the questions which the students found difficult are distributed throughout the cognitive and nature of science categories in the same proportions as these types of questions occur in the entire test. This would support the hypothesis that the difficulty of these items was not due to any item invalidities. Instead, students answered the questions incorrectly with a frequency that can be directly related to the frequency of occurrence of these types of items on the entire test. In conclusion, the analysis of the questions students answered incorrectly yielded no firm reasons for dropping items.

<u>Item Discriminatory Ability</u>. Very often books on designing tests will suggest that items should be chosen on the basis of their ability to discriminate between good and poor students. Therefore, items which are answered correctly by a greater proportion of poor students than

good students, and thus receive a negative index of discrimination, are discarded because they negatively discriminate. Items which are answered correctly by the same proportion of both good and poor students, and thus do not discriminate at all, are similarly discarded.

Considering the nature of the items in this test and the educational background of the students in the area examined by the test, it is conceivable that no students, either good or poor will get certain items correct. If no other reason can be found to discard these items, they will be retained, since a stated use for this test is to diagnose students' misconceptions rather than to differentiate between good and poor students. However, the criteria specified in Chapter III, p.68 of checking items with indices of discrimination below 0.20 will be followed.

By examining TABLE XXI, twelve items which discriminate negatively can be found: items 9, 15, 17, 25, 41, 46, 51, 56, 61, 66, 69, and 73. As an initial observation, TABLE XXI reveals that these items were also among the more difficult ones on the test, having an average difficulty level of 0.176. Whether or not this fact has anything to do with the negative discrimination indices can be only speculated.

Below, hypotheses are presented of why some items were answered correctly by a higher proportion of the poorer students than the better students. It must be remembered that these ideas are hypothesized and do not have empirical support. Yet, they are useful in that they would be helpful in devising any subsequent forms of the test.

Initially, it is to be noted that items 51 and 61 have, at this point, been dropped from the instrument because of reasons stated in the previous section. They will not be considered in the following discussion.

TABLE XXIV gives the distribution of answers among the four alternatives for each of the items listed above for both the top and bottom 20 percent groups. By examining Appendix E and TABLE XXIV simultaneously, it is plausible to state that in certain questions a student's intelligence would lessen his chances of getting the correct answer. This is so, because incorrect alternatives would seem plausible to him but would not even make sense to the less intelligent student. Item 15 might be just such a case because the brighter student may be familiar with all four words fact, law, theory, and hypotheses as they apply to science, while the poorer student may just have heard of facts and theories, which are the more everyday words applied to science.

In other cases, the brighter students picked alternatives ~~which~~, say, three of the scientists had chosen as the correct answer. However, this choice was considered wrong for the purposes of this test since seven scientists had chosen another answer. Again, paradoxically, the students' knowledge may be their downfall.

TABLE XXIV

Frequency that Each Alternative Was Chosen
by Both the Top and Bottom 20 Percent Groups

Item Number	Group	Alternative a	b	c	d
9	Top	4	18	38	6
	Bottom	8	17	11	5
15	Top	16	22	12	16
	Bottom	14	7	12	8
17	Top	13	1	1	51
	Bottom	9	7	2	23
25	Top	8	40	2	16
	Bottom	12	12	7	10
41	Top	6	4	4	52
	Bottom	6	11	9	14
46	Top	4	5	49	8
	Bottom	14	17	9	1
56	Top	2	0	16	48
	Bottom	4	6	13	18
66	Top	10	7	38	11
	Bottom	9	16	11	5
69	Top	15	14	7	29
	Bottom	8	9	7	17
73	Top	21	14	31	0
	Bottom	8	10	15	8

A further examination of TABLE XXIV yielded the following qualitative conclusion: except for the possible exception of item 15, it seems clear that, for each question, the answers given by the bottom 20 percent group are distributed much more randomly among the four alternatives. This would lead one to suspect that the brighter students are giving their choice of answer some thought and therefore, as a group, choosing only certain alternatives for each question. Why they picked incorrect alternatives could only be discovered through personal interviews. On the other hand, it seems reasonable to believe that the slower students are not giving their answers as much thought but are answering many questions by guessing. This type of procedure would also explain the type of response format found in TABLE XXIV. It could also explain why the slower students answered some questions better than the brighter students. While a low percentage of the brighter students chose the correct alternative because of some logical reasoning, a higher percentage of slower students chose the alternative because of chance.

Since none of the questions receiving the negative indices of discrimination were found to have any ambiguities which could lead the brighter students from the correct response, and since the hypotheses presented above are at least plausible, none of these questions were eliminated at this point.

Distribution of Answers Among Distractors. As was explained in Chapter III, p.68, each of the distractors to the correct answer on a multiple-choice test should function effectively. That is, each incorrect alternate should have misled a representative proportion of the total number of people answering the item incorrectly. Ideally, in a four-alternative multiple-choice test each distractor should receive about one third of the incorrect answers. However, no test can be expected to follow these criteria for every item.

For the purposes of this study, distractors which attracted 15 percent or more of the incorrect choices were accepted. Distractors which received a smaller proportion of the incorrect answers than 15 percent were dropped. If more than one distractor had to be dropped, then the entire item was eliminated from the test. The above procedure will lead to the situation where items on the same multiple-choice test will have either two or three distractors to the correct answer. Usually such a situation is frowned upon. However, in separate studies, Smith (1958) and Ebel and Williams (1957) have shown that there is no real reason why all items in a multiple-choice test should have the same number of alternatives and that it is quite possible to write good test items with two distractors and occasionally only one distractor.

TABLE XXI gives the distribution of responses throughout the four distractors for each test item. Any decisions concerning the elimination of distractors or of enti~~re items had to~~ rest, ultimately, not only upon the item analysis results but also upon the structure of the items and upon the structure of the test. In certain items distractors are linked to other distractors, for example, in questions with an alternative of the type "both a and b", and in other cases questions are related to other questions. In the first type of situation, decisions had to be made as to whether or not the entire item would be excluded or be left as it was. In the second type of situation, either the poor question had to be changed, left as it was, or the entire set of questions, related to the poor one, excluded. These judgments were made so as to include some of these dubious items while at the same time attempting to keep the test validity as high as possible.

The items which received consideration under this section are marked ** in TABLE XXI. There are twenty-eight of these items. Based on the previous discussion the decisions described in TABLE XXV were made. Three of the items, numbers 51, 61, and 65, had been excluded already for the reasons described in the section <u>Item Difficulty Level</u>. Two of the questions, numbers 20 and 56, were excluded at this point because they had two or more non-viable distractors, while ten of the items were left unchanged. The remaining thirteen items were changed by excluding one distractor from each, based on the criteria of 15 percent explained previously.

TABLE XXV

Results of Re-evaluating Items with Poor Distractors*

Item Number	Result of Re-evaluation
7	+
8	c
16	+
17	d
20	–
23	c
24	c
25	+
26	b
27	d
28	+
30	b
41	c
44	b
46	+
47	+
51	0
52	a
55	c
56	–
58	a
61	0
63	+
65	0

TABLE XXV (continued)

67	+
70	+
73	d
74	+

*Key + item included unchanged
- item excluded
0 item excluded already
a exclude alternate a
b exclude alternate b
c exclude alternate c
d exclude alternate d

<u>Summary</u>. It is believed by the researcher of this study that the item analysis results described in this section have increased the validity of the original Form B. While on several occasions compromises had to be made because the directions to be taken were not completely clear, it is believed that each decision has been supported by logical arguments.

The following section describes a factor analysis of the students' responses while the next section describes the results of administering Forms A and B to teachers. These sections will also address themselves to the validity of the instrument.

Factor Analysis of All Students' Responses

The factor analysis procedures used in this study were performed by two computer programs devised by the Division of Educational Research Services of the University of Alberta. The first program, FACTØ1, carries out a principal components factor analysis and applies Varimax, Quartimax, and Equamax orthogonal rotations to the principal axes factors (see Harman, 1960). The second program, FACTØ5, was used to perform an oblique transformation of the Varimax solution by the Procrustes method to match the theoretical factor structure hypothesized for the instrument (see Hurley and Cattell, 1962). The results of these analyses are given in the next two sections.

FACTØ1 Results. The program was instructed to extract four factors from the data. Four factors were chosen since the hypothesized structure, described in the following section, contained four factors. The results of the Varimax rotation indicated that the factors extracted could account for only about 20 percent of the variance in test scores. According to the factor analysis model, the remaining 80 percent of the variance is accounted for by certain unique factors, that is factors which the test items do not hold in common with one another, and by error. Most probably, the error variance contributes more to the 80 percent than the variance due to unique factors.

These results may have been caused by the low student mean which was not much more than would be expected by chance alone. However, the high inter-item consistency indicated by the KR-20 result of 0.70, would lead one to expect that distinct factors might not be found. Perhaps, also, the items can be classified according to a scheme which could not be detected in the results.

Despite these shortcomings, it was decided to continue the analysis as had been planned. The Varimax transformation factor matrix was examined and loadings greater than or equal to 0.300 were noted. It is significant to note that only sixty such loadings were found of the possible three hundred. The factors were then further studied to determine whether or not either of them resembled the hypothesized factors.

This examination yielded no discernible correlation between the extracted factors on which items loaded most heavily and the hypothesized factors for the items. This was true for both the cognitive and nature of science hypothesized factor structures. At this point, it was believed that since no qualitative explanation of the extracted factors had been found, the factor match procedures would be performed in search of some interpretable results.

<u>FACTØS Results</u>. The factor match procedures indicate to what extent the theoretical factors around which the instrument was constructed correlate with the factors emerging from the nature of the students' responses. The test developed for this study contains two underlying theoretical factor structures, one associated with the cognitive domain, the other with the nature of science domain. These two structures, derived from Appendix D, are given in TABLES XXVI and XXVII. In both structures, each item is hypothesized to have a loading of 1.00 on the factor with which it is associated. The factor match procedure was used to indicate to what degree this hypothesis was true.

TABLE XXVI

Hypothesized Cognitive Factor Structure

Item Number	Factor I	Factor II	Factor III	Factor IV
1	X			
2		X		
3		X		
4		X		
5		X		
6			X	
7			X	
8	X			
9			X	
10			X	
11			X	
12			X	
13				X
14	X			
15	X			
16	X			
17		X		
18	X			
19			X	
20	X			
21		X		
22		X		
23		X		
24				X
25	X			
26			X	
27		X		
28				X
29	X			
30			X	
31	X			
32		X		
33	X			
34	X			
35		X		
36	X			
37			X	
38			X	
39			X	
40				X
41		X		
42		X		
43	X			
44			X	
45	X			
46	X			
47	X			
48		X		
49			X	
50			X	
51				X
52				X
53		X		
54				X
55				X
56				X
57		X		
58		X		
59		X		
60				X
61				X
62	X			
63	X			
64		X		
65				X
66		X		
67		X		
68	X			
69			X	
70	X			
71		X		
72				X
73				X
74	X			
75	X			

The criteria for placing items into the various cognitive factors was based on the following scheme:

Factor	Bloom Level
I	all 1.00 and all 2.00
II	all 3.00 and 4.10
III	4.20 and 4.30
IV	all 5.00 and all 6.00

The division of the Analysis level into two factors is based on the assumption that the types of mental operations associated with the 4.20 and 4.30 levels, such as the ability to check the logical consistency of hypotheses, is significantly more complex than the operations associated with the 4.10 level, such as the ability to distinguish facts from hypotheses.

TABLE XXVII

Hypothesized Nature of Science Factor Structure

Item Number	Factor I	Factor II	Factor III	Factor IV	Item Number	Factor I	Factor II	Factor III	Factor IV
1	X				39		X		
2		X			40		X		
3		X			41		X		
4		X			42				X
5		X			43		X		
6	X				44	X			
7			X		45	X			
8	X				46	X			
9		X			47	X			
10		X			48		X		
11		X			49		X		
12		X			50		X		
13			X		51		X		
14			X		52			X	
15		X			53		X		
16				X	54			X	
17			X		55				X
18		X			56			X	
19				X	57			X	
20		X			58		X		
21		X			59		X		
22		X			60		X		
23		X			61			X	
24	X				62		X		
25		X			63			X	
26		X			64	X			
27		X			65		X		
28			X		66		X		
29			X		67		X		
30		X			68			X	
31				X	69	X			
32		X			70		X		
33	X				71		X		
34		X			72			X	
35			X		73			X	
36			X		74		X		
37		X			75		X		
38		X							

The FACTØ5 program was run to check the agreement between the factors extracted by the Varimax rotation and the two hypothesized structures: the nature of science factor structure, and the cognitive factor structure based on Bloom's Taxonomy. The results tabulated in TABLE XXVIII indicate that of the oblique factors produced in the nature of science match, factors one and two, one and three, and three and four were highly correlated with each other; factors two and three and two and four were moderately correlated; while only factors one and four had a low correlation. This observation in itself leads one to suspect that not too much success will result from an attempt to match four hypothesized factors to the four extracted factors, since the students' responses seem to contain only two distinct factors, in this case factors one and four.

Similarly undesirable results were obtained in the cognitive match, as TABLE XXVIII shows. Variables one and three were highly correlated, one and four, two and three, three and four were moderately correlated, while only factors one and two, and two and four were correlated slightly. These results indicated that probably no interpretable results will be found in this factor match either.

TABLE XXVIII A

Correlation Between Primaries for the Nature of Science Factor Match

Factor	1	2	3	4
1	1.00	-0.866	0.757	-0.086
2		1.00	-0.455	-0.315
3			1.00	-0.657
4				1.00

TABLE XXVIII B

Correlations Between Primaries for the Cognitive Factor Match

Factor	1	2	3	4
1	1.00	-0.125	-0.562	-0.244
2		1.00	-0.305	-0.064
3			1.00	-0.234
4				1.00

As was expected from the correlations between the factors reported in TABLE XXVIII, the pattern of factor loadings produced by the FACTØ5 program were largely uninterpretable. While the factor loadings were somewhat more structured than the almost random results of the Varimax rotation, one has to be careful when interpreting such results as significant. This is so, since the Procrustes method, by the connotations of the name itself, will brutally make almost any data fit almost any hypothesis. If an investigator is satisfied, he can report a good fit from mere visual judgement without testing the statistical significance of the match (Hurley and Cattell, p.260, 1962). It must be remembered that in this study the initial student results were somewhat random and, because of this, no pattern could be accepted as significant unless it was very clearly and unmistakenly a true pattern in the students' responses. No such patterns were found in the data collected for this research.

Summary. Overall, the results of the factor analysis were disappointing. The attempt to match the hypothesized factor structures to the factor analysis results derived from the students' responses yielded no discernible results. These results could be partially due to the fact that the students scored very poorly on the instrument, but more likely are due to the absence of distinct factors which the KR-20 of 0.70 would suggest.

It was believed that the teachers' responses which were significantly higher than those of the students, as reported in the following section, would have succumbed to factor analysis. However, these results were constrained by yet other limitations. A sample of only about forty teachers were used to write the test. The low numbers were unavoidable, since teachers could not be readily found who were willing to cooperate. Because of the small numbers, it was thought that a factor analysis would not be worthwhile.

The results of this section, then, were inconclusive. While the hypothesized nature of science and cognitive factor structures cannot be quantitatively supported, neither can they be rejected. The question of whether or not students can perceive these factors as underlying the test structure might be examined more fruitfully using students who had been exposed to a course designed to teach certain facets of the nature of science.

THE RESULTS DERIVED FROM ADMINISTERING FORMS A AND B TO TEACHERS

As was previously stated, the sample of teachers used in this study was indeed limited. Unfortunately, cooperation could not be received for whatever the reasons. A total of only forty-three teachers were used in the analysis of results, most of whom taught at the kindergarten to grade VI level.

Because of the limited number of teachers, it was thought inappropriate to follow the same detailed report of an item analysis as was performed with the student data. The numbers of teachers choosing certain distractors, for example, are so small in most cases as to be insignificant. However, the results were used in a more general sense, as is outlined in the following sections, to offer more support to the test's construct validity.

The Test Mean, Variance, and Reliability

The descriptive statistics produced by the TESTØ1 item analysis results of the teacher data are included in TABLE XXIX. Note that each of the statistics is larger than that obtained in the student sample. While no particular reasons exist for comparing the variance and reliability values for the two groups, there is a reason for comparing the means. It can be hypothesized that the teachers, because of their age superiority over the students, would be able to function at higher cognitive levels than the students. Also, it can be hypothesized that they will also possess a better knowledge of the nature of science because of their greater academic experience. Both of these factors would lead one to expect that teachers should score significantly higher than students. If, indeed, they do score significantly higher, support will be added to the instrument's construct validity. This hypothesis will be checked in the next section.

TABLE XXIX

Test Statistics Results (Teachers)

Statistic	Value
Number of Subjects	43
Test Mean	33.05
Test Variance	121.67
KR-20 Reliability	0.8874

Comparison of Student and Teacher Means

The hypothesis tested in this section is that the difference between the means of the student and teacher populations, $\mu_1 - \mu_2$, is equal to zero against the alternative hypothesis that it is different from zero:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

The test statistic used was the t described in Glass and Stanley, 1970, p.295.

Calculations yielded,

$$t = 3.80$$

Referring to TABLE D (Glass and Stanley, p.521), it can be seen that H_0 can be rejected at the .001 level of significance. Indeed, a value of t = 3.80 units has a probability of approximately .0005 if H_0 is true.

The rejection of the hypothesis that the means of the student and teacher samples come from the same population and the acceptance of the hypothesis that the teacher mean is significantly greater than the student mean at the .001 level of significance indicates that the instrument may be measuring what it purports to measure. The validity of the instrument is increased to the degree that a higher teacher mean is related to validity. The argument that they are related has been presented in the previous section. However, care must be taken in relying too heavily on these results since the difference in teacher and student means is not all that large. The previous discussion on educational and statistical significance (see p.19) warrants this caution.

SUMMARY

This Chapter has presented a detailed account of the procedures followed during the development of the "Test on the Nature of Science and Scientific Thinking". Each step has been analyzed with respect to its logicalness and significance to the test validity. While good factor analysis results, which would have given much support to the instrument's validity, were missing, it is believed that the limits to the scope of this study have been reached. The next logical step, if the development of this instrument were pursued, would be to administer the revised instrument to a wider base of students who had been instructed in the nature of science.

CHAPTER V

SUMMARY AND DISCUSSION OF FINDINGS

This study attempted to develop a valid, reliable, and usable instrument to measure certain aspects of the nature of science and scientific thinking. That students should be instructed in these areas of scientific knowledge is well documented in science education journals. Also found in the literature are instruments which have been designed to measure various facets of this area of knowledge. However, none of the reviewed tests dealt with quite the same domain as the one constructed in this research and nor were any of them developed along the same theoretical dimensions. To the extent that the instrument designed in this study is unique, as well as to the degree that it is valid, reliable, and usable, it is a contribution to the field of science education.

Summary of the Procedure

The initial stages in the development of the test were spent delineating the domain of knowledge which was to be examined. This consisted of developing an extensive model of the nature of science based upon the writings of several eminent writers in the philospophy of science. Only non-controversial areas in this domain were chosen to form part of the model. The model addressed itself to the question "What is science?", to the forms in which scientific knowledge is cast, to the manner in which scientific knowledge emerges, and to the characteristics of scientific knowledge. The model did not discuss such areas as the relationship between science and society or the funding of scientific research. It was believed that the inclusion and exclusion of certain areas of knowledge are largely subject to the uses for which the test developer wishes the instrument to be applied.

First Preliminary Form. Once the first draft of the model had been completed it was decided to attempt to construct a sample set of items which could be given to some grade X and XI students in order to obtain information on the feasibility of the test development. Since the instrument was to measure knowledge of the nature of science and scientific thinking, which supposedly is not dependent upon knowledge of any particular content area in science, it was decided that, in order not to bias the test in favour of any students because of their knowledge of content, the main body of test items would be constructed around content which would be largely unknown to all students.

The content which was chosen was ancient Greek astronomy because it was not included in any of the school curricula. Related to this topic several multiple-choice questions were constructed, some of which were grouped into five or six items centring around one particular description or problem. In total, fifty, four alternative multiple-choice items were designed for this preliminary form. The multiple-choice format was chosen in favour of the essay mainly because of the objectivity in scoring and the ability to cover a wide range of content.

This form of the test was administered to approximately thirty grade X and thirty grade XI male students. The results were encouraging in that nearly all the students completed the test in less than one hour and there were very few complaints received from the students when they were asked about the readability of the test. These results, although only qualitative, indicated that the project which was planned was at least feasible.

Form for Scientists. The next step in the test development was to extend the test beyond the area of astronomy while still keeping the content largely new to the students. It was decided to develop further questions based upon old and new ideas concerning both the evolution of living things and continental drift. The resulting version consisted of ninety-one, four alternative multiple-choice questions based on the three content areas and on more general facets of the nature of science.

This form was given to ten validators, consisting of practicing scientists, science educators, and a philosopher, who were asked to choose the answers they believed were correct and also to comment upon any question concerning its importance with respect to the nature of science, its readability, its clearness, etc. From these results, questions were rewritten and an answer key, based on 70 percent agreement among the validators, was constructed. Only seventy-five of the initial ninety-one items survived this stage of validation.

Forms A and B. Because this test was going to be a power test when administered to students, it was decided to construct two forms, one having its items in the reverse order to the other, from the seventy-five items. This technique was used to counteract any fatigue effects which might bias students' responses to the items nearer the end of the test.

These forms, Form A and B, were then administered in approximately equal numbers to about 250 grade X and grade XI students and about forty-five grade I to XI teachers. In order to analyse the results, the items of both forms were afterwards transformed into the same order.

From the students' and teachers' responses a detailed item analysis was performed, on the basis of which several items were either excluded from the test or rewritten. Also, the students' responses were used to carry out factor analysis procedures in search of any underlying patterns in the way students viewed the test, and to determine to what degree the students' view of the test structure, as exemplified in their answers, was similar to the theoretical structures underlying the test construction.

Summary. The methods followed in developing the instrument to the Form A and B stage were both arduous and time consuming. Before any attempt could be made at item construction, a detailed analysis and synthesis of the literature in the field had to be done. The next step required the assistance of a panel of validators to decide to what degree the test items were really measuring important facets of the nature of science and scientific thinking. Based upon the panel's analysis, the test was rewritten to comprise Forms A and B. These forms were then administered to a sample of teachers and students whose responses were used to do item and factor analyses.

It is believed that each of these steps contributed to the validity of the instrument. Basing the test on both an extensive model of the nature of science and on Bloom's model of cognitive abilities increases the test's content validity. Having a panel of people, who should be knowledgeable in the domain examined by the test, comment on the validity of the test items also contributes to the content as well as the construct validity of the test. Modifying the test on the basis of an item analysis of students' responses also increases the construct validity as the variance in scores of the revised version should be less attributable to error arising from grammatical and/or logical ambiguities. Finally, attempting to explain the students' responses to the test on the basis of a small number of factors or traits, and attempting to match the theoretical factor structure underlying the test to the factor structure as viewed by the students, also increases substantially the instrument's construct validity.

Summary of the Findings

Since the purpose of this thesis was to construct a valid, reliable, and usable testing instrument, the findings should address themselves to a discussion of the extent to which this goal has been fulfilled. The criteria for determining test validity, reliability, and usability are somewhat more than arbitrary but somewhat less than unalterable. Because of this, the burden is placed upon the test developer to decide, in many cases based upon incomplete information, to what degree certain decisions will affect test validity. This section treats the three characteristics of the test validity, reliability, and usability and the extent to which they are contained in the test.

Validity. Validity is an all-embracing test characteristic which includes the traits of reliability and usability as subsets. Validity is concerned with what a test measures. The validity of a test is a function of any variable that can affect the test outcome.

These may include such things as the subjects' knowledge, manipulative skill, visual acuity, or even degree of fatigue. Validity is concerned with the extent that the final score is dependent upon each variable affecting the score. For example, if in an extreme case 90 percent of the variance in scores on an arithmetic test is due to the subject's ability to follow instructions, then the test is not valid as a test of arithmetical knowledge but may well be considered valid as a measure of the subject's ability to follow directions.

A test developer can never be certain that his test is valid as there are no purely objective measures of test validity. However, test validity can be inferred if one follows certain procedures which are considered by experts in the field to increase test validity. In this test development, such inferences were made concerning the instrument's content validity and construct validity.

A test possesses content validity to the degree that it adequately samples a specified universe of content (Ebel, 1965, p. 380). The content validity of this test is supported by the fact that it was designed to examine as completely as possible a comprehensive model of the nature of science at various cognitive levels. The model of the nature of science was derived from an extensive review of the literature in the philosophy of science and from an exhaustive examination of the models used to construct previous instruments in this area of knowledge. The model of cognitive abilities used was that developed by Bloom and his associates, a model which has received widespread acclaim in educational circles. Any items which were considered by a majority of panel members, of a ten person validating panel, as not being important questions to ask in the area in question were dropped.

The construct validity of a test is concerned with the extent that the test is measuring some psychological construct or trait. An instrument is usually considered more valid if it can be demonstrated that certain constructs or traits possessed by the subjects account to some degree for performance on the test (Ebel, 1965, p. 380). The present instrument was designed around such traits, some of which were associated with the nature of science while others were related to the nature of thinking. To the degree that these theoretical models are valid and to the degree that the instrument reflects these models, the instrument possesses construct validity.

The procedures which are usually employed to demonstrate the construct validity of an instrument are those of factor analysis. The principal goal of factor analysis is to resolve a set of variables into a small number of categories or factors which convey all the essential information contained in the original set of variables. The chief aim is to achieve a more parsimonious structure than originally existed (Harman, 1960, p. 4).

Unfortunately, the factor analysis attempted of the students' responses to this instrument did not produce a simplified structure. Neither did a factor match attempt using the Procrustes method succeed in verifying the hypothesized factor structure on which the test was constructed. These results were not unexpected after a previously completed item analysis had shown that the students had responded to the test in an essentially random fashion. Trying to impose structure on randomness is a task hardly likely to be productive.

The factor analysis results showed that, for the sample of students who took the test, their responses could not be described by any logical set of factors, thus limiting the evidence for construct validity. It is believed that this outcome was partly a result of the students' inability to answer the test using sound knowledge because of their lack of instruction in the required area of knowledge. It is also believed that the theoretical factor structure of the instrument, based on the nature of science model and Bloom's cognitive model, are better than the factor analysis results indicate. The logical next step to be taken in the development of this instrument is to administer the test to students who have been instructed in the appropriate areas of knowledge. However, this move is beyond the scope of the present study.

Another procedure used to support a test's construct validity is to identify two groups of people who possess, in different amounts, certain abilities needed to do well on the test. If the group possessing the higher ability does significantly better on the test, the test's construct validity is given support.

This procedure was followed in the present study by comparing student and teacher mean scores. Teachers were assumed to possess higher mental abilities than students permitting them to operate more readily at higher levels of the cognitive domain. Also, they were assumed to have a greater knowledge of the nature of science, since most had completed more science courses than the students. It was hypothesized that the teachers' mean score would be significantly higher than the mean score of the students. Results showed that this hypothesis could not be rejected at the 0.001 level of significance.

The validity of the instrument is also dependent upon any other variables which may affect test scores. Two of these, reliability and usability, are discussed in the following sections.

Reliability. The reliability of a test is the consistency with which a test measures. The validity of a test is an obvious function of its reliability, because a test which yeilds inconsistent results can hardly be valid. The reliability of a test is measured by a coefficient which ranges from 0.0 for a totally unreliable test to 1.0 for a totally reliable instrument. Total unreliability would imply that people respond to the test in nothing but a random fashion, while total reliability implies that people would receive exactly the same score on a test every time they wrote it.

Reliability coefficients always lie somewhere in between 0.0 and 1.0. The ideal size desired depends very heavily on the uses to which the test results are to be put. The more important the decisions to be made using the test results, the higher the reliability coefficient needs to be. In this study, the reliability coefficients were 0.70 and 0.89 for the student and teacher scores respectively. The 0.89 coefficient is very acceptable thus allowing decisions concerning the teachers that are trustworthy. The student reliability coefficient, while much lower than the teachers', is surprisingly high considering the seemingly random responses which the students gave. This result can probably be explained if the item difficulty levels in TABLE XXI are examined. It can be seen that approximately one quarter of the items have difficulty levels of 0.50 or higher. These items are not typical of the whole test, since most of the items were much harder than this. It is believed that these items made the reliability coefficient spuriously high.

For the reasons cited above, it is believed that the reliability coefficient obtained for the students is misleadingly high in the sense that the instrument is not applicable to the sample of students to whom it was administered. Given to students who had been instructed in the domain of knowledge examined the instrument would have been more suitable, more reliable, and hence more valid.

Usability. The validity of a measuring instrument of any sort is contingent upon its usability. If the instrument is such that certain factors make it difficult or awkward to use, the instrument will not measure as effectively those things for which it was designed. The instrument developed in this study was designed to be easily administered, easily scored, and easily interpreted.

Experience with the test showed that the seventy-five items could be completed in less than one and one half hours by the students. Schools have been accustomed to administering tests of much longer duration. Also, it was found that minimal directions over and above those shown in APPENDIX E were needed. This was so because students were quite used to taking multiple-choice tests.

However, the effects of student fatigue in a test of such duration could possibly play a significant role in determining test scores. Therefore, it is suggested that in any further refinement of this instrument consideration be given to the idea of dividing the test into two parts to be administered at different times. Such a split would necessarily have to take into account the relationship between certain groups of items which make sense only when juxtaposed.

The test is easily scored, as are all multiple-choice tests, and the computer program described for item analyzing the results is easily set up. However, whether or not the instrument has applicability to the classroom situation is questionable. While it is ideal to expect test results to be interpretable by the regular classroom teacher, this investigator believes that the average high school science teacher may have great difficulty using this test for diagnostic purposes. It is believed that only a person extensively versed in the nature of science, examining the student results item-by-item, could make significant use of this instrument. In this respect, the instrument is not usable for the average teacher. To the degree that the instrument is not usable, this study is limited.

Implications for Education

This study has attempted to develop a valid instrument for measuring knowledge of the nature of science and scientific thinking. This is a difficult task since much of the content is concerned with "intangibles" with which even scientists do not unanimously agree. Nevertheless, a model of the nature of science has been proposed and a measuring instrument designed to test knowledge of the model has been presented. The strengths and limitations discovered in the instrument have been discussed at length.

While the instrument is not a finished product, it is believed that a contribution, however small, has been made to an area in which little work has been done. It is also believed that the nature of science model has implications for curriculum development as well as for test construction.

BIBLIOGRAPHY

BOOKS

Anastasi, Anne. Psychological Testing. Third Ed. New York: The MacMillan Company, 1968.

Anderson, Alan H. Jr. The Drifting Continents. New York: G. P. Putnam's Sons, 1971.

Barber, Bernard. Science and the Social Order. New York: Collier Books, 1962.

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

Bronowski, J. The Common Sense of Science. Cambridge: Harvard University Press, 1955.

Biological Sciences Curriculum Study. High School Biology. Chicago: Rand McNally and Company, 1963.

__________. Biological Science: Molecules to Man. Boston: Houghton Mifflin Company, 1963.

Campbell, William G. Form and Style in Thesis Writing. Boston: Houghton Mifflin Company, 1969.

Carter, G. S. A Hundred Years of Evolution. London: Sidgwick and Jackson, 1957.

Cattell, Raymond B., and Frank W. Warburton. Objective Personality and Motivation Tests. Chicago: University of Illinois Press, 1967.

Clodd, Edward. Pioneers of Evolution from Thales to Huxley. Freeport: Books for Libraries Press, 1972.

Conant, James B. On Understanding Science. London: Yale University Press, 1951.

Copi, Irving M. Introduction to Logic. London: The MacMillan Company, 1968.

Cronback, Lee J. Essentials of Psychological Testing. Second Ed. New York: Harper & Row, 1960.

DuToit, Alex. L. Our Wandering Continents. Edinburg: Oliver and Boyd, 1937.

Ebel, Robert L. Measuring Educational Achievement. New Jersey: Prentice Hall, 1965.

Eiseley, Loren. Darwin's Century. New York: Doubleday Anchor Books, 1958.

Gardner, Eldon J. History of Biology. Minneapolis: Burgess Publishing Company, 1972.

Gerberich, J. R. Specimen Objective Test Items. New York: Longmans, Green and Co., 1956.

Glass, Bentley, Owsic Temkin, and William L. Strauss, Jr. (eds.). Forerunners of Darwin: 1745 - 1859. Baltimore: The John Hopkins Press, 1959.

Glass, Gene V., And Julian C. Stanley. Statistical Methods in Education and Psychology. New Jersey: Prentice Hall, 1970.

Gronlund, Norman E. Measurement and Evaluating in Teaching. New York: The MacMillan Company, 1965.

Hanson, Norwood Russell. Patterns of Discovery. Cambridge University Press 1961.

Harman, Harry H. Modern Factor Analysis. Chicago: The University of Chicago Press, 1960.

Heath, Sir Thomas L. The Copernicus of Antiquity. New York: The MacMillan Company, 1920.

________. Greek Astronomy. New York: J. M. Dent & Sons, Ltd., 1932.

Heisenberg, Werner. The Physicists Conception of Nature. London: Hutchinson Scientific and Technical, 1958.

Hempel, Carl. Philosophy of Natural Science. New Jersey: Prentice-Hall, 1966.

Kaplan, Abraham. The Conduct of Inquiry. San Francisco: Chandler Publishing Company, 1964.

Kemeny, John G. A Philosopher Looks at Science. New York: Van Nostran Reinhold Company, 1959.

Krathwohl, David R., et.al. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, Inc., 1956.

Kuhn, Thomas S. The Structure of Scientific Revolutions. Second Ed. Chicago: The University of Chicago Press, 1970.

Ley, Willy. The Drifting Continents. New York: Weybright and Talley, 1969.

Moore, Patrick. The Development of Astronomical Thought. Edinburg: Oliver & Boyd, 1969.

Nagel, Ernest. The Structure of Science. New York: Harcourt, Brace & World, Inc., 1961.

Nagel, Ernest, S. Bromberger, and Aldof Grünbaum. Observation and Theory in Science. Baltimore: The John Hopkins Press, 1969.

Nunnally, J. C., Jr. Tests and Measurements. New York: McGraw-Hill Book Company, Inc., 1959.

Osborn, Henry Fairfield. From the Greeks to Darwin. New York: Charles Scribner's Sons, 1929.

Pannekock, A. A History of Astronomy. London: George Allen & Unwin Ltd., 1951.

Popper, Karl R. The Logic of Scientific Discovery. Toronto: University of Toronto Press, 1959.

Reichen, Robert S. The Stars and Serendipity. New York: Pantheon Books, 1971.

Sheffer, Isreal. Science and Subjectivity. Indianapolis: The Bobbs-Merrill Company, Inc., 1967.

Smallwood, William L., and Edna R. Green. Biology. Silver Burdett Company, 1968.

Stanley, Julian C., and Kenneth D. Hopkins. Educational and Psychological Measurement and Evaluation. New Jersey: Prentice-Hall, Inc., 1972.

Tabeuchi, H., S. Uyeda, and H. Kanomori. Debate about the Earth. San Francisco: Freeman, Cooper & Co., 1970.

Theory into Action. National Science Teachers Association. Washington, D. C., 1964.

Thorndike, Robert L. (ed.). Educational Measurement. Second Ed. Washington, D. C.: American Council on Education, 1971.

Thorndike, Robert L., and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. Third Ed. New York: John Wiley & Sons, Inc., 1969.

Toulmin, Stephen. The Philosophy of Science. London: Hutchinson's University Library, 1953.

Wrightstone, J. Wayne, et.al. Evaluation in Modern Education. New York: American Book Company, 1956.

PERIODICALS

Aikenhead, Glen S. "The Measurement of High School Students' Knowledge About Science and Scientists", Science Education, 57(4): 539-49, 1973.

Baumel, Howard B., and J. Joel Berger. "An Attempt to Measure Scientific Attitudes," Science Education, 49(3): 267-9, 1965.

Carey, Russell L., and Nyles G. Strauss, "An Analysis of the Understanding of the Nature of Science by Prospective Secondary Science Teachers," Science Education, 52(4): 358-63, 1968.

Cooley, William W., and Leopold E. Klopfer, "The Evaluation of Specific Educational Innovations," Journal of Research in Science Teaching, 1(1): 73-80, 1963.

Crumb, Glenn H. "Understanding of Science in High School Physics," Journal of Research in Science Teaching, 3(3): 246-50, 1965.

Ebel, Robert L., and Bob J. Williams. "The Effect of Varying the Numbers of Alternatives per Item on Multiple-Choice Vocabulary Test Items," The 14th Yearbook of the National Council on Measurements used in Education, 63-65, 1957.

Hurley, J. R., and R. B. Cattell. "The Procrustes Program," Behavioral Science, 7: 258, 1962.

Issertedt, Glen A., and Donald J. Schmidt. "Attitudes Concerning Scientists: A Study of a Selected Group of High Ability Students," School Science and Mathematics, 71(2): 116-20, 1971.

Jones, Katherine M. "The Attainment of Understandings About the Scientific Enterprise, Scientists, and the Aims and Methods of Science by Students in a College Physical Science Course," Journal of Research in Science Teaching, 6(1): 47-9, 1969.

Kahn, Paul. "An Experimental Study to Determine the Effect of a Selected Procedure for Teaching the Scientific Attitude to Seventh and Eight Grade Boys through the Use of Current Events in Science," Science Education, 46(2): 115-27, 1962.

Kennedy, Keith. "Reading Level Determination for Selected Tests," The Science Teacher, 41(3): 26-27, 1974.

Kimball, Merritt E. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers," Journal of Research in Science Teaching, 5(2): 110-20, 1967.

Klinchman, E. "The BSCS Grid for Test Analysis," BSCS Newsletter, 19: 17-21, 1964.

Klopfer, Leopold E., and William W. Cooley. "The History of Science Cases for High Schools' in the Development of Student Understandings of Science and Scientists," Journal of Research in Science Teaching, 1(1): 33-47, 1963.

Kropp, R. P., H. W. Stoker, and W. L. Bashaw. "The Validation of the Taxonomy of Educational Objectives," The Journal of Experimental Education, 34(3): 1966.

Lowery, Lawrence F. "An Experimental Investigation into the Attitudes of Fifth Grade Students Toward Science," School Science and Mathematics, 67: 569-79, 1967.

McQuitty, L. "Hierarchical Syndrome Analysis," Educational and Psychological Measurement, 20: 293-304, 1960.

_________. "Improved Hierarchical Syndrome Analysis of Discrete and Continuous Data," Educational and Psychological Measurement, 26: 577-82, 1966.

Moore, Richard W. "A Profile of the Scientific Attitudes of 672 Ninth Grade Students," School Science and Mathematics, 71(3): 229-32, 1971.

Moore, Richard W., and Frank X. Sutman. "The Development, Field Test and Validation of an Inventory of Scientific Attitudes," Journal of Research in Science Teaching, 7(2): 85-94, 1970.

Nelson, Clarence H., and John M. Mason. "A Test of Science Comprehension for Upper Elementary Grades," Science Education, 47(4): 319-30, 1963.

Olstad, Roger G. "The Effect of Science Teaching Methods on the Understanding of Science," Science Education, 53(1): 9-11, 1969.

Pella, Milton O., George T. O'Hearn, Calvin W. Gale. "Referents to Scientific Literacy," Journal of Research in Science Teaching, 4: 199-208, 1966.

Poole, Richard L. "Characteristics of the Taxonomy of Educational Objectives: Cognitive Domain," Psychology of Schools, 8(3): 379-85, 1971.

_________. "Characteristics of the Taxonomy of Educational Objectives: Cognitive Domain. A Replication," Psychology of Schools, 9(1): 83-8, 1972.

Schwirian, Patricia M. "On Measuring Attitudes Toward Science," Science Education, 52: 172-9, 1968.

Schwirian, Patricia M., and Barbara Thomson. "Changing Attitudes Toward Science: Undergraduates in 1967 and 1971," Journal of Research in Science Teaching, 9(3): 253-9, 1972.

Simpson, Ronald D., John W. Shrum, and R. Robert Rentz. "The Science Support Scale: Its Appropriateness with High School Students," Journal of Research in Science Teaching, 9(2): 123-6, 1972.

Smith, Kendon. "An Investigation of the Use of 'Double Choice' Items in Testing Achievement," Journal of Educational Research, LI: 387-89, 1958.

Smith, Richard B. "An Empirical Examination of the Assumptions Underlying the Taxonomy of Educational Objectives: Cognitive Domain," Journal of Educational Measurement, 5(2): 1968.

________. "An Empirical Investigation of Complexity and Process in Multiple-Choice Items," Journal of Educational Measurement, 7(1): 1970.

Trent, John. "The Attainment of the Concept 'Understanding Science' Using Contrasting Physics Courses," Journal of Research in Science Teaching, 3(3): 224-9, 1965.

Welch, Wayne W., and Milton O. Pella. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," Journal of Research in Science Teaching, 5(1): 64-8, 1967.

Wood, Roger L. "University Education Students' Understanding of the Nature and Processes of Science," School Science and Mathematics, 72(1): 73-9, 1972.

DISSERTATIONS AND STUDIES

Cooley, William W., and Leopold E. Klopfer. "Manual: Test on Understanding Science," Educational Testing Service, New Jersey, 1961.

Klein, M. "Evaluation of Instruction: Measurement of Cognitive Behaviors as Defined by the Taxonomy of Educational Objectives," unpublished doctoral dissertation, University of California, Los Angeles, 1965.

Welch, Wayne W. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," unpublished doctoral dissertation, University of Wisconsin, 1966.

APPENDICES

APPENDIX A

Form of Test Given to
Research Scientists, Science Educators, and Philosophers
for Answers and Commentary

MEMORIAL UNIVERSITY OF NEWFOUNDLAND
St. John's, Newfoundland, Canada A1C 5S7

Department of Curriculum and Instruction

Telex: 016-4101
Telephone: (709) 753-1200

Box 18
July 8, 1974

Dear Scientist,

Please excuse the impersonal flavour of the salutation of this letter but it was unavoidable, since I did not know which scientists I would be able to contact and receive cooperation from for my study.

The content of this communication restates much of our conversation with regards to the purpose of my study and the role you will be playing in it. Your participation in this work is deeply appreciated, and it is hoped that your time and energy will be rewarded with improvements in the Newfoundland school science curriculum.

The purpose of the study is to develop and validate an evaluation instrument for grades 10 and 11 high school students and for grades 1 to 11 science teachers. The test will assess the examinees' knowledge of certain aspects of the nature of science and scientific thinking.

It is realized that a science course teaching these facets of science does not exist in our schools at the present time. However, it is believed that the model of the nature of science on which the instrument is based could also be used as a guideline for developing such a curriculum.

Your role, as well as that of other cooperating scientists who have been approached, will be to answer the questions on the test and to pass comments where you think appropriate. Such comments should primarily be concerned with the content of the test items, with regards to their actually testing an important component of the nature of science or scientific thinking, and secondarily, with the structure of the items, with reference to their clarity, historical accuracy, grammar, etc.

A compilation of your comments and answers and those of the other scientists will be used as a validation check. First of all, answers will be compared to see if all or most scientists agree. Any large disagreement on any particular item will result in the elimination of the question, since only items on which there is agreement will be considered, for the purposes of this study, to be testing valid components of the nature of science. Second, your comments and those of the other scientists will be used to rewrite the items. The revised test will be administered to a sample of students and teachers in the Fall of 1974, and hopefully will eventually be standardized for use in Newfoundland schools.

It is also intended that no extensive amount of time be spent answering these questions. They were designed so that even a person who had never heard of the particular content being discussed, but who has a generalized knowledge about science, would still be able to answer the questions intelligently.

Yours, truly,

Stephen Norris

A PRELIMINARY FORM OF A TEST ON THE NATURE OF SCIENCE AND SCIENTIFIC THINKING

Information

Please indicate your subject speciality. ___________

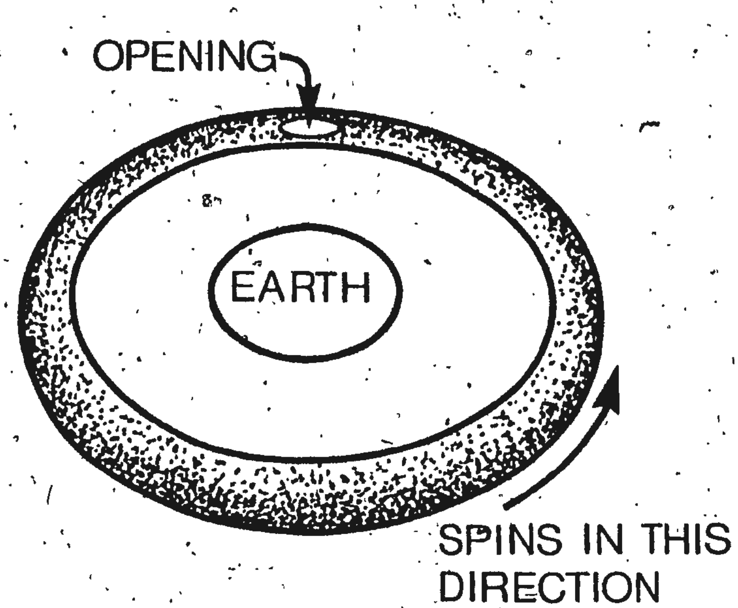
DIRECTIONS

This is a multiple choice test. You are given four answers from which to choose in each question. Put the letter of the answer you choose in the correct place on the answer cheet.

1. Scientific facts are discovered
   (a) in experiments which have been repeated with the same results very many times
   (b) while looking for predictions that have been made using scientific explanations
   (c) in observations of nature which have been seen many, many times
   (d) all of the above

2. One of the well supported theories of biology is the Theory of Evolution. It would be true to say of this theory, or indeed any well supported theory in science, that
   (a) it is not a belief nor an observational fact
   (b) it is useful in that it can explain a body of facts
   (c) both a and b
   (d) neither a nor b

3. Anaximander, a Greek who lived from 611 BC to 547 BC, had some ideas which he used to explain the behavior of the moon. He claimed that the moon is a spinning, doughnut-shaped object, surrounding the earth. (see diagram below) He said the "doughnut" is full of fire, having just one round opening for the light to escape. This opening is what we see in the night sky. He also claimed that the opening had a shutter, which could be adjusted to any position from fully opened to fully closed. Anaximander's ideas would be called a scientific
   (a) principle
   (b) postulate
   (c) neither a nor b
   (d) both a and b



4. If no other ideas explaining the behavior of the moon were available, scientists would
   - (a) accept Anaximander's ideas, because there is no other explanation available explaining how the moon behaves
   - (b) accept Anaximander's ideas, because they do explain some of the moon's behavior
   - (c) reject Anaximander's ideas, because they cannot explain all of the moon's behavior
   - (d) none of the above

5. Anaximander used only the observations of the moon he had made, plus his imagination, to come up with his explanation of the moon's behavior. In science, using one's imagination is
   - (a) not all right, because too often false explanations, like Anaximander's, are the result
   - (b) all right, because very often an answer cannot be reached using any other method
   - (c) not all right, because explanations should be reached using the most logical procedures known
   - (d) all right, because this is the best way to reach explanations as quickly as possible

6. Very often in science an old theory is replaced by a new theory which scientists believe is better. What characteristics must the new theory have before this can be done?
   - (a) It should be able to account for all the facts that the old theory accounts for, plus some more.
   - (b) If it cannot account for more facts than the old theory, then it should be simpler and more convenient.
   - (c) It should be able to predict phenomena that were not even known when the theory was made up.
   - (d) any of the above

7. Which of the following statements about science and technology is true?
   - (a) Science and technology are different names for the same area of work.
   - (b) Without science, technology could not advance as rapidly; but without technology, science could advance just as rapidly.
   - (c) Both the aims and products of science differ from the aims and products of technology.
   - (d) both b and c

8. According to the Bible, everything on the earth was created in a single week: man, rock, and tree, and has remained nearly unchanged ever since. However, observations of the earth show that it is always changing. Some mountains are washing away while others have grown. Some types of animals and plants have died out while others have been born.

   James Ussher, an Irish bishop, used clues given in the Bible to calculate 4004 BC as the date of creation. However, evidence gathered by scientists give clues that the earth is billions of years old.

In general, scientists are not satisfied with the explanation of creation given in the Bible because
(a) they want to have explanations which agree with what is observed
(b) their curiosity forces them to look for a better explanation
(c) they do not consider the Bible a scientific book, and therefore do not use it for scientific purposes
(d) all of the above

9. Before a scientist would even consider accepting an explanation as true, the explanation must be able
(a) to make some correct predictions of what will happen in certain situations
(b) to explain correctly all the facts related to the problem at hand
(c) neither a nor b
(d) both a and b

10. The ancient Greeks knew that the sun did not travel the same path across the sky every day. Day by day, from June to December, they observed that the noon day height of the sun above the southern horizon became less and less. As this was happening, they also observed that the number of hours of daylight became less and less. The statement relating these two observations is known as a scientific
(a) law
(b) hypothesis
(c) fact
(d) theory

11. This relationship between the height of the sun above the horizon and the number of hours of daylight could be used to tell
(a) from the position of the sun when to plant or harvest crops
(b) the approximate time of year from the sun's position
(c) both a and b
(d) neither a nor b

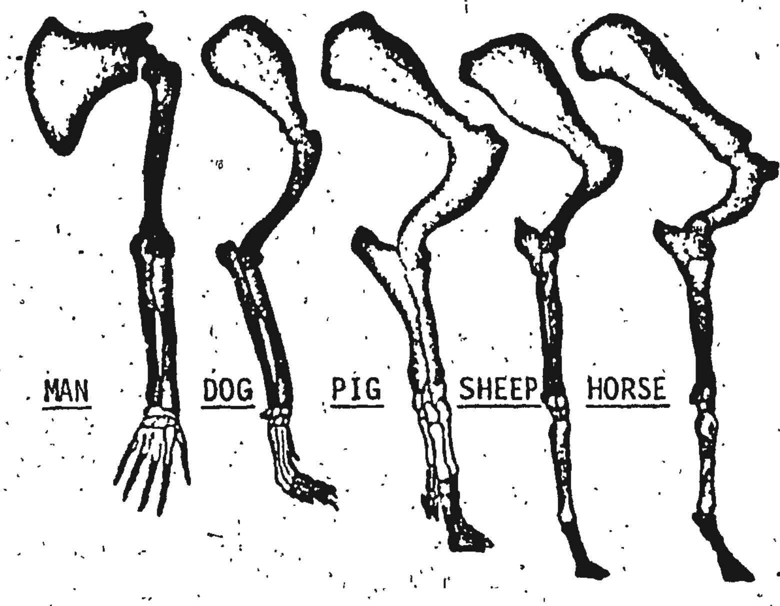
12. Many biologists believe that forms of life can change over many years to become new forms. Two of the pieces of evidence they supply for this belief are given below.

EVIDENCE X

When the poison DDT was first used to control insects, it killed nearly all the houseflies it contacted. Now, after years of use, there are types of houseflies that strongly resist DDT.

EVIDENCE Y

If one compares the front limbs of many animals, such as dogs, pigs, sheep, and horses, they are found to be very similar. The whole pattern or structure is the same. (see diagram next page)



Which of the following statements is true?
(a) Evidence X is an example of indirect evidence.
(b) Evidence Y is an example of direct evidence.
(c) Evidence X gives more support to the biologists' belief than Evidence Y.
(d) Both Evidence X and Evidence Y give about the same support to the biologists' belief.

13. The acceptance of an explanation in science is based upon
(a) many observations which provide conclusive evidence for the truth of the explanation
(b) successful predictions made using the explanation
(c) both a and b
(d) neither a nor b

14. Tycho Brahe, a Danish astronomer, improved the design of many of the instruments used to observe the sky. He used these new instruments to make the most accurate records of the heaven's motions up to his time. Which of the following is correct?
(a) Improvement of the instruments illustrates a main concern of science.
(b) Production of the records illustrates a main concern of science.
(c) Both are main concerns of science.
(d) Neither are main concerns of science.

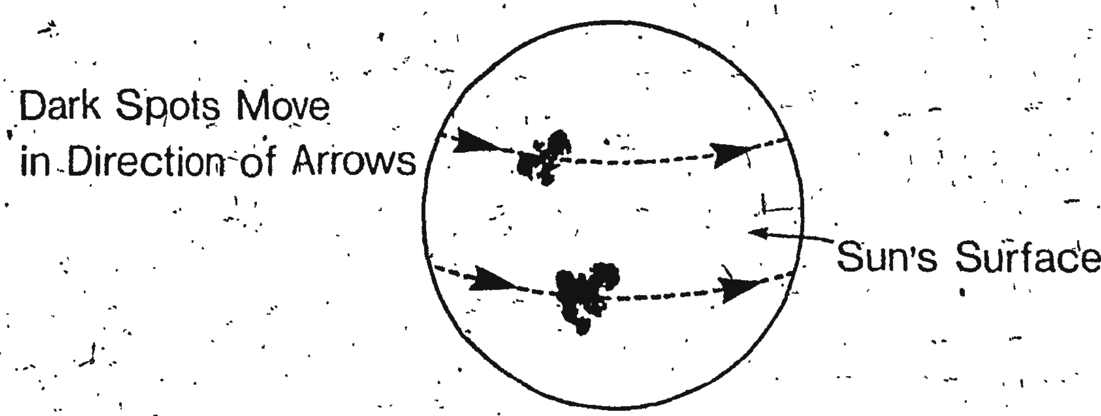
15. The scientific method can be most accurately described as including
(a) the five steps of stating the problem, collecting data, forming a hypothesis, testing the hypothesis, and making conclusions
(b) a demand for describing accurately, arguing logically, and explaining logically
(c) an assurance that the investigator will be successful, if he follows the steps outlined in the method
(d) both a and c

16. From the list of characteristics below, pick any that are characteristic of scientific theories.
    (a) They always contain an element of doubt concerning their truth.
    (b) They are not of use to scientists until they have been proven correct.
    (c) Before accepted, a theory explains all the facts important to the problem at hand.
    (d) none of the above

17. There are statements in science describing certain regularities that have been observed in nature. They say that when a certain event happens, a certain other event always happens afterwards. Such statements are called scientific
    (a) facts
    (b) theories
    (c) hypotheses
    (d) laws

18. In trying to find evidence for the evolution of life on earth, biologists should follow
    (a) any method whatsoever, whether logical or imaginative, which can produce testable hypotheses
    (b) the truly scientific method of observing all the facts first and then drawing conclusions
    (c) the scientific method of defining the problem, collecting data, forming hypotheses, testing the hypotheses, and drawing conclusions
    (d) both b and c

19. Galileo discovered that the sun had many dark spots on its surface. He also found that these dark spots moved across the face of the sun from one side to the other. (see diagram below) This observation could be explained by assuming that the sun
    (a) is spinning and carrying the dark spots around with it
    (b) is not moving and the dark spots are spinning around the sun's surface
    (c) neither a nor b
    (d) both a and b



20. Steno, a seventeenth century geologist, noticed that the rocks of the earth seemed to lie in orderly layers. He stated that if one layer was on top of another layer, then the bottom layer was older. In science, this statement would be considered a
    (a) law
    (b) hypothesis
    (c) theory
    (d) fact

21. Another geologist, William Smith, noticed that trapped in the different layers of rocks were the imprints of small sea animals and plants. He claimed that the animals and plants in the upper layers must be younger than those in the lower layers. This statement would be considered in science to be a
    (a) fact
    (b) theory
    (c) hypothesis
    (d) principle

22. Many of the ancient Greeks believed that the earth was unmoving and at the centre of the universe. This does not agree with what we believe today, because
    (a) the Greeks were not aware of the scientific method and, therefore, made mistakes in their explanations
    (b) since we have become aware of the scientific method, our explanations have come closer to the truth
    (c) since we have learned to use the scientific method, we have been able to show the Greeks' ideas to be inaccurate
    (d) none of the above

23. The Greek astronomer, Ptolemy, gave a very complicated explanation of how the universe worked. The mathematics needed to make predictions using this explanation was very involved. However, his explanation made very accurate predictions. Theories like Ptolemy's are likely to be
    (a) pleasing to scientists, because they make good predictions.
    (b) pleasing to scientists, because they are so complex
    (c) annoying to scientists, because they are so complex
    (d) both a and c

24. The Polish astronomer, Copernicus, also gave an explanation of how the universe worked. His explanation could make as many and as accurate predictions as Ptolemy's. However, predictions were easier to make using Copernicus' ideas because his explanation was simpler. Most scientists would
    (a) accept the explanation that is held by most people to be true
    (b) accept Ptolemy's explanation because it is more complex, and therefore more pleasing to scientists
    (c) accept Copernicus' explanation because it is simpler, and therefore more pleasing to scientists
    (d) accept either explanation, because both make equally good predictions

25. The astronomer, Copernicus, tried to explain the motions of the planets. To do this, he assumed that the sun was at the centre of the universe. Using his explanation, he could predict where the planets would be in say a year's time. Observations would show that the planets were where he said they would be. These observations show that
    (a) his assumption about the sun's position was true
    (b) his assumption about the sun's position was probably true
    (c) the ancient Greeks, who said the earth was at the centre of the universe, were wrong
    (d) both a and c

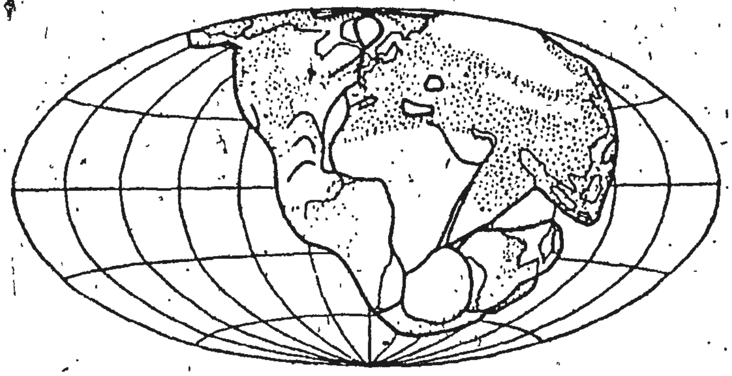
26. Starting with any two living things, some biologists have postulated that the population of these living things would increase in the following manner: 2, 4, 8, 16, 32, 64, 128, etc.; that is, doubling each generation. If this statement was true, it would have the form of a scientific
    (a) law
    (b) theory
    (c) fact
    (d) hypothesis

For many years, geologists have tried to explain why the same types of plant and animal fossils, minerals, mountains, etc., occur in different parts of the world. In many cases, these different parts of the world are thousands of miles apart.

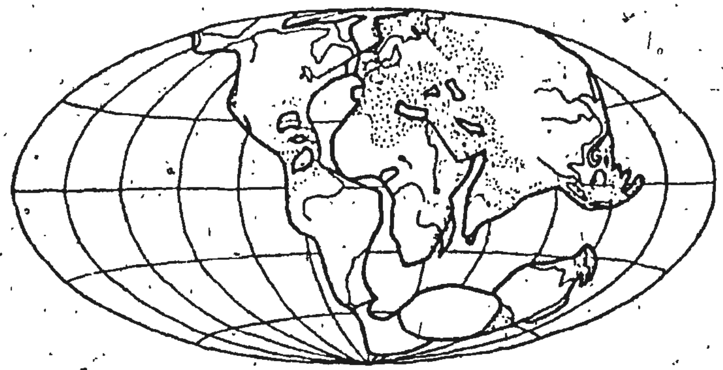
TWO EXPLANATIONS WHICH HAVE BEEN GIVEN FOR THESE OBSERVATIONS ARE DESCRIBED BELOW. READ THESE EXPLANATIONS AND USE THEM TO ANSWER THE NEXT SIX QUESTIONS.

A "DRIFTING" EXPLANATION

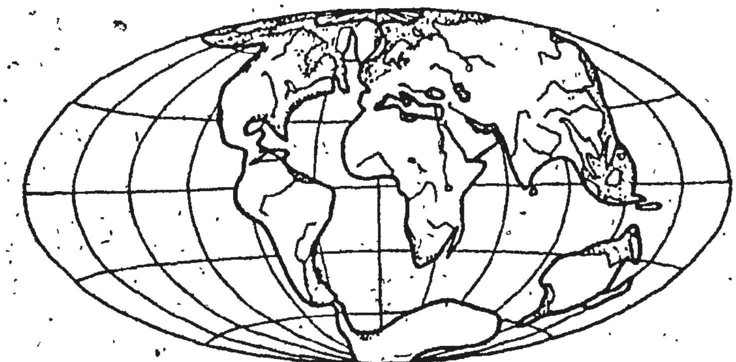
This explanation, given by the German, Alfred Wegener, claims that about 300 million years ago all the continents on earth were joined together to form one land area. Between then and about 50 million years ago, this land area split in many places. The pieces that were formed then drifted apart to make continents as we know them today. (see diagram next page)



300 MILLION YEARS AGO



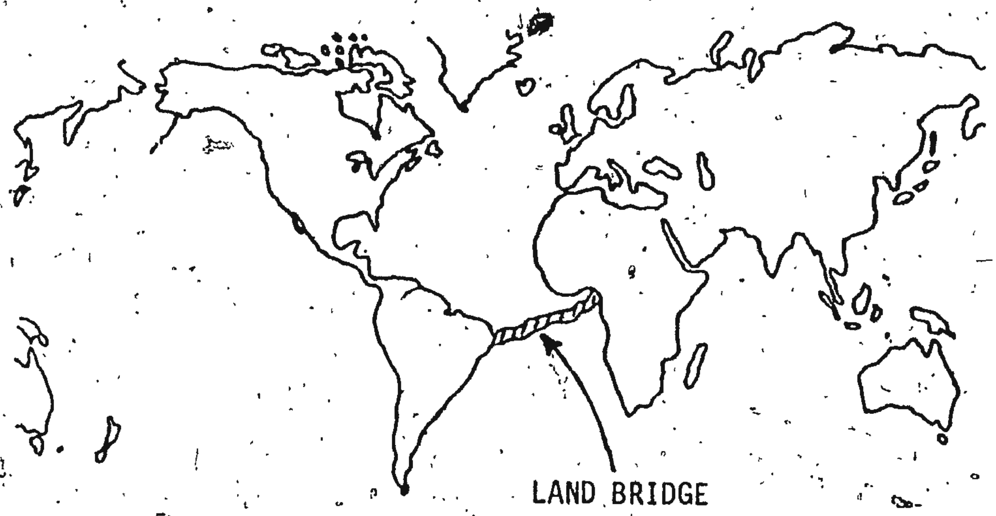
50 MILLION YEARS AGO



1 MILLION YEARS AGO

A "LAND BRIDGE" EXPLANATION

This explanation, supported by many geologists, claims that the continents were always in the same places that they are today. From time to time "land bridges", or narrow strips of land, rose out of the ocean. These strips of land sometimes joined two continents together like a bridge. They would then sink again beneath the ocean. (see diagram next page)



27. Many geologists would not believe the Drifting Explanation, because they thought the continents were too big and heavy to move. However, it has been observed that the whole northern part of North America is slowly rising out of the sea. Let's call this observation Evidence A.

Other geologists would not believe the Land Bridge Explanation, because they could not believe that land rose from and sank into the sea. However, there is evidence to say that this does happen. The island of Iceland is believed to have been formed by rising out of the ocean. Let's call this Evidence B.

Science would say that:
- (a) Evidence A supports the Drifting Explanation
- (b) Evidence B supports the Land Bridge Explanation
- (c) both a and b
- (d) neither a nor b

28. Scientists have discovered that many of the same types of worms and snails live on the east coast of South America as on the west coast of Africa. This observation can be explained using
- (a) the Drifting Explanation
- (b) the Land Bridge Explanation
- (c) either the Drifting Explanation or the Land Bridge Explanation
- (d) neither the Drifting Explanation nor the Land Bridge Explanation

29. Scientists have also discovered that certain mountain ranges and ore deposits found on the east coast of South America are continued on the west coast of Africa. This observation can be explained using
    (a) the Drifting Explanation
    (b) the Land Bridge Explanation
    (c) either the Drifting Explanation or the Land Bridge Explanation
    (d) neither the Drifting Explanation nor the Land Bridge Explanation

30. The observation that the South American east coast and the African west coast match almost exactly tends to
    (a) prove the Drifting Explanation
    (b) give a little support to the Drifting Explanation
    (c) give very strong support to the Drifting Explanation
    (d) none of the above

31. Eventually, evidence for the Drifting Explanation increased and evidence against the Land Bridge Explanation increased. However, many good scientists continued to believe the Land Bridge Explanation. This type of disagreement
    (a) happens often in science, because no explanation can be proven true
    (b) happens often in science, because no explanation can be proven false
    (c) does not usually happen in science, because explanations are ordinarily obviously true or false
    (d) both a and b

32. After much evidence for the Drifting Explanation had been gathered, many people still opposed it. They did not believe the explanation because it could not explain what made the continents move. In science, it is better to
    (a) reject an explanation which cannot explain some important details, although no better explanation exists
    (b) accept an explanation which cannot explain some important details, if no better explanation exists
    (c) accept only explanations which can explain all the important details
    (d) none of the above

33. Very often, biologists will guess at the size and general appearance of a type of animal no longer alive. They often do this using small amounts of information, such as a single leg bone. This type of work in science can be described as forming a
    (a) hypothesis
    (b) theory
    (c) neither a nor b
    (d) both a and b

34. The biologist, Charles Darwin, visited a series of islands, called the Galápagos Islands. He observed that there were different types of tortoises on each island. This information puzzled him so much he sought an explanation for it.

Many sailors had also visited these islands and had observed the same things as Darwin. However, they sought no explanation of the phenomena. It is true that advances in science will only occur when there are people like Darwin
   (a) who make the same observations as other people, but who see them as having a deeper meaning behind their otherwise ordinary appearances
   (b) who seek explanations for observations in nature
   (c) both a and b
   (d) neither a nor b

35. Scientists try to explain the behavior of nature principally to satisfy their
   (a) curiosity, or desire to know how nature functions
   (b) desire to solve the practical problems of the world
   (c) both a and b
   (d) neither a nor b

36. Geologists believe that the occurrence of earthquakes and volcanoes can be explained if they assume the continents are moving. They also believe that more knowledge about the movement of the continents will help man predict when earthquakes and volcanoes will occur. This information would save many lives.

The study of the movement of the continents is a task of science mainly because
   (a) the outcome may be of practical benefit to man
   (b) a basic objective of science is to explain how nature works
   (c) it satisfies man's curiosity about why things behave as they do
   (d) both b and c

37. A main goal of science is to
   (a) make discoveries that have practical uses
   (b) show the use of discoveries made about nature
   (c) provide explanations for events in nature
   (d) improve human welfare as much as possible

38. In science, there are many statements that explain, or try to explain, why nature behaves as it does. These statements are known as scientific
   (a) theories
   (b) facts
   (c) principles
   (d) formulas

39. When designing explanations of the universe to describe the positions of the planets, the ancient Greek astronomers used data that had been recorded by other people. This data gave the positions of the planets as they were years beforehand. When basing their explanations on such old, but accurate, data, the astronomers
   (a) had to take into account the possibility that nature might have behaved differently in the past
   (b) had to assume that nature behaved the same in the past as it does in the present
   (c) were making an explanation which could only describe how the planets behaved in the past
   (d) both a and c

40. Many biologists claim that the complex forms of life found on earth today are descended from simpler forms of life. These simpler forms have gradually changed into more and more complicated forms. Many other people very strongly object to this point of view. They claim, on religious grounds, that all living things were created as they are today, and have not changed into other types. Which of the following is true?
   (a) Controversies such as this have often occured between science and religion.
   (b) Such controversies have often arisen because people did not see that their beliefs could exist side by side with the beliefs of science.
   (c) both a and b
   (d) neither a nor b

41. Throughout the centuries, man has given many explanations of how the universe works. These have given better and better predictions of what will happen in the heavens. For example, predictions of the place a planet will be at a certain time in the future have become more and more accurate. Science expects that someday
   (a) there will be even better explanations, giving even better predictions of the planets' positions
   (b) someone will eventually find the true explanation giving the best possible predictions of the planets' positions
   (c) the point will be reached when perfect predictions of planetary positions will be made
   (d) all of the above

For hundreds of years, the ancient Greeks tried to explain the observations they made of the heavens. For example, they wondered what made the sun, moon, and stars behave as they did.

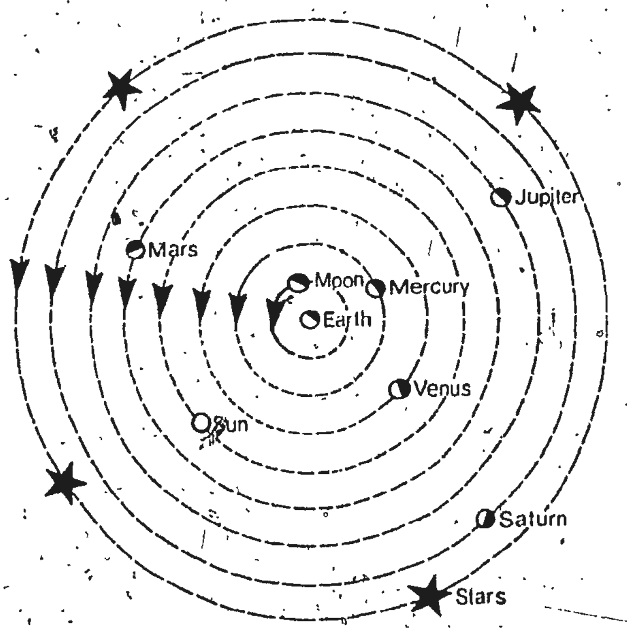
ON THE NEXT PAGE, YOU WILL FIND DESCRIPTIONS OF TWO DIFFERENT IDEAS PUT FORWARD BY THE EARLY GREEKS TO EXPLAIN THEIR OBSERVATIONS OF THE HEAVENS. YOU WILL NEED TO USE THESE DESCRIPTIONS TO ANSWER THE FOLLOWING SEVEN QUESTIONS.

AN "EARTH-CENTRED" EXPLANATION

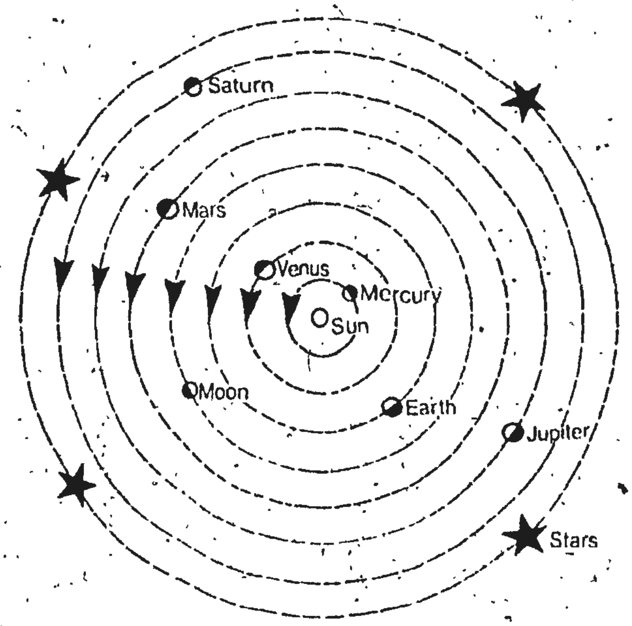
This idea, presented by Anaximenes and others, said that the earth was motionless at the centre of the universe. The stars, moon, sun, and planets travelled around the earth in circular orbits, each at its own speed. (see diagram below) The stars made their orbit once every 24 hours.

A "SUN-CENTRED" EXPLANATION

This second idea, described first by Aristarchus, said that the sun was motionless at the centre of the universe. The earth revolved around the sun with the moon and planets. They all moved in circular orbits, each at a different speed. The earth travelled completely around the sun in one year. The stars were joined to a huge, clear ball, which was also motionless. (see diagram below)



AN "EARTH-CENTRED" EXPLANATION



A "SUN-CENTRED" EXPLANATION

42. The supporters of the Earth-centred Explanation observed that the earth was obviously large, solid, and unchanging. On the other hand, the heavens were filled with small, far away objects, in constant motion. They concluded that, naturally, our big, heavy earth was at the centre of the universe.

The supporters of the Sun-centred Explanation had concluded from some calculations using geometry that the sun was really much larger than the earth. They believed that the sun should, naturally, be at the centre of the universe, because it was larger.

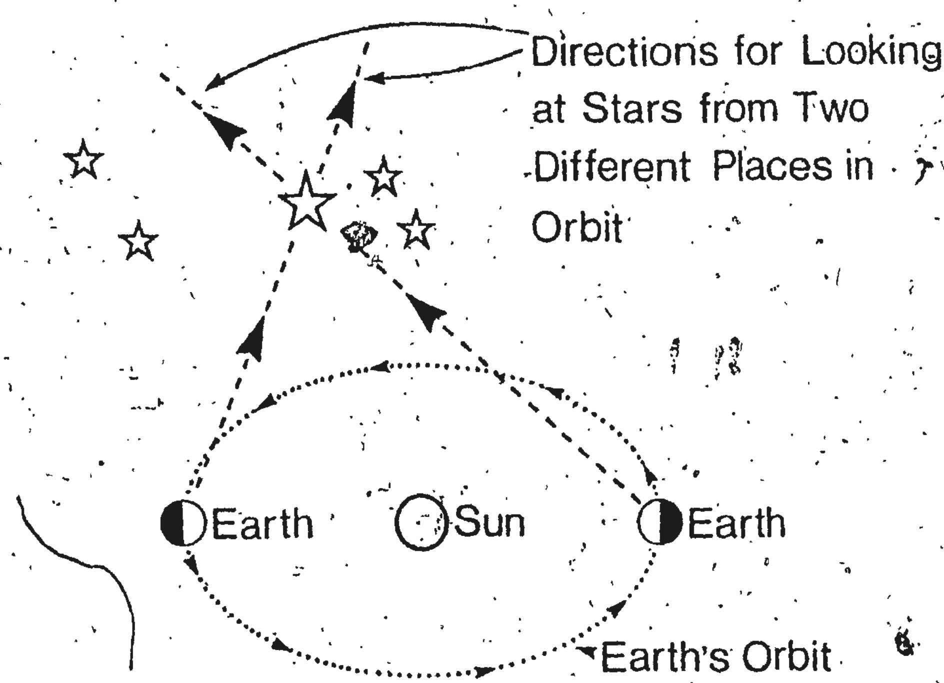
Which of the above arguments makes a defensible conclusion about whether the earth or the sun is at the centre of the universe?
- (a) both the Earth-centred Explanation and the Sun-centred Explanation
- (b) neither the Earth-centred Explanation nor the Sun-centred Explanation
- (c) the Earth-centred Explanation
- (d) the Sun-centred Explanation

43. One observation that ancient Greek astronomers tried to explain was that the stars travelled in circular paths, in a counter-clockwise direction, across the night sky. To account for this observation, the Earth-centred Explanation claims that the earth is stationary, while the sphere of the stars turns. The Sun-centred Explanation claims that the stars are stationary, while the earth spins as it travels around the sun. Which idea can explain why the stars travel in circular paths?
    - (a) the Earth-centred Explanation
    - (b) the Sun-centred Explanation
    - (c) neither the Earth-centred Explanation nor the Sun-centred Explanation
    - (d) both the Earth-centred Explanation and the Sun-centred Explanation

44. The Greeks knew that the planets changed in brightness and size from one time of year to another. This observation is explained by
    - (a) both the Earth-centred Explanation and the Sun-centred Explanation
    - (b) neither the Earth-centred Explanation nor the Sun-centred Explanation
    - (c) the Earth-centred Explanation
    - (d) the Sun-centred Explanation

45. The Earth-centred Explanation made predictions, which were only approximately accurate, about the future positions of the planets. The Sun-centred Explanation was not used to make any predictions about the planets' future positions. This information tends to support
    - (a) the Earth-centred Explanation more than the Sun-centred Explanation
    - (b) the Sun-centred Explanation more than the Earth-centred Explanation
    - (c) neither the Earth-centred Explanation nor the Sun-centred Explanation
    - (d) the Earth-centred Explanation and the Sun-centred Explanation about the same

46. Many Greeks argued that if the earth travelled around the sun below motionless stars, then to see a particular star at the same time every night, one would have to look in different directions as the earth moved in its orbit. (see diagram below) That this changing direction was not observed is evidence
    (a) for the Earth-centred Explanation and against the Sun-centred Explanation
    (b) against both the Earth-centred Explanation and the Sun-centred Explanation
    (c) for the Sun-centred Explanation and against the Earth-centred Explanation
    (d) against the Sun-centred Explanation but not for the Earth-centred Explanation



47. It took until 1838 for astronomers to find out that one did have to look in different directions to observe the same star at the same time each night. This difference was so small that the Greeks had not noticed it. Which of the following is correct?
    (a) The Greeks did not see the difference in direction because they did not know the rules for making accurate measurements.
    (b) We can see the difference in direction today, because we can make exact measures with our more accurate instruments.
    (c) This is an example of an important discovery in science, which had to wait to be uncovered until we could make measurements free of error.
    (d) none of the above

48. Using information from the preceding six questions, it would seem reasonable to
    (a) accept the Sun-centred Explanation, because we know that the planets revolve around the sun
    (b) reject the Sun-centred Explanation, because there is more evidence against it than for it, and because it makes no predictions
    (c) reject the Earth-centred Explanation, because we know that the planets do not revolve around the earth
    (d) both a and c

49. In developing an explanation for some observations, a scientist might
    (a) use his imagination to make up an explanation that agrees with the observations
    (b) change someone else's explanation so that it better explains the observations
    (c) derive an explanation, based on what he already knows, using some mathematical ideas
    (d) all of the above

50. When stars are observed through even the most powerful telescope, they still look like points of light. But, when a planet or a comet is observed, it looks larger, like a disc. On March 13, 1781, an amateur astronomer, Herschel, reported seeing a new object in the night sky. It appeared as a disc, and Herschel claimed it was a comet. Later, this object was found to be another planet, Uranus. Which of the following is true?
    (a) Surprising observations, such as Herschel's, have played an important role in the advance of science.
    (b) Some scientific discoveries are the result of a "lucky break" or chance discovery.
    (c) An unexpected observation will have no impact on the advance of science, unless someone is able to recognize the importance of the observation.
    (d) all of the above

51. Biologists have classified or grouped living things since the time of Aristotle. Which of the following statements is true?
    (a) The ways in which certain living things appear to be similar to or different from one another helps biologists to group them.
    (b) Systems of classification can be based entirely upon observation.
    (c) Systems of classification can be based on certain assumptions, for example, assumptions of how living things are thought to have evolved or developed.
    (d) all of the above

52. Charles Darwin, an English biologist, looked for an explanation of how all the different types of living things came to be. If Darwin was like most scientists, he probably looked for this explanation mainly because
    (a) he thought that this information might be of use to man, in the field of medicine, say
    (b) he was curious and wanted to satisfy his desire to know how nature behaved
    (c) he wanted to verify what scientists already knew about living things
    (d) he wanted to demonstrate the wonder and orderliness that exists in the universe

53. Biologists have found the fossils or remains of once living things all over the earth. Many of these remains are of types of plants and animals that do not live on the earth now. That there were types of living things on earth in the past that are not on earth now can be considered a scientific
    (a) theory
    (b) fact
    (c) hypothesis
    (d) principle

54. Which of the following statements must science assume, if it is to try to explain anything in nature?
    (a) The laws of nature remain the same in the past, present, and in the future.
    (b) The laws of nature may change from time to time depending upon the age of the universe.
    (c) Some of the laws of nature may be too difficult for us to understand.
    (d) none of the above

55. Many geologists believe that the continents were once joined together and have slowly drifted apart. Using this idea, they have correctly predicted where to find minerals on one continent by examining where these minerals are on another continent. This prediction
    (a) proves that their idea about drifting continents is correct
    (b) proves that other ideas about how the earth has formed are wrong
    (c) shows that their ideas are probably correct
    (d) none of the above

READ THE NEXT TWO PARAGRAPHS AND ANSWER THE FOUR QUESTIONS WHICH FOLLOW. EACH PARAGRAPH GIVES A DIFFERENT EXPLANATION OF HOW MOUNTAINS WERE FORMED.

THE "FOLDED EDGE" EXPLANATION

Alfred Wegener, a German geologist, claimed that the continents have moved or "drifted" about the earth. The front edge of a moving continent pushes against the ocean bottom. While doing this, it folds so as to form mountain ranges.

THE "DRIED APPLE" EXPLANATION

Other geologists claimed that the earth was at its beginning very hot and was slowly cooling down. Most objects in nature contract, or get smaller, when they cool. These geologists said that the surface of the earth had wrinkled to form mountains as it cooled and contracted. They compared the earth to an old apple whose skin had wrinkled after it had dried and contracted.

56. Comparing the earth to an apple is an example of a scientific
    (a) model
    (b) hypothesis
    (c) theory
    (d) principle

57. As evidence in support of the Dried Apple Explanation, geologists discovered that the earth was always giving off amounts of heat. This made them believe that the earth was cooling off.

    However, as evidence against the Dried Apple Explanation, other geologists discovered that the earth was much older than predicted by the explanation.

    On the basis of the above statements, science would
    (a) know the Dried Apple Explanation is not completely right because there is at least one piece of evidence against it
    (b) not know whether the Dried Apple Explanation is right or wrong because there is one piece of evidence for it and one piece against it
    (c) know the Dried Apple Explanation may be right because there is evidence in favor of it
    (d) none of the above

58. Geologists have found out that the continents are indeed moving at the present time. North America is moving farther away from Europe year by year. This evidence proves that
    (a) the Folded Edge Explanation is correct
    (b) the Dried Apple Explanation is not correct
    (c) both a and b
    (d) neither a nor b

59. While first coming up with the Folded Edge Explanation, Wegener most likely
    (a) had done much study in areas related to his explanation
    (b) used his imagination to come up with the explanation
    (c) both a and b
    (d) neither a nor b

60. Scientists very often find it useful to group objects in nature into different classes. For example, in astronomy the following groups of objects are found: stars, planets, moons, comets, and galaxies. In biology, one finds mammals, birds, reptiles, and fish. Which of the following is true?
   (a) The same materials can be logically grouped into classes in only one way.
   (b) Once shown to be true, classification schemes are not modified further.
   (c) Such classification schemes are based on observed differences and similarities among the materials.
   (d) all of the above

61. Today, scientists do not believe a great deal of what the ancient Greeks said about the structure of the universe. Many of the Greeks' explanations have been replaced with new ideas. In science,
   (a) sometimes explanations are used for the time being, even though their truth is seriously doubted
   (b) explanations are replaced by new ones, as soon as they are proven wrong
   (c) it is believed that once an explanation has been proven, it is no longer subject to change
   (d) both b and c

62. In order for a scientific hypothesis to be good, it should
   (a) be able to predict exactly how certain events will happen
   (b) be a clearly thought out statement rather than a guess or a hunch
   (c) in principle, be able to be proven wrong
   (d) all of the above

63. Which of the following is true scientific work?
   (a) Galileo designing and building a telescope so that far away objects could be seen easier.
   (b) Jean Leverrier using a scientific explanation to predict the position of a then undiscovered planet, Neptune.
   (c) Eudoxus trying to logically explain the motion of the planets.
   (d) both b and c

64. Look at the map of the world on the next page. Look especially closely at the east coast of South America and the west coast of Africa. The observation that the shape of these two shorelines match each other very well is considered a scientific
   (a) hypothesis
   (b) fact
   (c) theory
   (d) principle

65. Alfred Wegener, a German geologist, tried to explain this likeness between the African and the South American shores. His explanation said that Africa and South America were once joined together to form a single continent. At some time in the past, this single continent split in two and the pieces drifted apart. Wegener's explanation would be considered a scientific
    - (a) principle
    - (b) phenomenon
    - (c) fact
    - (d) hypothesis

66. When trying to explain why the continents fit together so exactly, geologists believe that
    - (a) they will eventually find the correct explanation of why this is so
    - (b) they will never find an explanation which does not need to be changed in some way
    - (c) it is not likely that many of the explanations they believe today will be changed in the future
    - (d) none of the above

67. Imagine that all astronomers, Aristotle, Copernicus, Galileo, etc., up to our time, had never lived. Also, imagine that other astronomers had taken their place. If this was true,
    - (a) many explanations of the behavior of the universe would probably be different than the ones we now know
    - (b) the explanations given for the universe's behavior would be the same as those we now know
    - (c) we would have reached the same level of knowledge as we have today
    - (d) both b and c

68. In trying to explain what the universe was like the Greek, Ptolemy, assumed certain things. One thing he assumed was that the earth does not move. Call this assumption "Statement X".

Ptolemy also used observations of the heavens when explaining what the universe was like. One such observation was that the planet Venus is brighter sometimes than at others. Call this observation "Statement Y".

In science,
   (a) both Statement X and Statement Y are hypotheses
   (b) both Statement X and Statement Y are facts
   (c) Statement X is a hypothesis and Statement Y is a fact
   (d) neither Statement X nor Statement Y is a fact

69. The Greek astronomer, Ptolemy, gave an explanation for the workings of the universe in the year 150 AD. His ideas included the assumption that the earth was at the centre of the universe and did not move. His ideas were believed until around 1650 AD. One of the reasons why his assumption about the earth's position was believed so long was that the Christian Church also said it was true. Anyone who disagreed with it, by saying the sun was at the centre of the universe, was going against the Church. In this quarrel, science would be more likely to support
   (a) Ptolemy, because there are enough places in the Bible (for example, Joshua 10:12-14) saying that the sun moved around the earth
   (b) the explanation that best explains the known observations of the universe
   (c) the Church, because many of the most educated people in the world are the Church's leaders
   (d) the people who say the sun is at the centre of the universe, because they know this to be true

70. It is possible for a piece of scientific evidence to support
   (a) only one explanation
   (b) more than one explanation
   (c) none of the available explanations
   (d) all of the above

71. Biologists have made records of the types of plants and animals that are living or have lived in different parts of the world. Geologists have made widespread use of these records when testing their explanations about how the continents have formed. Which of the following is correct?
   (a) The different branches of science are separate from each other. Information from one area is very rarely of use in another area.
   (b) Geologists would have more support for their explanations using evidence from their own field, rather than from the field of biology.
   (c) neither a nor b
   (d) both a and b

72. Astronomers use models when describing the universe mainly because
   (a) they are convenient ways of describing the universe understandably
   (b) they are pictures of the universe as they know it to actually be
   (c) they represent what they see when looking through powerful telescopes
   (d) the universe is so large it needs to be scaled down for them to understand

73. Scientists carry our experiments to
   (a) test whether hypotheses they have made are false
   (b) prove that the laws of nature are true
   (c) have a situation where measurements can be made free of error
   (d) all of the above

74. In science,
   (a) scientists do very careful experiments so that other scientists will not have to repeat them
   (b) scientists are convinced that the results of a scientist who does his work extremely carefully are true
   (c) results are not believed unless they can be repeated time and time again
   (d) both a and b

75. There are events in nature that many people have seen happening a large number of times. There are other events that people can make happen if they have the wish and the material to do so. In science, these events are known as
   (a) theories
   (b) laws
   (c) facts
   (d) hypotheses

Biologists have found evidence suggesting that the complex forms of life found on earth today have developed from simpler forms of life which lived in the past.

ON THE NEXT PAGE, YOU WILL FIND DESCRIPTIONS OF TWO EXPLANATIONS OF HOW THE FORMS OF LIFE ON EARTH HAVE DEVELOPED OR EVOLVED. YOU WILL NEED TO USE THESE DESCRIPTIONS TO ANSWER THE FOLLOWING <u>SEVEN</u> QUESTIONS.

THE "USE/DISUSE" EXPLANATION

This first explanation was put forward in 1809 by the French biologist, Jean Baptiste Lamarck. As a basis for his explanation, he reasoned that any great change in an environment can produce a need for change in the plants and animals living there.

This idea led him to make two major assumptions. He called his first assumption the "law of use or disuse". He assumed that as any particular part of the body is used more and more, it develops and grows larger. Those parts not being used grow smaller and can even disappear.

His second assumption said that any living thing could pass on to its offspring those characteristics which had grown larger through much use or grown smaller through much disuse. He claimed that new types of living things develop after many generations, because of new characteristics gained or old characteristics lost.

THE "VARIATION/NATURAL SELECTION" EXPLANATION

This second explanation was proposed in 1859 by the English biologist, Charles Darwin. He made several assumptions. His first assumption was that living things tend to multiply so fast that, if they were not destroyed, the whole earth would soon be covered by the offspring of a single pair. He then assumed that although living things tend to increase in numbers, the number of individuals of any particular type stays about the same.

To explain this, he used the observation that there is variation in every type of living thing. This means that individuals differ slightly from one another within the same basic type. For example, no two German Shephards are exactly alike. He assumed that some variations would help individuals survive, but other variations would not be helpful. Those members with helpful variations would survive, or be selected by nature, and have offspring. The helpful variations would be passed on to their offspring. After many generations, so many small variations would be passed on that a new type of living thing would be formed, completely different from the old type.

Those members with variations that are not helpful would die without having offspring. Thus, these types of variations do not get passed on.

76. Imagine a place whose yearly rainfall began to grow less and less. As the area became more like a desert, the plants, which normally needed large amounts of water, began to develop water-saving characteristics. This observation can be explained
   (a) by neither the Use/Disuse Explanation nor by the Variation/ Natural Selection Explanation
   (b) by either the Use/Disuse Explanation or by the Variation/ Natural Selection Explanation
   (c) by only the Use/Disuse Explanation
   (d) by only the Variation/Natural Selection Explanation

77. An experimenter cut off the tails of two white mice, one male and one female, and then mated them. All the offspring were born with tails. The tails were then removed from the mice of this second generation, and they were then mated. This procedure was continued for twenty generations. However, the mice of the twenty-first generation had tails just as long as those of the original two mice. The result of this experiment is evidence
    (a) against the Use/Disuse Explanation
    (b) against the Variation/Natural Selection Explanation
    (c) in favor of the Use/Disuse Explanation
    (d) in favor of the Variation/Natural Selection Explanation

78. Athletes develop strong muscles and greater staying power by practising long hours. This is evidence
    (a) against the Use/Disuse Explanation
    (b) against the Variation/Natural Selection Explanation
    (c) in favor of the Use/Disuse Explanation
    (d) in favor of the Variation/Natural Selection Explanation

79. While exploring the Galapagos Islands near South America, Darwin dug up fossil remains of large animals. These remains were of much different animals than those living when the remains were found. This is evidence in favor of
    (a) the Use/Disuse Explanation
    (b) the Variation/Natural Selection Explanation
    (c) both explanations
    (d) neither explanation

80. The Variation/Natural Selection Explanation has some major faults. For example, it cannot explain how a small variation could be of so much use that it could enable its possessor to live, while those without it die.

    The Use/Disuse Explanation also has major flaws. For example, no one has ever observed a characteristic gained through use or disuse that has been passed on to offspring.

    When all the available explanations contain faults like those above, scientists usually
    (a) reject all explanations and begin to look for another one
    (b) reject the explanation which they believe is wrong and try to correct the other one
    (c) keep both explanations for the time being and try to gain more evidence on both
    (d) none of the above

81. Through the years, animal breeders have noticed that some horses are born with special characteristics. For example, some horses are better climbers than others, while some are faster runners. By selecting those animals with special characteristics and by breeding them with each other, herds of horses with excellent climbing ability or running ability have been developed. This information supplies evidence in favour of
    (a) the Use/Disuse Explanation
    (b) the Variation/Natural Selection Explanation
    (c) both explanations
    (d) neither explanation

82. Today, biologists believe that the Use/Disuse Explanation is wrong. They believe that, with some changes, the Variation/Natural Selection Explanation is closer to the truth. In science, after an explanation has been rejected,
    (a) it may still be brought back and used again some time in the future
    (b) it is known to be false and will not be brought back into use
    (c) it is not known for sure to be false but will not be used again in the future
    (d) none of the above

83. Around 400 BC, the Greek philosopher, Plato, was trying to think of an explanation to account for what was seen in the heavens. This type of work is often done by
    (a) philosophers
    (b) applied scientists
    (c) pure scientists
    (d) any of the above groups

84. Very often scientists will continue to support an explanation which they doubt, if
    (a) no other acceptable explanation is proposed to take its place
    (b) it is the simplest of a number of equally doubtful explanations, which all have equal power of explanation
    (c) both a and b
    (d) neither a nor b

85. Many people claim that science is going beyond its limits when it tries to give natural explanations of how living things have come about. They say that the Bible gives a perfectly good supernatural account of how this happened. Concerning science and the supernatural, it may be said that science
    (a) denies the existence of supernatural phenomena and claims that science can eventually explain everything using the knowledge of matter and energy
    (b) can neither affirm nor deny supernatural phenomena but only say that it has not observed them
    (c) affirms the existence of supernatural phenomena, because it is a way of explaining things it, itself, cannot
    (d) none of the above

86. Science can best be described as

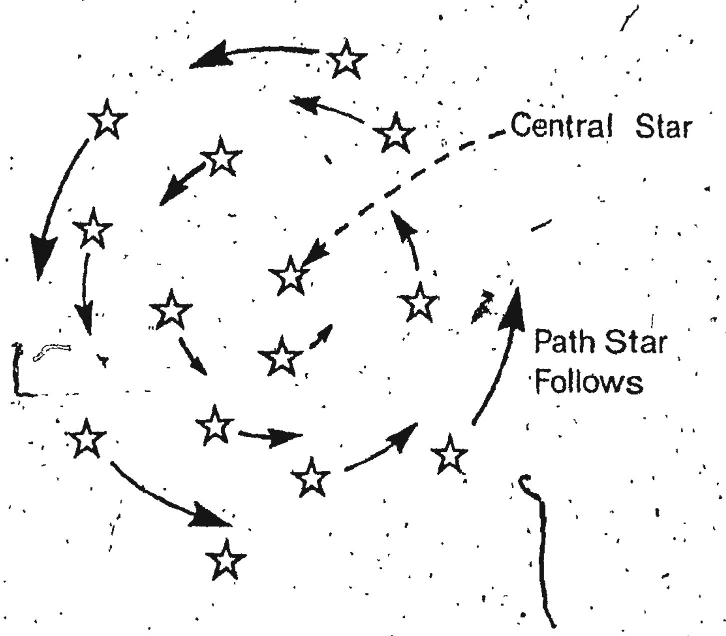
(a) a body of knowledge to which scientists add more and more accurate facts
(b) an activity of people whose discoveries make possible modern medicine and technology
(c) a series of uncertain explanations of nature which are good only if they lead to further thought and experiments
(d) a way of describing the universe in which we are living

87. According to many of the ancient Greeks, for example Plato, scientific explanations could be based on "self-evident" ideas. By this, they meant ideas which were obviously true. One such idea was that all heavenly bodies must have perfectly circular motions. In science, these ideas would be called

(a) facts
(b) theories
(c) hypotheses
(d) none of the above

88. Around 1000 BC, some Greek astronomers described the motion of the stars in the night sky. They observed that during each night all the stars travelled in counter-clockwise circles around one star at the centre. (see diagram below) In science, this observation would be considered a

(a) fact
(b) hypothesis
(c) principle
(d) theory



89. The Greek, Anaximenes, tried to explain this movement of the stars described above. His explanation said that all the stars were stuck on to a clear ball. The earth was at the centre of this ball. The ball turned around the earth carrying the stars with it. This description would be considered a scientific
   (a) principle
   (b) theory
   (c) postulate
   (d) theorem

90. Another Greek, Heraclides, claimed that while the stars were stuck on to a clear ball with the earth at the centre, the ball did not turn. He said that the earth did the turning, which made the stars appear to move around the sky. This description would be considered a scientific
   (a) model
   (b) principle
   (c) theory
   (d) both a and c

91. When trying to devise explanations of how the earth has taken shape, geologists are studying things that happened many millions of years ago. Because of the type of work they are doing, geologists believe that
   (a) they will never be sure that their explanations are correct
   (b) explanations can be useful even though their truth is uncertain
   (c) both a and b
   (d) neither a nor b

APPENDIX B

Table Giving Validators' Responses to the Ninety-one Item Form for Scientists

APPENDIX B

Item Number	Validators									
	A	B	C	D	E	F	G	H	I	J
1	d	d	d	d	d	d	d	d	d	d
2	c	b	c	b	b	c	c	c	c	b
3	b	b	b	b	c	b	b	b	b	b
4	b	d	b	b	d	b	b	b	b	d
5	b	-	-	-	b	b	b	b	b	b
6	d	d	d	d	d	d	d	-	d	d
7	c	c	c	-	c	c	b	c	c	-
8	d	b	a	-	b	d	a	d	d	-
9	d	d	a	d	b	d	d	c	d	d
10	a	a	a	a	c	c	a	a	a	b
11	c	c	b	c	c	c	c	c	c	c
12	c	c	c	d	c	c	d	c	d	d
13	b	c	b	c	c	b	c	b	b	c
14	c	d	d	d	c	c	c	c	c	d
15	b	a	a	a	a	a	a	b	b	a
16	a	d	a	a	c	d	d	a	a	d
17	d	d	d	d	d	d	d	d	d	a
18	a	c	a	c	a	d	c	a	a	d
19	d	a	d	d	d	d	d	d	d	d
20	b	b	b	d	b	c	b	c	b	b
21	c	c	c	c	c	b	c	b	c	c
22	d	d	c	c	d	d	d	d	d	d
23	d	d	d	d	d	d	d	d	c	-
24	c	c	c	c	c	c	c	c	c	-
25	b	c	b	d	b	b	b	b	b	d
26	a	a	a	a	a	a	a	a	a	d
27	c	d	b	b	c	b	c	d	c	-
28	c	c	c	c	c	c	c	c	c	a
29	a	a	a	a	a	c	c	a	a	a
30	b	b	c	c	c	b	b	c	b	c
31	a	d	a	d	d	a	a	d	a	-
32	b	c	b	b	b	b	b	b	d	d
33	a	c	a	a	a	b	a	b	c	c
34	c	c	c	c	c	c	c	c	b	c
35	a	a	a	c	a	a	c	a	a	c
36	d	d	d	-	d	d	d	d	d	a
37	c	c	c	c	c	c	c	c	c	a
38	a	a	a	a	a	a	a	a	a	d
39	b	b	b	b	c	d	-	b	b	b
40	c	a	-	c	a	c	c	c	a	c
41	a	a	a	a	a	a	d	a	a	d
42	b	b	b	a	a	b	a	-	a	d

APPENDIX B (continued)

Item Number	Validators									
	A	B	C	D	E	F	G	H	I	J
43	d	d	d	d	d	d	d	d	d	d
44	d	d	a	a	d	d	d	d	d	-
45	a	c	a	a	-	a	c	a	a	-
46	a	-	d	d	d	b	d	a	a	-
47	d	b	b	b	b	d	b	-	d	a
48	b	d	-	b	b	d	d	-	-	d
49	d	d	d	d	d	d	d	d	d	d
50	d	d	d	d	d	d	d	d	d	d
51	d	d	d	-	-	d	d	d	d	d
52	b	b	b	b	b	b	b	b	b	b
53	b	b	b	b	b	b	b	b	d	b
54	a	a	a	d	-	a	a	a	a	a
55	c	c	c	c	c	c	c	c	c	-
56	a	a	a	a	a	a	-	a	a	a
57	a	a	a	a	a	c	b	c	d	d
58	d	d	d	d	d	d	d	b	d	d
59	c	c	c	a	c	c	c	c	a	c
60	c	c	c	c	c	c	c	c	c	c
61	a	a	a	a	a	a	a	a	a	b
62	d	b	c	d	d	c	d	c	c	d
63	d	b	d	d	-	d	-	d	d	d
64	b	b	b	b	-	b	b	b	b	a
65	d	d	d	d	d	d	d	d	d	d
66	b	b	c	a	-	b	a	b	a	a
67	d	a	-	b	a	a	d	d	a	d
68	c	c	c	c	d	c	c	c	c	c
69	b	b	b	b	b	b	c	b	b	b
70	d	d	-	b	-	d	d	d	d	b
71	c	b	-	c	c	c	c	c	c	c
72	a	a	a	a	a	d	a	a	a	d
73	a	a	a	a	a	d	d	a	a	d
74	c	d	c	-	c	c	-	-	c	c
75	c	c	c	-	b	c	c	c	c	b
76	b	b	b	b	b	b	b	b	b	b
77	a	a	a	-	a	a	a	a	a	d
78	c	c	c	c	c	c	c	c	-	d
79	c	c	c	c	c	b	b	d	c	b
80	c	c	c	c	b	c	c	c	c	c
81	b	b	-	c	b	c	c	a	b	b
82	a	d	a	d	a	a	a	a	a	a
83	d	c	c	d	d	c	d	c	d	d
84	c	c	c	b	c	c	c	c	c	c

APPENDIX B (continued)

Item Number	Validators									
	A	B	C	D	E	F	G	H	I	J
85	b	b	b	d	b	b	b	b	a	b
86	c	-	c	c	a	d	c	d	d	a
87	d	d	c	d	d	a	c	b	d	c
88	a	a	a	c	a	a	a	a	a	-
89	b	b	b	b	b	b	b	c	b	c
90	c	d	d	c	c	c	d	d	d	-
91	c	d	b	b	c	c	c	c	c	d

APPENDIX C

A Description of the Taxonomy of Educational Objectives

Cognitive Domain

APPENDIX C

Cognitive Level	General Description
KNOWLEDGE	
1.00 KNOWLEDGE	This level involves little more than the recall of previously stored information. Although some alterations may have to be made in the material, the processing of information forms a relatively minor part of this task.
1.10 Knowledge of Specifics	The recall of specific and isolated bits of information of a concrete nature.
1.11 Knowledge of Terminology	Knowledge of the different referents used for specific symbols, such as being able to define terms by giving their attributes.
1.12 Knowledge of Specific Facts	Knowledge of very specific information such as dates, events, names of people and places, etc.
1.20 Knowledge of the Ways and Means of Dealing with Specifics	Knowledge of the methods of inquiry and standards of judgment within a field. No demand is made to be able to use these inquiry methods at this level.
1.21 Knowledge of Conventions	Knowledge of the usual preferred ways of presenting or defending ideas.
1.22 Knowledge of Trends and Sequences	Knowledge of the ways phenomena evolve with respect to time.

APPENDIX C (continued)

Cognitive Level	General Description
1.23 Knowledge of Classifications and Categories	Knowledge of the classes, sets, divisions and arrangements which are regarded as fundamental for a given subject field.
1.24 Knowledge of Criteria	Knowledge of the standards by which facts, principles, opinions, and conduct are judged.
1.25 Knowledge of Methodology	Knowledge of the methods of inquiry, techniques, and procedures employed in a particular subject field.
1.30 <u>Knowledge of the Universals and Abstractions in a Field</u>	Knowledge of the large structures, theories, and generalizations in a subject area.
1.31 Knowledge of Principles and Generalizations	Knowledge of the abstractions which are of value in explaining, describing, predicting, or determining the most appropriate direction to be taken.
1.32 Knowledge of Theories and Structures	Knowledge of a body of principles and generalizations which can be used to organize and interrelate a large number of specifics.

APPENDIX C (continued)

Cognitive Level	General Description
INTELLECTUAL ABILITIES AND SKILLS	
2.00 COMPREHENSION	This represents the lowest level of understanding, where what is being communicated can be apprehended without necessarily being able to relate it to other material.
2.10 Translation	The ability to faithfully and accurately render one form of language or communication into another.
2.20 Interpretation	The ability to summarize or explain material by reordering or giving a new view of the original material.
2.30 Extrapolation	The ability to extend the trends or sequences of the given data to make predictions of what will happen if these trends continue.
3.00 APPLICATION	The ability to use abstractions in particular cases which may differ situationally from those in which the abstractions were first learned.
4.00 ANALYSIS	The ability to clarify a communication by pointing out how its ideas are related, how it is organized, or how it conveys its effects.
4.10 Analysis of Elements	The ability to recognize unstated assumptions or the capability of distinguishing fact from hypothesis.

APPENDIX C (continued)

Cognitive Level	General Description
4.20 Analysis of Relationships	The ability to recognize the connections between elements and parts of a communication.
4.30 Analysis of Organizational Principles	The ability to point out the organization, systematic arrangement, and structure which hold the communication together.
5.00 SYNTHESIS	The arranging and combining of elements and parts to form a structure that clearly was not present before.
5.10 Production of a Unique Communication	The development of a communication in which the writer expresses his own feelings and thoughts.
5.20 Production of a Plan, or Proposed Set of Operations	The development of a plan or proposal to solve a problem or to get a specific task completed.
5.30 Derivation of a Set of Abstract Relations	The development of a set of relationships or generalizations from a previously unstructured set of data.
6.00 EVALUATION	The ability to qualitatively and quantitatively judge material and methods with respect to certain criteria.
6.10 Judgments in Terms of	The judgment of the accuracy of a communication with respect to its logical accuracy and internal consistency.

APPENDIX C (continued)

Cognitive Level	General Description
6.20 Judgments in Terms of External Evidence	The ability to judge a communication with respect to the highest known standard in its field.

APPENDIX D

Classification of Form B Test Items into Bloom and Nature of Science Categories

APPENDIX D

Item Number	Bloom Category	Nature of Science Category
1	1.31	I-3
2	4.10	II-1
3	4.10	II-2
4	4.10	II-2
5	4.10	II-1
6	4.20	I-3
7	4.30	III-1
8	1.23	I-1
9	4.20	II-2
10	4.20	II-8
11	4.20	II-8
12	4.20	II-8
13	5.20	III-5
14	1.24	III-5
15	1.11	II-1
16	1.24	IV-2
17	4.10	III-3
18	1.10	II-7
19	4.20	IV-3
20	1.24	II-8
21	4.10	II-3
22	4.10	II-1
23	4.10	II-3
24	6.20	I-1
25	1.24	II-2
26	4.30	II-6
27	4.10	II-7
28	6.10	III-5
29	1.24	III-7
30	4.20	II-8
31	1.24	IV-2
32	4.10	II-1
33	1.24	I-2
34	1.24	II-6
35	3.00	III-6
36	1.25	III-2
37	4.20	II-2
38	4.20	II-2
39	4.20	II-8
40	6.20	II-2
41	3.00	II-2
42	3.00	IV-2
43	1.11	II-2
44	4.30	I-1

APPENDIX D (continued)

Item Number	Bloom Category	Nature of Science Category
45	1.24	I-1
46	1.24	I-2
47	2.30	I-2
48	4.10	II-3
49	4.20	II-2
50	4.20	II-2
51	6.10	II-8
52	6.10	III-5
53	4.10	II-4
54	6.20	III-5
55	6.20	IV-1
56	6.20	III-5
57	4.10	III-2
58	4.10	II-3
59	4.10	II-3
60	5.30	II-3
61	5.20	III-2
62	1.11	II-4
63	1.25	III-2
64	3.00	I-1
65	6.20	II-8
66	4.10	II-4
67	3.00	II-4
68	1.24	III-5
69	4.30	I-4
70	1.24	II-2
71	4.10	II-2
72	6.10	III-5
73	6.10	III-7
74	2.20	II-2
75	1.31	II-1

APPENDIX E

A Test on the Nature of Science
and Scientific Thinking

Form B

A TEST ON THE NATURE OF SCIENCE AND SCIENTIFIC THINKING

FORM B

DIRECTIONS

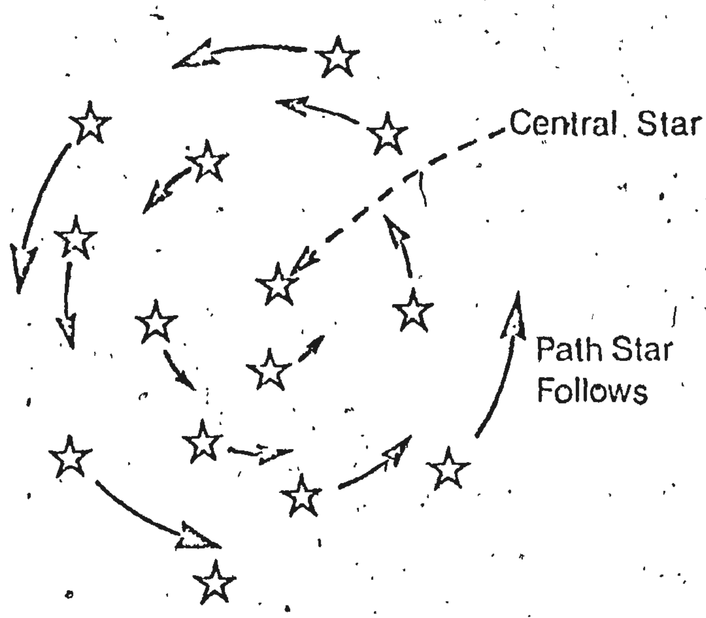
DO NOT PUT ANY MARKS ON THIS TEST. OTHER PEOPLE WILL USE IT AFTER YOU.

This is a multiple choice test. You are given four answers from which to choose in each question. Put the letter of the answer you think is correct in the appropriate place on the answer sheet.

Some questions on the test are very short. For other questions you will have to do some reading and careful thinking. Do not rush. You may have as much time as you need to finish.

1. When trying to devise explanations of how the earth has taken shape, geologists are studying things that happened many millions of years ago. Because of the type of work they are doing, geologists believe that
    (a) they will never be sure that their explanations are correct
    (b) explanations can be useful even though their truth is uncertain
    (c) both a and b
    (d) neither a nor b

2. Around 1000 BC, some Greek astronomers described the motion of the stars in the night sky. They observed that during each night all the stars travelled in counter-clockwise circles around one star at the centre. (see diagram below) In science, this observation would be considered a
    (a) fact
    (b) hypothesis
    (c) principle
    (d) theory



3. The Greek, Anaximenes, tried to explain this movement of the stars described above. His explanation said that all the stars were stuck on to a ~~clear ball~~. The earth was at the centre of this ball. The ball turned around the earth carrying the stars with it. This description would be considered a scientific
    (a) principle
    (b) theory
    (c) postulate
    (d) theorem

4. Another Greek, Heraclides, claimed that while the stars were stuck on to a clear ball with the earth at the centre, the ball did not turn. He said that the earth did the turning, which made the stars appear to move around the sky. This description would be considered a scientific
    (a) model
    (b) principle
    (c) theory
    (d) either a or c

5. According to many of the ancient Greeks, for example Plato, scientific explanations could be based on "self-evident" ideas. By this, they meant ideas which were obviously true. One such idea was that all heavenly bodies must have perfectly circular motions. In science, these ideas would be called
   (a) facts
   (b) principles
   (c) laws
   (d) none of the above

6. Many people claim that science is going beyond its limits when it tries to give natural explanations of how living things have come about. They say that the Bible gives a perfectly good supernatural account of how this happened. Concerning science and the supernatural, it may be said that science
   (a) denies the existence of supernatural phenomena and claims that science can eventually explain everything using the knowledge of matter and energy
   (b) can neither affirm nor deny supernatural phenomena but only say that it has not observed them
   (c) affirms the existence of supernatural phenomena, because it is a way of explaining things it, itself, cannot
   (d) none of the above

7. Very often scientists will continue to support an explanation which they doubt, if
   (a) no other acceptable explanation is proposed to take its place
   (b) it is the simplest of a number of equally doubtful explanations, which all have equal power of explanation
   (c) both a and b
   (d) neither a nor b

8. Around 400 BC, the Greek, Plato, was trying to think of an explanation to account for what was seen in the heavens. This type of work is characteristic of
   (a) philosophy
   (b) science
   (c) technology
   (d) any of the above

Biologists have found evidence suggesting that the complex forms of life found on earth today have developed from simpler forms of life which lived in the past.

BELOW, YOU WILL FIND DESCRIPTIONS OF TWO EXPLANATIONS OF HOW THE FORMS OF LIFE ON EARTH HAVE DEVELOPED OR EVOLVED. YOU WILL NEED TO USE THESE DESCRIPTIONS TO ANSWER THE FOLLOWING SIX QUESTIONS.

THE "USE/DISUSE" EXPLANATION

This first explanation was put forward in 1809 by the French biologist, Jean Baptiste Lamarck. As a basis for his explanation, he reasoned that any great change in an enviornment can produce a need for change in the plants and animals living there.

This idea led him to make two major assumptions. He called his first assumption the "law of use or disuse". He assumed that as any particular part of the body is used more and more, it develops and grows larger. Those parts not being used grow smaller and can even disappear.

His second assumption said that any living thing could pass on to its offspring those characteristics which had grown larger through much use or grown smaller through much disuse. He claimed that new types of living things develop after many generations, because of new characteristics gained or old characteristics lost.

THE "VARIATION/NATURAL SELECTION" EXPLANATION

This second explanation was proposed in 1859 by the English biologist, Charles Darwin. He made several assumptions. His first assumption was that living things tend to multiply so fast that, if they were not destroyed, the whole earth would soon be covered by the offspring of a single pair. He then assumed that although living things tend to increase in numbers, the number of individuals of any particular type stays about the same.

To explain this, he used the observation that there is variation in every type of living thing. This means that individuals differ slightly from one another within the same basic type. For example, no two German Shephards are exactly alike. He assumed that some variations would help individuals survive, but other variations would not be helpful. Those members with helpful variations would survive, or be selected by nature, and have offspring. The helpful variations would be passed on to their offspring. After many generations, so many small variations would be passed on that a new type of living thing would be formed completely different from the old type.

Those members with variations that are not helpful would die without having offspring. Thus, these types of variations do not get passed on.

9. Imagine a place whose yearly rainfall began to grow less and less. As the area became more like a desert, the plants, which normally needed large amounts of water, began to develop water-saving characteristics. This observation can be explained
   (a) by neither the "use/disuse" explanation nor by the "variation/natural selection" explanation
   (b) by either the "use/disuse" explanation or by the "variation/natural selection" explanation
   (c) by only the "use/disuse" explanation
   (d) by only the "variation/natural selection" explanation

10. An experimenter cut off the tails of two white mice, one male and one female, and then mated them. All the offspring were born with tails. The tails were then removed from the mice of this second generation, and they were then mated. This procedure was continued for twenty generations. However, the mice of the twenty-first generation had tails just as long as those of the original two mice. The result of this experiment is evidence
(a) against the "use/disuse" explanation
(b) against the "variation/natural selection" explanation
(c) in favor of the "use/disuse" explanation
(d) in favor of the "variation/natural selection" explanation

11. Athletes develop stronger muscles and greater staying power by practising long hours. This is evidence
(a) against the "use/disuse" explanation
(b) against the "variation/natural selection" explanation
(c) in favor of the "use/disuse" explanation
(d) in favor of the "variation/natural selection" explanation

12. While exploring the Galápagos Islands near South America, Darwin dug up fossil remains of large animals. These remains were of much different animals than those living when the remains were found. This is evidence in favour of
(a) the "Use/Disuse" explanation
(b) the "Variation/Natural Selection" explanation
(c) both explanations
(d) neither explanation

13. The "Variation/Natural Selection" explanation has some major faults. For example, it cannot explain how a small variation could be of so much use that it could enable its possessor to live, while those without it die.

The "Use/Disuse" explanation also has major flaws. For example, no one has ever observed a characteristic, gained through use or disuse, that has been passed on to offspring.

When all the available explanations contain faults like those above, scientists usually
(a) reject all the explanations and begin to look for another one
(b) reject the explanation which they believe is wrong and try to correct the other one
(c) keep both explanations for the time being and try to gain more evidence on both
(d) none of the above

14. Today, biologists believe that the "use/disuse" explanation is wrong. They believe that, with some changes, the "variation/natural selection" explanation is closer to the truth. In science, after an explanation has been rejected,
(a) it may still be brought back in modified form and used again some time in the future
(b) it is known to be false and will not be brought back into use
(c) it is not known for sure to be false but will not be used again in the future
(d) none of the above

15. There are events in nature that many people have seen happening a large number of times. There are other events that people can make happen if they have the wish and the material to do so. In science, these events are known as
(a) theories
(b) laws
(c) facts
(d) hypotheses

16. In science,
(a) scientists do very careful experiments so that other scientists will not have to repeat them
(b) scientists are convinced that the results of a scientist who does his work extremely carefully are true
(c) results are not believed unless they can be repeated time and time again

17. Scientists carry out experiments to
(a) test whether hypotheses they have made are false
(b) prove that the laws of nature are true
(c) have a situation where measurements can be made free of error
(d) all of the above

18. Astronomers use models when describing the universe mainly because
(a) they are convenient ways of describing the universe understandably
(b) they are pictures of the universe as they know it to actually be
(c) they represent what they see when looking through powerful telescopes
(d) the universe is so large it needs to be scaled down for them to understand

19. Biologists have made records of the types of plants and animals that are living or have lived in different parts of the world. Geologists have made widespread use of these records when testing their explanations about how the continents have formed. Which of the following is correct?
(a) The different branches of science are separate from each other. Information from one area is very rarely of use in another area.
(b) Geologists would have more support for their explanations using evidence from their own field, rather than from the field of biology.
(c) neither a nor b
(d) both a and b

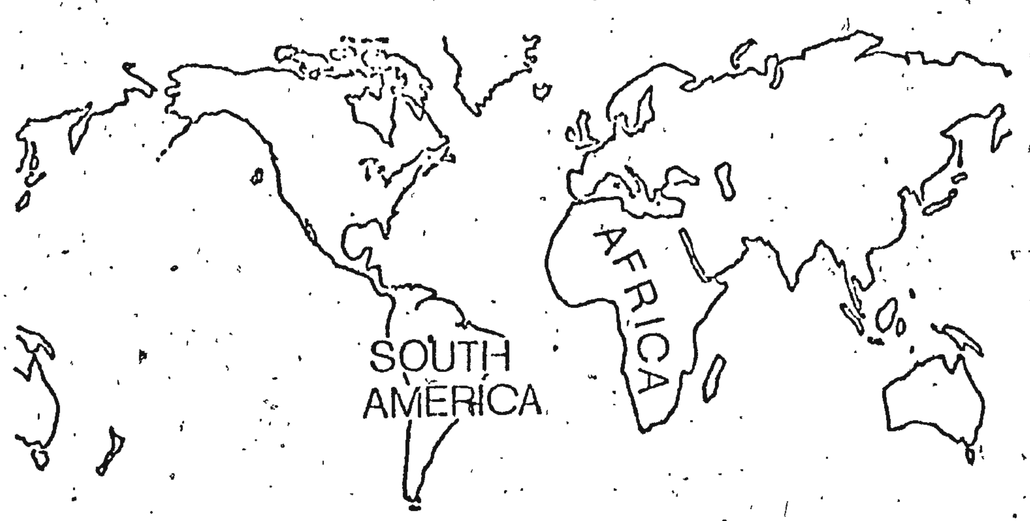
20. It is possible for a piece of scientific evidence to support
(a) only one explanation
(b) more than one explanation
(c) none of the available explanations
(d) any of the above

21. In trying to explain what the universe was like the Greek, Ptolemy, assumed certain things. One thing he assumed was that the earth does not move. Call this assumption "Statement X".

Ptolemy also used observations of the heavens when explaining what the universe is like. One such observation is that the planet Venus is brighter sometimes than at others. Call this observation "Statement Y".

In science,
(a) both Statement X and Statement Y are hypotheses
(b) both Statement X and Statement Y are facts
(c) Statement X is a hypothesis and Statement Y is a fact
(d) neither Statement X nor Statement Y is a fact

22. Look at the map of the world below. Look especially closely at the east coast of South America and the west coast of Africa. The observation that the shape of these two shorelines match each other very well is considered a scientific
(a) hypothesis
(b) fact
(c) theory
(d) principle



23. Alfred Wegener, a German geologist, tried to explain this likeness between the African and the South American shores. His explanation said that Africa and South America were once joined together to form a single continent. At some time in the past this single continent split in two and the pieces drifted apart. Wegener's explanation would be considered a scientific
(a) principle
(b) phenomenon
(c) fact
(d) hypothesis

24. Which of the following is true scientific work?
    (a) Galileo designing and building a telescope so that far away objects could be seen easier.
    (b) Jean Leverrier using a scientific explanation to predict the position of a then undiscovered planet, Neptune.
    (c) Eudoxus trying to logically explain the motion of the planets.
    (d) all of the above

25. Today, scientists do not believe a great deal of what the ancient Greeks said about the structure of the universe. Many of the Greek's explanations have been replaced with new ideas. In science,
    (a) sometimes explanations are used for the time being, even though their truth is seriously doubted
    (b) explanations are replaced by new ones, as soon as they are proven wrong
    (c) it is believed that once an explanation has been proven, it is no longer subject to change
    (d) both b and c

26. Scientists very often find it useful to group objects in nature into different classes. For example, in astronomy the following groups of objects are found: stars, planets, moons, comets, and galaxies. In biology, one finds mammals, birds, reptiles, and fish. Which of the following is true?
    (a) The same materials can be logically grouped into classes in only one way.
    (b) Once shown to be true, classification schemes are not modified further.
    (c) Such classification schemes are based on observed differences and similarities among the materials.
    (d) all of the above

READ THE TWO PARAGRAPHS BELOW AND ANSWER THE THREE QUESTIONS WHICH FOLLOW. EACH PARAGRAPH GIVES A DIFFERENT EXPLANATION OF HOW MOUNTAINS WERE FORMED.

THE "FOLDED EDGE" EXPLANATION

Alfred Wegener, a German geologist, claimed that the continents have moved or "drifted" about the earth. The front edge of a moving continent pushes against the ocean bottom. While doing this, it folds so as to form mountain ranges.

THE "DRIED APPLE" EXPLANATION

Other geologists claimed that the earth was at its beginning very hot and was slowly cooling down. Most objects in nature contract, or get smaller, when they cool. These geologists said that the surface of the earth had wrinkled to form mountains as it cooled and contracted. They compared the earth to an old apple whose skin had wrinkled after it had dried and contracted.

27. Comparing the earth to an apple is an example of a scientific
    (a) model
    (b) hypothesis
    (c) theory
    (d) principle

28. Geologists have found out that the continents are indeed moving at the present time. North America is moving farther away from Europe year by year. This evidence proves that
(a) the "folded edge" explanation is correct
(b) the "dried apple" explanation is not correct
(c) both a and b
(d) neither a nor b

29. While first coming up with the "folded edge" explanation, Wegener most likely
(a) had done some study in areas related to his explanation
(b) used his imagination to come up with the explanation
(c) both a and b
(d) neither a nor b

30. Many geologists believe that the continents were once joined together and have slowly drifted apart. Using this idea, they have correctly predicted where to find minerals on one continent by examining where these minerals are on another continent. This prediction
(a) proves that their idea about drifting continents is correct
(b) proves that other ideas about how the earth has formed are wrong
(c) shows that their ideas are probably correct
(d) none of the above

31. Which of the following statements must science assume, if it is to try to explain anything in nature?
(a) The laws of nature remain the same in the past, present, and in the future.
(b) The laws of nature may change from time to time depending upon the age of the universe.
(c) Some of the laws of nature may be too difficult for us to understand.
(d) none of the above

32. Biologists have found the fossils or remains of once living things all over the earth. Many of these remains are of types of plants and animals that do not live on earth now. That there were types of living things on earth in the past that are not on earth now can be considered a scientific
(a) theory
(b) fact
(c) hypothesis
(d) principle

33. Charles Darwin, an English biologist, looked for an explanation of how all the different types of living things came to be. If Darwin was like most scientists, he probably looked for this explanation mainly because
(a) he thought that this information might be of use to man, in the field of medicine, say
(b) he was curious and wanted to satisfy his desire to know how nature behaved
(c) he wanted to verify what scientists already knew about living things
(d) he wanted to demonstrate the wonder and orderliness that exists in the universe

34. Biologists have classified or grouped living things since the time of Aristotle. Which of the following statements is true?

(a) The ways in which certain living things appear to be similar to or different from one another helps biologists to group them.

(b) Systems of classification can be based entirely upon observation.

(c) Systems of classification can be based on certain assumptions, for example, assumptions of how living things are thought to have evolved or developed.

(d) all of the above

35. When stars are observed through even the most powerful telescope, they still look like points of light. But, when a planet or a comet is observed, it looks larger, like a disc. On March 13, 1781, an amateur astronomer, Herschel, reported seeing a new object in the night sky. It appeared as a disc, and Herschel claimed it was a comet. Later, this object was found to be another planet, Uranus. Which of the following is true?

(a) Surprising observations, such as Herschel's, have played an important role in the advance of science.

(b) Some scientific discoveries are the result of a "lucky break" or chance discovery.

(c) An unexpected observation will have no impact on the advance of science, unless someone is able to recognize the importance of the observation.

(d) all of the above

36. In developing an explanation for some observations, a scientist might

(a) use his imagination to make up an explanation that agrees with the observations

(b) change someone else's explanation so that it better explains the observations

(c) derive an explanation, based on what he already knows, using some mathematical ideas

(d) all of the above

For hundreds of years, the ancient Greeks tried to explain the observations they made of the heavens. For example, they wondered what made the sun, moon, and stars behave as they did.

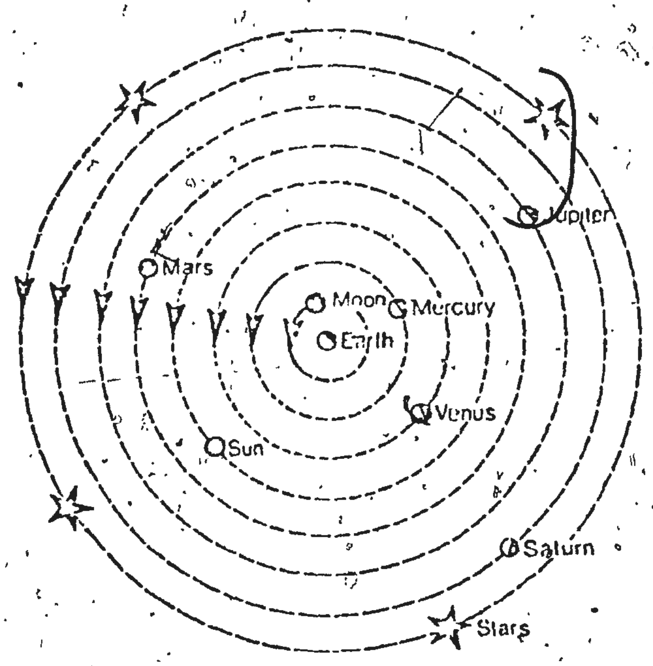
BELOW, YOU WILL FIND DESCRIPTIONS OF TWO DIFFERENT IDEAS PUT FORWARD BY THE EARLY GREEKS TO EXPLAIN THEIR OBSERVATIONS OF THE HEAVENS. YOU WILL NEED TO USE THESE DESCRIPTIONS TO ANSWER THE FOLLOWING FOUR QUESTIONS.

AN "EARTH-CENTRED" EXPLANATION

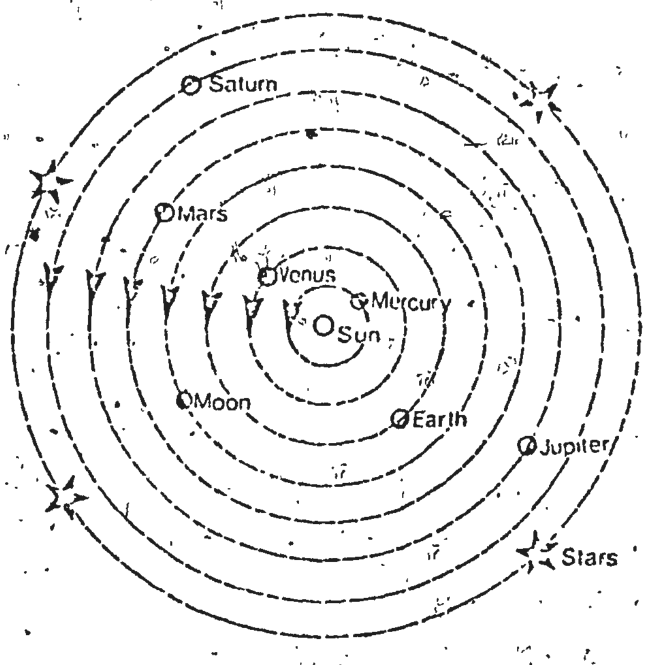
This idea, presented by Anaximenes and others, said that the earth was motionless at the centre of the universe. The stars, moon, sun, and planets travelled around the earth in circular orbits, each at its own speed. (see diagram below) The stars made their orbit once every 24 hours.

A "SUN-CENTRED" EXPLANATION

This second idea, described first by Aristarchus, said that the sun was motionless at the centre of the universe. The earth revolved around the sun with the moon and planets. They all moved in circular orbits, each at a different speed. The earth travelled completely around the sun in one year. The stars were joined to a huge, clear ball, which was also motionless. (see diagram below)



AN "EARTH-CENTRED" EXPLANATION

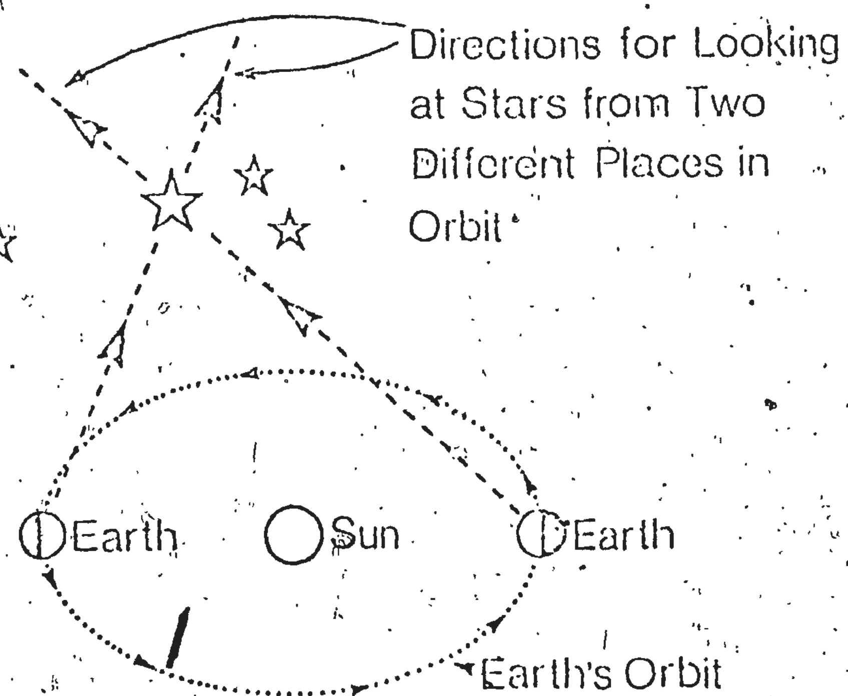


A "SUN-CENTRED" EXPLANATION

37. One observation that ancient Greek astronomers tried to explain was that the stars travelled in circular paths, in a counter-clockwise direction, across the night sky. To account for this observation, the "earth centred" explanation claims that the earth is stationary, while the sphere of the stars turns. The "sun-centred" explanation claims that the stars are stationary, while the earth spins as it travels around the sun. Which idea can explain why the stars travel in circular paths?
   (a) the "earth-centred" explanation
   (b) the "sun-centred" explanation
   (c) neither the "earth-centred" explanation nor the "sun-centred" explanation
   (d) both the "earth-centred" explanation and the "sun-centred" explanation

38. The Greeks knew that the planets changed in brightness and size from one time of year to another. This explanation is explained by
   (a) both the "earth-centred" explanation and the "sun-centred" explanation
   (b) neither the "earth-centred" explanation nor the "sun-centred" explanation
   (c) the "earth-centred" explanation
   (d) the "sun-centred" explanation

39. Many Greeks argued that if the earth travelled around the sun below motionless stars, then to see a particular star at the same time every night, one would have to look in different directions as the earth moved in its orbit. (see diagram below) That this changing direction was not observed is evidence
   (a) against the "sun-centred" explanation
   (b) for the "sun-centred" explanation
   (c) against the "earth-centred" explanation
   (d) none of the above



40. The "earth-centred" explanation made predictions which were only approximately accurate about the future positions of the planets. The "sun-centred" explanation was not used to make any predictions about the planets' future positions. This information tends to support
    (a) the "earth-centred" explanation more than the "sun-centred" explanation
    (b) the "sun-centred" explanation more than the "earth-centred" explanation
    (c) neither the "earth-centred" explanation nor the "sun-centred" explanation
    (d) the "earth-centred" explanation and the "sun-centred" explanation about the same

41. Throughout the centuries, man has given many explanations of how the universe works. These have given better and better predictions of what will happen in the heavens. For example, predictions of the place a planet will be at a certain time in the future have become more and more accurate. Science expects that someday
    (a) there will be even better explanations, giving even better predictions of the planets' positions
    (b) someone will eventually find the true explanation giving the best possible predictions of the planets' positions
    (c) the point will be reached when perfect predictions of planetary positions will be made
    (d) all of the above

42. When designing explanations of the universe to describe the positions of the planets, the ancient Greek astronomers used data that had been recorded by other people. These data gave the positions of the planets as they were years beforehand. When basing their explanations on such old, but accurate, data, the astronomers
    (a) had to take into account the possibility that nature might have behaved differently in the past
    (b) had to assume that nature behaved the same in the past as it does in the present
    (c) were making an explanation which could only describe how the planets behaved in the past
    (d) both a and c

43. In science, there are many statements that explain, or try to explain, why nature behaves as it does. These statements are known as scientific
    (a) theories
    (b) facts
    (c) principles
    (d) formulas

44. A main goal of science is to
    (a) make discoveries that have practical uses
    (b) show the use of discoveries about nature
    (c) provide explanations for events in nature
    (d) improve human welfare as much as possible

45. Geologists believe that the occurrence of earthquakes and volcanoes can be explained if they assume the continents are moving. They also believe that more knowledge about the movement of the continents will help man predict when earthquakes and volcanoes will occur. This information could save many lives.

The study of the movement of the continents is a task of science mainly because
(a) the outcome may be of practical benefit to man
(b) a basic objective of science is to explain how nature works
(c) it satisfies man's curiosity about why things behave as they do
(d) both b and c

46. Scientists try to explain the behavior of nature principally to satisfy their
(a) curiosity, or desire to know how nature functions
(b) desire to solve the practical problems of the world
(c) both a and b
(d) neither a nor b

47. The biologist, Charles Darwin, visited a series of islands, called the Galápagos Islands. He observed that there were different types of tortoises on each island. This information puzzled him so much he sought an explanation for it.

Many sailors had also visited these islands and had observed the same things as Darwin. However, they sought no explanation of what they observed. It is true that advances in science will only occur when there are people like Darwin
(a) who make the same observations as other people, but who see them as having a deeper meaning behind their otherwise ordinary appearances
(b) who seek explanations for observations in nature
(c) both a and b
(d) neither a nor b

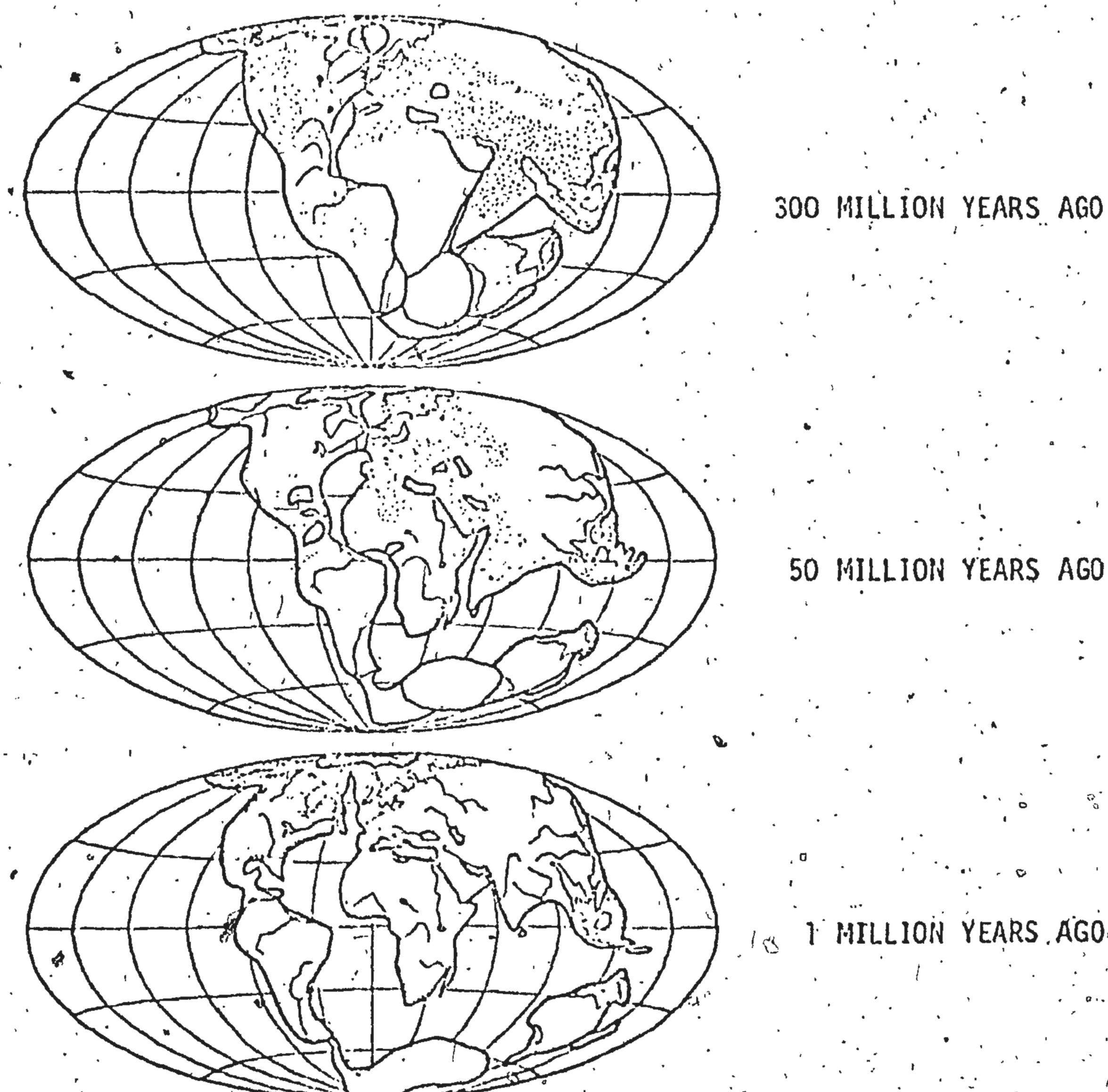
48. Very often, biologists will guess at the size and general appearance of a type of animal no longer alive. They often do this using small amounts of information, such as a single leg bone. This type of work in science can be described as
(a) hypothesizing
(b) experimenting
(c) neither a nor b
(d) both a and b

For many years, geologists have tried to explain why the same types of plant and animal fossils, minerals, mountains, etc., occur in different parts of the world. In many cases, these different parts of the world are thousands of miles apart.

TWO EXPLANATIONS WHICH HAVE BEEN GIVEN FOR THESE OBSERVATIONS ARE DESCRIBED BELOW. READ THESE EXPLANATIONS AND USE THEM TO ANSWER THE NEXT FOUR QUESTIONS.

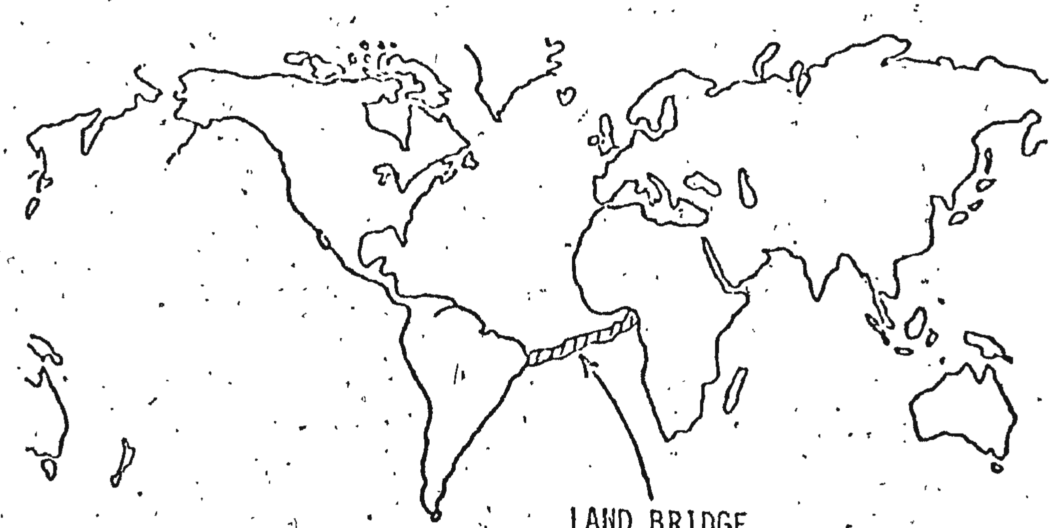
A "DRIFTING" EXPLANATION

This explanation, given by the German, Alfred Wegener, claims that about 300 million years ago all of the continents on earth were joined together to form one land area. Between then and about 50 million years ago, this land area split in many places. The pieces that were formed then drifted apart to make continents as we know them today. (see diagram below)



A "LAND BRIDGE" EXPLANATION

This explanation, supported by many geologists, claims that the continents were always in the same places that they are today. From time to time "land bridges", or narrow strips of land, rose out of the ocean. These strips of land sometimes joined two continents together like a bridge. They would then sink again beneath the ocean. (see diagram below)



49. Scientists have discovered that many of the same types of worms and snails live on the east coast of South America as on the west coast of Africa. This observation can be explained using
   (a) the "Drifting" explanation
   (b) the "Land Bridge" explanation
   (c) either the "Drifting" explanation or the "Land Bridge" explanation
   (d) neither the "Drifting" explanation nor the "Land Bridge" explanation

50. Scientists have also discovered that certain mountain ranges and ore deposits found on the east coast of South America are continued on the west coast of Africa. This observation can be explained using
   (a) the "Drifting" explanation
   (b) the "Land Bridge" explanation
   (c) either the "Drifting" explanation or the "Land Bridge" explanation
   (d) neither the "Drifting" explanation nor the "Land Bridge" explanation

51. The observation that the South American east coast and the African west coast match almost exactly tends to
   (a) prove the "Drifting" explanation
   (b) give a little support to the "Drifting" explanation
   (c) give very strong support to the "Drifting" explanation
   (d) none of the above

52. After much evidence for the "Drifting" explanation had been gathered, many people still opposed it, because it could not explain what made the continents move. In science, it is better to
   (a) reject an explanation which cannot explain some important details, although no better explanation exists
   (b) accept an explanation which cannot explain some important details, if no better explanation exists
   (c) accept only explanations which can explain all the important details
   (d) none of the above

53. Starting with any two living things, some biologists have postulated that the population of these living things would increase in the following manner: 2, 4, 8, 16, 32, 64, 128, etc.; that is, doubling each generation. If this statement was true, it would have the form of a scientific
   (a) law
   (b) theory
   (c) fact
   (d) hypothesis

54. The astronomer, Copernicus, tried to explain the motions of the planets. To do this, he assumed that the sun was at the centre of the universe. Using his explanation, he could predict where the planets would be in say a year's time. Observations would show that the planets were where he said they would be. These observations show that
   (a) his assumption about the sun's position was true
   (b) his assumption about the sun's position was probably ture
   (c) the ancient Greeks, who said the earth was at the centre of the universe, were wrong
   (d) both a and c

55. The Greek astronomer, Ptolemy, gave a very complicated explanation of how the universe worked. The mathematics needed to make predictions using this explanation was very involved. However, his explanation made very accurate predictions. A theory like Ptolemy's is likely to be
   (a) pleasing to scientists, because it makes good predictions
   (b) pleasing to scientists, because it is so complex
   (c) annoying to scientists, because it is so complex
   (d) both a and c

56. The Polish astronomer, Copernicus, also gave an explanation of how the universe worked. His explanation could make as many and as accurate predictions as Ptolemy's. However, predictions were easier to make using Copernicus' ideas because his explanation was simpler. Most scientists would accept
   (a) the explanation that is held by most people to be true
   (b) Ptolemy's explanation because it is more complex and, therefore, more pleasing to scientists
   (c) Copernicus' explanation because it is simpler and, therefore, more pleasing to scientists
   (d) either explanation, because both make equally good predictions

57. Many of the ancient Greeks believed that the earth was unmoving and at the centre of the universe. This does not agree with what we believe today, because

(a) the Greeks were not aware of the scientific method and, therefore, made mistakes in their explanations
(b) since we have become aware of the scientific method, our explanations have come closer to the truth
(c) since we have learned to use the scientific method, we have been able to show the Greeks' ideas to be inaccurate
(d) none of the above

58. Steno, a seventeenth century geologist, noticed that the rocks of the earth seemed to lie in orderly layers. He stated that if one layer was on top of another layer, then the bottom layer was older. In science, this statement would be considered a

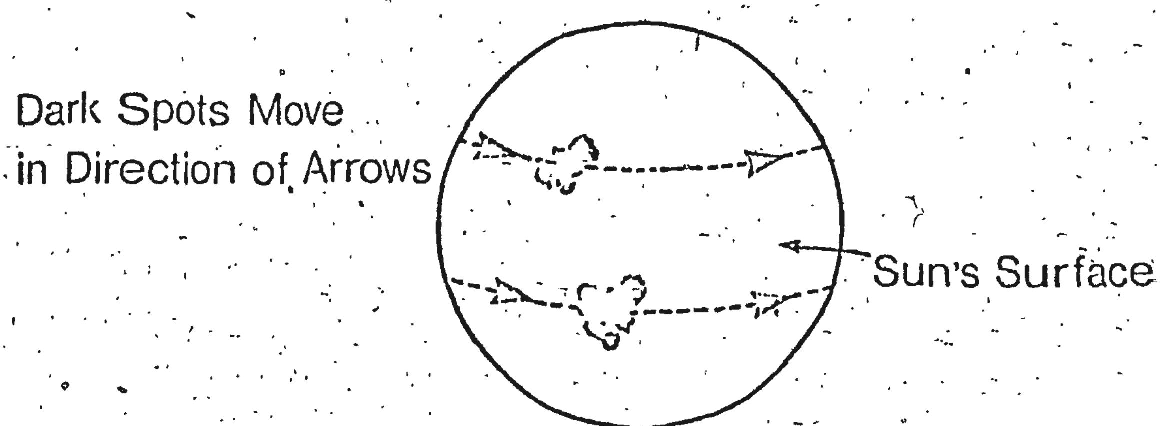
(a) law
(b) hypothesis
(c) theory
(d) fact

59. Another geologist, William Smith, noticed that trapped in the different layers of rocks were the imprints of small sea animals and plants. He claimed that the animals and plants in the upper layers must be younger than those in the lower layers. This statement would be considered in science to be a

(a) fact
(b) theory
(c) hypothesis
(d) principle

60. Galileo discovered that the sun had many dark spots on its surface. He also found that these dark spots moved across the face of the sun from one side to the other. (see diagram below) This observation could be explained by assuming that the sun

(a) is spinning and carrying the dark spots around with it
(b) is not moving and the dark spots are spinning around the sun's surface
(c) neither a nor b
(d) both a and b



61. In trying to find evidence for the evolution of life on earth, biologists should follow
   (a) any method whatsoever, whether logical or imaginative, which can produce testable hypotheses
   (b) the truly scientific methods of observing all the facts first and then drawing conclusions
   (c) the scientifio method of defining the problem, collecting data, forming hypotheses, testing the hypotheses and drawing conclusions
   (d) both b and c

62. There are statements in science describing certain regularities that have been observed in nature. They say that when a certain event happens, a certain other event always happens afterwards. Such statements are called scientific
   (a) facts
   (b) theories
   (c) hypotheses
   (d) laws

63. The scientific method can be most accurately described as including
   (a) the five steps of stating the problem, collecting data, forming a hypothesis, testing the hypothesis, and making conclusions
   (b) a demand for describing accurately, arguing logically, and explaining logically
   (c) an assurance that the investigator will be successful, if he follows the steps outlined in the method
   (d) both a and c

64. Tycho Brahe, a Danish astronomer, improved the design of many of the instruments used to observe the sky. He used these new instruments to make the most accurate records of the heaven's motions up to his time. Which of the following is correct?
   (a) Improvement of the instruments illustrates a main concern of science
   (b) Production of the records illustrates a main concern of science
   (c) Both are main concerns of science
   (d) Neither are main concerns of science

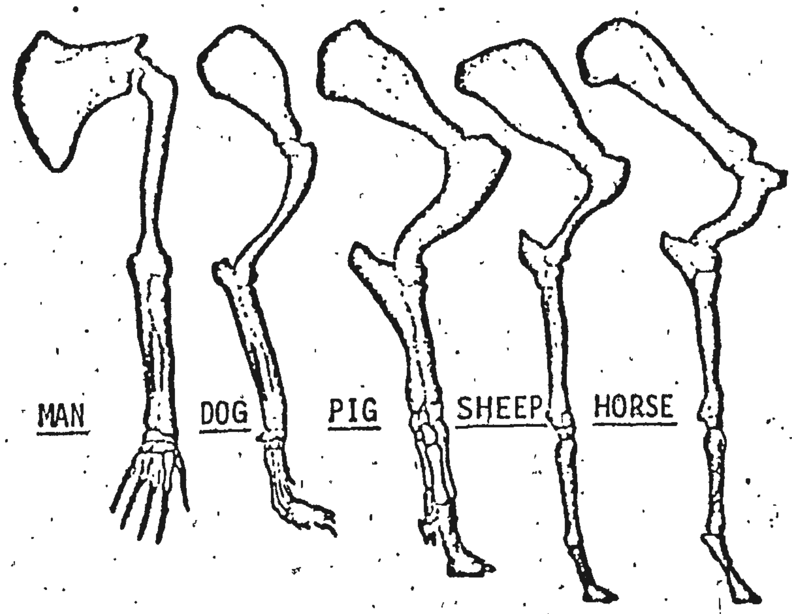
65. Many biologists believe that forms of life can change over many years to become new forms. Two of the pieces of evidence they supply for this belief are given below.

EVIDENCE X

When the poison DDT was first used to control insects, it killed nearly all the houseflies it contacted. Now, after years of use, there are types of houseflies that strongly resist DDT.

EVIDENCE Y

If one compares the front limbs of many animals, such as dogs, pigs, sheep, and horses, they are found to be very similar. The whole pattern or structure is the same. (see diagram below)



Which of the following statements is true?

(a) Evidence X is an example of indirect evidence.
(b) Evidence Y is an example of direct evidence.
(c) Evidence X gives more support to the biologists' belief than Evidence Y.
(d) Both Evidence X and Evidence Y give about the same support to the biologists' belief.

66. The ancient Greeks knew that the sun did not travel the same path across the sky every day. Day by day, from June to December, they observed that the noon day height of the sun above the southern horizon became less and less. As this was happening, they also observed that the number of hours of daylight became less and less. The statement relating these two observations is known as a scientific
(a) law
(b) hypothesis
(c) fact
(d) theory

67. This relationship between the height of the sun above the horizon and the number of hours of daylight could be used to tell
(a) from the position of the sun when to plant or harvest crops
(b) the approximate time of year from the sun's position
(c) both a and b.
(d) neither a nor b

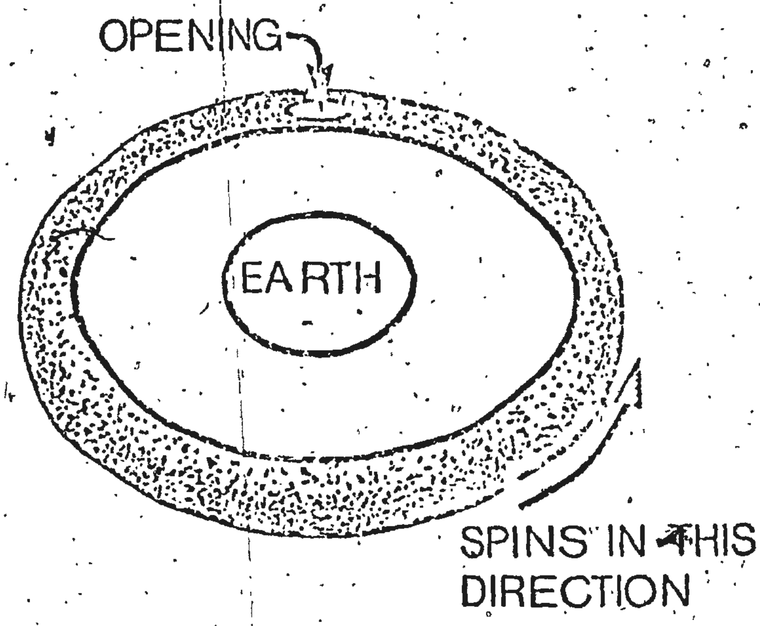
68. Before a scientist would even consider accepting an explanation as true, the explanation must be able
(a) to make some correct predictions of what will happen in certain situations
(b) to explain correctly all the facts related to the problem at hand
(c) neither a nor b
(d) both a and b.

69. Which of the following statements about science and technology is true?
(a) Science and technology are different names for the same area of work.
(b) Without science, technology could not advance as rapidly; but without technology, science could advance just as rapidly.
(c) Both the aims and products of science differ from the aims and products of technology.
(d) both b and c

70. Very often in science an old theory is replaced by a new theory which scientists believe is better. What characteristics must the new theory have before this can be done?
(a) It should be able to account for all the facts that the old theory accounts for, plus some more.
(b) If it cannot account for more facts than the old theory, then it should be simpler and more convenient.
(c) It should be able to predict phenomena that were not even known when it was invented.
(d) any of the above

71. Anaximander, a Greek who lived from 611 BC to 547 BC, had some ideas which he used to explain the behavior of the moon. He claimed that the moon is a spinning, doughnut-shaped object, surrounding the earth. (see diagram below) He said the "doughnut" is full of fire, having just one round opening for the light to escape. This opening is what we see in the night sky. He also claimed that the opening had a shutter, which could be adjusted to any position from fully opened to fully closed. Anaximander's ideas would be called a scientific

(a) law
(b) hypothesis
(c) neither a nor b
(d) both a and b



72. If no other ideas explaining the behavior of the moon were available, scientists would

(a) accept Anaximander's ideas, because there is no other explanation available concerning how the moon behaves
(b) accept Anaximander's ideas, because they do explain some of the moon's behavior
(c) reject Anaximander's ideas, because they cannot explain all of the moon's behavior
(d) none of the above

73. Anaximander used only the observations of the moon he had made, plus his imagination, to come up with his explanation of the moon's behavior. In science, using one's imagination is

(a) not all right, because too often false explanations, like Anaximander's, are the result
(b) all right, because very often an answer cannot be reached using any other method
(c) not all right, because explanations should be reached using the most logical procedures known
(d) all right, because this is the only way to reach explanations as quickly as possible

74. One of the well supported theories of biology is the Theory of Evolution. It would be true to say of this theory, or indeed any well supported theory in science, that

(a) it is not an observational fact
(b) it is useful in that it can explain a body of facts
(c) both a and b
(d) neither a nor b

75. Scientific facts are discovered

(a) in experiments which have been repeated with the same results very many times
(b) while testing predictions that have been made using scientific explanations
(c) in observations of nature which have been seen many, many times
(d) all of the above

APPENDIX F

Answer Key for Form B

Item Number	Answer	Item Number	Answer	Item Number	Answer
1	C	33	B	65	C
2	A	34	D	66	A
3	B	35	D	67	C
4	D	36	D	68	D
5	D	37	D	69	C
6	B	38	D	70	D
7	C	39	A	71	B
8	B	40	A	72	B
9	B	41	A	73	B
10	A	42	B	74	C
11	C	43	A	75	D
12	C	44	C		
13	C	45	D		
14	A	46	A		
15	C	47	C		
16	C	48	A		
17	A	49	C		
18	A	50	A		
19	C	51	B		
20	D	52	B		
21	C	53	A		
22	B	54	B		
23	D	55	D		
24	D	56	C		
25	A	57	D		
26	C	58	B		
27	A	59	C		
28	D	60	D		
29	C	61	A		
30	C	62	D		
31	A	63	A		
32	B	64	C		