69-4052

STORY, William Emerson, 1927-
A COMPARATIVE STUDY OF A READING AND NONREADING SCIENCE TEST AT THE SIXTH-GRADE LEVEL.

University of Arizona, Ph.D., 1968
Education, general

University Microfilms, Inc., Ann Arbor, Michigan

A COMPARATIVE STUDY OF A READING AND NONREADING SCIENCE TEST AT THE SIXTH-GRADE LEVEL

by

William Emerson Story

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ELEMENTARY EDUCATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 6 8

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my direction by William E. Story entitled A Comparative Study of a Reading and Nonreading Science Test at the Sixth-Grade Level be accepted as fulfilling the dissertation requirement of the degree of Doctor of Philosophy

July 11, 1968

Dissertation Director — Date

After inspection of the final copy of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:*

7-11-68

7/11/68

7/11/68

7/15/68

*This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: William E. Stoney

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
ABSTRACT	vii
CHAPTER	
I. THE PROBLEM, THE HYPOTHESES AND DEFINITION OF TERMS	1
The Problem	4
Statement of the Problem	4
Basic Assumptions	5
Importance of the Study	5
The Hypotheses	6
Statement of the Hypotheses	6
Limitations of the Study	8
Definition of Terms	8
Readability	8
Reading Comprehension	9
II. REVIEW OF THE LITERATURE	10
Rationale of a Nonreading Test	10
Language and Vocabulary	11
Oral Language and Reading	13
Listening	15
Application of the Rationale to Science	17
Summary of the Rationale	22
III. RESEARCH PROCEDURES	24
The Subjects	25
Selection of the Subjects	25
The Reading Measure	27
Description of the Test	28
The Readability Measure	32
Validity of the Readability Formula	33
Other Studies Involving the Dale-Chall Formula	34

TABLE OF CONTENTS--Continued

CHAPTER	Page
Construction and Design of the Nonreading Science Test	36
Description of the Test	37
Pilot Item Analysis	38
Final Item Analysis	42
Research Design and Procedures	46
IV. RESEARCH FINDINGS	50
Analysis Techniques	50
Description of Findings Pertinent to the Hypotheses	52
Secondary Findings	60
Summary of Findings	66
V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS . .	69
Summary	69
Conclusions	71
Recommendations	74
Indications of Further Studies	75
APPENDIX A: THE DALE-CHALL READABILITY FORMULA . .	79
APPENDIX B: NONREADING SCIENCE TEST ORAL INSTRUCTIONS, ORAL QUESTIONS, AND TEST ITEMS	80
LIST OF REFERENCES	133

LIST OF TABLES

Table		Page
I.	PILOT TEST-ITEM STATISTICS	40
II.	FINAL TEST-ITEM STATISTICS	44
III.	DALE-CHALL READABILITY FORMULA AS APPLIED TO THE STANFORD SCIENCE TEST--INTERMEDIATE II LEVEL	54
IV.	SIGNIFICANCE OF THE DIFFERENCE BETWEEN NONREADING SCIENCE--READING AND READING SCIENCE--READING CORRELATION COEFFICIENTS	56
V.	SIGNIFICANCE OF THE CORRELATION COEFFICIENTS INVOLVED IN THE COMPUTATION PERTAINING TO HYPOTHESIS TWO	61
VI.	CORRELATIONS BETWEEN NONREADING SCIENCE, READING SCIENCE, AND READING TESTS	63

ABSTRACT

Preliminary to this study, an informal survey was taken of many educators and testing officers. The concensus of opinion was that most of the more commonly used standardized science tests at the elementary-school level were little more than reading tests. To determine the effect of reading on science achievement, a nonreading science test was developed with which to compare a standardized reading science test. Forty-eight nonreading science test items were constructed, sixteen designed to parallel items of similar concept in the reading science test. Due to wide usage, the Stanford Science Test, Intermediate II level of the Stanford Achievement Battery, 1964 revision was chosen as the reading science test.

After construction of the nonreading science test, a pilot testing was conducted to determine item difficulty, item discrimination index, and test reliability. The pilot sample consisted of seventeen seventh-grade students considered representative of the larger population of seventh-grade students. Because the nonreading test was constructed to assess science achievement at the completion of the sixth grade, the pilot testing was conducted on seventh-grade students having recently completed sixth-grade studies. The pilot study, completed shortly after

the beginning of the school year, revealed several ambiguous and poorly discriminating items. These were discarded, resulting in a final forty-seven item test. To insure the best comparison possible with the reading science test, the nonreading science test was administered immediately following testing with the Stanford Achievement Battery. In the school district where this study was conducted, the Stanford testing was scheduled for March, 1968. The sample of 277 sixth-grade students was tested with the nonreading science test within two weeks after administration of the Stanford Battery. Final item analysis on the nonreading science test determined item difficulty, item discrimination, and test reliability. Also, an analysis was made of the data collected pertaining to the reading science and the nonreading science tests using a correlation-relationship design. The two science tests were analyzed to determine if any significant difference existed between them as each correlated with a reading test, the Stanford Paragraph Meaning Test. Also, the Stanford Science Test was analyzed for readability using the Dale-Chall Readability Formula to determine if the reading science test could be read with comprehension by those sixth-grade students reading at grade level expectations.

Findings pertaining to this study were:

1. The Stanford Science Test was graded in reading difficulty with the first two-thirds (approximately) of the test suitable to those students reading at grade-level expectations. The last one-third of the test (approximately) was better suited to students reading at the eleventh- or twelfth-grade level. The average readability was more appropriate for eighth- or ninth-grade level.
2. A significant difference existed between the reading science and the nonreading science tests as each correlated with a reading test. Significance was tested at the .05 level. The reading science test was more highly related to reading achievement than to science achievement as measured by the nonreading science test. These two science tests were apparently measuring different areas of achievement.
3. Secondary findings indicated that when the influence of intelligence was partialled out, the relationship between reading and nonreading science achievement was no greater than could be expected from chance. However, the relationship between reading and reading science was substantial (beyond

the .01 level of significance) when intelligence was partialled out.

It was concluded that the Stanford Science Test was far above the average reading ability of sixth-grade students. Also, this reading science test was found to be measuring essentially reading ability rather than science learnings.

CHAPTER I

THE PROBLEM, THE HYPOTHESES AND DEFINITION OF TERMS

To evaluate achievement adequately in science education in the elementary school, a testing program of sufficient breadth and depth is required. In order to insure as thorough a testing program as possible, all relevant factors must be considered which may have an influence on the stated objectives of the science curriculum. There is some question as to the sufficiency of elementary science evaluation which is employed in the public schools today. Many educators and testing officials believe that the current science tests used in the elementary schools are not considering all the pertinent factors, such as knowledge of science vocabulary and reading ability as it pertains to comprehending the test item. There is some evidence that science evaluation, especially in the elementary schools, is based heavily on a reading ability which most students apparently do not possess. Ragan (29: 251) states that there is ample evidence that many children are failing to develop sufficient reading ability to meet the demands of the school curriculum and that approximately one-fourth of the failures in the elementary school are caused by a lack of reading ability. Malkin (27:5) states that the lack of reading achievement probably accounts for a considerable lack of success in those curriculum areas which depend on reading. Shores and Saupe (32:149-158) point out that there is little doubt that reading comprehension is affected by the kind of material being read and that it is reasonable to expect that the reading skills required for science material will differ from those required for other content areas. This difference is primarily due to the complexity of the vocabulary and the more abstract nature of the concepts involved in science when compared to reading classroom stories. There also seems to be a need to separate tests of ability to read in the various content areas, including science education, again due to the more complex nature of science reading material.

Due to the many technical terms, science vocabulary differs markedly from that required in other content areas. Since evaluation of elementary-school science depends almost exclusively upon some form of written instrument heavily loaded with the reading and comprehension of this technical vocabulary, the question arises whether evaluation in science is actually measuring science understandings and applications or is in fact assessing the student's reading comprehension and knowledge of vocabulary. Regardless of the type of science test used, standardized or teacher-made, the reading comprehension factor, which includes vocabulary understanding, is present to a considerable degree.

Since the validity of any test is based on the extent to which the behavior elicited in the testing situation actually represents the behavior being evaluated, current science tests used in the elementary school may not be valid. Gronlund (17:74-76) states that lower validity results when anything in the construction or administration of the test causes the results to be unrepresentative of the characteristics of the person being tested. Such factors as reading vocabulary and sentence structure being too difficult, test items of inappropriate level of difficulty, and test items which are inappropriate for the outcome being measured can lower the validity of the test results. The question logically arises as to just how valid a science test is for evaluating science achievement at the elementary level, especially when no apparent consideration is given to comprehension of the written material involved in the test. It is questionable if any science test at this level, which is based on reading comprehension, actually measures what the test constructor had in mind at the time the test items were formulated.

The Problem

Statement of the Problem

The main contention of this study is that regardless of how well current science tests at the elementary level are constructed or have been constructed, as long as these tests require reading on the part of the student, then the reading comprehension of the students taking the tests must be considered. Reading comprehension would seem to be one of the most prominent factors influencing validity, but its impact on the student's science achievement has been largely disregarded. Since current science tests, such as the Stanford Science Test, seem to assume a certain competence in reading, reading must be a factor in science achievement and evaluation. If a science test is to measure what it purports to measure, then the reading factor must receive considerable consideration in order to assess adequately the attainment of those objectives deemed desirable in science education.

Based upon the above consideration, this study attempted to assess science achievement by eliminating the direct influence of reading. This necessitated the development of a nonreading science test using pictures and oral instructions. It was the purpose of this study to compare science achievement on this nonreading measure with one of the more commonly used reading science tests, the Stanford Science Test.

Basic Assumptions

This study was based upon the following assumptions:

1. Reading comprehension is an important factor in science achievement and subsequent evaluation.
2. Reading comprehension can influence the results of elementary science tests.
3. Reading comprehension varies in degree from child to child.
4. Science education involves reading skills peculiar only to science.
5. The readability formula used in this study can accurately measure the readability of science tests.
6. The reading and nonreading science tests used in this study can measure desired science outcomes.
7. Children selected from one particular geographic area will be representative of all children falling within the larger geographic area chosen for this study.

Importance of the Study

The reading comprehension factor seems to play a prominent role in performance on the elementary level standardized science tests. Through an informal poll conducted as a preliminary to this study, the consensus of opinion of educators and testing personnel was that these tests are essentially reading tests. If this is indeed the case, then the validity of science tests requiring a high degree of competency in reading is in serious doubt and the score which a child receives is not a true indication of achievement in science. This results in an unrealistic appraisal of the child's mastery of science concepts, understandings, and applications.

To plan the science curriculum better and to provide the student with current and future guidance in his pursuit of science studies, it is essential to obtain the best possible measure of his knowledge and understanding of science concepts and applications. If reading comprehension really has such a profound effect on the results of current science tests, then it is important to recognize this influence and to develop testing instruments which eliminate or minimize the reading factor.

The Hypotheses

Statement of the Hypotheses

For the purpose of this study, the following hypotheses were advanced:

1. The Science Test of the Stanford Achievement Battery, Intermediate II level, 1964 revision,

will not differ appreciably in reading difficulty from the level of reading expectancy for sixth-grade students.

Since the Intermediate II level Battery was designed for sixth-grade students, the reading difficulty should be no greater than what the sixth-grade student is capable of comprehending as determined by the Dale-Chall Readability Formula. This hypothesis was tested only by determining from the readability formula the average grade equivalent of the Stanford Science Test.

2. There will be a significant difference between the correlations of scores on a reading science test and a reading test, and a nonreading science test and a reading test when applied to the same selected group. The null hypothesis of no difference was tested at the .05 level for significance.

   Although there was some empirical evidence to indicate that a difference did exist, the null hypothesis was tested in order to make evaluation of the research statistically more precise (16:212-213).

Limitations of the Study

The limitations of this study were the following:

1. The ability of the author to design a nonreading science test.
2. The validity and reliability of the Dale-Chall Readability Formula in determining the reading difficulty level of a standardized test.
3. The visual and auditory perceptual abilities of the testees in viewing the pictures of the nonreading science test and in hearing the oral instructions.
4. The limited scope of the study, which included only sixth-grade students enrolled in the public schools of one district.

Definition of Terms

Readability

That level of reading difficulty, stated in grade equivalents, as determined by applying the Dale-Chall Readability Formula, will herein be termed "readability." Readability or reading difficulty, considered to have the same or nearly the same meaning, is based on the difficulty of the vocabulary and the length of the sentences found in the reading material.

Reading Comprehension

This term refers to the student's ability to understand and comprehend what he reads in science. Recognition of and understanding the meaning of words peculiar to science education are especially important when considering reading ability in science, and constitute the essential aspect of the reading factor considered in this study. Reading comprehension and reading ability will be considered hereafter to have the same or nearly the same meaning and are understood to involve more than just the mere recognition and reading of words without understanding.

CHAPTER II

REVIEW OF THE LITERATURE

A considerable amount has been written concerning achievement testing in the content areas, and much has been written about science testing specifically. However, there apparently has been little research concerning achievement testing in science utilizing any type of nonreading science test. Although the use of pictorial representation of test items is certainly not new, especially in the area of intelligence testing, this study apparently presents the first rather comprehensive attempt at developing a nonreading science achievement test based exclusively on pictures and oral instructions.

Considering the lack of literature which pertains directly to the development of a nonreading achievement test in science, the review of the literature has been developed to lend support to the reasoning behind the current attempt to design such an achievement test for the sixth-grade testing program.

Rationale of a Nonreading Test

In the development of any nonreading instrument, the primary consideration is to control the influence of reading, and often oral language. Whatever the cause of the deficit in reading and language, it seems quite certain that either one or both can depress a child's score on an otherwise valid testing instrument. To understand better why the nonreading instrument was developed for this study, as many of the relevant factors as possible have been discussed from the standpoint of related literature.

Language and Vocabulary

Ames (1:64) has pointed out that the words that a child has heard and understands form the basis for the speaking, writing, and reading that the child will do. He concluded that the larger the size of this understanding vocabulary, the richer the foundation on which other vocabularies can be built. On the other hand, many excellent studies have been conducted, such as that by Edmonds (13:61-63), which indicated rather significantly a vast difference in verbal ability between students from high and low socioeconomic backgrounds, but that this difference no longer exists when socioeconomic status is held constant. Strang (33:503) stressed the fact that a cultural background devoid of meaningful experiences leads to inappropriate speech. A child's background of experience is the critical element here. All children, who are considered "normal" at least, do not lack language ability and facility, regardless of their background. It is more a case of whether or not the appropriate language has been acquired which will "fit" the child into the currently accepted curriculum. Other studies, conducted by Davidson and Balducci (9:476-480), Saltzman (31:71-81), and Davis (10:65), which claimed to have held IQ constant while investigating the influence of experiential background, found that children from environments which afforded the child many and varied experiences were superior in verbal ability when compared with the child from an impoverished environment. Regardless of the child's background before he comes to school, there is always the possibility that the school environment can provide the necessary experiences which will aid the child in developing language and vocabulary.

Although, according to others, a child's comprehension vocabulary differs from his speaking vocabulary since he hears and understands words he never uses, Deutsch (11:1-32) found that children who are impoverished in language when they enter school will fall relatively further behind as they continue in school. If this is the case, then the development of a nonreading evaluative instrument should be more valid than a reading test in assessing achievement in content areas other than the language arts. Perhaps it is impossible ever to develop a nonlanguage or nonreading test which is fair to all students, but the effort should be made to minimize the influences of reading.

Oral Language and Reading

There can be little doubt that reading achievement depends greatly upon language facility and vocabulary development. Strickland (35:106) states that the oral language children use is far more advanced than the language of the books in which they are taught to read, but many students who can use oral language well, may read poorly because the set of basic responses required for oral language is not the same as for reading. The use of oral language is a natural skill whereas reading is acquired. Carroll (5:337-339) states that learning to read a language depends upon the ability to understand the spoken form of the language as well as upon the ability to reconstruct the spoken forms of written messages. Since children vary widely in their speech repertoires, difficulty in language development generally results in some difficulty in learning to read until the handicaps are no longer present. He continues by stating that many children simply do not speak the same language as the teacher and that restriction of understanding and vocabulary development is often quite prevalent. Again, this is no doubt largely a result of experiential background.

Teachers should bear in mind that students possess quite different patterns of aptitude for reading. Strang (34:239-245) has stated that teachers do not often recognize the complexity of the reading situation and expect a certain procedure to produce good results with all students. Level of difficulty must be considered in any reading experience, and whether or not reading meets a need or is vitally interesting to the student must also be considered as influencing a child's reading. Often a student who has experienced unpleasantness with reading will find ways of avoiding reading and learns little in any reading situation. It is essential to realize that the child must react favorably with the reading situation in order to learn. Thus, regardless of the child's background or his intellectual capacity, if he does not find reading a pleasant experience, he will not profit from reading. But, this does not mean that he has failed to profit from other classroom activities or that he has not learned by other means, such as listening and observing.

In a rather comprehensive treatment concerning reading in science, Tannenbaum, Stillman, and Piltz (36: 311-312) state that slow learners are typically poor readers due to their lack of understanding of ideas and concepts. These students need to use their senses to a greater degree, such as observing and describing simple phenomena. Too much reading and writing can become a barrier to their learning. In addition, Koehn (25:133-134) found that although there is a high relation between general reading ability and other content areas, the relationship with reading in science is very low.

Listening

As long ago as 1928, Russell (30:238) concluded that students in grade five learn more from having material read to them than from reading it themselves. He also found that the relative effectiveness of the two methods (reading and listening) became practically equal by grade seven, with learning by reading slightly favored in grade nine.

According to Taylor (38:15-18), listening abilities are more advanced than reading skills in the primary and intermediate grades, and children prefer to listen rather than read when offered a chance. Listening is preferred because it is a more usual act in which the child has had many years of experience. He concluded that reading is a slower and less efficient process both in terms of word recognition and rate of thinking, and as a result, listening usually makes possible better comprehension and retention than does reading.

The amount of learning through listening is not at all surprising when the amount of time spent in listening is considered. A study in 1950, conducted by Wilt (40:626-636), found that elementary school students listened 158 minutes or 57.5 per cent of each school day, which means that more than half the school day was spent in listening while the remainder was devoted to reading, speaking, and writing. Although children actually spent much more time in listening than in any other activity, it was found that teachers considered listening less important than reading and speaking.

The continued importance of listening has been emphasized by Brown (4:316-321) in a study of listening and academic success. In this study, the students involved rated listening more important than reading as a factor in achieving academic success. Also, for students graduating "with high distinction" and "with distinction," the average percentile rank in reading of those graduating with distinction was seventy-eight, and in listening it was ninety-two. For those graduating with distinction, the average percentile rank in reading was again seventy-eight and in listening eighty-one.

In a study conducted by the Central New York Study Council (6:58-62), it was concluded that listening is a good technique for teaching slow learners. It was also concluded that while sixth graders were both better readers and listeners than fourth graders, there was greater improvement in listening than in reading at the sixty-grade level.

Similarly, Hampleman (20:49-53), in a rather comprehensive study of fourth- and sixth-grade students, found that listening comprehension was significantly superior to reading comprehension for both sexes.

Tiedt and Tiedt (39:87-89) state that listening is very selective and that a person may hear many things but listens to only what he wants to. Each person hears, comprehends, and interprets the spoken word in terms of his unique background of experience. Purposive listening is certainly in direct relationship to the degree of interest in the content of the spoken word. On the other hand, successful listening also depends on the ability of the listener to comprehend what is being said. If a child is unable to understand, he will quickly "tune out" the speaker. The authors conclude that if the content is of interest to the listener, and the material presented is not beyond his capacity to comprehend, considerable effective learning can result.

Application of the Rationale to Science

Science study can be, and usually is, a most interesting study to the majority of children. Due to the impact of radio and television, very few children come to school today without some knowledge of scientific principles and applications. They see and use science every day in their homes and are frequently seen to be playing with toys and games of a scientific nature. Children like science and are usually very willing to participate in, and pursue, the study of science. The burden of maintaining the child's interest in science, in teaching the subject, and in evaluating the learning outcomes rests with the teacher.

An obvious conclusion was reached by Lewis and Potter (26:5) when they stated that understanding the meanings of words in science is essential to correct concept formation. Words have many different meanings and depending upon the context in which they are found, can be quite misleading and confusing. Frequently elementary science textbooks express ideas in words that children use in a sense quite different from the author's intent. In order to form accurate concepts in science, teachers must attempt to assure correct meanings and understandings of the words used in science education.

In this regard, Tannenbaum et al. (36:104) state that children should not be presented with science words merely for the sake of verbalization. These words should be meaningful and understood. In order to have meaning, science words need to be brought into the experience range of the child. Understanding follows experience with what the words represent. They also stress the fact that teachers must recognize that children have different basic learning styles. While some children learn best from reading, others learn best by seeing, hearing, or doing. Although some minimal reading skill is required in science study, an understanding of other learning styles would seem to indicate that rather than attempting to transform all students into reading-style learners, the teacher should utilize all the different learning styles present in the classroom. Along with reading, opportunities for seeing, hearing, doing, and manipulating should be utilized in the classroom.

Ragan (29:367) fairly well condemns the textbook-centered science curriculum and the heavy emphasis placed on reading when he states that there is overwhelming evidence that curiosity and the urge to explore, which are powerful drives in children, are largely lost as children grow older. This loss of interest is largely due to the school's emphasis upon accepting the word of the teacher and the textbook-centered curriculum. There should be greater emphasis upon doing and observing.

Malkin (27:6-7) suggests that a science program in the elementary school should allow poor readers to participate and feel a sufficient degree of success in all areas of science. This can be done by using the child's language skills, other than reading, such as listening, speaking, reporting, and observing. The use of filmstrips and other pictorial representations can often be more effective in conveying many science concepts than can textbooks. Another highly recommended procedure, which can be quite effective without high reading attainment, consists of handling materials and equipment by participating in demonstrations and experiments.

Since the study of science is intrinsically interesting because of its appeal to the seemingly natural desire of children to do and explore, it is surprising that greater advantage is not taken of this fact in the classroom. Over and above this, the effort is made by the teacher in the elementary classroom to present as much science content as he deems necessary or is capable of presenting. Based on the amount of science instruction and the preceding factors related to science education, each teacher must to some degree attempt to evaluate what teaching he has done and the amount of instruction which has preceded in other grades, as well as those learnings which students have acquired incidental to formal instruction.

In the past, great emphasis has been placed on evaluating each student's performance in science in relation to his peers by using a standardized instrument. This type of evaluation pertains mostly to testing the more long range objectives of instruction. One of the most frequently used science tests is contained in the Stanford Achievement Battery. The unfairness of any type of test which is heavily loaded in vocabulary, including the Stanford Achievement Battery, has been demonstrated by many studies. Eells et al. (14:14, 254-265) reported on several studies which indicated that test items essentially linguistic or scholastic in nature showed comparatively large differences in favor of children from better experiential backgrounds, while test items which were primarily perceptual or practical in nature showed either smaller differences or differences in favor of children from poorer experiential backgrounds.

In further support of the contention that reading tests are unfair to many children, Davis (10:87-88) reported on a study involving 516 students, nine to ten years old, who were divided into equal groups and matched for IQ, chronological age, and socioeconomic index. The essence of the study was that the children from the poorer experiential backgrounds performed much better on oral items than on reading items, whereas those from the better experiential backgrounds did slightly better on the regular reading items. It was interesting to note that those items which were considered fairer culturally yielded higher scores for both groups of children.

The connection between reading ability and science achievement evaluation seems rather obvious. If experiential background reflects on language development, and language development influences reading ability, then any test which depends upon reading ability must be unfair to many children in the public schools today. The cause is not lost, however. Tannenbaum, Stillman, and Piltz (37: 214) state that a variety of objectives in science can be measured by objective tests which use pictures. Such objectives as ability to recall information; interpret data, graphs, and charts; and the application of principles can be evaluated by pictorial materials, especially for students with limited reading comprehension. They state that test items based on pictures can provide the student with clear and unambiguous problems that are very realistic, interesting, and novel.

Summary of the Rationale

After a rather thorough search of the literature, it was determined that research in the area of nonreading achievement tests either does not exist or is very meager. Only a few remote suggestions were found to exist relative to the possibility of utilizing pictorial material in the construction of nonreading science tests, which is the basis of this study. For this reason, the literature was reviewed only to the extent of establishing a rationale for the development of a nonreading science instrument for the evaluation of sixth-grade achievement in science education.

The literature was rather conclusive in regard to the relationship which exists between experiential background, language development, and the more complex reading task. Since most achievement tests are based on an assumed competency in reading, it can be readily noted that these tests are grossly unfair to those children who are restricted or handicapped in their reading ability.

However, the background from which the child comes does provide certain opportunities for incidental learnings in science. Likewise, throughout the child's formal education there are numerous opportunities to listen, observe, and do things which result in learning science, without the benefit of formal reading instruction or reading competency. Since children apparently learn a considerable amount via these nonreading means, and since children from the poorer experiential backgrounds have been found to perform better on tests of an oral nature, it is deemed sufficiently justifiable, for the purposes of this study, to develop a nonreading evaluative instrument in assessing learning outcomes in elementary science at the sixth-grade level.

CHAPTER III

RESEARCH PROCEDURES

In considering the research procedures as they applied to this study, it was necessary to consider each of the following:

1. The subjects and their selection.
2. The pilot and final sample.
3. The reading measure, its description, standardization, validity, and reliability.
4. The readability measure, its validity, and related studies.
5. Construction and design of the nonreading science test, its description, validity, pilot item analysis, item difficulty index, item validity, and the pilot test reliability.
6. The final item analysis, item difficulty index, item validity, and the final test reliability.
7. The research design and procedures.

The above listed factors were considered in detail, and the following information pertains to each in the order given.

The Subjects

The information contained herein pertains to the selection and description of the subjects involved in this study and to the limitations placed upon those subjects.

Selection of the Subjects

With the procedures of this study in mind, the selection of the subjects was left to the discretion of the school district in which the study was conducted. A pilot group was necessary for individual testing with the non-reading science test and a final group was selected later to fulfill the requirements of sample size as indicated in the initial format.

The Pilot Sample. The sample consisted of twenty seventh-grade students selected from three grouped science classes attending the same junior high school. These subjects were selected to undergo individual testing to determine the pilot item analysis. Since the pilot item analysis was conducted at the beginning of the school year, and since the nonreading science test was constructed to assess science achievement at the conclusion of the entire sixth year, it was necessary to conduct the pilot work on this group of seventh-grade students.

The pilot sample was selected at random by the principal of the junior high school with full awareness of the objectives of the author of this study. The sample was reduced from twenty to seventeen due to student absences at the time of testing.

The pilot sample consisted of ten boys and seven girls, representing three science-ability groups. The three groups, previously determined by testing procedures, were a high group, an average group, and a low ability group. Six students were selected from the high science group, six students from the average science group, and five students from the low science group. No limitations were placed on the sample.

Although the sample was small and from only one school, it was considered to be fairly representative of the larger population from which the selection was made. Since this group of students was selected only for a trial run to determine item discrimination and item difficulty, the above sample was considered sufficient for this purpose.

<u>The Final Sample</u>. The subjects included in the final sample were selected by the school district central office and comprised what they considered a representative sample of the larger district population. This sample consisted of 317 students and included all sixth-grade students attending four elementary schools within the district. Ninety-eight students were selected from one elementary school, ninety-six from a second school,

fifty-eight students from a third school, and sixty-five students from a fourth elementary school.

Although 317 subjects were selected for the final sample and the scores of this number were included in the final item analysis of the nonreading science test, the sample was reduced to 277 for final consideration in this study. The reduction in sample size was due to student absences. Since the essence of this study was to compare achievement on the nonreading science test with achievement on the Stanford Science Test (the reading science test), it was thus necessary to compare only those students who had performed on both tests.

No limitations were placed on the final sample other than the fact that each subject must have had scores available for the three measures considered in this study, a nonreading science score, a reading science score, and a reading test score. All subjects for whom these three scores were available were included in this study. Since each subject performed on all three measures listed above, it was considered unnecessary to further refine the sample by age, score extremes, or sex.

The Reading Measure

The 1964 revision of the Stanford Achievement Battery, Intermediate II level, Form W was used in the study to obtain the reading science test scores and the reading scores. The reading science test scores were obtained from the use of the Stanford Science Test, while the reading scores were obtained from the test of Paragraph Meaning.

Description of the Test

The Stanford Achievement Test consists of a series of comprehensive achievement tests developed to measure the important knowledges, understandings, and skills commonly accepted as the desirable outcomes of the major branches of the elementary curriculum. Although no elaboration will be made concerning the revisions, the current edition, comprising Forms W, X, Y, and Z, is the fifth in the series of Stanford Achievement Tests. The revisions have been necessary to insure that the content of the tests may continue to be closely related to what is actually being taught in the schools; that the normative data may reflect accurately the current achievement of students of various grades and ages; that the tests may keep current on improvements in measurement theory and techniques; and that overfamiliarity of test content may be minimized as a result of repeated use in the schools.

The four forms, W, X, Y, and Z, of the various tests are matched for content and difficulty and represent equally good measures of the included subjects while yielding directly comparable results. Although almost all tests are timed tests, the time limits are considered administrative convenience and do not place any premium upon speed of work. The time limits are considered generous and permit practically all students sufficient time to attempt all items which they are capable of answering correctly. Thus, the tests are essentially power tests and not speed tests.

Only two of the subtests of the Intermediate II level battery were used in this study. The following information represents a more detailed description of these two subtests, the Stanford Science Test and the Paragraph Meaning Test.

The Paragraph Meaning Test consists of a series of paragraphs which are graduated in reading difficulty. Since one or more words have been omitted from each paragraph, the student's task is to demonstrate comprehension of the paragraph by selecting from four alternatives the proper word for each omission. The test also includes complete paragraphs about which questions are asked, and answered by selecting one of four possible alternatives. This part of the test provides a functional measure of the student's ability to comprehend connected discourse involving levels of comprehension varying from very simple recognition to the making of inferences from what is stated in several sentences. The areas covered in the paragraphs include miscellaneous items from general reading material, life science, physical science, literature, geography, history, other social sciences, and the fine arts. The items have been constructed to place a premium on actual comprehension of the material read.

The Science Test of the Intermediate II battery has been designed to measure the ability to see the application of the principles of science in the environment and everyday activities, to measure knowledge of facts and generalizations from the various branches of the natural sciences, and to measure some knowledge of the scientific method (22:3-7).

Standardization Procedure. The entire Stanford Achievement Battery for all grades was standardized by testing 850,000 students in a total of 264 school systems drawn from the fifty states. All participating school systems administered Form Xr of the Stanford Achievement Battery to all the students in at least six consecutive grades between February 25 and March 16, 1963. In Grades four to nine, all students completing each battery were included. The number of students to be included for each region of the United States was determined by comparison with the 1960 census data. This was done to insure a representative distribution from the greater population.

Validity. According to the test manual of the Stanford Achievement Battery, validity of the entire battery was described as content or curricular validity, which was assessed through a careful analysis of the actual content of each subtest in relation to the objectives of instruction in the various areas. In order to insure this validity, the authors examined appropriate courses of study and textbooks as a basis for determining the knowledges, understandings, and skills to be measured.

Reliability. The present edition of the Intermediate II level battery is organized in nine subtests designed for use from the middle of Grade five to the end of Grade six. Since only two of these subtests actually pertain to this study, the following reliability information concerns only the Stanford Science Test and the Paragraph Meaning Test.

The reliability coefficients on the Stanford Achievement Battery were based on odd-even split-half reliability coefficients and Kuder-Richardson Formula 20 reliability coefficients. The split-half coefficients were corrected by the Spearman-Brown Prophecy Formula. For the Science Test of the Intermediate II level, a split-half reliability coefficient of 0.90 was reported and a Kuder-Richardson 20 coefficient of 0.89. The Paragraph Meaning Test of the Intermediate II level yielded a split-half reliability coefficient of 0.93 and a Kuder-Richardson 20 coefficient of 0.92 (22:24-27).

The Readability Measure

The Dale-Chall Readability Formula (7:11-20), used to determine level of difficulty of reading materials, was developed in 1948 to provide a rather simple and fast method of determining the grade-level score equivalent of reading material. This formula is based on a factor of vocabulary loading and a factor of sentence structure (average sentence length) found in the reading material. The vocabulary load factor is based on the number of words outside the Dale list of three thousand words. Words not included or found in the Dale list are considered unfamiliar and not readily comprehended by the reader. This list of three thousand words was developed by testing fourth-graders on their knowledge in reading of a list of approximately ten thousand words, which had previously appeared on commonly used word lists for students. An attempt was made to include all words that fourth-graders would possibly know. A word was considered as known when at least 80 per cent of the fourth-graders indicated they knew the word. The authors readily admit that some words actually known to fourth-graders may have been left out, but the Dale list is considered to represent a fairly complete list of familiar and simple words.

Validity of the Readability Formula

The validity of the Dale-Chall Readability Formula was established by correlating the formula predictions with external criteria, such as the judgments of experienced teachers, the judgments of readability experts, and the actual comprehension scores of readers on passages.

When applied to fifty-five passages of health-education materials, the formula predictions correlated 0.92 with the judgments of readability experts, and 0.90 with the reading grades of children and adults who were able to answer at least three questions out of four on thirty of the passages. On seventy-eight passages on foreign affairs from current-events magazines, government pamphlets, and newspapers, the correlation between the formula predictions and judgments of difficulty by expert teachers in the social studies was 0.90.

Dale and Chall (7:20) point out that the nature of the difficulty of a given piece of writing depends greatly on its interest appeal to the reader as well as his background in the subject-matter being read. Although these two factors enter into readability, it is not possible to make allowances for them in the formula. In addition, the various uses made of many words can render them either easy or difficult. When simple and familiar words are used in a symbolic or metaphoric sense, these words can become extremely difficult to comprehend. The formula is not sensitive to this type of variation.

Other Studies Involving the Dale-Chall Formula

In a study conducted by Bormuth (3:131-134), the Dale-Chall Readability Formula was utilized in order to test his hypothesis concerning what he termed the "cloze procedure." The formula provided the criterion for writing several passages included in the study. The formula was used in order to insure that certain materials, devised for the testing situation, did not exceed the level of difficulty specified for Grade four, five, and six.

Arnsdorf (2:243-246), in determining the readability of basal social studies materials, also used the Dale-Chall Readability Formula. In this study, the formula was used only for the intermediate level. It was found that the readability level of the social studies series under study, determined by the application of the formula, generally progressed according to the publisher's recommended sequence. However, the general progression was marked by irregularities both within and between the texts. In general, each of the texts proved too difficult for the particular grade level for which the book was intended.

The Dale-Chall Readability Formula was utilized by Miller (28:205-209) in studying the difficulty of textbooks designed for use by ninth-grade industrial arts students.

It was found that the difficulty of the textbooks varied, with some samples so difficult that they were above the reading ability of over 86 per cent of the students involved in the testing situation. For a second textbook, 67 per cent of the samples were within the reading abilities of a majority of students in the group. For a third textbook, less than half of the samples were rated within the reading abilities of 50 per cent of ninth-grade students, and 20 per cent of these samples were beyond the measured reading ability of 99 per cent of the group.

Klare, Mabry, and Gustafson (24:287-295) applied the formula to determine the effect of readability upon immediate retention (comprehension) and delayed retention. Results showed that with easier readability, greater and more complete retention was effected, a greater amount was read, and material was judged more acceptable to the testees.

Guckenheimer (18:231-238) applied the formula, and the judgment of experts in the field, to the difficulty of international affairs pamphlets. It was found that 75 per cent of the material was at or above the college level, 14 per cent at the senior high school level, and 11 per cent at high school freshman level.

In a study comparing the estimates of 92 reading experts with scores from the Flesch, Lorge, and Dale-Chall Formulas, Herrington and Mallinson (21:385-390) found that the formulas judged the grade level of science texts used in grades four to eight much more consistently than reading experts.

Forbes and Cottle (15:185-189) evaluated five of the more popular techniques for evaluating the reading difficulty of printed matter in relation to standardized tests. The Dale-Chall, Flesch, Lorge, Lewerenz, and Yoakam Formulas were applied to twenty-seven selected standardized tests. It was found that the Dale-Chall Formula gave a more realistic and practical interpretation of difficult words than the other formulas. Based on the intercorrelations for the five formulas applied to twenty-seven tests in this study, it was found that the Dale-Chall Formula correlated highest with the other formulas, and correlated highest with the mean of the five formulas.

Based on the extensive use of the Dale-Chall Readability Formula, Klare (23:22-23) concluded that this formula is consistently more accurate than the others with which it was compared. He also stated that this formula is one of the most popular and can be applied equally well from the third-grade level to adult level.

Construction and Design of the Nonreading Science Test

The nonreading science test developed for this study was based on the content of textbooks for the sixth year of the elementary school in the Tucson, Arizona public schools. The following information, pertaining to the description of the test, the sample, and the pilot item analysis, was determined by closely following established patterns for the development of an achievement instrument.

Description of the Test

The nonreading science test was developed utilizing oral instructions and questions accompanied by the visual presentation of pictures. The picture or pictures pertaining to each test item were drawn on individual sheets (see Appendix B). Some test items have three alternatives, only one of which is the correct response. Other items contain a key picture and three alternatives. In responding to this type of item, the testee is required to refer to the key picture in order to select the correct alternative. Each item is answered simply by crossing out on the accompanying answer sheet the letter corresponding to the selected alternative. No reading ability is needed or assumed for successful solution of any item.

Test Validity. Content validity is of primary concern in achievement testing and is concerned with the extent to which the test items actually call forth the responses represented in the table of specifications (17: 62-63). With this in mind, a table of specifications was developed for the nonreading test. This utilized the content of the three state-adopted textbooks currently in use in the public schools and the course of study which has been developed as the most comprehensive and inclusive guide to the accomplishment of the learning outcomes deemed desirable by the local school district in which this test was to be administered.

Pilot Item Analysis

Although this test was designed for the evaluation of those desired learning outcomes in science ordinarily assessed at the end of the sixth year of elementary school, the pilot item analysis was accomplished by administering the complete test to seventeen seventh-grade students during the first four months of the 1967-1968 school year. This was considered necessary since the sixth-grade students later included in this study had not completed a sufficiently broad coverage of the science content for which this test was designed. The items of the original test were presented to each testee individually, one item at a time. Then, he was asked why he responded as he did to the particular item. This procedure was utilized to insure that the student perceived the pictures as they were intended to be perceived. Fortunately, there were no ambiguities in the way each student saw the pictures, and in only two cases was it necessary to repeat the oral question.

The pilot item analysis was performed in order to appraise the effectiveness of each test item by determining item difficulty and item discriminating power. On the basis of this item analysis, several of the items were found to be misleading while some others failed to discriminate satisfactorily. These items were subsequently revised or removed. Sixteen items were designed to parallel items in the Stanford Science Test and provided the basis of this study. Although some of these items did not discriminate well when presented as nonreading items, it was considered necessary to retain these items intact lest the comparison with the Stanford Science Test be negated.

Item Difficulty Index. Since item difficulty has been considered an important feature in item analysis, each item in the pilot testing situation was subjected to this procedure and listed in Table I. Although there is not universal agreement on the range of acceptable item difficulty, Garrett and Woodworth (16:363) have stated that other things being equal, items of moderate difficulty (index of 0.40, 0.50, and 0.60) are to be preferred to those which are much harder or easier. Gronlund (17:112) has stated that due to errors in judgment, some items will be easier than the desired index of approximately 0.50, while other items will be more difficult. He continued by stating that except for a few items at the beginning of

TABLE I

PILOT TEST-ITEM STATISTICS

Item	Difficulty Index	Validity (Discrimination)	Item	Difficulty Index	Validity (Discrimination)
1	0.94	0.26	25[a]	0.65	0.33
2	0.94	0.26	26	0.59	0.70
3[a]	0.94	0.26	27	0.29	0.43
4	0.29	0.30	28	0.53	0.42
5	0.65	0.42	29	0.47	0.56
6[a]	0.88	0.45	30	0.76	0.21
7[a]	0.94	-0.05	31[a]	0.88	0.54
8	0.82	0.08	32	0.76	0.72
9[a]	0.53	0.60	33	1.00	0.00
10[a]	0.82	0.13	34	0.35	0.39
11[a]	0.76	-0.08	35	1.00	0.00
12[a]	0.88	0.54	36	0.82	0.21
13[a]	0.65	0.19	37	0.53	0.22
14[a]	0.59	0.37	38	0.71	0.73
15[a]	0.59	0.24	39	0.59	0.55
16	0.53	0.48	40[a]	0.94	0.26
17[a]	1.00	0.00	41	0.47	0.64
18	0.47	0.52	42	0.88	-0.13
19[a]	0.82	-0.01	43	0.24	0.52
20	0.24	0.30	44	0.35	0.16
21[a]	0.88	0.46	45	0.88	-0.13
22	0.71	0.64	46	0.24	0.54
23	0.88	0.28	47	0.59	0.12
24	0.82	0.51	48	0.94	0.26

[a]Items constructed to parallel Stanford Science Test items.

the test, for motivational purposes, no item should be so easy that everyone answers it correctly. Likewise, no item should be so difficult that everyone misses it. In the pilot item analysis, items with a difficulty index of 0.20 or more were considered acceptable. All forty-eight items met this criterion.

Item Validity (Discrimination). Garrett and Woodworth (16:365) have stated that the validity index of an item, or how well it discriminates, is determined by the extent to which the given item is capable of distinguishing between examinees who differ markedly in the function measured by the test as a whole. The point biserial *r* is considered especially useful in the analysis of the items of a test relative to item-test correlations. Also, it is generally a more dependable statistic for this purpose (16:383). Therefore, this measure was utilized in determining those item validities found in Table I.

Ebel (12:364) has raised the question of how high an item validity should be to insure the best discrimination of any given item. Generally speaking, the test which has the highest average item validity or index of item discrimination would be the most reliable. In Ebel's terms, items showing a validity of 0.20 to 0.29 would be considered "marginal," from 0.30 to 0.39 would be considered "reasonably good," and from 0.40 and up would be considered "very good" items. These ranges would, of course, vary with the method of determining the validity of each item. In the pilot item analysis, thirty-six items were found to be within the range of "marginal to very good." Those items found to be below these limits were revised or discarded.

Pilot Test Reliability. After the completion of the pilot item analysis, the reliability of the nonreading test was determined by utilizing the split-half method. This method involves dividing the entire test into two equivalent halves and then determining the correlation for these half-tests. Since the original order of presentation of the items was selected by randomly arranging the items, the two half-tests were determined simply by placing all the odd items into one half-test and the even numbered items into the other half-test. From the reliability of the half-test, the self-correlation of the whole test was estimated by applying the Spearman-Brown Prophecy Formula. The pilot reliability coefficient was found to be 0.86, which was considered to be a satisfactory estimate of the internal consistency of the test.

Final Item Analysis

The final item analysis was accomplished by administering forty-seven of the items to 317 sixth-grade students. Although all forty-eight items were administered to the first group of students, immediately subsequent to this first group testing item number twenty was determined to be highly ambiguous and was deleted from all further testing. This first group consisted of three sixth grades totaling ninety-seven students, and the testing was completed on March 18, 1968. The second group, consisting of two sixth grades, totaled sixty-three, and the testing was completed on March 19. The third group, consisting of ninety-six sixth-graders from three classrooms was tested on March 20; and the fourth group, consisting of fifty-nine sixth-grade students from two classrooms was tested on March 22, 1968. Item difficulty and item discriminating power were then determined from the remaining forty-seven items, based on the responses of these 317 sixth-grade students.

Item Difficulty Index. Item difficulty was completed in the same manner as was the pilot item difficulty index. These indices are listed in Table II. Except in some cases, the final item difficulty indices did not differ greatly from the pilot difficulty indices.

Item Validity (Discrimination). Because of the large number included in the final sample, and the amount of time which would have been necessary, it was decided to compute the individual item validity indices using the biserial *r* rather than the point biserial *r* as was done with the pilot sample. This was done because use of the biserial *r* has the advantage over point biserial *r* in that

TABLE II

FINAL TEST-ITEM STATISTICS

Item	Difficulty Index	Validity (Discrimination)	Item	Difficulty Index	Validity (Discrimination)
1	0.94	0.26	25[a]	0.70	0.28
2	0.90	0.50	26	0.54	0.05
3[a]	0.89	0.45	27	0.30	-0.06
4	0.09	0.49	28	0.42	0.50
5	0.62	0.18	29	0.31	0.45
6[a]	0.69	0.31	30	0.50	0.16
7[a]	0.78	0.32	31[a]	0.77	0.50
8	0.53	0.15	32	0.61	0.30
9[a]	0.75	0.29	33	0.87	0.21
10[a]	0.57	0.51	34	0.40	0.07
11[a]	0.76	0.20	35	0.92	0.39
12[a]	0.87	0.40	36	0.58	0.42
13[a]	0.56	0.36	37	0.37	0.27
14[a]	0.73	0.45	38	0.62	0.48
15[a]	0.65	0.21	39	0.63	0.39
16	0.52	0.26	40[a]	0.88	0.56
17[a]	0.91	0.33	41	0.52	0.53
18	0.39	0.36	42	0.82	0.40
19[a]	0.85	0.38	43	0.44	0.16
20[b]	deleted		44	0.30	-0.03
21[a]	0.67	0.39	45	0.76	0.23
22	0.75	0.28	46	0.28	0.34
23	0.82	0.64	47	0.63	0.59
24	0.70	0.46	48	0.70	0.57

[a]Items constructed to parallel Stanford Science Test items.

[b]Deleted due to failure to discriminate.

tables are available from which values of biserial r can be read quickly and with sufficient accuracy. Garrett and Woodworth (16:383-384) have stated that although the biserial r yields higher correlations than the point biserial r, it is somewhat less precise. These limitations were considered relatively minor since the indices obtained were still to be considered only as reasonable estimates. The item validity indices were computed for all items and were included in Table II.

Final Test Reliability. After the completion of the final item analysis, the reliability of the nonreading test was again determined by utilizing the split-half method, in the same manner as described under the heading Pilot Test Reliability (p. 42). The final reliability coefficient was found to be 0.66, which was considered to be a satisfactory estimate of the internal consistency of the test as reflected by the characteristics of the final sample and the difficulty of the items included.

Ebel (12:339) has stated that the reliability coefficient for a set of scores depends to a large degree upon the range of talent (heterogeneity) in the group, and that test items of middle difficulty from 0.25 to 0.75 contribute much to test reliability. The heterogeneity of the pilot sample, when compared with the final sample, appeared to offer the most logical explanation for the difference in the two test reliabilities. The pilot sample, although consisting of students from the same grade level, were divided into three distinct groups by the school (high, average, low) by means of gradepoint average and standardized tests. The final samples were not divided by any means other than attendance at different, but closely situated, schools within the same district. It appeared, therefore, that the pilot sample of seventeen students was in fact a more heterogeneous group than the final sample of 317.

Research Design and Procedures

Due to the nature of this study, a correlational-relationship design was necessary in order to compare the two different science test scores for the measured variable, reading comprehension. Since each student was tested on each type of measure involved in this study, no restrictions were necessary for the sample.

In applying each of the science measures and the reading measure to the 317 subjects in this sample, great care was taken to maintain standard conditions during each testing episode.

In order to minimize any possible effect of time-of-day and day-of-the-week, and to avoid possible testing errors due to conflict with recesses and lunch periods, a schedule was formulated in which it was possible to control the administration of the nonreading science test relative to the district-scheduled Stanford Science and Paragraph Meaning Tests. All testing was conducted during regular school hours, Monday through Friday. Each subject's test was scored immediately in order to minimize the amount of recall on the part of the examiner.

After final compilation of the test results of the 317 subjects, item analysis was again completed on the nonreading science test, which included item difficulty and item validity (discrimination index). At this time, the reliability of the nonreading science test was again determined by the split-half method, estimating the reliability coefficient of the whole test by using the Spearman-Brown Prophecy Formula.

Upon completion of the item analysis and the recomputation of the reliability of the nonreading science test, an analysis of the data was undertaken to determine if any significant difference existed between the two science measures as they correlated with the reading measure. The _t_-test was used to determine the significance between the two correlations. The following statistical analysis techniques were used:

1. The relationship between variables is often expressed as the correlation between variables. The degree of correlation is expressed as a correlation coefficient, which indicates the extent to which the scores on one of the variables bear

a systematic linear relation to the scores on the other. The product-moment coefficient of correlation expresses the extent to which changes in one variable are accompanied by, or are dependent upon, changes in a second variable. Since only raw scores were collected for the variables under consideration, the following formula was used to calculate the correlation coefficients for the relationship between the Stanford Science Test and the Stanford Reading Test, the nonreading science test and the Stanford Reading Test, and the Stanford Science Test and the nonreading science test:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}} .$$

2. The essence of this study was to determine if the Stanford Reading Test was more highly correlated with the Stanford Science Test than with the nonreading science test. Since three variables were involved and each variable pertained to only one population, the following equation was used in determining the significance of the relationship between the Stanford Science Test and the

nonreading science test as each correlated with the Stanford Reading Test:

$$t = (r_{xz}-r_{yz})\sqrt{\frac{(N-3)\ (1+r_{xy})}{2(1-r^2_{xy}-r^2_{xz}-r^2_{yz}+2r_{xy}r_{xz}r_{yz})}}\ .$$

CHAPTER IV

RESEARCH FINDINGS

The most important aspect of any research study is the analysis of the research findings as they pertain to the stated hypotheses. The type of analysis depends upon the design of the study, the type of data collected, and the hypotheses to be tested. Educational research involves the study of samples of a defined population, and the purpose of research is to make observations of the selected sample in order to apply the results of that research to the population.

Analysis Techniques

In this study, three methods of analysis were applied to the testing of the hypotheses. The first method, which applied to the first hypothesis only, involved the simple computation of a grade-level of readability using the Dale-Chall Readability Formula. This grade-level of readability, which was found for the reading science test, was then compared with the level of reading expectancy for sixth-grade students in order to determine if any difference existed.

The analysis of the second hypothesis consisted of (a) finding the relationship between the nonreading science variable, the reading science variable, and the reading variable expressed as the correlation between variables; and (b) applying the *t*-test to determine the significance of the relationship between the nonreading science variable and the reading science variable as each correlated with the reading variable.

The correlational technique was used in this study because it has the principal advantage of permitting the measurement of a number of variables and their interrelationships simultaneously. Another advantage of this type of design is that it provides information concerning the degree of relationship between the variables being studied. In addition, the ability of the correlational technique to yield the degree to which the different variables concerned are related often yields an understanding of the way in which the variables are operating.

Although all correlational studies are concerned with the discovery and clarification of relationships, they can be broadly classified as either relationship studies or prediction studies, depending upon the particular emphasis. This study was considered to fall within the category of a correlational-relationship study and the data were treated accordingly.

The final step in the analysis procedure, as it related to this study, was to determine the significance of the difference between the two correlation coefficients found in the step above. The two correlation coefficients involved were the correlation coefficient of the relationship between the nonreading science test and the reading test, and the correlation coefficient of the relationship between the reading science test and the reading test. The significance of the difference was determined by applying the t-test or t ratio. Since there were N - 3 degrees of freedom involved in this calculation, and significance was tested at the .05 level, a t ratio of 1.968 or greater was required in order to state that the difference between the two correlation coefficients was significant.

Description of Findings Pertinent to the Hypotheses

The first hypothesis advanced in this study was: "The Science Test of the Stanford Achievement Battery, Intermediate II level, 1964 revision, will not differ appreciably in reading difficulty from the level of reading expectancy for sixth-grade students." In reflecting upon this hypothesis, it was necessary to apply the Dale-Chall Readability Formula in order to determine the grade level best suited for the level of difficulty of the Stanford Science Test. In order to accomplish this, it was necessary to follow the general rules prescribed by the authors (8:37-54). Essentially, the instructions state that when reading material is rather lengthy, it is necessary to select at least three samples, from the beginning, approximately the middle, and at the end. It is also necessary to select those samples approximating one hundred words per sample. In applying the formula to the Stanford Science Test, the instructions were followed as closely as possible and the samples selected were considered adequate and in keeping with the Dale-Chall requirements.

By referring to Table III, it can be seen that the average corrected grade-levels (readability) best suited for the Stanford Science Test, Intermediate II level, are grades seven to eight. This indicates that those students reading up to grade-level expectations at the time this test was administered could not have been expected to read this material with understanding. Since the entire Stanford Achievement Battery, including the Science Test, was administered during the month of March 1968, their reading expectancy should have been at the sixth-year, seventh-month. Had these subjects been reading at this level, considerable difficulty could have been expected because the average readability of the Stanford Science Test was one to two grade-levels above sixth-grade reading expectancy. However, by referring again to Table III, it can be seen that the Stanford Science Test varies considerably in readability. This test has apparently been constructed so that the further one progresses into the test material, the more difficult the reading becomes. Samples 1 and 2 indicate that the readability is within

TABLE III

DALE-CHALL READABILITY FORMULA AS APPLIED
TO THE STANFORD SCIENCE TEST--
INTERMEDIATE II LEVEL

	Sample 1	Sample 2	Sample 3
1. Number of words in the sample	153	147	121
2. Number of sentences in the sample	8	10	4
3. Number of words not on Dale List	5	10	28
4. Average sentence length (1 ÷ 2)	19.1	14.7	30.25
5. Dale score (3 ÷ 1 x 100)	3.27	6.80	23.14
6. Average sentence length (4) x 0.0496	0.9474	0.7291	1.5004
7. Dale score (5) x 0.1579	0.5163	1.0737	3.7538
8. Constant	3.6365	3.6365	3.6365
9. Formula raw score (6 + 7 + 8)	5.1002	5.4393	8.7907
Corrected grade-level	5-6th	5-6th	11-12th
Average raw score of three samples		6.4434	

the sixth-grade level of reading expectancy for approximately the first two-thirds of the test. Conversely, Sample 3 shows the readability to be eleventh- to twelfth-grade level for the remainder of the test. This is far above the reading expectancy of those sixth-grade students who are reading up to grade level, and even further for those who are not. Therefore, it was concluded that the first hypothesis should be rejected. Considering the average readability of the Stanford Science Test, it could be assumed that students reading at the sixth-grade expectancy level could not adequately read the science materials contained in this test.

The second hypothesis advanced in this study was: "There will be a significant difference between the correlations of scores on a reading science test and a reading test, and a nonreading science test and a reading test when applied to the same selected group." The null hypothesis of no difference was tested at the .05 level for significance. From Table IV it can be seen that the scores on the reading science test (Stanford Science Test) correlated 0.64 with the scores on the reading test (Stanford Reading Test). It can also be seen that the nonreading science test correlated 0.43 with the scores on the same reading test. The nonreading science test correlated 0.45 with the reading science test. Since all scores resulting in the three correlation coefficients were obtained from the same

TABLE IV

SIGNIFICANCE OF THE DIFFERENCE BETWEEN NONREADING SCIENCE--READING AND READING SCIENCE--READING CORRELATION COEFFICIENTS

Tests	Correlation Coefficient	*t* ratio
Nonreading science--reading science	0.45	
Nonreading science--reading	0.43	
Reading science--reading	0.64	
Nonreading science--reading and reading science--reading		4.429

To be significant at the .05 level for N - 3 or 274 degrees of freedom, a *t* ratio of 1.968 would have to be reached. Since the *t* ratio for the comparison is 4.429, the significance is beyond the .05 level and also beyond the .01 level.

sample of individuals, formula two, page 49 was applied to test the significance of the difference between the reading science test scores and the nonreading science test scores as each correlated with the reading test. With N - 3 or 274 degrees of freedom, a *t* ratio of 1.968 was necessary in order to reject the null hypothesis at the .05 level of significance. Table IV shows a *t* ratio of 4.429 for the comparison, which is far beyond the .05 level of significance. The null hypothesis was rejected and it was concluded that the difference between the reading science test--reading test correlation coefficient and the nonreading science test--reading test correlation coefficient was highly significant.

In interpreting the significance of this finding, it appeared that the nonreading science test, as it correlated with the reading test, did not prove to be highly related to reading achievement. This was expected since the nonreading science test was designed specifically to eliminate the influence of reading in science achievement. The correlation between the reading science test (Stanford Science Test) and the reading test (Stanford Reading Test) indicated that both of these tests seemed to be measuring reading achievement. From Table IV the correlation coefficient indicates a substantial relationship between the reading science test and the reading test. Again, this was expected and seemed to substantiate the findings of the previously mentioned opinion survey that the Stanford Science Test does measure, to a great extent, reading ability on the part of the student taking the test.

The correlation coefficient of 0.45 indicated that the relationship between the nonreading science test and the reading science test was substantial. However, the relationship between these two science tests involved in this study was not as great as the relationship between the reading science test and the reading test. It appeared that the two science tests were both measuring science achievement, but the reading science test appeared to be measuring reading achievement more than science achievement.

Upon comparing the correlation of the nonreading science test with the reading test, it was found that the coefficient of 0.43 showed nearly the same degree of relationship as was found between the nonreading science test and the reading science test. Since the reading science test correlated 0.64 with the reading test, it was inferred from these relationships that the reading science test was much better suited to the evaluation of reading achievement than science achievement. This finding supported the acceptance of hypothesis number two.

To lend further support to the findings that the two science tests were not measuring the same thing, the means of the two tests were compared in order to determine if there was a significant difference. It was found that the mean of the nonreading science test was 12.03, and the mean of the reading science test was 11.62. In computing the significance of the difference between these two means, a t ratio of 1.968 would have to be obtained for the difference to be significant at the .05 level. Since the t ratio was found to be 2.6511, it was determined that the difference was significant beyond the .05 level. Based on this significance of the difference between the means, it was concluded that the mean performance on the nonreading science test was significantly different from the mean performance on the reading science test.

In considering any given coefficient of correlation, Guilford (19:145) states that the strength of relationship can be described roughly as follows for various coefficients of correlation:

Less than 0.20 . . .	Slight; almost negligible relationship
0.20-0.40 . . .	Low correlation; definite but small relationship
0.40-0.70 . . .	Moderate correlation; substantial relationship
0.70-0.90 . . .	High correlation; marked relationship
0.90-1.00 . . .	Very high correlation; very dependable relationship

Referring again to hypothesis two, it was determined that the intercorrelations used in computing the significance of the difference between the nonreading science--reading correlation and the reading science--reading correlation were highly significant. The significance of the correlation coefficients used in the calculation involved in hypothesis two was summarized in Table V.

Secondary Findings

Perhaps of equal interest and importance were those secondary findings which were based on a limited number of the total sample. The question arose as to the influence of intelligence as it related to performance on the non-reading science test, the reading science test, and the reading test. Although the sample included only thirty-five sixth-graders, the findings were considered important enough to include within this section of the study.

Garrett and Woodworth (16:403) state that the correlation between two variables is often misleading and may be erroneous if there is little or no correlation between the variables other than that brought about by their common dependence upon a third variable. Partial correlation deals with the residual relationship between two variables where the common influence of one or more variables has been removed. In this part of the study, it was decided to statistically remove or partial out the influence of intelligence, in order to obtain a measure of correlation with the effect of intelligence removed.

TABLE V

SIGNIFICANCE OF THE CORRELATION COEFFICIENTS INVOLVED IN THE COMPUTATION PERTAINING TO HYPOTHESIS TWO

Correlations Involved	Correlation Coefficient	*t* ratio
Nonreading science--reading science	0.45	8.356
Nonreading science--reading	0.43	7.898
Reading science--reading	0.64	13.813

To be significant at the .05 level for N - 2 or 275 degrees of freedom, a *t* ratio of 1.968 would have to be reached. Since the *t* ratios as listed above are 8.356, 7.898, and 13.813, each correlation coefficient is significant beyond the .05 level.

From the results of the several correlations and the partial correlation summarized in Table VI, it can be seen that two correlation coefficients lack significance. For this limited sample, although there appeared to be a positive but low correlation between the nonreading science test and the reading test, this correlation was not quite significant at the .05 level. On this basis, it could be assumed that the relationship between nonreading science achievement and reading achievement for this limited group could have resulted by chance alone, or that the relationship between the two variables was in fact nonexistent.

On the other hand, the partial correlation between nonreading science, reading, and intelligence was also below the .05 level of significance. The partial correlation coefficient, with intelligence removed, was only 0.11, which indicated an almost negligible overlapping between nonreading science achievement and reading achievement. Another way of stating this finding would be to state that 89 per cent of the relationship between nonreading science achievement and reading achievement resulted from the influence of intelligence, while the remaining 11 per cent of the total relationship resulted from the influence of other factors, possibly reading ability. The degree of this relationship was borne out by the significance of the partial correlation coefficient, which was practically zero. This result could possibly be explained due to the

TABLE VI

CORRELATIONS BETWEEN NONREADING SCIENCE, READING SCIENCE, AND READING TESTS

Tests	Correlation Coefficient	*t* ratio
Nonreading science--intelligence	0.39	2.433
Nonreading science--reading	0.33	2.008
Reading--intelligence	0.63	4.660
Reading science--intelligence	0.65	4.914
Reading science--reading	0.70	8.044
Nonreading science--reading--intelligence (intelligence removed)	0.11	0.636
Reading science--reading--intelligence (intelligence removed)	0.49	3.229

To be significant at the .05 level for N - 2 or 33 degrees of freedom, a *t* ratio of 2.035 would have to be reached. All correlation coefficients are significant beyond the .05 level with the exception of nonreading science--reading and the partial correlation nonreading science--reading--intelligence (intelligence removed).

heavy reliance of intelligence testing upon verbal ability and reading ability. This seems quite reasonable since the intelligence scores used in this study were based on the Lorge-Thorndike Verbal Battery, which is a group test requiring each student to be able to read the appropriate questions. It is also quite reasonable to assume that the high degree of the influence of intelligence on the total relationship should be expected. Not only should an intelligent student be able to read adequately, he should also be able to understand learned concepts when presented in pictorial form. An intelligent student who does not read adequately for whatever reason, may still have learned the particular concepts and be able to score well on a non-reading science test but not score well on a reading science test. This seemed to be the case in this study.

In considering the correlation between reading and intelligence, between reading science and intelligence, and between reading science and reading, it was found that these relationships were quite high. According to Guilford's (19:145) categories of correlation coefficients, all of these correlations showed a "substantial relationship." Again, considering the influence of reading in the reading test, the reading science test, and the intelligence test, this high degree of relationship was expected.

In addition, when considering the partial correlation between the above three variables, a coefficient of

0.49 was found after the removal of intelligence. This was considered a substantial relationship (overlapping) between the reading test and the reading science test. This partial correlation coefficient was found to be highly significant, actually beyond the .01 level of significance. Again, this relationship would seem to be obvious and expected. This relationship could be considered to be beyond the realm of chance.

The above findings were considered to provide further support for the acceptance of the second hypothesis. This hypothesis, which stated that there would be a significant difference between the correlations of scores on a reading science test and a reading test, and a nonreading science test and a reading test, was realized. Since all of the reading measures, including intelligence, were found to be so highly correlated, and since the nonreading science test was found to be correlated to a far lesser degree with the reading measures, it was again inferred that the nonreading science test was in fact <u>not</u> measuring the same achievement as the reading science test was assumed to measure. It appeared rather conclusively, after considering all the data, that the reading science test was in fact measuring reading achievement and not science achievement.

Summary of Findings

By applying the Dale-Chall Readability Formula to the Stanford Science Test, Intermediate II level, which was the reading science test used in this study, it was found that the average reading difficulty was beyond the expected reading level for sixth-grade students. Also, it was determined that this science test was graded in difficulty with the last part of the test far beyond the reading expectancy level for sixth-grade students. On the basis of the average reading difficulty level, which was found unsuited to sixth-grade reading achievement, the first hypothesis was rejected. This hypothesis stated that the Stanford Science Test would not differ appreciably in reading difficulty from the level of reading expectancy for sixth-grade students.

By applying the correlational-relationship design to the second hypothesis, it was found that the reading science test was much more highly related to reading achievement than to science achievement as measured by a nonreading science test. It was found that the nonreading science test correlated to approximately the same degree with the reading test as it did with the reading science test, which was considerably lower than the relationship between the reading test and the reading science test.

It was found from the above correlations that a significant difference (beyond the .01 level) existed between the correlation coefficients for the nonreading science--reading and reading science--reading tests. The above findings indicated that the nonreading science test and the reading science test were actually measuring different areas of achievement. On the basis of these findings, the second hypothesis was accepted. This hypothesis stated that there would be a significant difference between the correlations of scores on a reading science test and a reading test, and a nonreading science test and a reading test.

In addition, secondary findings showed that the nonreading science test correlated to a low degree with intelligence, while the reading science test correlated to a much higher degree with intelligence. The reading science test and the reading test each correlated to approximately the same high degree with intelligence. When the influence of intelligence was partialled out of the relationship reading--nonreading science--intelligence, the relationship between reading and nonreading science achievement was no greater than could be reasonably expected from chance factors alone. When intelligence was removed from the relationship reading--reading science--intelligence, the resulting relationship between reading and reading science was substantial, beyond the .01 level of significance.

These findings contributed additional support to the acceptance of the second hypothesis, and it was concluded that the reading science test and the nonreading science test were in fact measuring different achievements.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

There have been numerous undocumented comments by educators and testing officers that many of the current instruments used in assessing science achievement in the elementary school are essentially tests of reading ability. Many of these comments have been directed to the Stanford Science Test, which is included in the Stanford Achievement Battery. It is a reading science test commonly used in the assessing of achievement in the elementary schools. Therefore, it was partially the purpose of this study to investigate the reading aspect of this science test. In order to accomplish this, it was necessary to design a nonreading science test with certain test items paralleling items in the Stanford Science Test. It was also decided to direct the entire study to the sixth-grade level.

After the development of the nonreading test, it was necessary to complete a pilot study to determine the difficulty level and discriminating ability of each of the items included in the test. This was done to insure the selection of the best items for the final testing of the selected sample of sixth-grade students. The findings of this study were based on the testing of this final sample, as well as utilizing the Dale-Chall Readability Formula to determine the level of reading necessary to perform satisfactorily on the Stanford Science Test. The findings were as follows:

1. The first hypothesis stated was: "The Science Test of the Stanford Achievement Battery, Intermediate II level, 1964 revision, will not differ appreciably in reading difficulty from the level of reading expectancy for sixth-grade students." By applying the Dale-Chall Readability Formula, it was found that the average reading level of this science test was beyond that of the average reading expectancy of sixth-grade students. This hypothesis was rejected.
2. The second hypothesis stated was: "There will be a significant difference between the correlations of scores on a reading science test and a reading test, and a nonreading science test and a reading test when applied to the same selected group." By applying correlational techniques to the relationship between nonreading science achievement, reading science achievement, and reading achievement, it was found that there was a significant difference between the relationship of reading science--reading and nonreading science--reading.

The difference between these two relationships was found to be highly significant, beyond the .01 level. This hypothesis was accepted.

Secondary findings indicated that intelligence was much more highly related to reading science achievement and reading achievement than it was to nonreading science achievement. When intelligence was removed from the relationship of nonreading science--reading--intelligence, the resulting relationship was very insignificant--no more than could be expected by chance factors alone.

Conclusions

Relative to the first hypothesis, it must be concluded that the average reading difficulty of the Stanford Science Test was above the reading comprehension ability of those sixth-grade students who were reading at grade-level expectations. Due to the graded nature of the test, which increases in reading difficulty with progression of items, the last portion of the test was far beyond the reading comprehension ability of sixth-grade students. On this basis, it appeared that reading achievement was a greater factor in determining success on this science test than a knowledge of the science concepts being assessed. It would appear from this conclusion that the validity of the Stanford Science Test is questionable as a measure of science achievement for those sixth-grade students reading up to grade-level expectations and more questionable for those students reading below grade-level expectations. This test seems to measure little more than reading ability.

Dale and Chall (7:19-20) have stated that no reading material should be more difficult than it need be, and materials with a high readability score should be simplified by substituting more concrete, familiar words for the unfamiliar and abstract. On the other hand, where some reading material is hard to understand because the ideas are hard and complicated, it may be difficult or impossible to simplify this type of writing. This appears to be especially true of the Stanford Science Test. However, if the reading material is too difficult to be read and if it is impossible to simplify the material, then perhaps the solution is a nonreading science test such as that developed in this study.

These authors also stated that the reader's purpose in reading, his interest in the subject, and his background in the subject-matter must be considered when determining reading difficulty. Although the Dale-Chall Readability Formula may indicate a given level of difficulty, lack of interest, motivation, and background may increase the actual level of difficulty by a considerable amount. Apparently the subject-matter background of these students was sufficient to the task on the nonreading science test, but seemed to be lacking when reading was required. Perhaps the lack was only in reading ability and not in subject-matter background, interest, or motivation.

Concerning the second hypothesis, it must be concluded that there was a significant difference between the relationship between a reading science test and a nonreading science test as each correlated with a measure of reading achievement. Due to the high correlation found between the Stanford Science Test and the Stanford Reading Test, and the low correlation between the Stanford Reading Test and the nonreading science test, it was concluded that each science test was measuring something different. In the case of the Stanford Science Test, it appeared that this test was measuring reading ability to a greater extent than it was measuring science achievement. Since this appeared to be the case, then it would seem that the Stanford Science Test is not an adequate instrument for assessing those outcomes of science instruction which have been deemed important to science achievement at the sixth-grade level.

Since the science items of the nonreading science test were considered adequately parallel to those of the Stanford Science Test, and since the nonreading science test did not require reading ability for test performance, it also appeared that the nonreading science test was more adequate to the task of assessing science achievement at the sixth-grade level. This could be an important consideration for the future development of instruments which would be more valid for those students lacking a high degree of reading ability.

Based on the findings of this study, it was likewise concluded that reading ability is an essential factor in the assessing of science achievement whenever the Stanford Science Test is used to assess this achievement.

Finally, it was concluded that the nonreading science test was better suited to the evaluation of science achievement as it did not penalize the student lacking a high degree of reading proficiency.

Recommendations

Since little work has been done in the area of developing a nonreading science test, the recommendations contained herein cannot be supported by prior studies, but must be, of necessity, based on the findings of this study and common sense considerations. The following recommendations seemed appropriate:

1. It is recommended that the schools re-evaluate the use of the Stanford Science Test as a measure of science achievement, especially in those cases where reading ability of the student is below grade level. From the findings of this study, it appeared that many students had in fact learned

much about the science concepts assumed to be covered in the sixth grade, but that they were unable to adequately demonstrate their knowledge on the basis of a reading science test.

2. It is further recommended that those teachers and principals who place great faith in standardized tests, such as the Stanford Science Test, should seriously question the results before concluding that their students lack the knowledge of science which is to be learned at the sixth-grade level.
3. Also, it is recommended that those students who read poorly be encouraged and given the opportunity to attain science learnings through the development of listening, observing, and manipulating skills.
4. Finally, it is recommended that those students with poor reading ability be given a more individualized program of science instruction so that each student may realize the greatest measure of success possible.

Indications of Further Studies

Based on the final item analysis of the nonreading science test and the findings of this study, the following suggestions are recommended for further exploration:

1. Since the final item analysis indicated that many of the nonreading items were actually too easy for

the majority of the sixth-grade students, it would be of value to administer this same nonreading science test to fifth-grade students and possibly fourth-grade students in an attempt to determine to what extent these science learnings have been attained at the respective grade levels.

The importance of this type of study could result in the determination that many of the science learnings which are supposed to be taught in the sixth-grade may have in fact already been learned through some other means not associated with the classroom. If this proved to be the case, then perhaps the current courses of study and the currently used textbooks would need revising in order to "catch up" with the knowledge of the students.

2. Since the nonreading science test seemed to be possibly more attuned to a laboratory-type curriculum, it would be of interest to administer this test to some of those schools which profess to be more laboratory oriented. These results could then be compared with those obtained from a school which is more textbook oriented in order to determine any difference in achievement.

3. Considering the fact that the sample included in this study appeared to be quite homogeneous, as far

as the nonreading science test indicated, it would be of great value to compare two groups which were very definitely heterogeneous in science ability. However, this type of study might be impossible.

4. Since a child gains understanding from the way he perceives a picture or object, it could be of value to attempt a study relating achievement on the non-reading science test to perceptual ability. This type of study might indicate that the student who reads poorly actually is superior at perceiving those events which do not require a high degree of reading proficiency, and those students who read very well may have developed what is sometimes referred to as a "reading set."

This reading set may be caused by an over-emphasis on learning through the written word. A predisposition toward this type of set may begin in the early years when a child is made aware of the importance of the written word. Parents and teachers who appreciate only the worth of learning from the written word may possibly discourage learning through other approaches. This child may become rigid in his attitude toward the importance of other media and resistant to those things he perceives which are not expressed as written words. When a child places undue emphasis on the importance of reading words as a way of learning, it is possible that he may become unable to perceive pictures and objects as having any real importance. On this basis, he may fail to appreciate or to learn from what he perceives.

Now that the trend seems to be toward greater use of all of the senses in the learning process, a child who depends exclusively upon the written word for meaning and experience is handicapped. This child could also be at a disadvantage in a laboratory setting where the use of a textbook is secondary to the direct experience. To this child, therefore, pictures and oral instructions, and a nonreading test of any type may have little meaning.

APPENDIX A

THE DALE-CHALL READABILITY FORMULA

1.	Number of words in the sample	________
2.	Number of sentences in the sample	________
3.	Number of words not on Dale list	________
4.	Average sentence length (1 ÷ 2)	________
5.	Dale score (3 ÷ 1 x 100)	________
6.	Average sentence length (4) x .0496	________
7.	Dale score (5) x .1579	________
8.	Constant	3.6365
9.	Formula raw score (6 + 7 + 8)	________
	Average corrected grade-level	________

APPENDIX B

NONREADING SCIENCE TEST ORAL INSTRUCTIONS, ORAL QUESTIONS, AND TEST ITEMS

Oral Instructions

In the following items, cross out with an X the letter which you believe best represents the correct answer:

Oral Questions

1. In this picture we have three candles under glass jars. Which one of the candles is most likely to go out first?

2. In this picture we have three different containers. If we put an equal amount of water into each container and place all three outside in the sun, which one will lose the most water?

3. In this picture we see three different types of science equipment. Which one is a terrarium?

4. In this picture we see a boy holding a toy balloon. If he loses the balloon and it goes high up over the mountains, which one of the following pictures best shows what will happen to the actual size of the balloon?

5. In this picture we have a copper wire hanging between two poles. Which one of the following pictures best shows what will happen to this wire if we heat it?

6. In this picture we have a nut which is screwed down very tightly. If all three wrenches fit the nut, which one would be best to use to loosen it?

7. In this picture we have a rabbit. Which one of the following pictures best shows where this rabbit would live?

8. In this picture we see two solid lead balls. If we drop these balls at exactly the same time, which one of the following pictures best shows how these balls will hit the floor?

9. Here is a picture of a toy boat floating in a tank of pure water. If we add salt to this water, which of the following pictures best shows how the boat would float in the salt water?

10. In this picture we see a telescope. In order to get the best possible view of the stars, which picture shows the best place to put the telescope?

11. Which one of these pictures best shows where the richest soil may be found?

12. In this picture, we see the foot of a bird. Which one of the following pictures best shows where this bird would live?

13. Here is a picture of a sliced potato. If we put a drop of iodine on the slice, it will most likely change to which color?

14. In this picture we see an electromagnet. In which one of the following pictures would the electromagnet be the strongest?

15. Here is a picture of an eye showing how light is focused into it. Which one of the following pictures best shows a near-sighted eye?

16. In this picture, which type of food would contain the <u>most</u> calories?

17. Which one of these pictures best shows the life cycle or life story of a plant?

18. In this picture we see a piece of lemon and a piece of blue litmus paper. If we squeeze a few drops of the lemon onto the litmus paper, which one of the following pictures best shows what will happen to the litmus paper?

19. In this picture we see a jar of jam with its lid screwed down very tightly. Which one of the following pictures shows the safest way to loosen the lid?

20. In this picture we have a high hill. At which place would it most likely be coldest on a windy night?

21. In this picture we see a globe representing the earth, and a flashlight representing the sun. Which of the following pictures best shows the winter season in the United States?

22. In this picture we have three pieces of the same kind of wire stretched between posts. If we pluck each wire exactly the same, which one will have the lowest pitch?

23. In this picture we see a bare hillside. In order to keep the soil from running off, we might decide to plant some grass. Which one of the following pictures shows the best way to plant the grass?

24. In this picture we see three jackets, all made of the same kind of material. Which jacket would be best to wear on a cold, sunny day?

25. Here is a picture of a bar magnet. Which one of the following pictures best shows where the magnet is strongest in attracting a tack?

26. In this picture we have two balloons, exactly alike, filled with air and balanced on a lever. Which one of the following pictures best shows what would happen if we let the air out of one balloon?

27. In this picture we are going to boil water on a camp stove. In the following picture are three camp sites on a high mountain. At which camp site would the water be most likely to boil quickest?

28. Which one of these pictures best shows how an iceberg would appear to float in the ocean?

29. In this picture we see someone holding a balloon closed. If he lets the balloon go, it will fly away. Which one of the following pictures involves the same principle or idea?

30. In this picture we have three automobile tires, all made exactly alike. If friction causes heat, which one of the tires would most likely be the hottest when you are driving?

31. Here are some different types of foods. In the pictures below, which group of foods would give you the best balanced diet?

32. In this picture we have a glass filled with water, and a bottle of red ink. If the water is warm and the ink cold, which one of the following pictures best shows what will happen when we put a drop of the cold ink into the warm water?

33. Here is a picture of two flashlights pointed in such a way that when they are turned on, their beams will meet. Which of the following pictures best shows what will happen to the two beams of light when they meet?

34. In this picture we have a weight hanging on a spring balance. If we put the weight into water, which one of the following pictures best shows what the <u>reading</u> on the spring balance will be?

35. In this picture we have three slopes or inclines, all exactly alike. One is a plain board, the next is painted, and the other board is covered with a piece of rug. If we roll the same ball down each slope, which one would cause the ball to roll slowest?

36. In this picture we see three thermometers hanging at different locations on the wall of a room. If the heater is on, which one of the thermometers is most likely to show the <u>highest</u> temperature?

37. In this picture we see a rock on the bottom of an aquarium. If we want to take a rod and touch the rock, which one of the following pictures best shows where we would aim the rod?

38. In this picture we have a round disc, divided into three equal parts. One part is colored blue, another red, and another is colored yellow. If we spin this disc around and around very, very fast, which of the following pictures best shows what color the spinning disc will become?

39. In this picture we have a glass of water, filled right up to the top. Which one of the following pictures best shows what will happen if we freeze this glass of water?

40. In this picture we see the head and foot of a bird. Which one of the following pictures best shows the type of food this bird likes to eat?

41. In this picture we have some water in one glass and some oil in the other glass. If we pour the water into the oil, which one of the following pictures best shows what will happen after a few minutes?

42. In this picture we see a bar graph which scientists often use to show information. In the United States, if heart disease most often causes death when compared with two other diseases, which bar <u>best</u> shows this fact?

43. In this picture, we see some sugar, a nail, and some salt. If we put each one into water, in which one of the following pictures will a chemical change result?

44. In this picture we see two fully grown corn plants of the same type or variety. If we cross these two plants, which one of these three <u>best</u> shows how tall a new plant would be after this cross?

45. In this picture, we have a very heavy box that must be lifted onto the platform. In the following pictures, which one of these simple machines would be best to use to get this heavy load onto the platform?

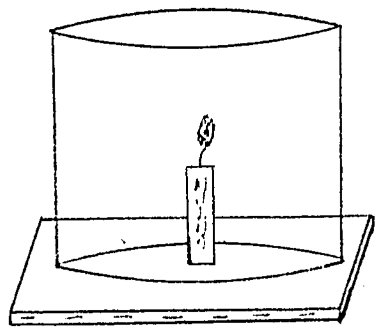
46. In this picture we have two metal bars which we are going to bang together. When we bang them together, which one of the following pictures <u>best</u> shows where the sound will travel the fastest--high on a mountain, on a flat plain, or under water?

47. In this picture we have a flashlight shining against a mirror. Which picture best shows how the light will be reflected?

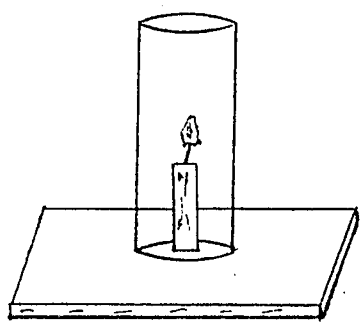
48. In this picture we have a diagram of a lever with a weight on one end. In which one of the following pictures would you have to push down the <u>hardest</u> in order to lift the weight?

1.

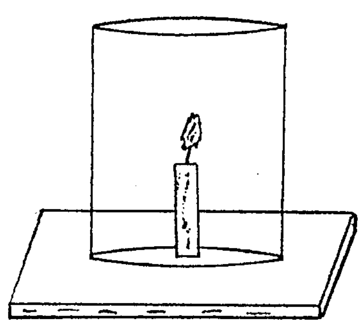
a.



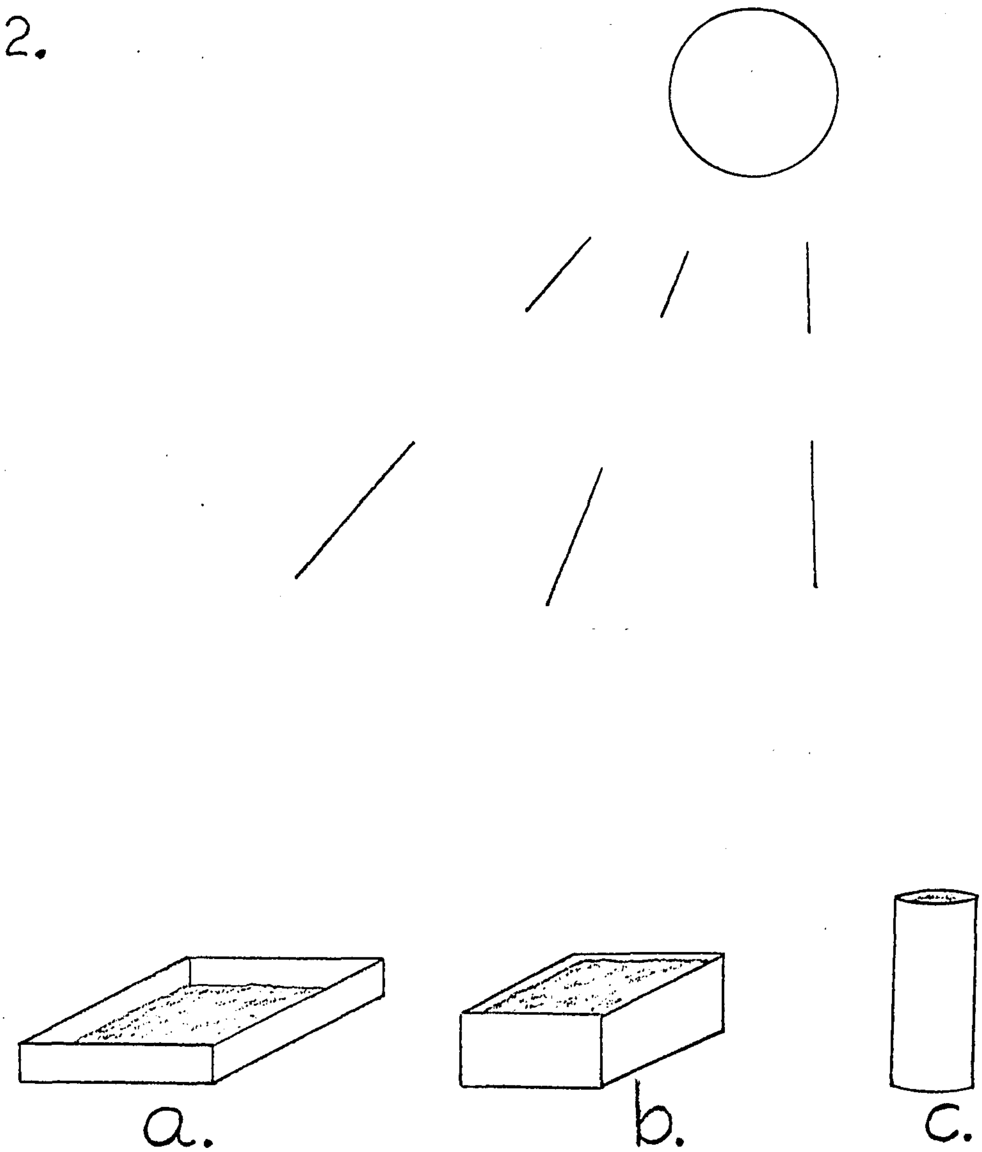
b.



c.

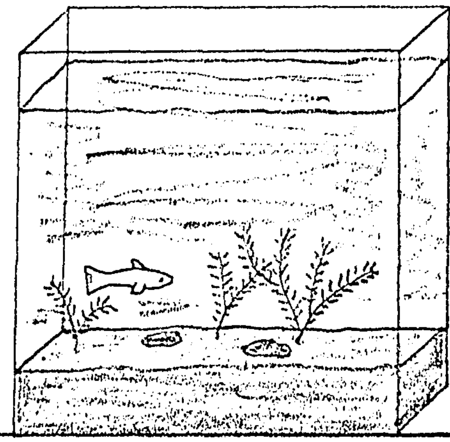


2.

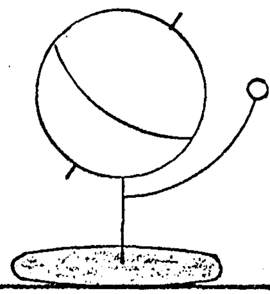


3.

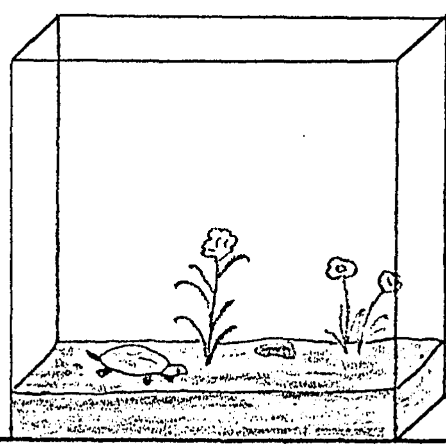
a.



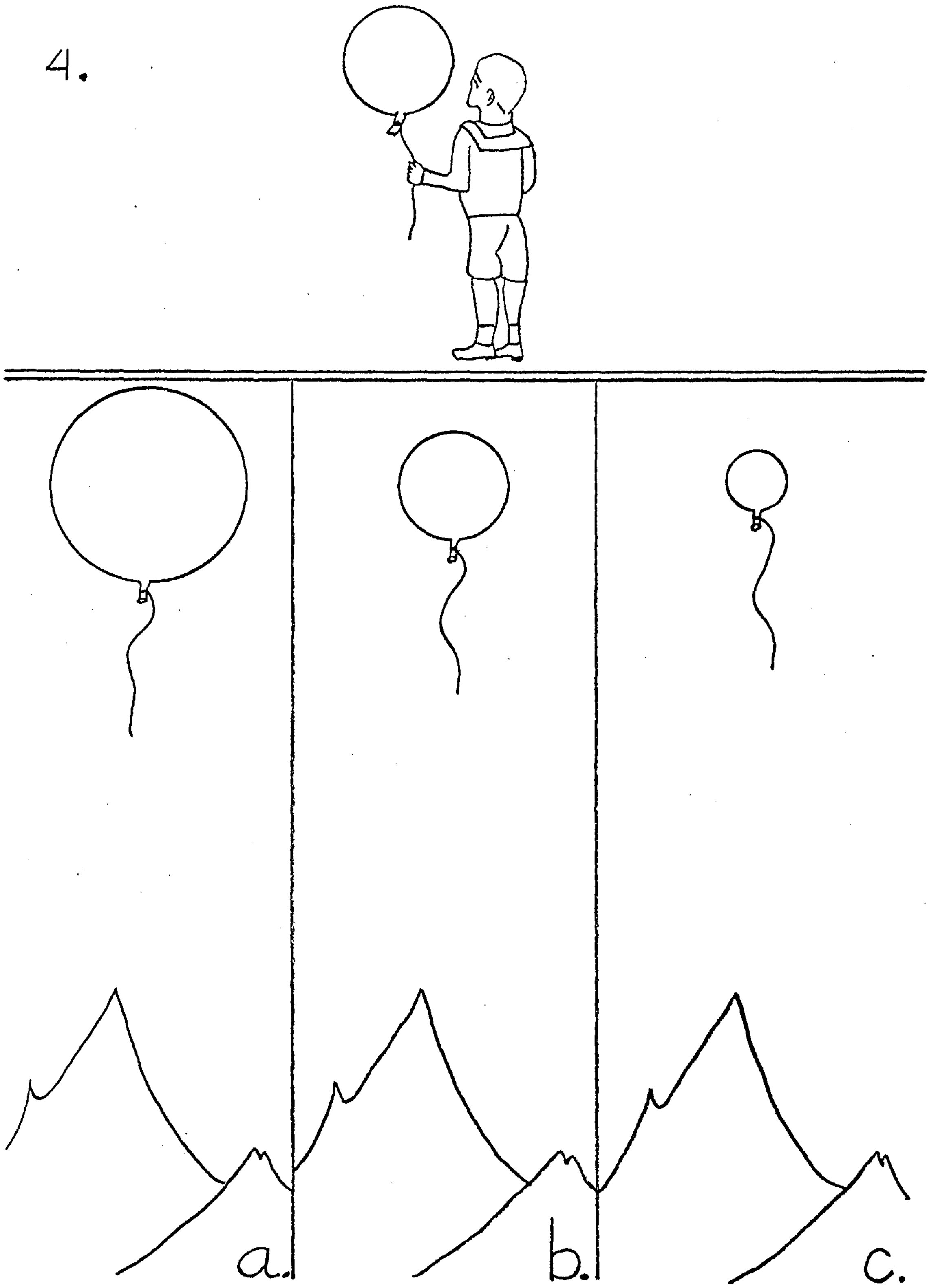
b.



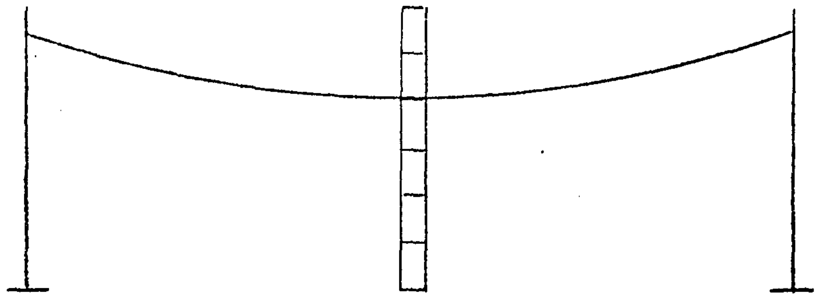
c.



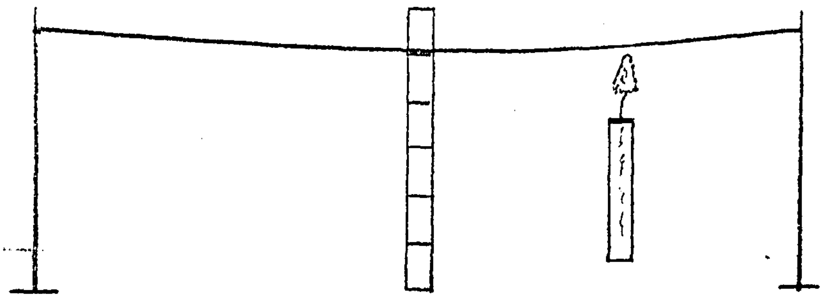
4.



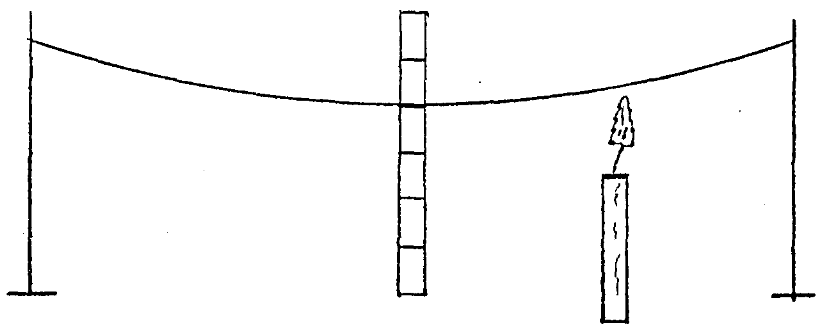
5.



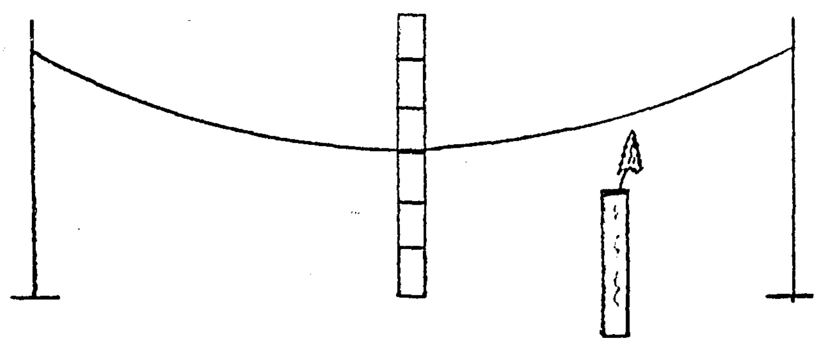
a.

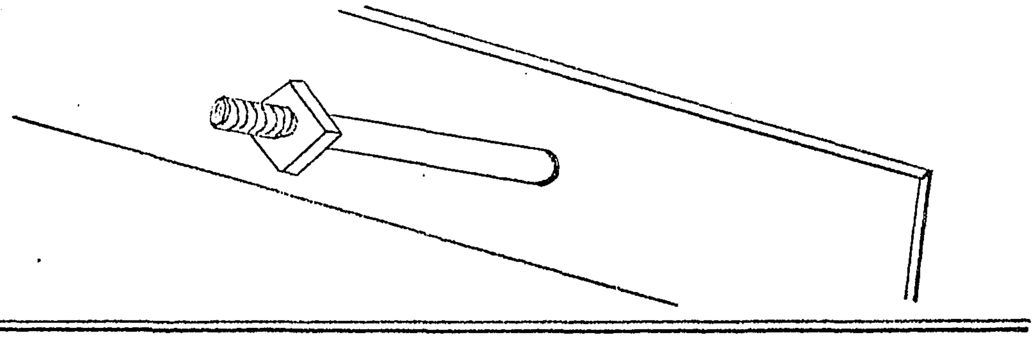


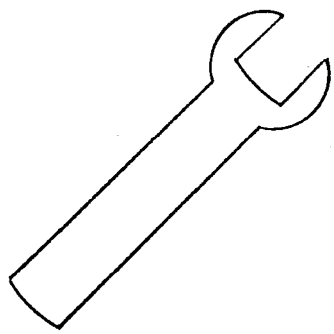
b.

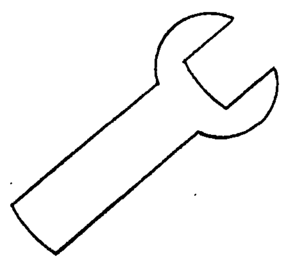


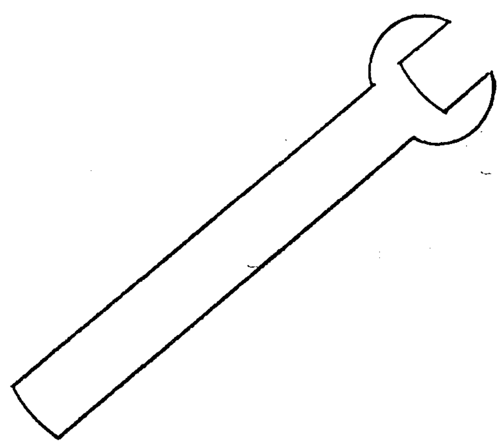
c.



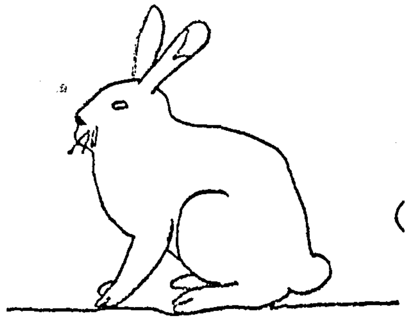
6. 

a. 

b. 

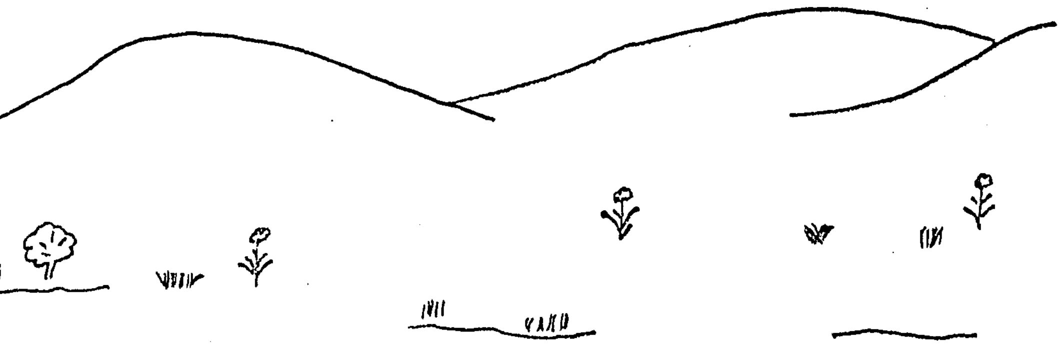
c. 

7.

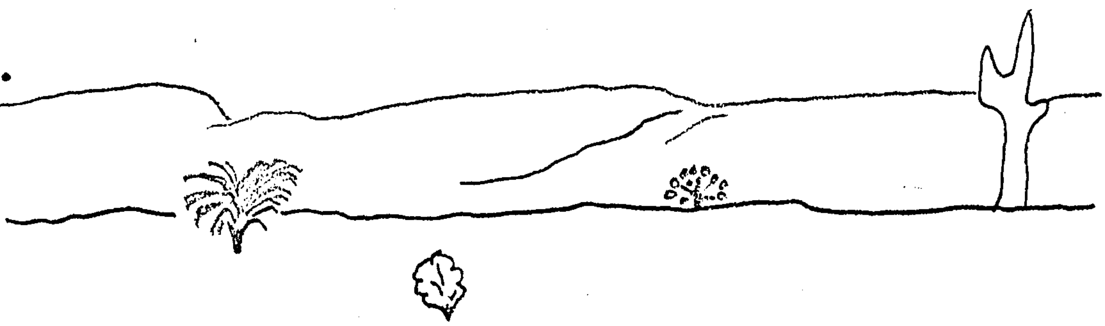


(tan rabbit)

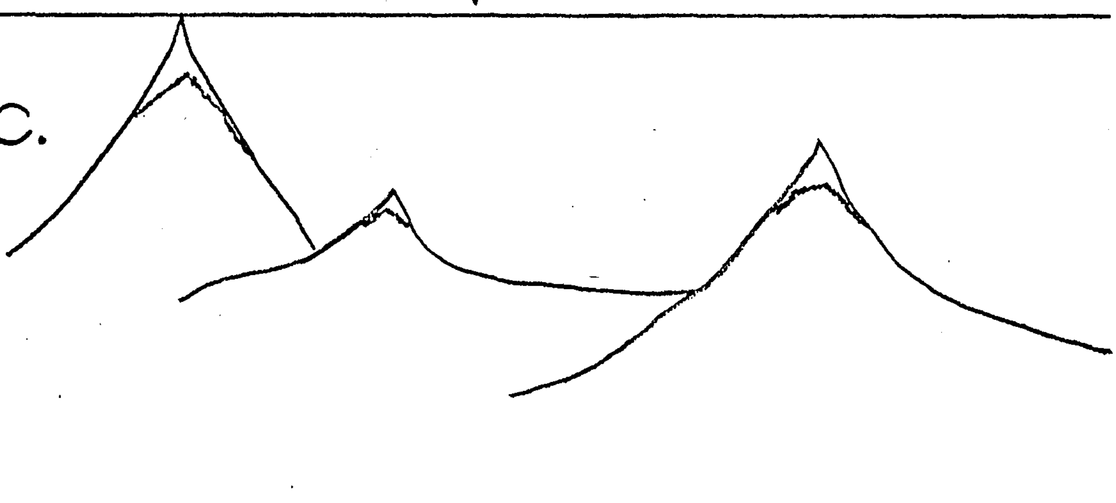
a.



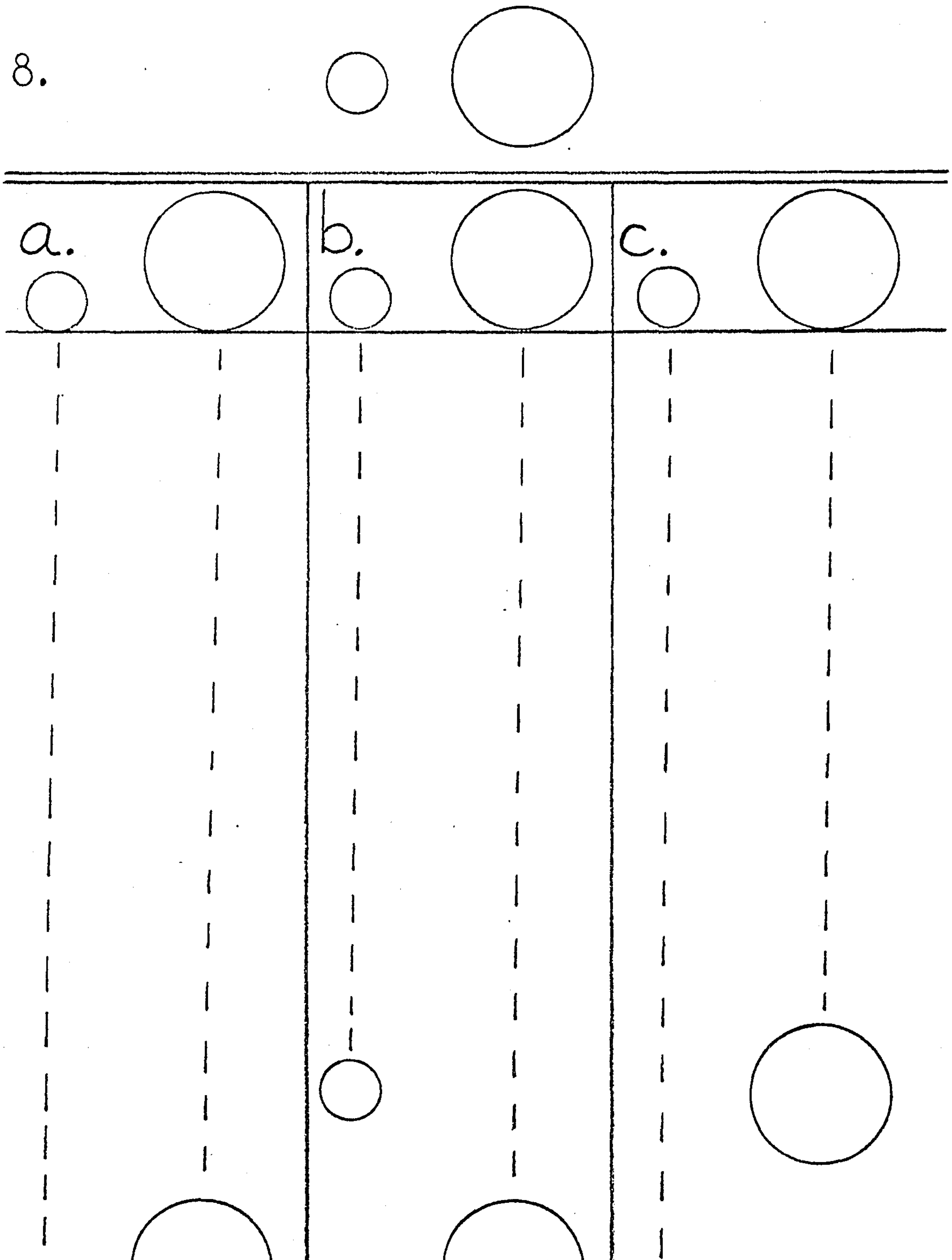
b.



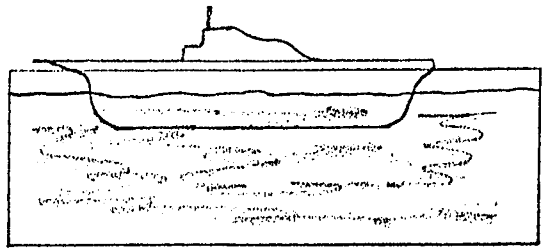
c.



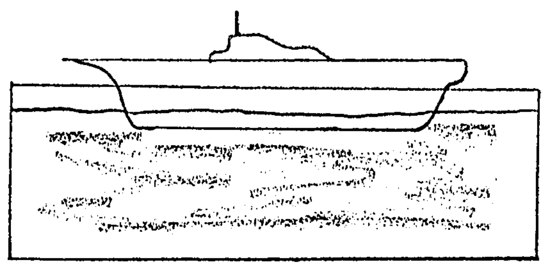
8.



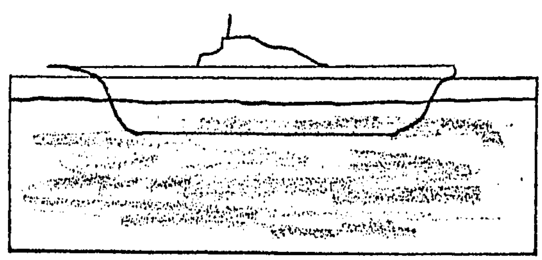
9.



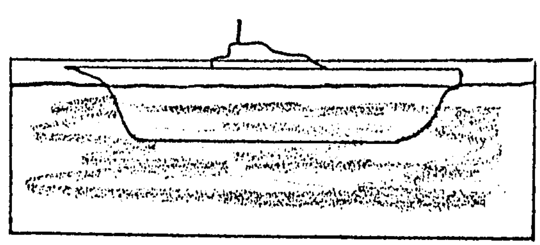
a.



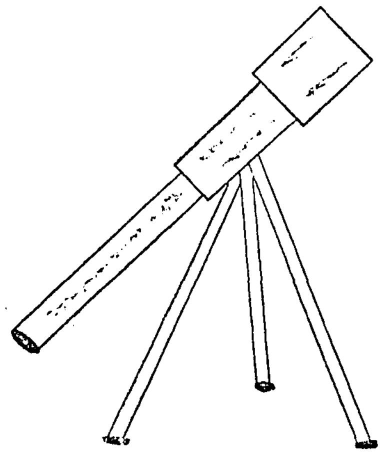
b.



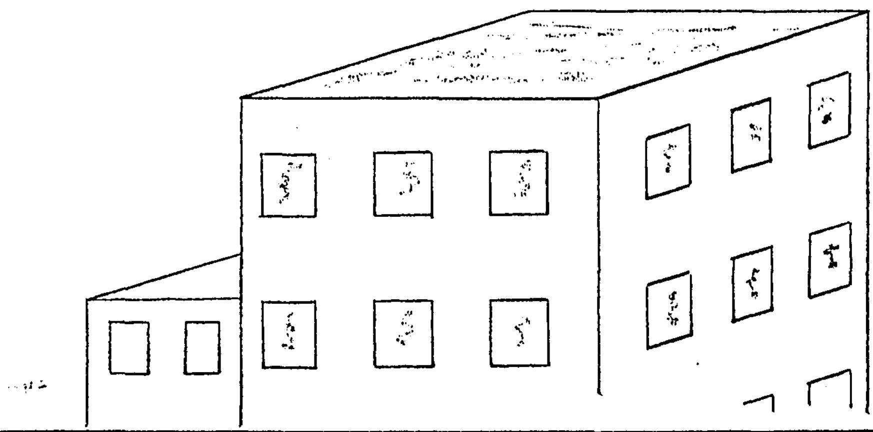
c.



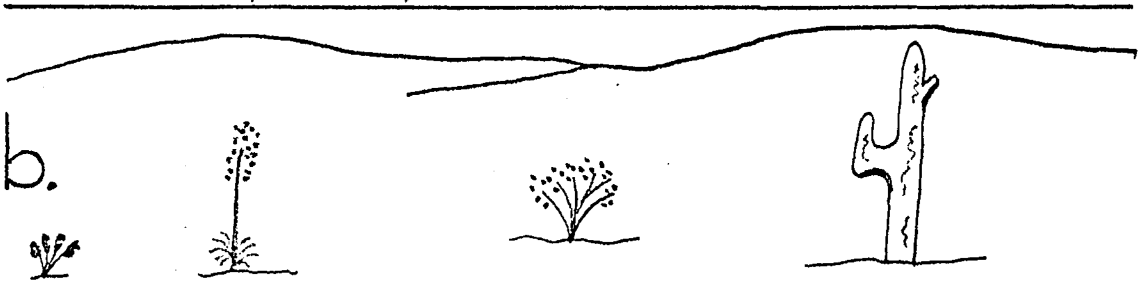
10.



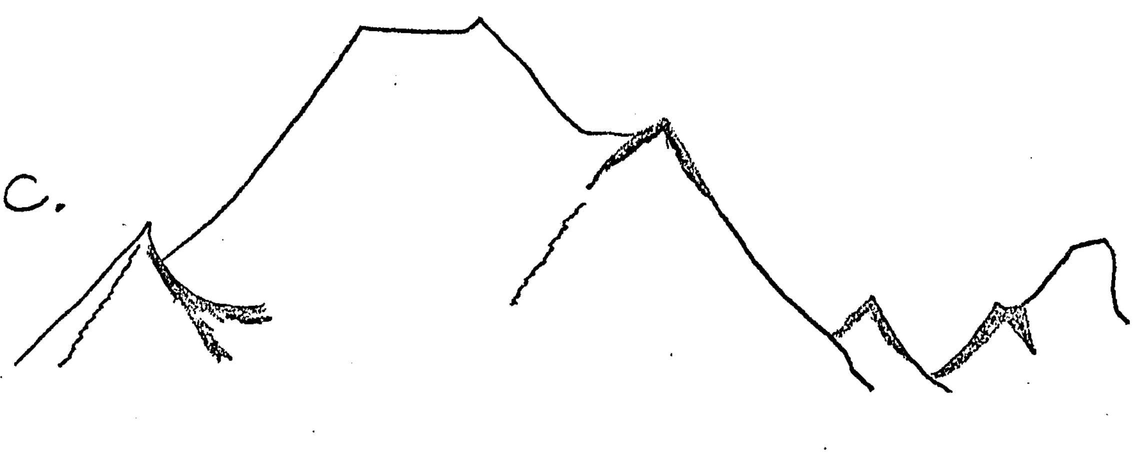
a.



b.



c.

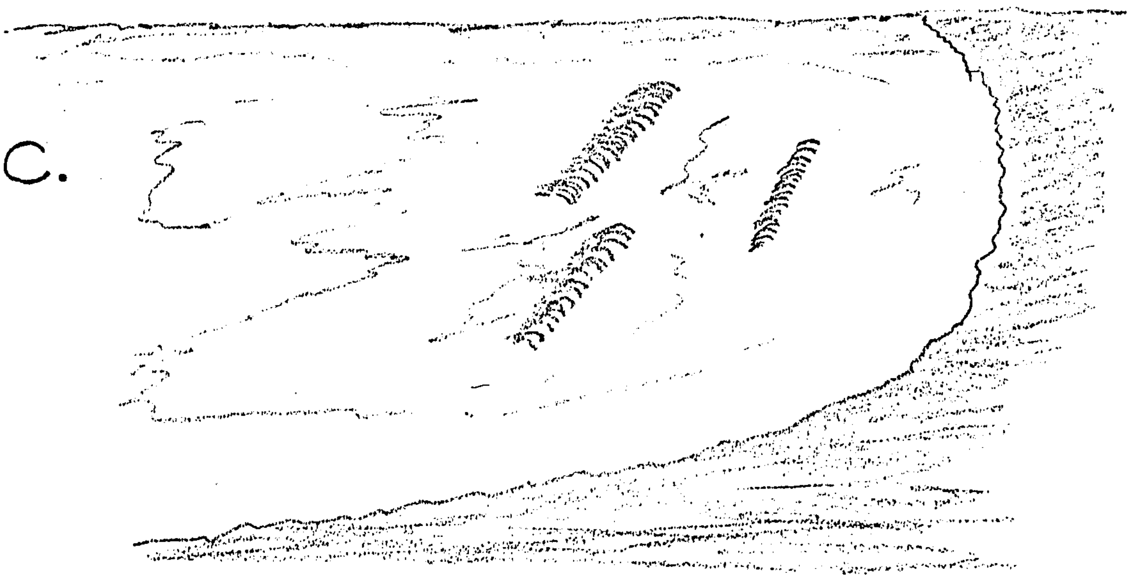


11.

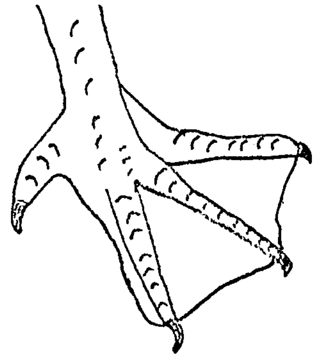
a.

b.

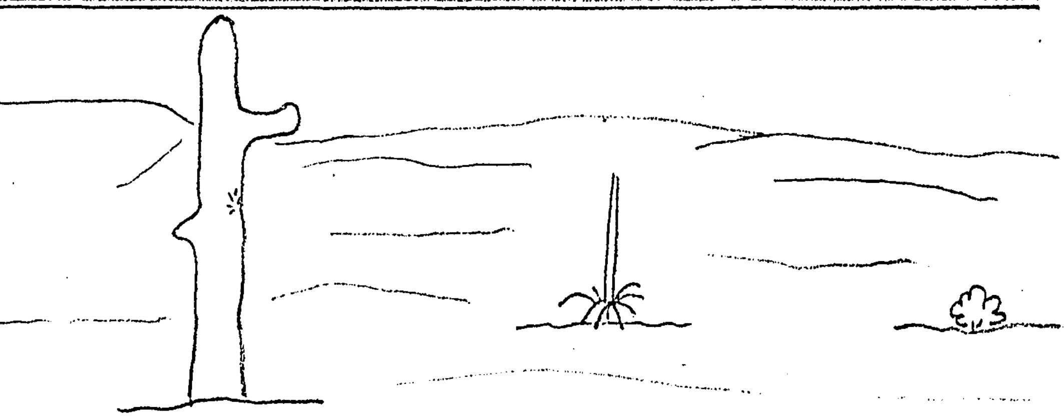
c.



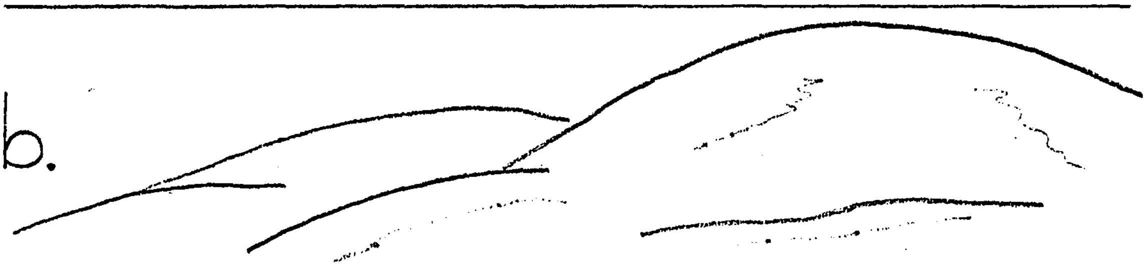
12.



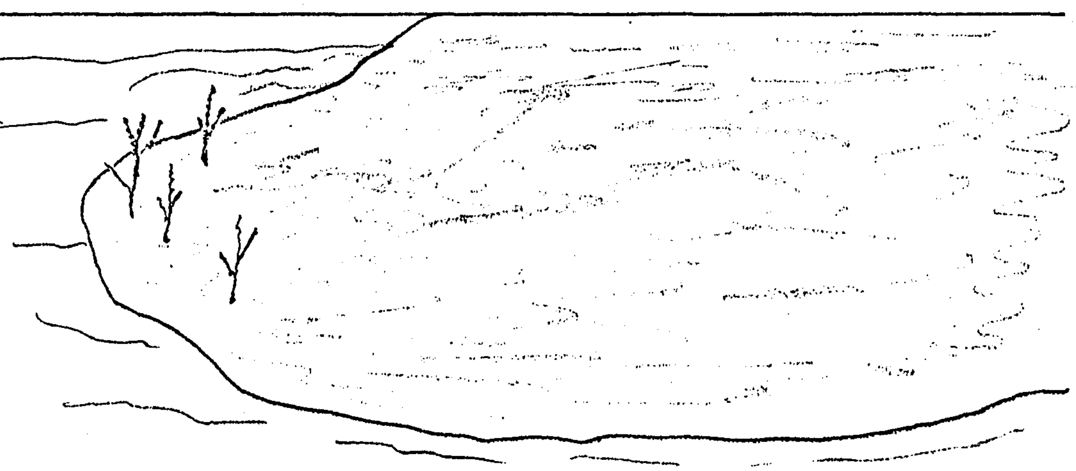
a.



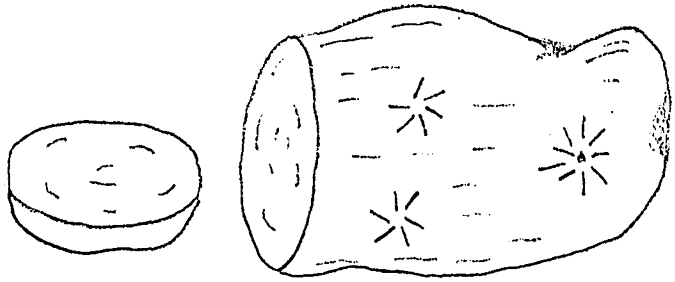
b.



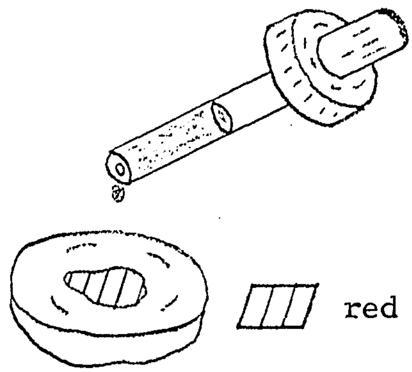
c.



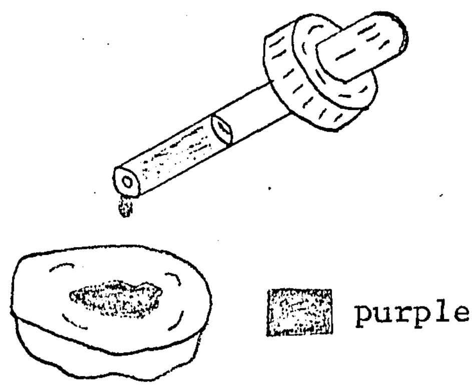
13.



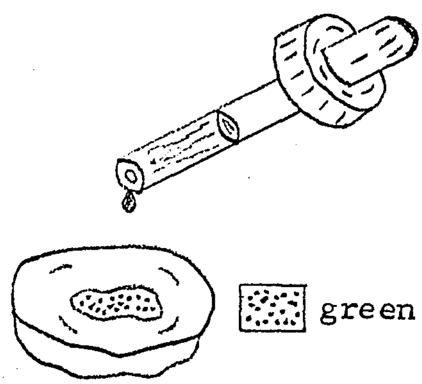
a.



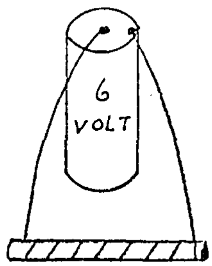
b.



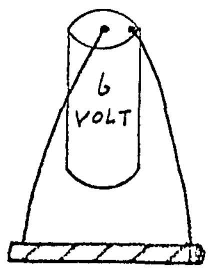
c.



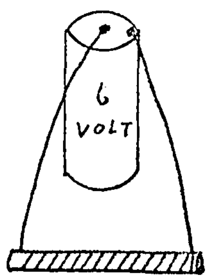
14.



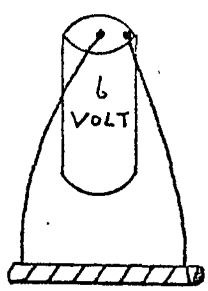
a.



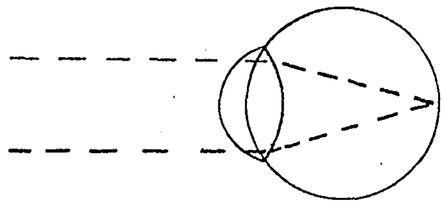
b.



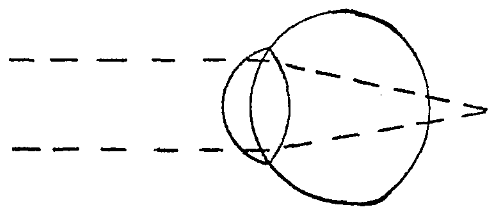
c.



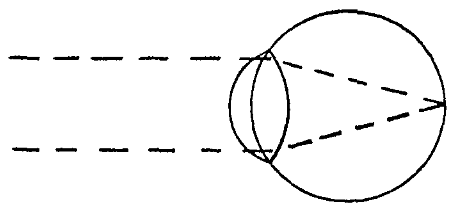
15.



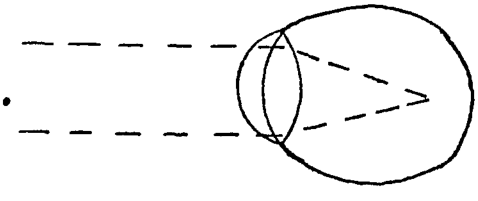
a.



b.

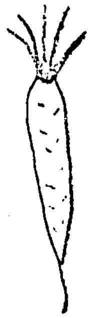


c.



16.

a.

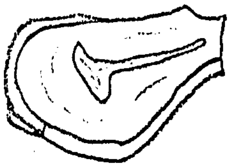


carrot

b.

potato

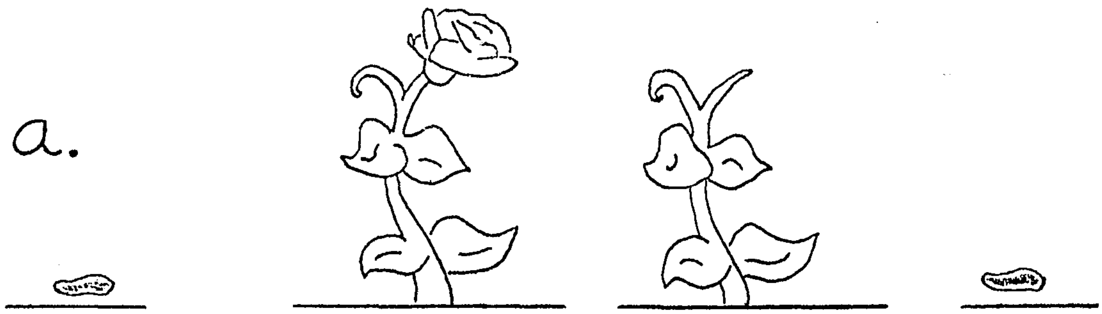
c.



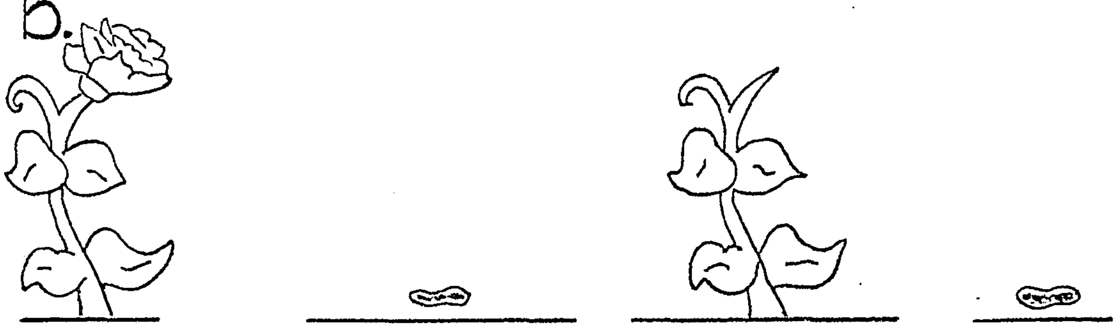
meat

17.

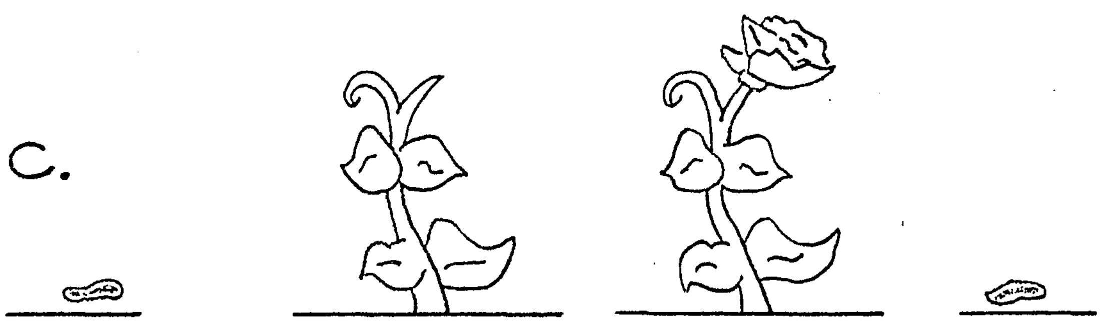
a.



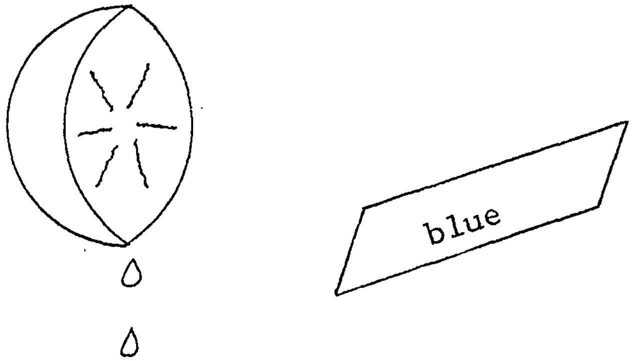
b.



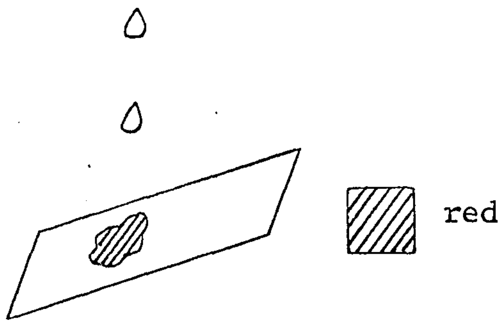
c.



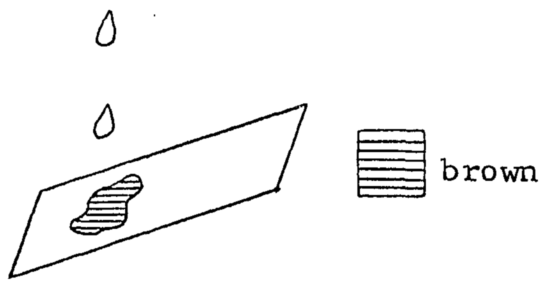
18.



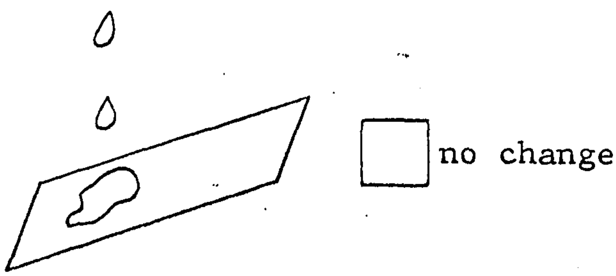
a.



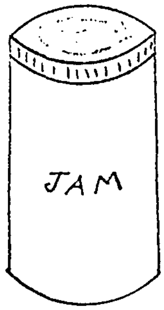
b.



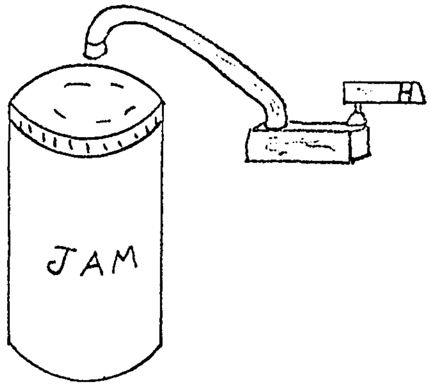
c.



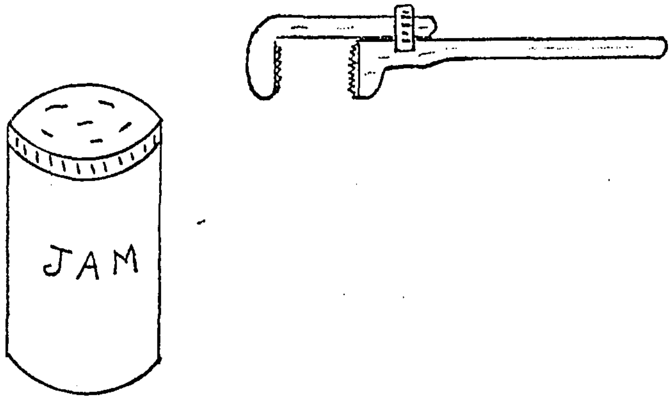
19.



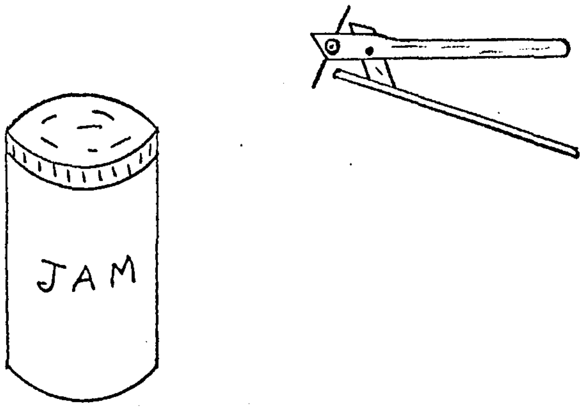
a.



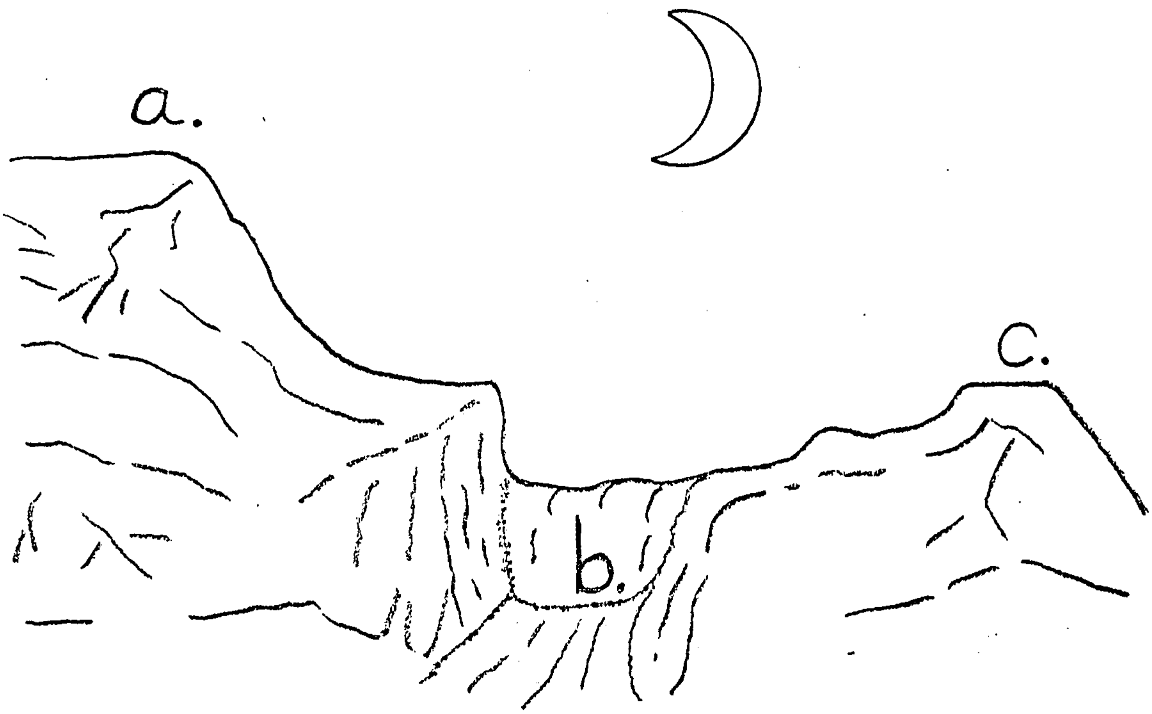
b.



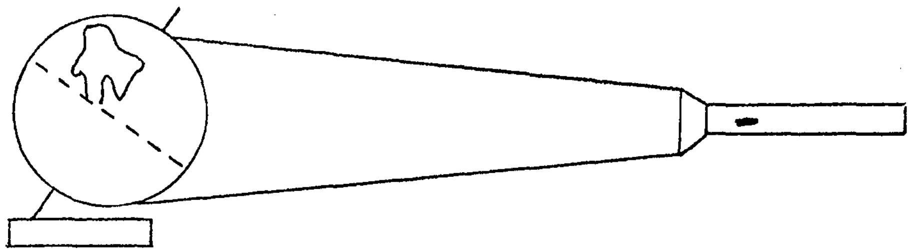
c.

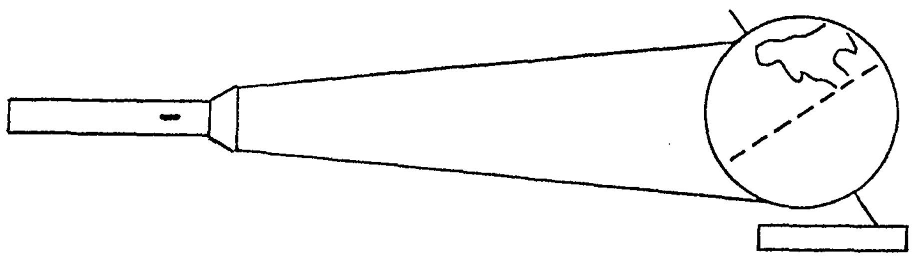


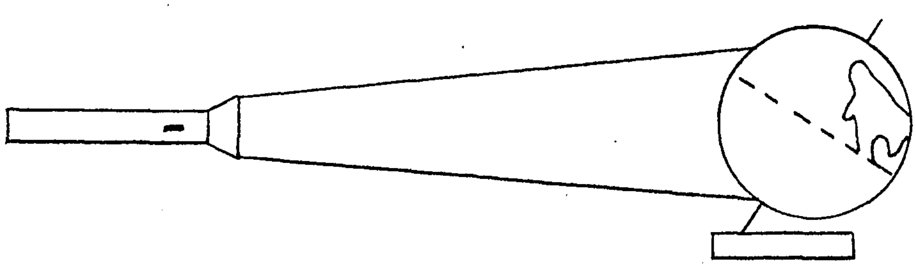
20.



21. 

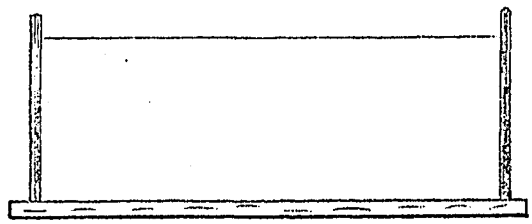
a. 

b. 

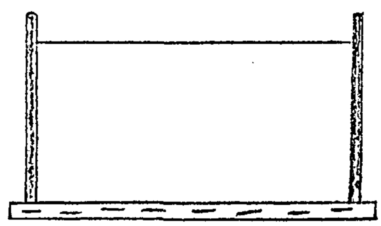
c. 

22.

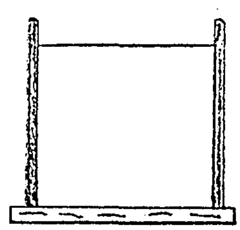
a.



b.



c.

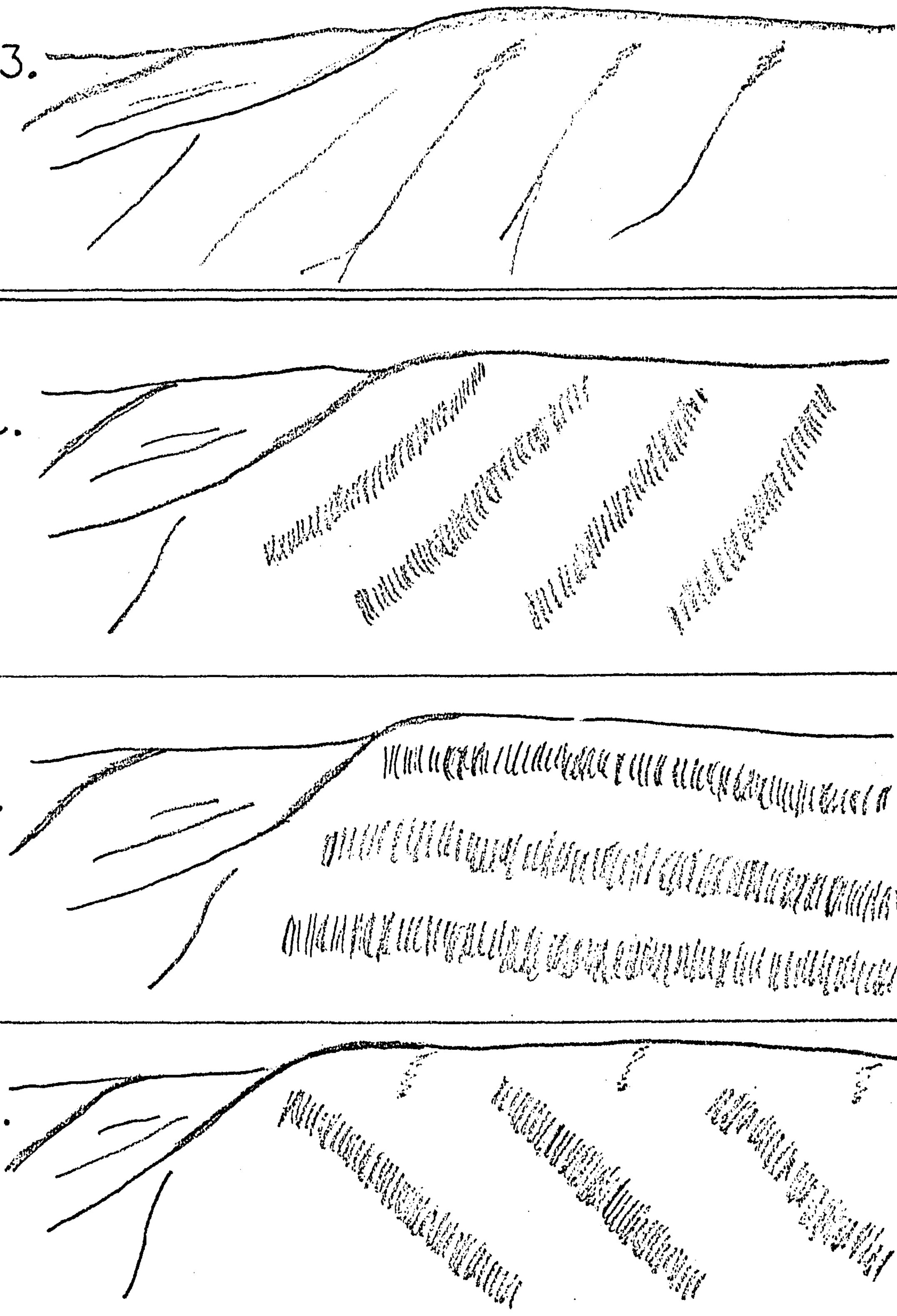


23.

a.

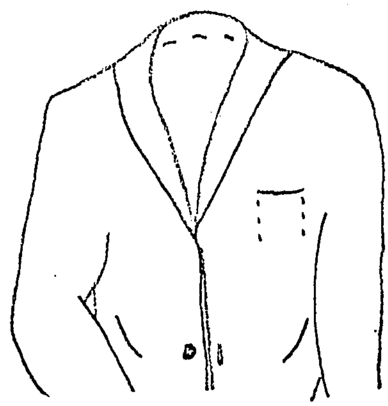
b.

c.



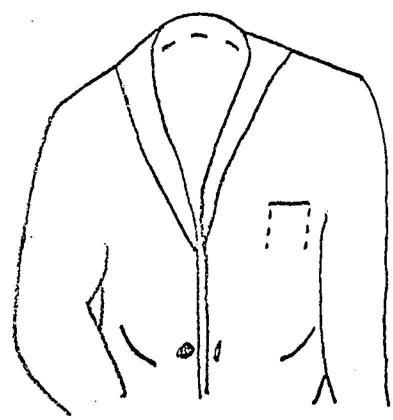
24.

a.



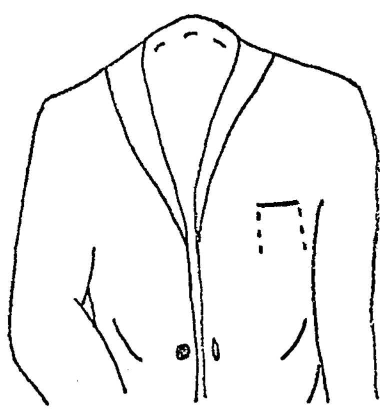
white

b.



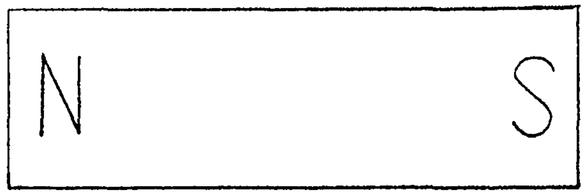
blue

c.

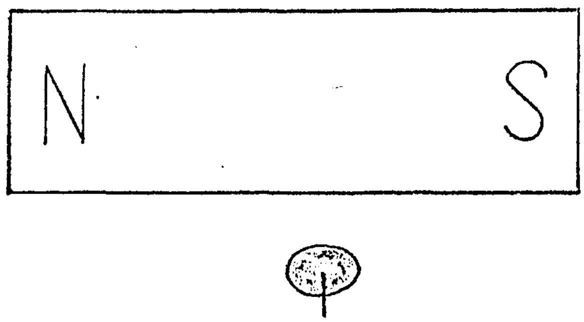


black

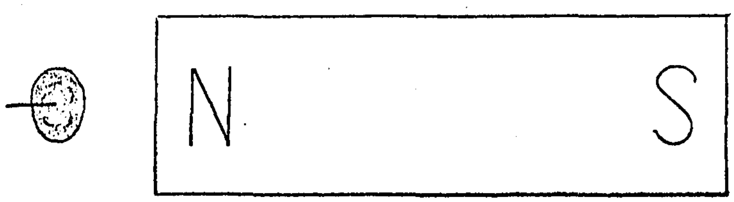
25.



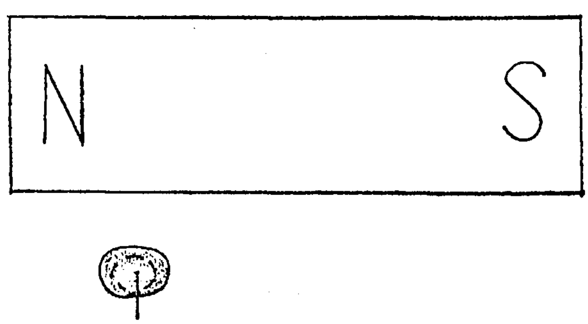
a.



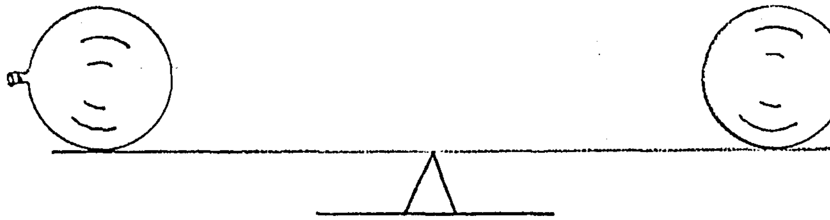
b.



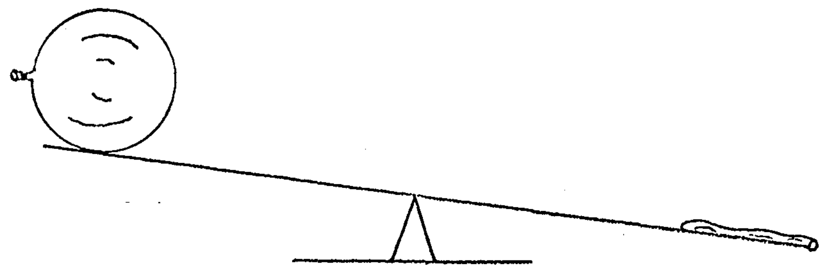
c.



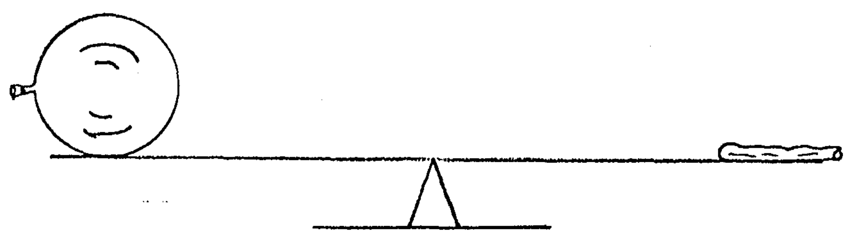
26.



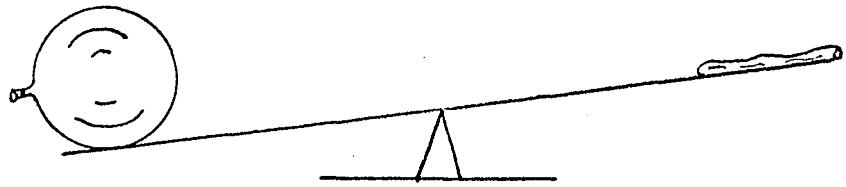
a.

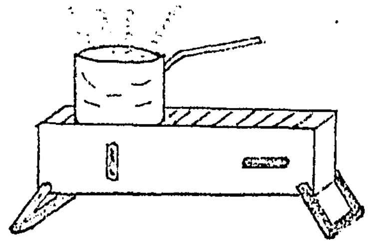


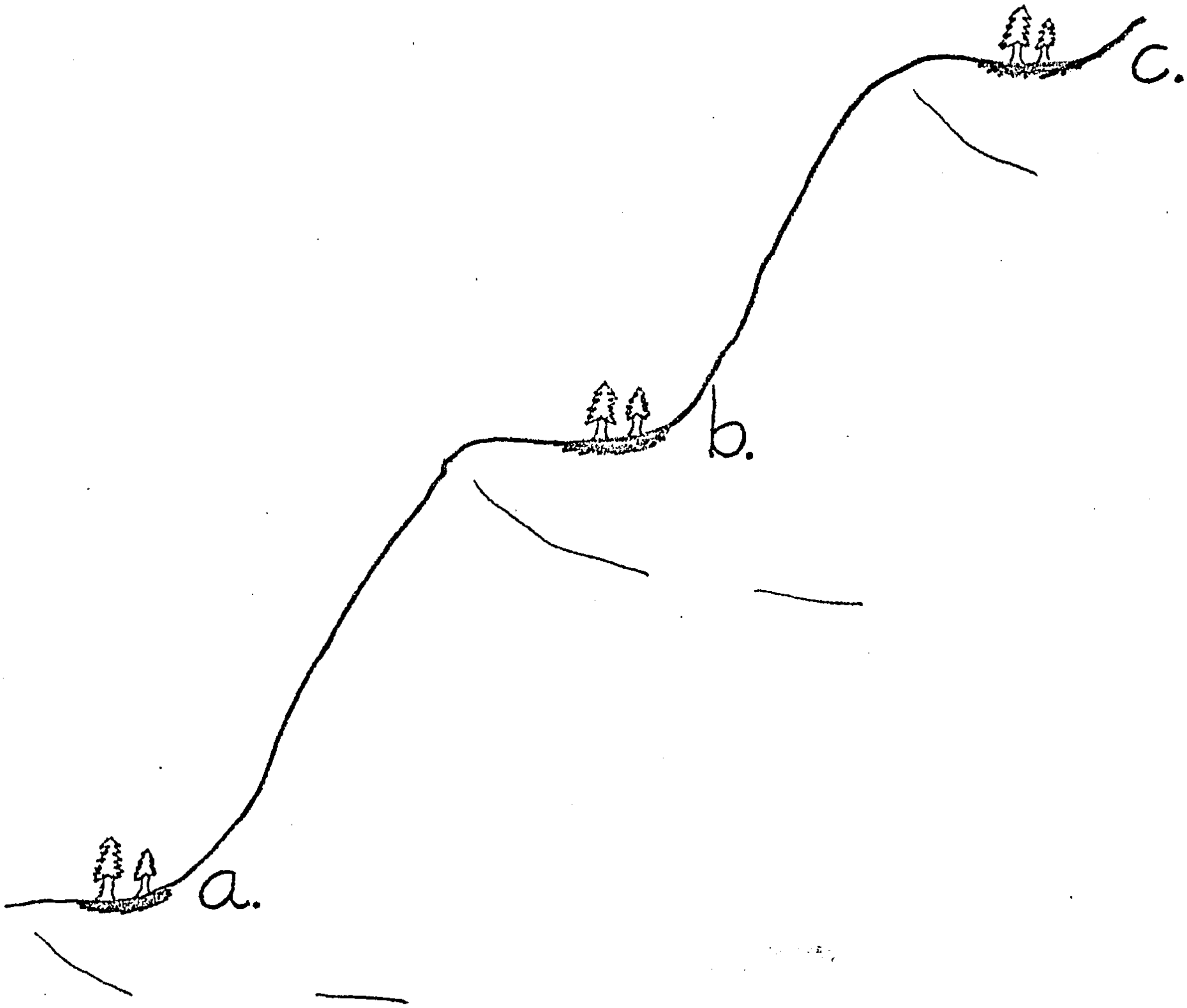
b.



c.

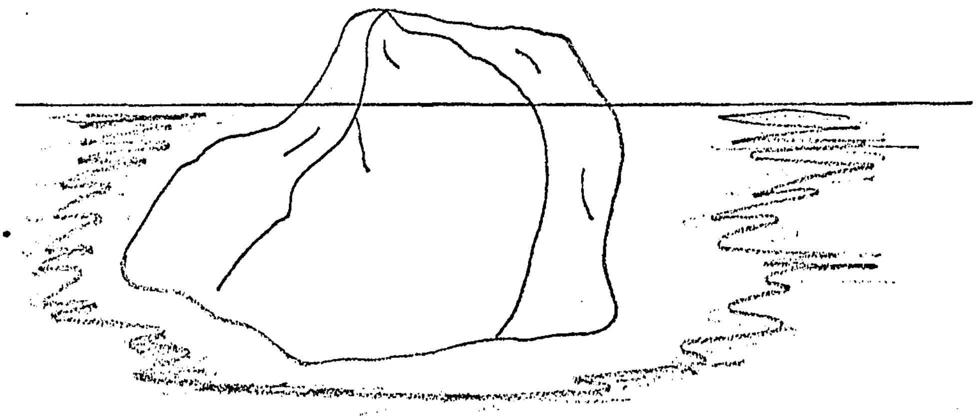


27. 

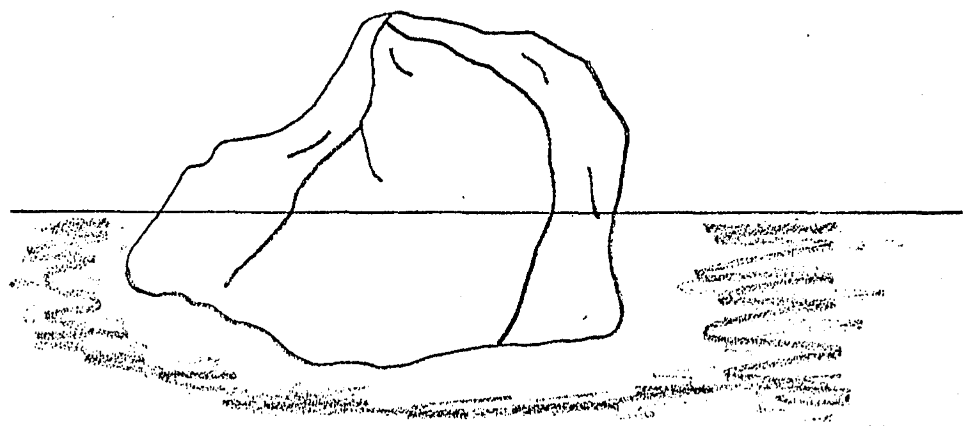


28.

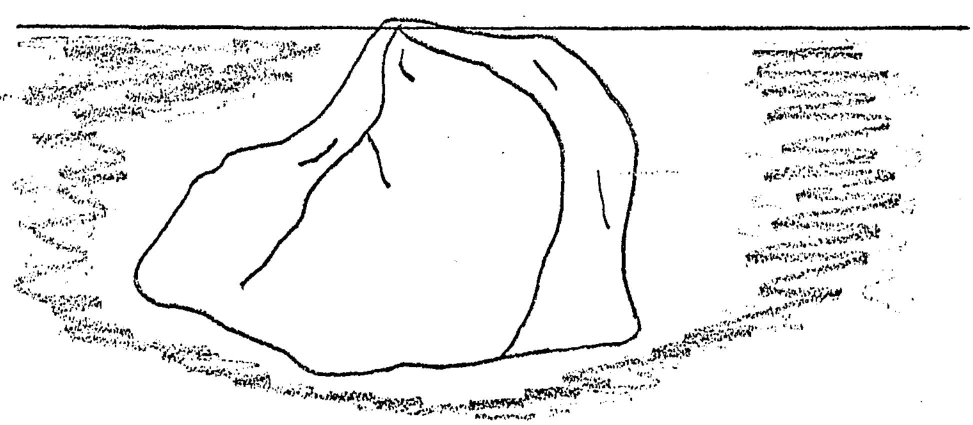
a.



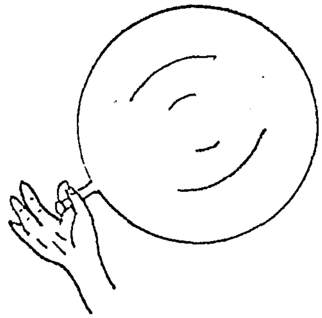
b.



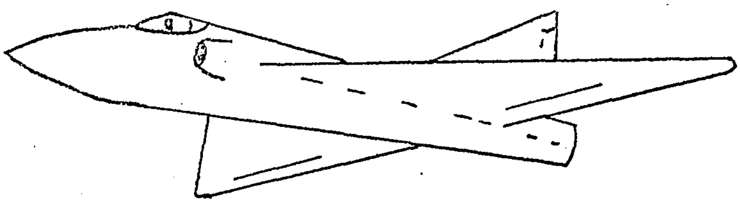
c.



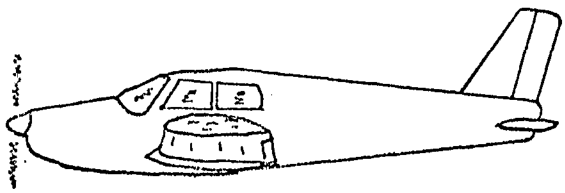
29.



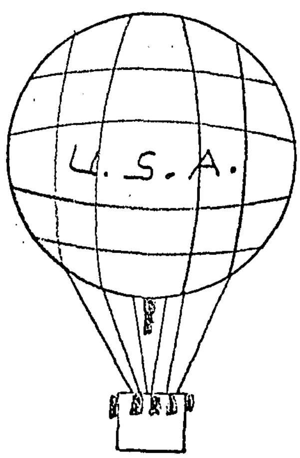
a.



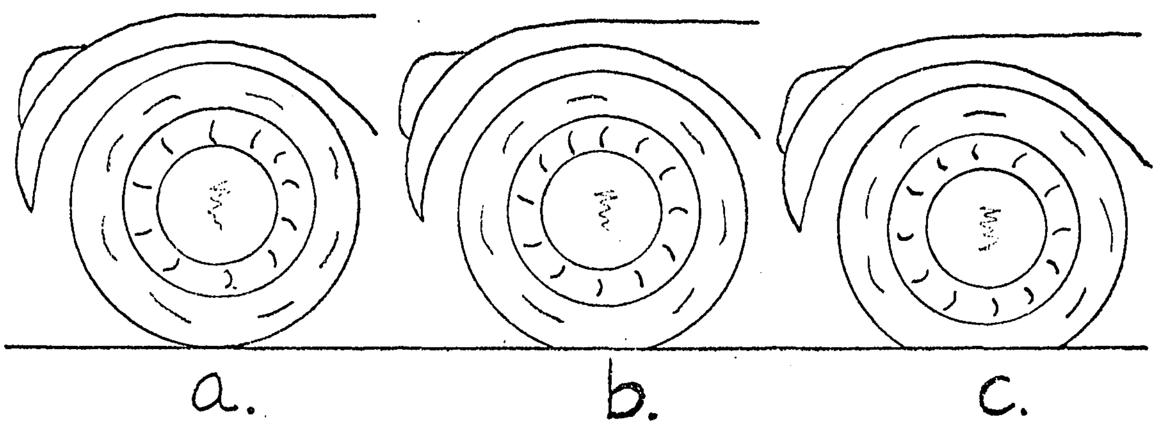
b.



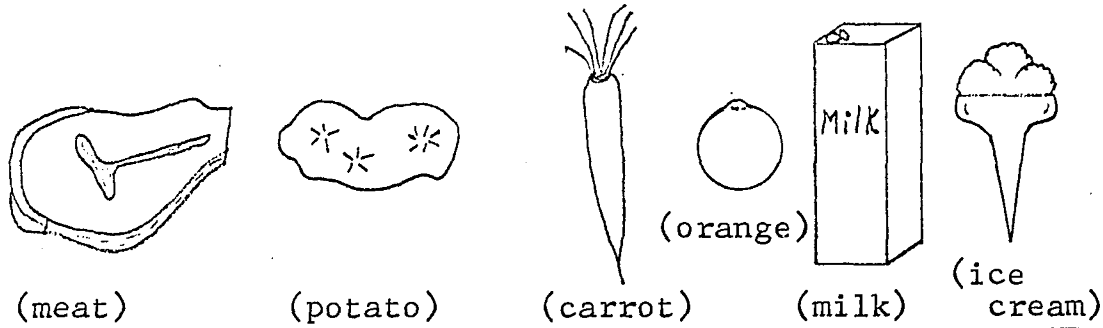
c.



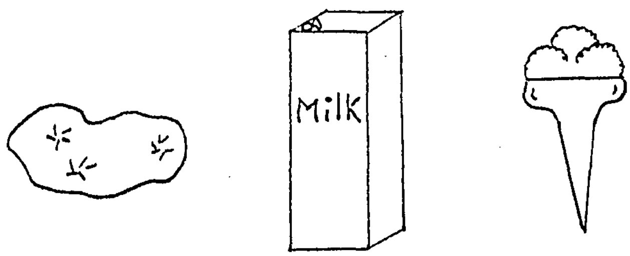
30.



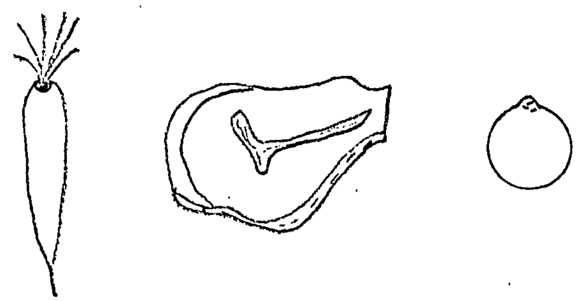
31.



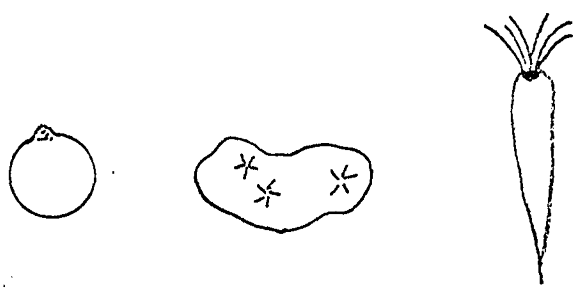
a.



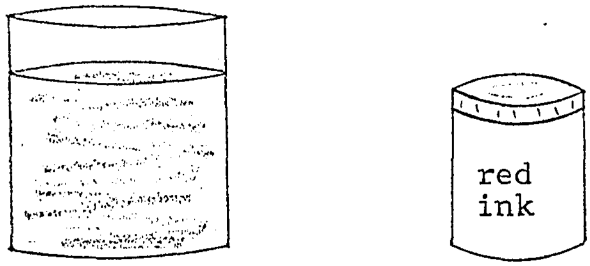
b.



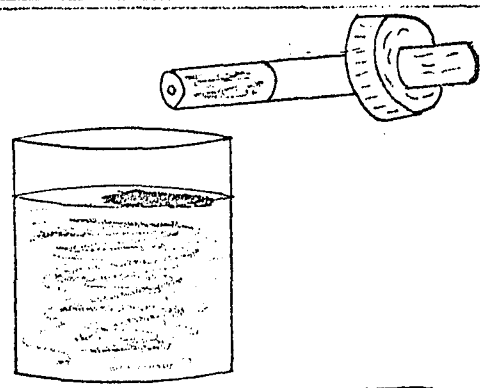
c.



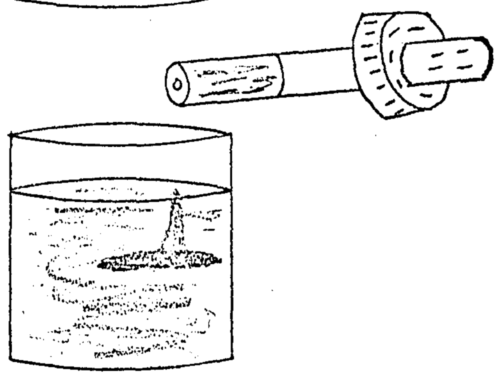
32.



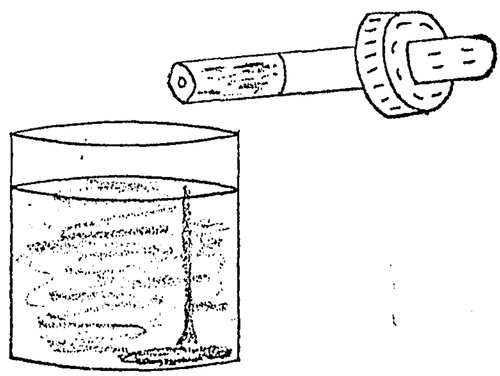
a.



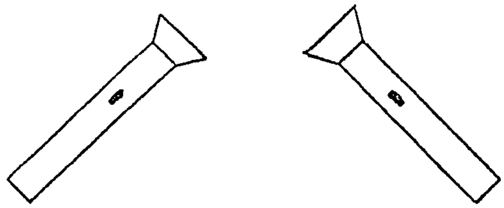
b.



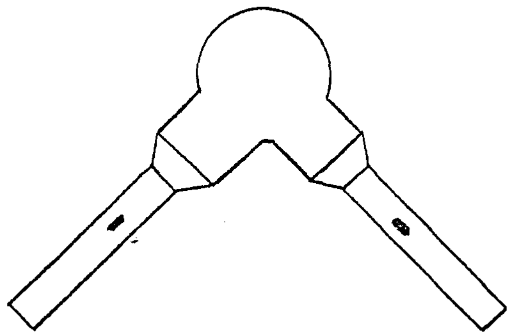
c.



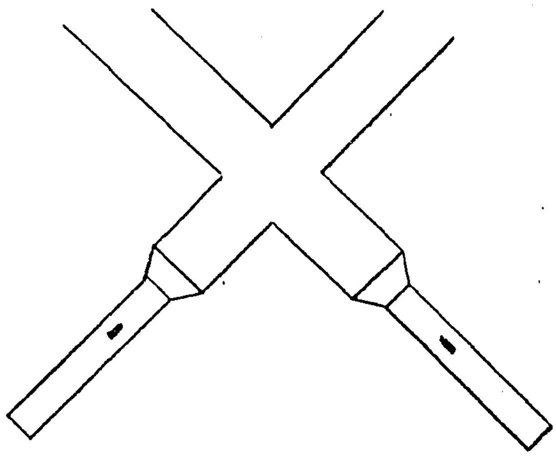
33.



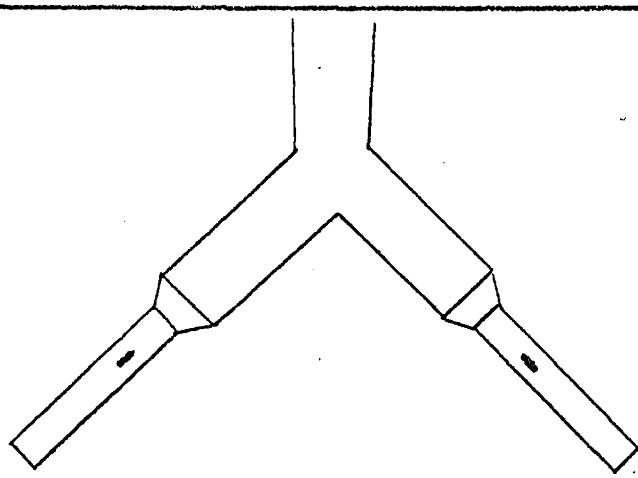
a.



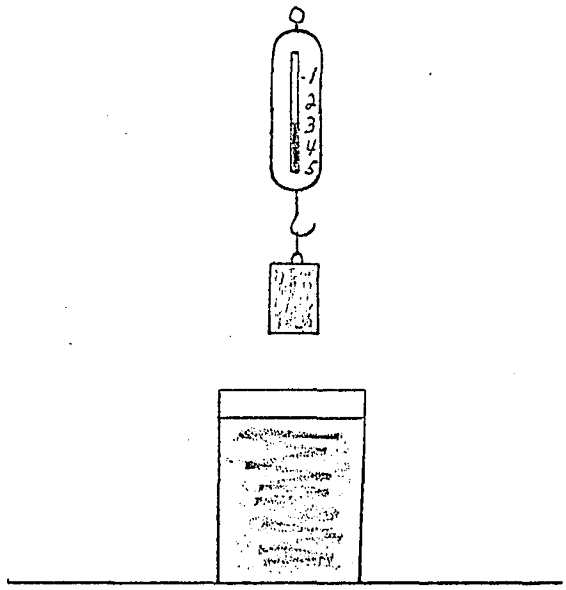
b.



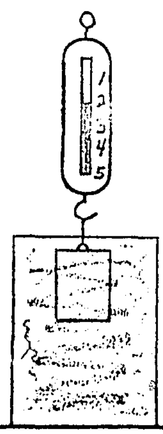
c.



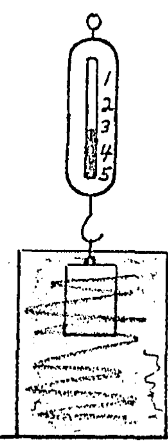
34.



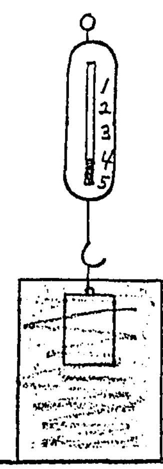
a.



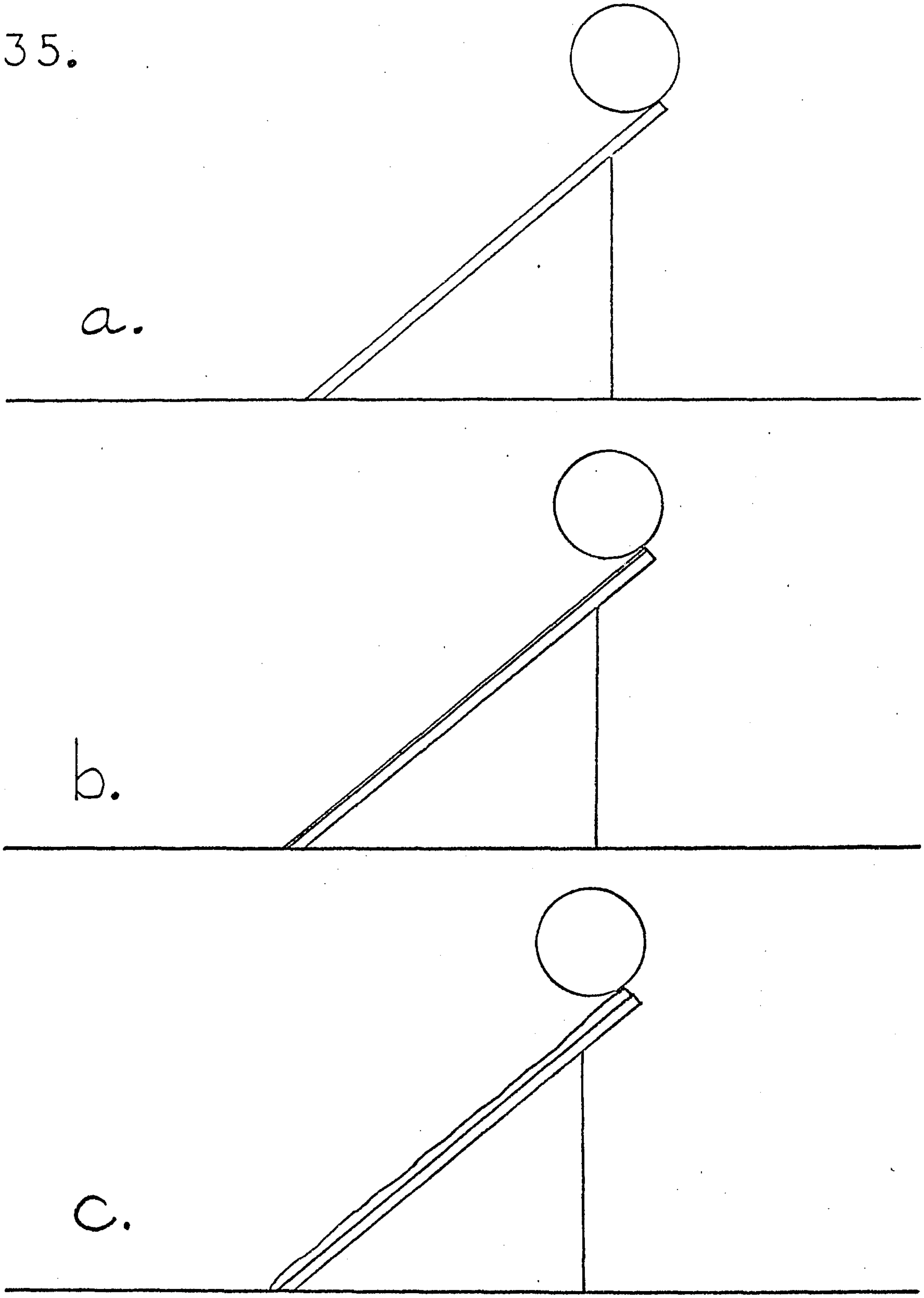
b.



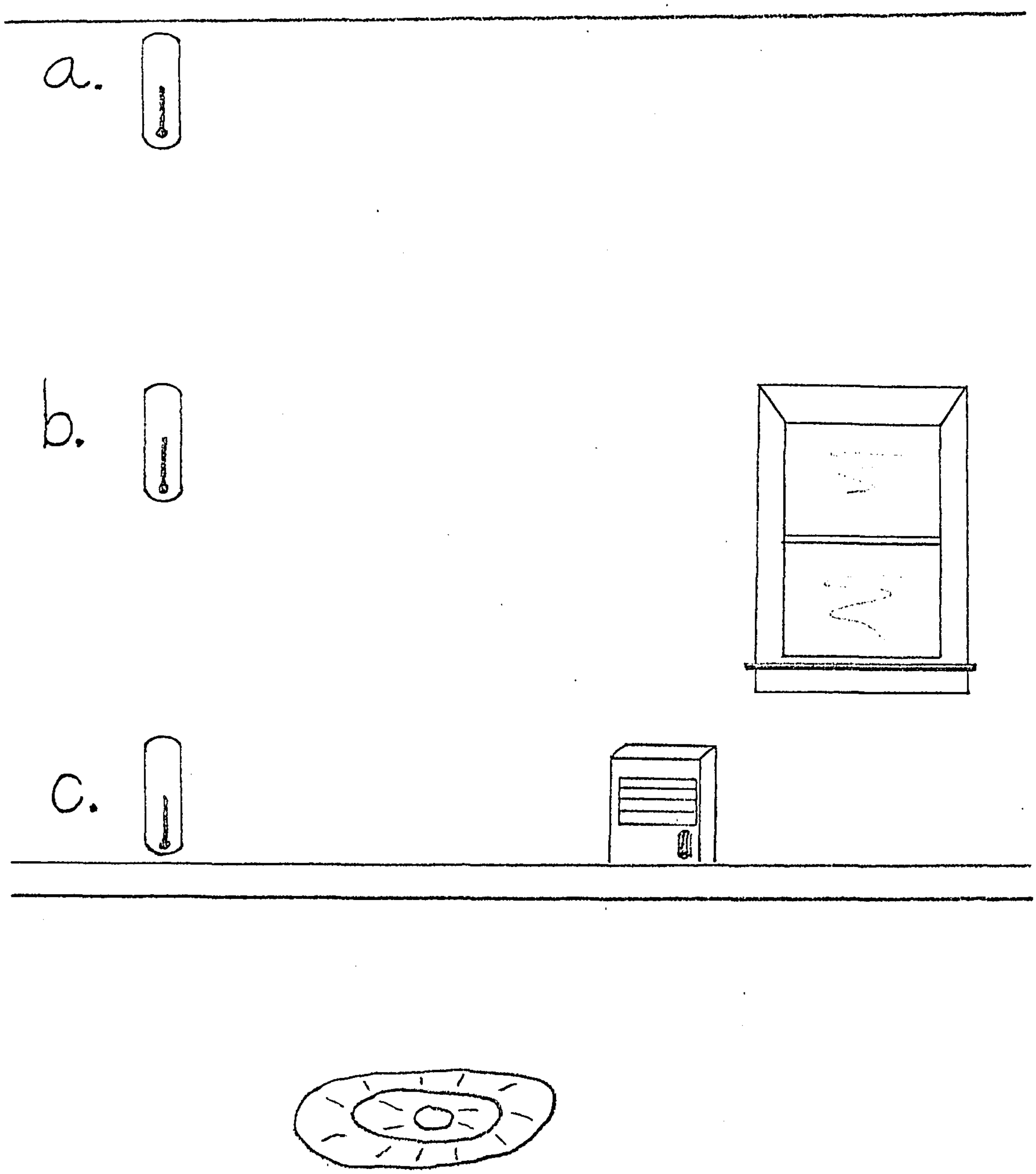
c.



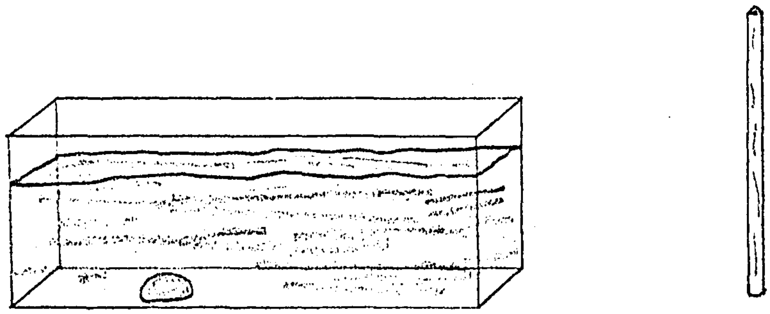
35.



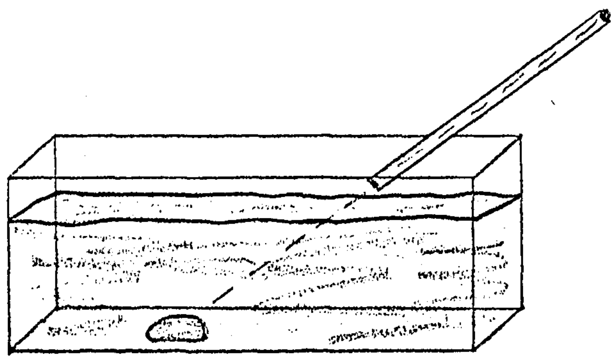
36.



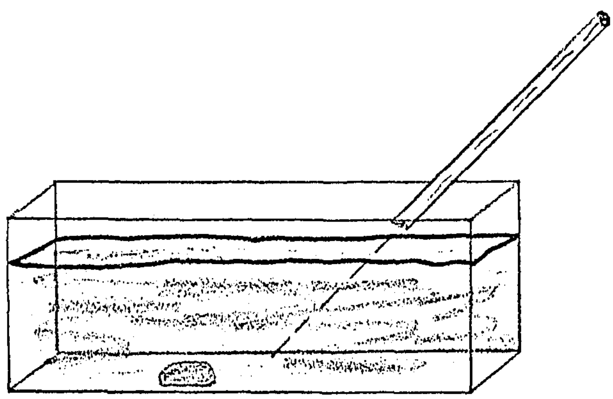
37.



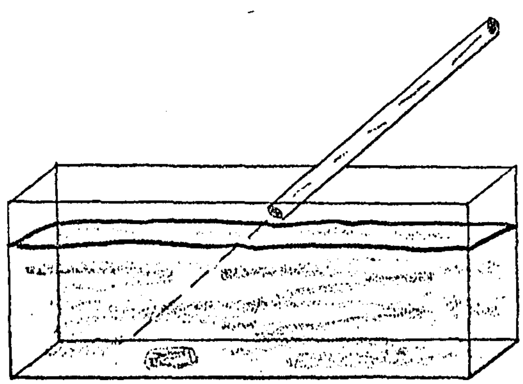
a.



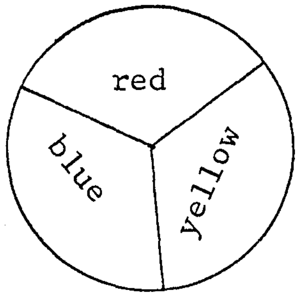
b.



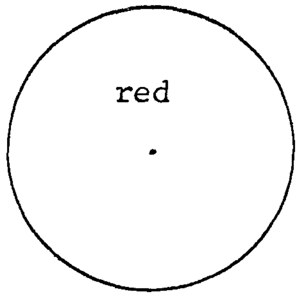
c.



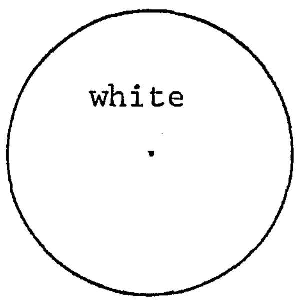
38.



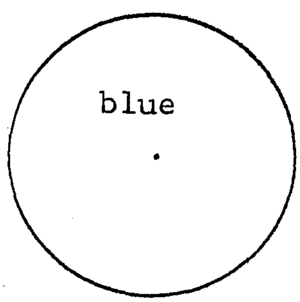
a.

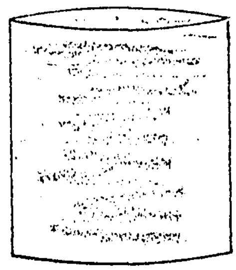


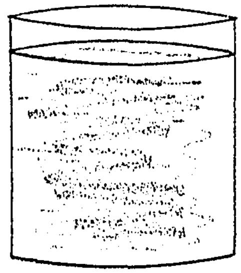
b.

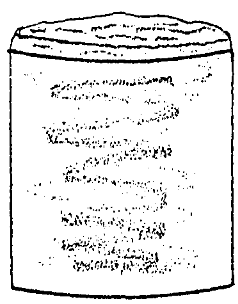


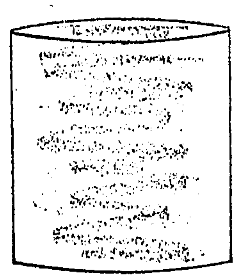
c.

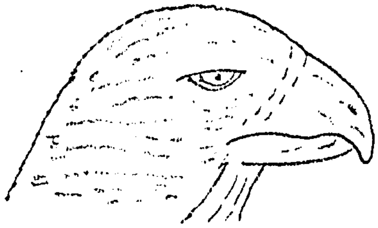
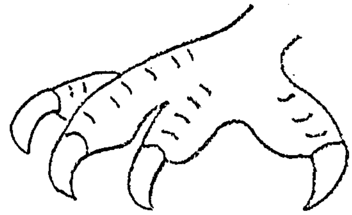


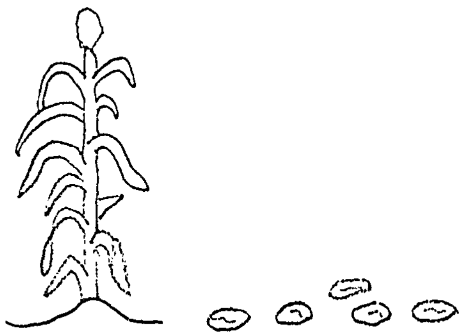
39. 

a. 

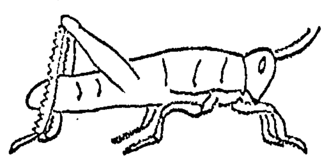
b. 

c. 

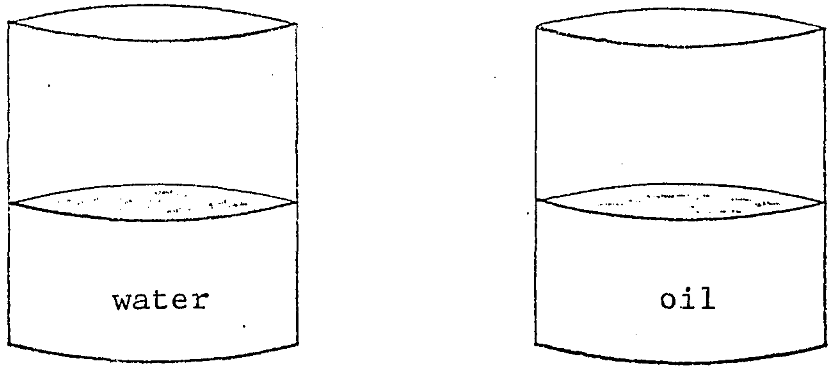
40.  

a. 

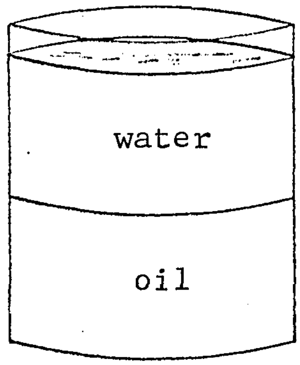
b. 

c. 

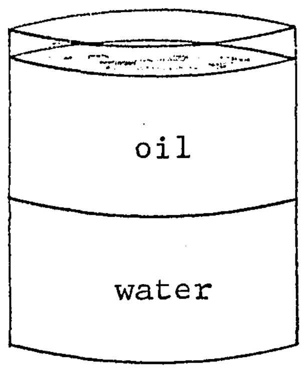
41.



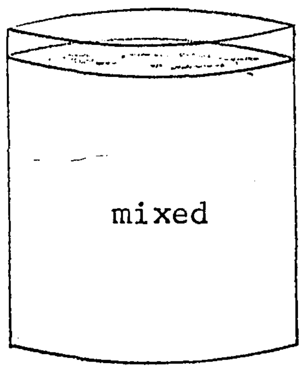
a.



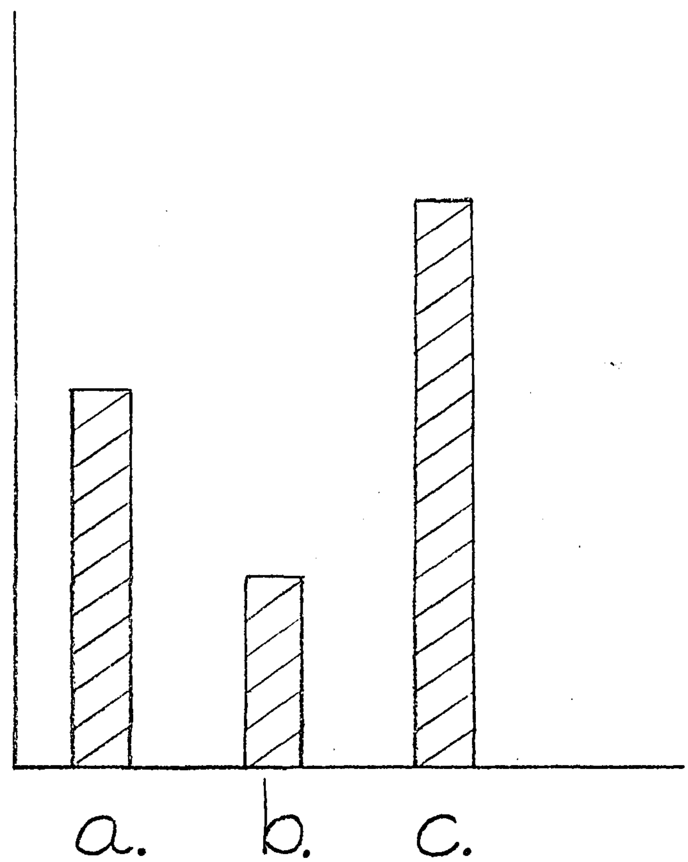
b.



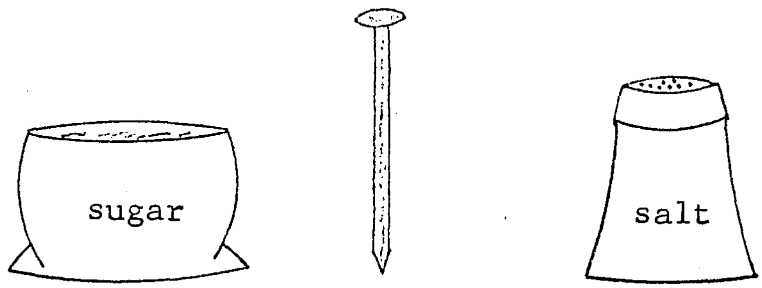
c.



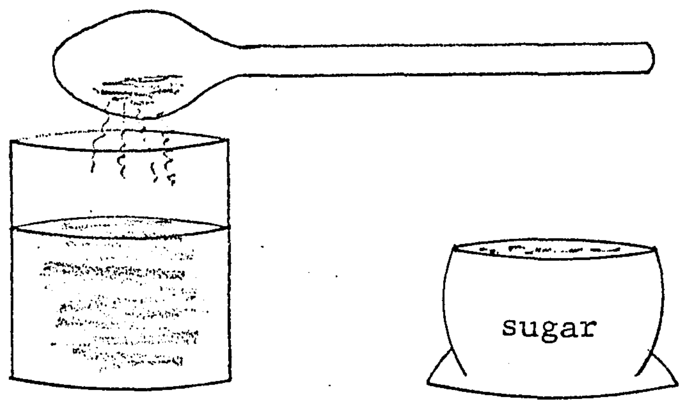
42.



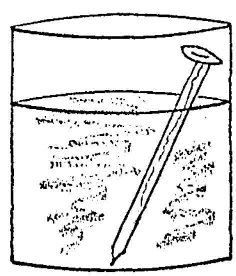
43.



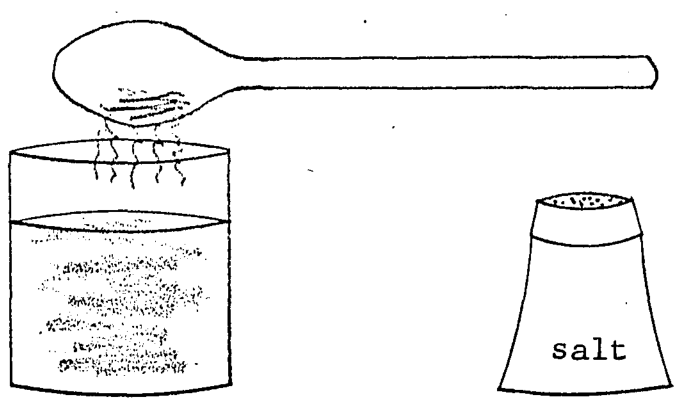
a.



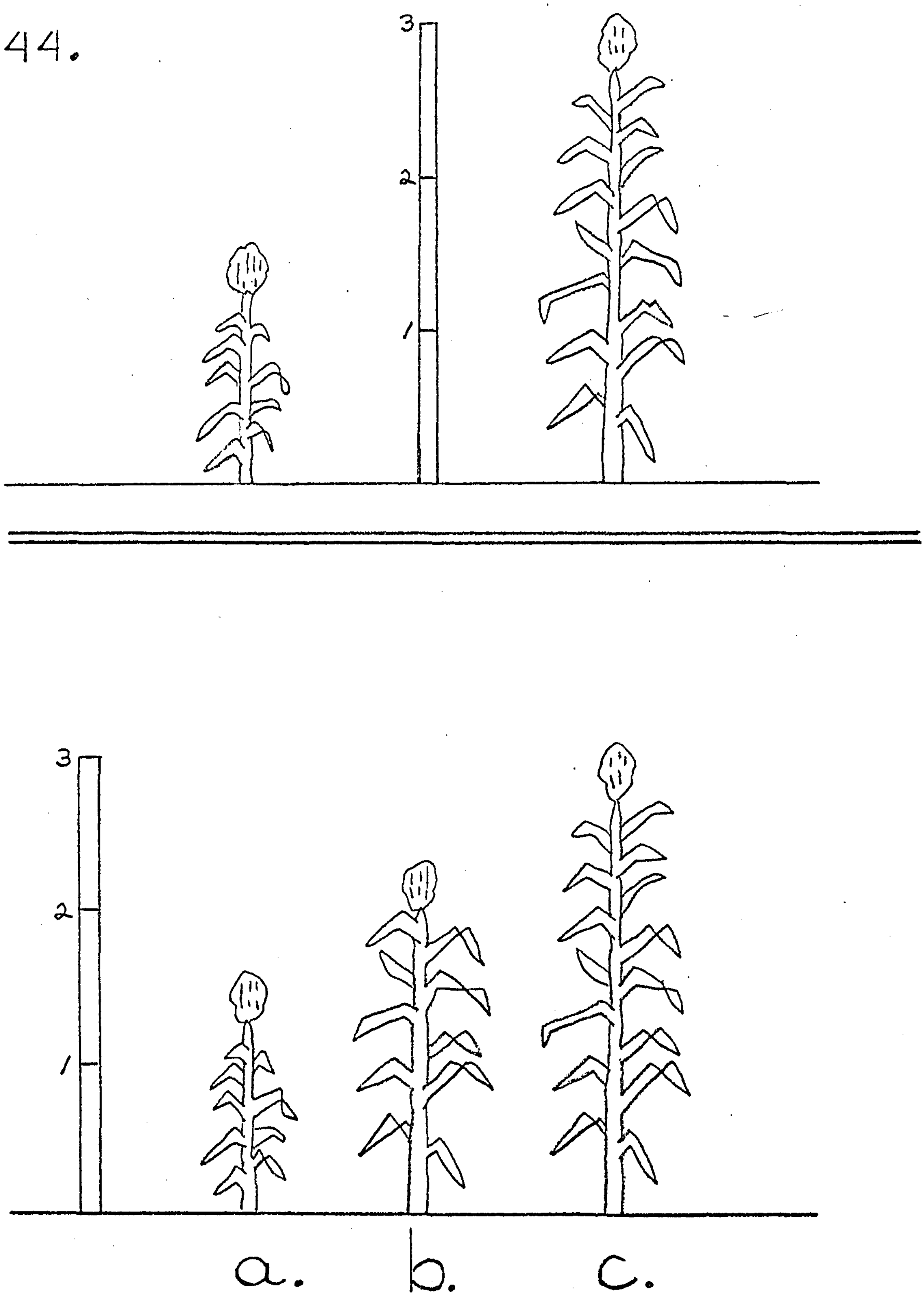
b.



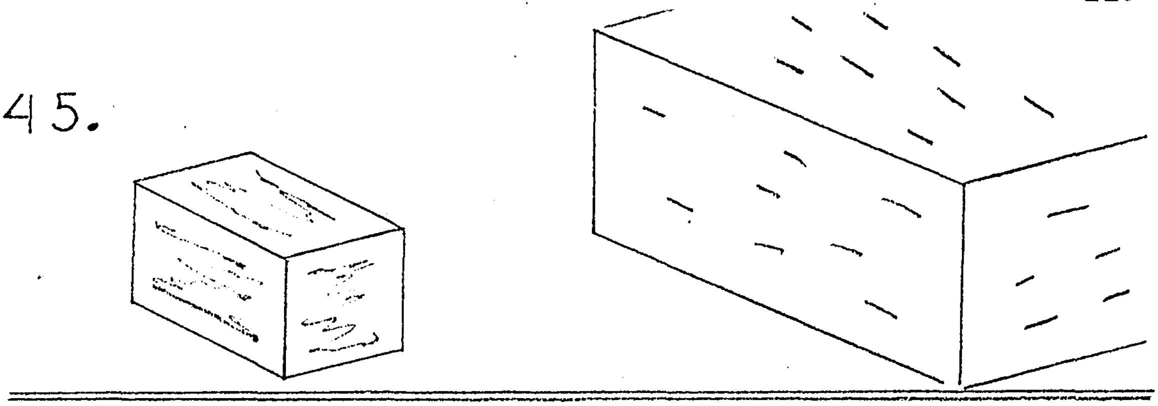
c.



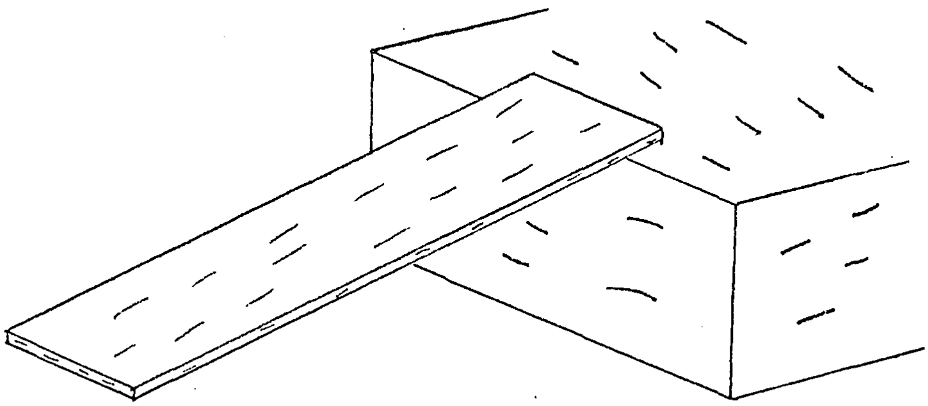
44.



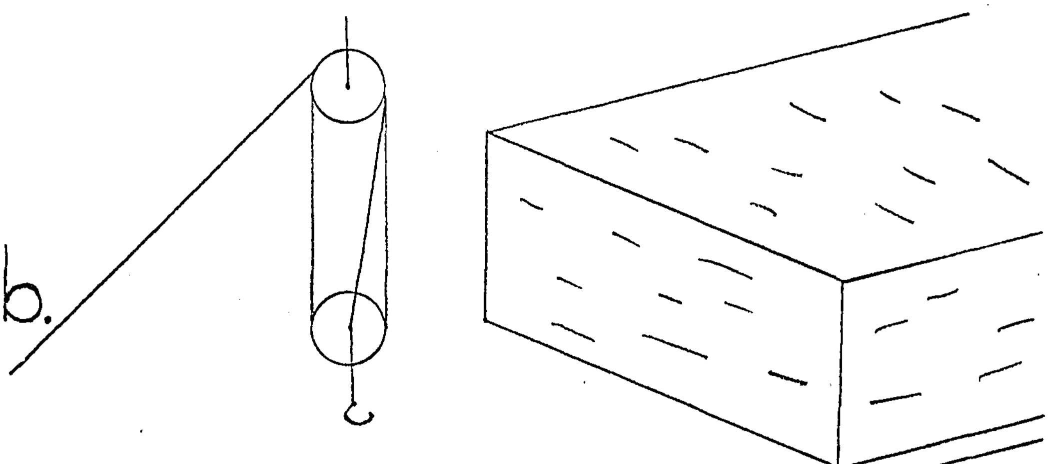
45.



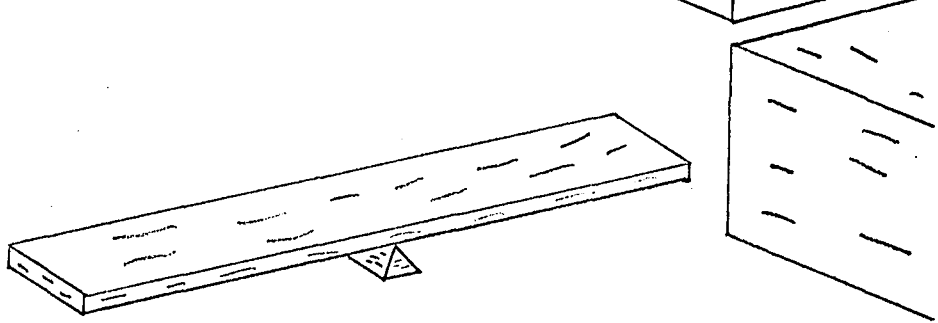
a.



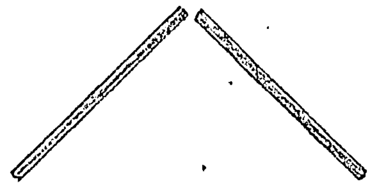
b.



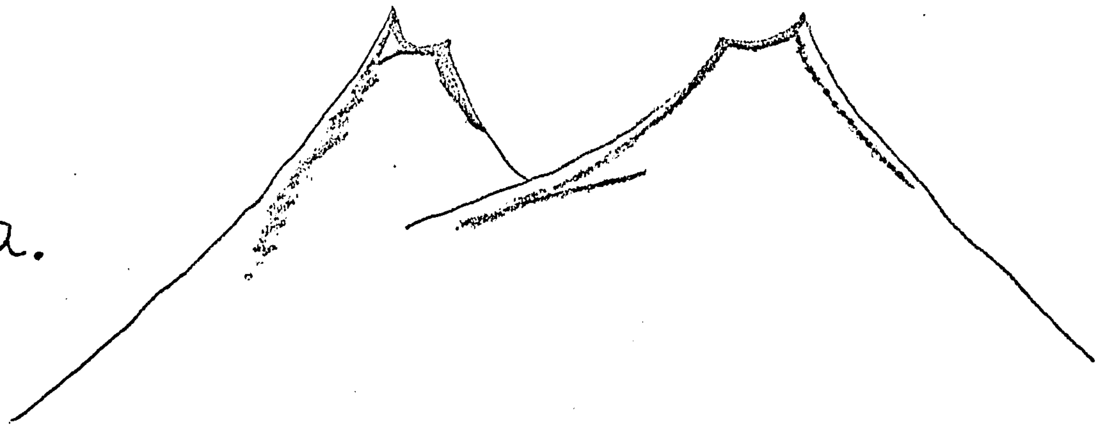
c.



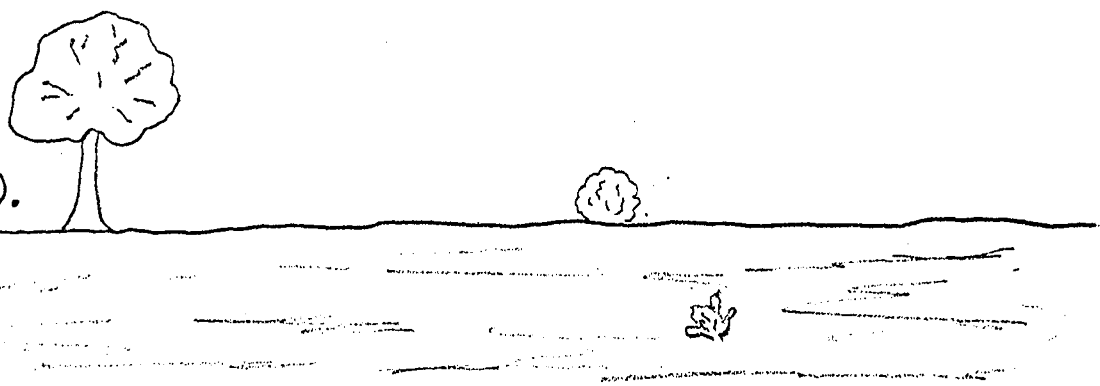
46.



a.



b.



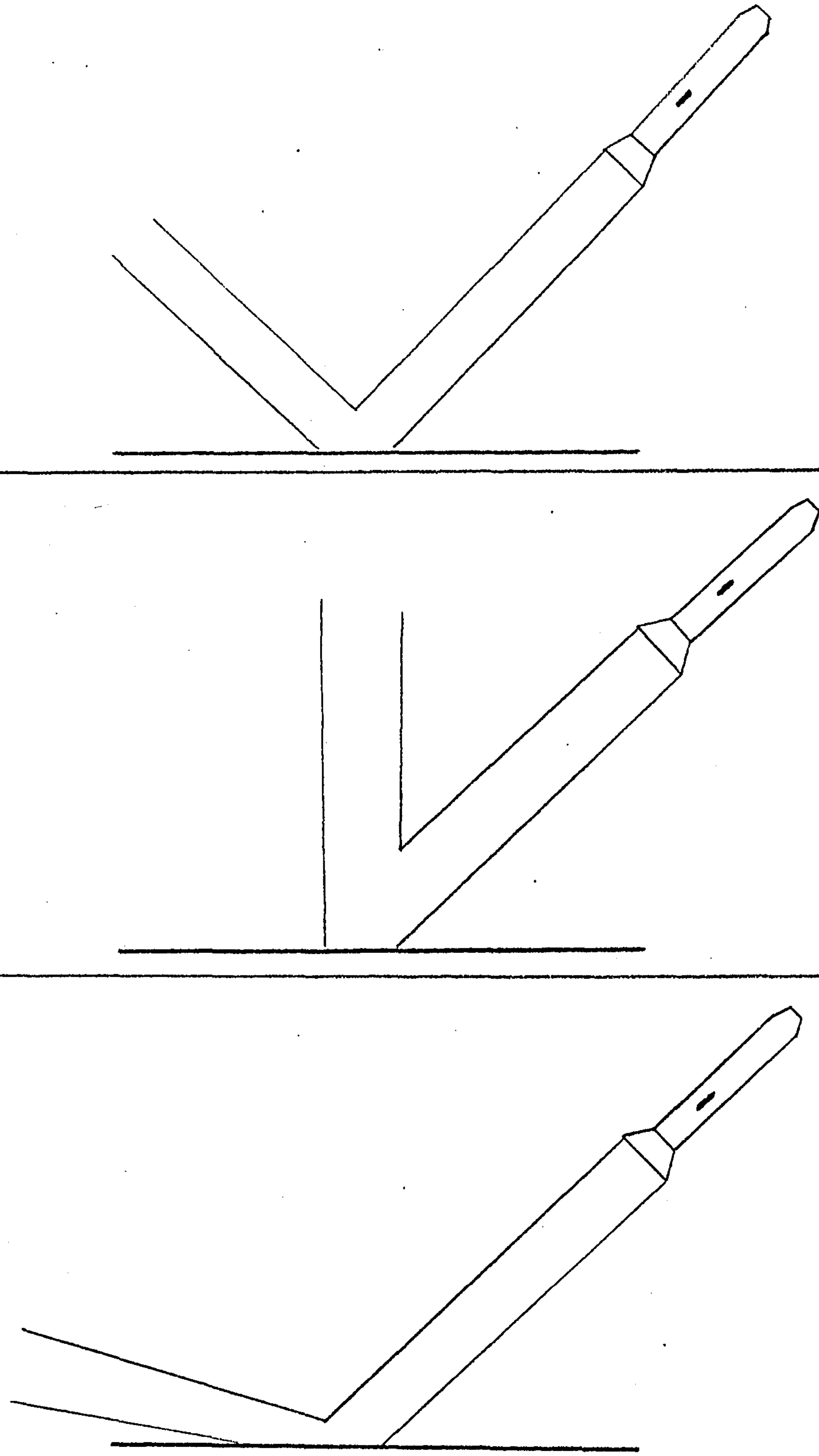
c.

47.

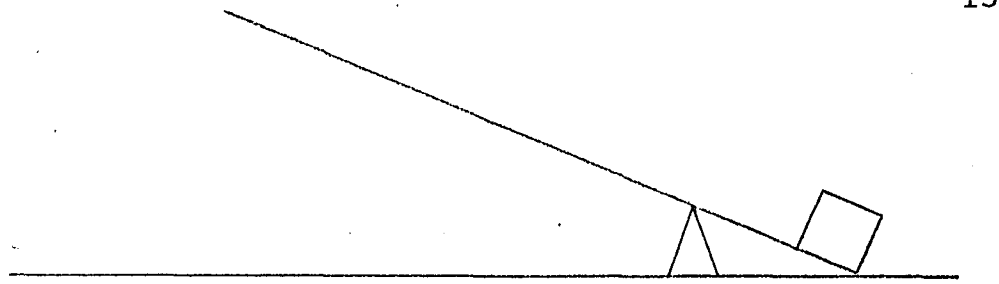
a.

b.

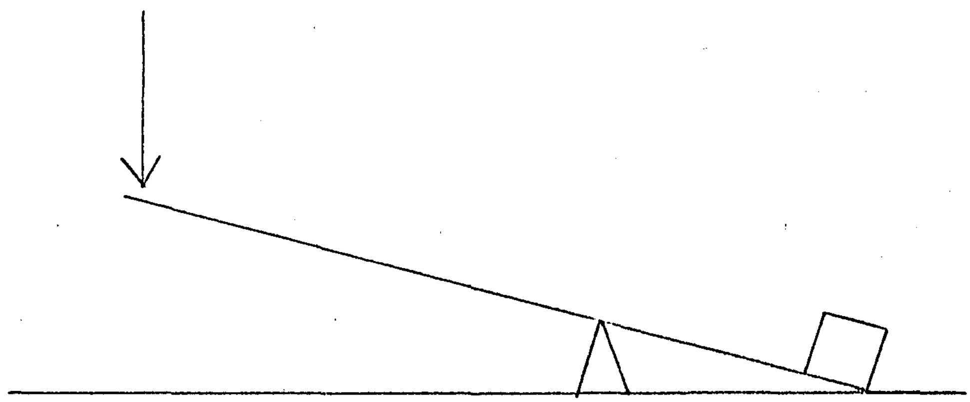
c.



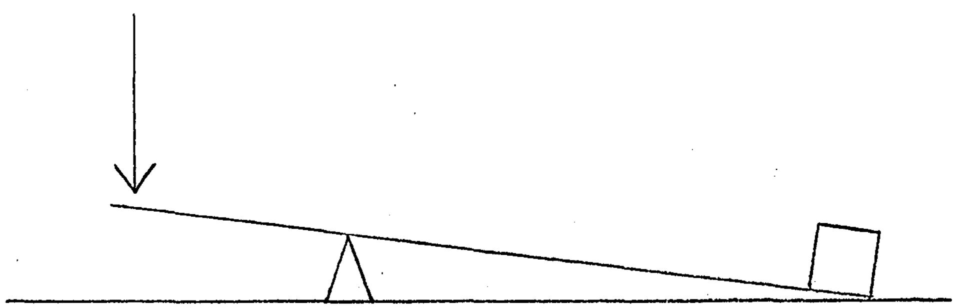
48.



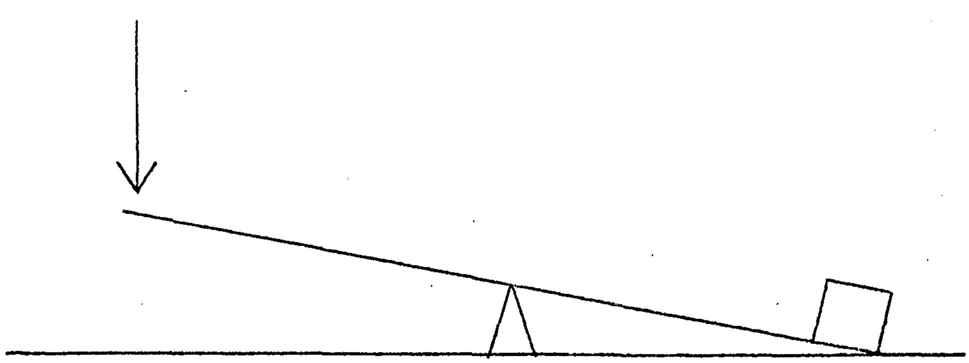
a.



b.



c.



LIST OF REFERENCES

1. Ames, Wilbur S. "The Understanding Vocabulary of First-Grade Pupils," Elementary English, XLI (January 1964), 64-68.

2. Arnsdorf, Val E. "Readability of Basal Social Studies Materials," Reading Teacher, XVI (January 1963), 243-246.

3. Bormuth, John. "Cloze as a Measure of Readability," Reading as an Intellectual Activity. Conference Proceedings of the International Reading Association, VIII. New York: Scholastic Magazines, 1963.

4. Brown, James I. "Evaluating Student Performance in Listening," Education, LXXV (January 1955), 316-321.

5. Carroll, John B. "The Analysis of Reading Instruction: Perspectives from Psychology and Linguistics," Theories of Learning and Instruction. Chicago: The University of Chicago Press, 1964. Pp. 336-353.

6. Central New York Study Council. Some Helps for Building Guides for Skill Development in the Language Arts. Report No. 7. Syracuse: Syracuse University, 1957.

7. Dale, Edgar and Jeanne S. Chall. "A Formula for Predicting Readability," Educational Research Bulletin, XXVII (January 1948), 11-20.

8. Dale, Edgar and Jeanne S. Chall. "A Formula for Predicting Readability: Instructions," Educational Research Bulletin, XXVII (February 1948), 37-54.

9. Davidson, Helen and Dom Balducci. "Class and Sex Differences in Verbal Facility of Very Bright Children," Journal of Educational Psychology, XLVII (December 1956), 476-480.

10. Davis, Allison. Social-Class Influences Upon Learning. Cambridge: Harvard University Press, 1951.

11. Deutsch, Martin. Minority Group and Class Status as Related to Social and Personality Factors in Scholastic Achievement. Ithaca: Cornell University Press, 1960.

12. Ebel, Robert L. Measuring Educational Achievement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

13. Edmonds, William S. "Sex Differences in the Verbal Ability of Socioeconomically Depressed Groups," Journal of Educational Research, LVIII (October 1964), 61-63.

14. Eells, Kenneth, and others. Intelligence and Cultural Differences. Chicago: University of Chicago Press, 1951.

15. Forbes, Fritz W. and William C. Cottle. "A New Method for Determining Readability of Standardized Tests," Journal of Applied Psychology, XXXVII (June 1953), 185-189.

16. Garrett, Henry E. and R. S. Woodworth. Statistics in Psychology and Education. New York: David McKay Company, Inc., 1964.

17. Gronlund, Norman E. Measurement and Evaluation in Teaching. New York: The Macmillan Company, 1965.

18. Guckenheimer, S. N. "The Readability of Pamphlets on International Relationships," Educational Research Bulletin, XXVI (December 10, 1947), 231-238.

19. Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, Inc., 1956.

20. Hampleman, Richard S. "Comparison of Listening and Reading Comprehension Ability of Fourth and Sixth Grade Pupils," Elementary English, XXXV (January 1958), 49-53.

21. Herrington, R. L. and G. G. Mallinson. "An Investigation of Two Methods of Measuring the Reading Difficulty of Materials for Elementary Science," Journal of Applied Psychology, XLII (December 1958), 385-390.

22. Kelley, Truman L., Richard Madden, Eric F. Gardner, and Herbert C. Rudman. Stanford Achievement Test: Directions for Administering Intermediate II Battery. New York: Harcourt, Brace and World, Inc., 1964.

23. Klare, George R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963.

24. Klare, George R., James E. Mabry, and Levarl M. Gustafson. "The Relationship of Style Difficulty to Immediate Retention and to Acceptability of Technical Material," Journal of Educational Psychology, XLVI (May 1955), 287-295.

25. Koehn, Edna. "The Relationship of the Basic Skill Development of Sixth Grade Gifted Children to Ninth Grade Achievement in the Content Fields," Dissertation Abstracts, XXI (July 1960), 133-134.

26. Lewis, June E. and Irene C. Potter. The Teaching of Science in the Elementary School. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1961.

27. Malkin, Samuel. "The Culturally Deprived Child and Science," Science and Children, VII (April 1964), 5-7.

28. Miller, Wilbur R. "Readability versus Reading Ability," Journal of Educational Research, LVI (December 1962), 205-209.

29. Ragan, William B. Modern Elementary Curriculum. New York: Holt, Rinehart and Winston, 1966.

30. Russell, R. D. "A Comparison of Two Methods of Learning," Journal of Educational Research, XVIII (October 1928), 235-238.

31. Saltzman, Sara. "The Influence of Social and Economic Background on Stanford-Binet Performance," Journal of Social Psychology, XII (August 1940), 71-81.

32. Shores, J. Harlan and J. L. Saupe. "Reading for Problem-Solving in Science," Journal of Educational Psychology, XLIV (1953), 149-158.

33. Strang, Ruth. An Introduction to Child Study. New York: The Macmillan Company, 1951.

34. Strang, Ruth. "A Dynamic Theory of the Reading Process," Merrill-Palmer Quarterly, VII (October 1961), 239-245.

35. Strickland, Ruth G. The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children. Bulletin of the School of Education, Vol. XXXVIII, No. 4. Bloomington: Indiana University Press, 1962.

36. Tannenbaum, Harold E., Nathan Stillman, and Albert Piltz. Science Education for Elementary School Teachers. Boston: Allyn and Bacon, Inc., 1965.

37. Tannenbaum, Harold E., Nathan Stillman, and Albert Piltz. "Evaluation in Elementary Science by Classroom Teachers and Their Supervisors," Readings in Science Education for the Elementary School. New York: The Macmillan Company, 1967. Pp. 192-220.

38. Taylor, Stanford E. What Research Says to the Teacher: Listening. National Education Association Bulletin No. 29. Washington, D. C.: National Education Association, 1964.

39. Tiedt, Iris M. and Sidney W. Tiedt. Contemporary English in the Elementary School. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967.

40. Wilt, Miriam E. "A Study of Teacher Awareness of Listening as a Factor in Elementary Education," Journal of Educational Research, XLIII (April 1950), 626-636.