5-1966

An Exploration of Test Taking Strategy with Sixth Grade Students

Mary Bates
Utah State University

AN EXPLORATION OF TEST TAKING STRATEGY WITH SIXTH GRADE STUDENTS

by

Mary Bates

A seminar report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF EDUCATION

in

Guidance

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1966

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The present study was an attempt to evaluate the benefits of certain cues in reaching correct answers to test questions given to sixth grade students. While it seems reasonable to assume that test-taking strategy is not as well developed in sixth grade children as in older students, this study was intended to discover if test-taking strategy would be apparent to a measurable degree at this age level. Before such an observation could be undertaken, a technique had to be devised to measure the behavior in question (in this case the use of cues). The technique used was in the form of two tests, both of which are included in the Appendix.

Need for developing new measuring techniques for test taking behavior

Whether or not one likes tests, trusts test results, or feels that tests actually measure what they are supposed to measure, it seems highly likely that tests are here to stay. In support of this statement are the ever increasing organizations whose sole purpose is that of test construction, distribution, and correction. Buros' first edition of the Mental Measurement Yearbook (1938) contained reviews of an impressive number of standardized tests; but it is interesting to note that 957 new ones were added within the six year interval between the 1953 and the 1959 editions.

Meanwhile, controversy has mounted. Those persons who are in the business of testing, and those in favor of testing continue to support their beliefs (Beitner, 1964; Miller, 1964). Opponents pile up

evidence of inadequacy of certain tests in general and certain items in particular (Hoffman, 1964; Carroll, 1960). The crux of the matter is that attempts have been made to reduce the complex workings of the human brain into mere numbers on a piece of paper. The test makers have been unable to prove empirically that their tests do measure the particular isolated quality they claim to measure, and the opposition has been unable to prove that they do not. Hoffman (1964) says, "Except in the simplest situation, there is no satisfactory method of testing--nor is there likely to be. Human abilities and potentialities are too complex, too diverse, and too intricately interactive to be measured satisfactorily by present techniques." While Hoffman is referring to the technique of testing itself as lacking in perfection, Ebel (1963) expresses the need for technology to verify empirically the recommendations given by authorities in both test making and test taking.

Since tests are subject to the errors of their human authors, there is much literature directed to students in the matter of test taking strategy. For the most part, the advise given is based on common sense and professional judgment (Thorndike and Hagen, 1962; Ebel, 1963). Experimental proof of the use of the advise is relatively sparce. With present limitations, it would seem that most test taking techniques are not reducable to observable behavior. But Moore et al. (1966), for example, were able to reduce one type of test taking strategy to observable behavior, e.g. guessing on tests where there was a penalty for guessing.

A condensation of the above quotations centers on three issues: (1) that we must live with tests whether we like it or not, (2) that

present techniques are inadequate to verify empirically the recommendations given by authorities in both test making and test taking; and (3) that there is a need for this type of technology. This study is aimed at filling that need in one small capacity. While the references cited thus far have meant their comments to include standard published tests, the present study was concerned with informal teacher-made tests primarily.

Operational definition of test taking strategy

Test taking strategy is felt to be largely a matter of educated guessing on items where the subject matter is not actually known or where it is questionable which answer the tester wants. One's ability to guess right more often than wrong is often related to his perception of cues, which apparently can be and often are inadvertantly written into tests by hurried or unwitting teachers. In many cases these cues can pinpoint the correct answer or reduce the odds against selecting a wrong answer. Therefore, the extent to which students recognize and utilize such cues is referred to in this study as "test taking strategy". Typical kinds of cues and those utilized in the testing for this study can be found in the Review of the Literature section.

Statement of the purpose

It was the purpose of this study to determine the extent to which sixth grade students would make use of cues that had purposely been written into a test.

In this study student response was measured on a test in which every item contained a cue. Since knowledge was considered the strongest

factor in test taking (Huff, 1961), an attempt was made to eliminate that variable by constructing an equivalent test without cues. The difference in the two tests was used as the criterion for estimating the extent to which the S used cues. No attempt was made to actually determine, statistically, the equivalency of the two tests nor of the test-retest reliability. However, the two tests were assumed to be comparable, since the items were constructed to be similar except for the insertion of cues in the items of Test 2. The test was refined by administering it to a pilot group before giving it to the study group.

REVIEW OF THE LITERATURE

Measurements of various kinds have either touched, or will touch, upon all humans in all walks of life both past and present. Even the ancients had tests of skill and dexterity, contests, tournaments, and tests of courage. Today, measurements of a new-born baby's weight and temperature are routine practices in many places; at death the deceased is measured for a casket. During the space between birth and death, measurements and tests take place almost daily. Courtship may be considered a period of testing another for the purpose of selecting a spouse. There is the job interview where the applicant is being tested for possible placement. The list is endless. In fact, if all our various measuring devises were suddenly destroyed, contemporary civilization would collapse like a house of cards (Ross and Stanley, 1954).

While it is well to keep in mind the wide scope of tests and measurements, this paper is concerned with that part of psychological measurement which tests knowledge of a particular subject matter.

Knowledge, being the elusive quality that it is, poses some problems in being represented by precise scores. Yet a test for knowledge is useful and necessary, especially in the field of education. Teachers have a need for testing to determine whether individual students and the class as a whole are doing satisfactory work and whether they as teachers are successful in putting across the subject matter. Tests assist the teacher in deciding what grades should be given to students

for their accomplishments. Both teacher and learner stand to benefit by accurate testing programs (Weitzman and McNamara, 1949).

Accuracy is the controversial issue in the matter of testing. For if the test is not valid, if it does not measure the knowledge intended to be measured, if the measurement is contaminated by factors other than that which is to be measured, then the result is a misrepresentation of the facts and can be harmful and unfair in many instances (Hoffman, 1964).

In this review, two avenues toward the accuracy of mental measurement will be explored. One is a brief look at the evolution of testing knowledge from the days of Socrates to the present. The other is the problems that still plague test makers today. The latter will lead into a discussion of the type of errors in test making that can contaminate the results. Errors which result largely from the careless wording of a test item may inadvertantly indicate the answer wanted. In other words, certain methods of test construction provide cues, or clues, to the right answer. These cues are one of the variables effecting test scores that form the basis of the present investigation.

Evolution of testing procedures and knowledge about tests

The oral examination is the oldest method known for assessing knowledge. It was used by Socrates with his pupils. (Green, 1963). It was used in institutions of learning almost exclusively until about a hundred years ago, when it came into disfavor partly as the result of the efforts of Horace Mann and, later, Emerson White. Both of these men attested to the superiority of written examinations (Ross and Stanley, 1954). Oral examinations are still in use today, especially at the graduate level in colleges. Job interviews can be considered

a type of oral examination. The oral examination has the advantage of flexibility. If the student lacks knowledge of a specific fact, the examiner can change subjects to find out what the student does know. The possibility of misunderstanding or misinterpretation are practically eliminated with the person-to-person flow of conversation. However, the disadvantages appear to outweigh the advantages for many practical purposes. First, it is too time consuming in a classroom situation to examine each student individually for the length of time necessary for adequate coverage of the subject matter. Second, the oral examination is too subjective, often turning out to be a blending or conflict of personalities. Third, the quality of the examination is contingent upon the thought and preparation put into it ahead of time by the teacher. Therefore, for reasons of expediency and fairness the oral exam gave ground to the written exam.

Written examinations are classified as either subjective or objective. Subjective examinations are usually known as essay examinations. According to Ross and Stanley (1954) Stalnaker (1951) compares the merits of essay and objective tests in a thorough and impartial manner. The essay type, though often maligned, they quote, "has potential values for measuring outcomes of learning not yet otherwise measured." It merits further development and research. However, essay tests do have low validity, low reliability, and low usability. Like the oral examination the essay type is subject to the mood or personality of the reader. It is possible to overrate the importance of knowing how to say a thing rather than having something to say. Teachers do not agree with each other on scoring a paper and are not always able to duplicate their own evaluation of the same paper

on two different readings of it. Neatness and spelling may influence the grade given. Lastly, the time the student spends in writing limits the amount of knowledge that can be covered during a testing period.

In contrast, the objective or short-answer test has proven to have two distinct advantages over the essay test in areas where writing ability is not an issue. It eliminates the subjectivity of the scoring, and it permits a larger coverage of the subject matter by eliminating the time it takes for writing long passages. It makes possible a longer test, in a sense, than could be given by essay in the same length of time. This was a move toward an increase in reliability according to Remmers et al. (1960), who say, "The reliability of a test is a function of its length, or longer tests tend to be more reliable than shorter tests."

Objective tests are of two types, i.e. recall and recognition types. The recall types are simple-recall or short answers and completion. The more common recognition types are the alternative-response (which includes true-false), multiple-choice, and matching. Of these, the multiple-choice type is usually regarded as the most valuable and most generally applicable of all test forms (Ross and Stanley, 1954). Other comments in support of the multiple-choice test are: Lindquist (1936) who asserts that it is "definitely superior to other types for measuring such educational objectives as inferential reasoning, sound judgment, and discrimination on the part of the pupil." Lee (1936) regards it as "one of the best means for testing judgment that is available." Wrightstone (1956) says that "such tests are generally superior to essay examinations in their sampling course content, reliability of scoring and ease of scoring." Green (1963) states that "multiple-choice tests are considered by most test experts to be the best type of objective test for measuring a variety of educational objectives."

The advantages of the multiple-choice test over the other objective tests are:

1. Ease of scoring--It is possible to score the tests rather quickly without benefit of a devise for that purpose. However, some teachers find it easier and quicker to use a key with holes punched out where the correct answer should fall. Still others, especially in large universities and colleges have access to scoring machines which process an entire classroom's answer sheets in a matter of minutes. Since most people would agree that a teacher's time is best spent teaching, relief from the tedious chore of correcting papers appears to be a real contribution to the profession.

2. Versatility--The multiple-choice test can be used for a wide variety of subjects from arithmetic answers and simple definitions to complex discriminatory thinking. The test can be very easy or extremely difficult. Increasing the homogeneity of the alternative responses makes the test item more difficult (Green, 1963).

3. Reduction of guessing--One early objection to the objective tests was that they permitted guessing. But the possibility of correct guessing is reduced when the number of options is increased. Therefore, on items containing six or more options the element of guessing is negligible. On items containing less than six options a correction can be made for guessing by using the following formula (Ross and Stanley, 1954):

$$S = R - \frac{W}{n - 1}$$

S - corrected score
R - number of right answers
W - number of wrong answers
n - number of options

It can be noted that the evolution of tests and testing procedures through the years has been stream-lined, with emphasis on efficiency of administering and grading, as well as on validity and reliability.

Although there is an abundance of formal, standardized tests which use multiple-choice items, it is with informal, teacher-made, multiple-choice tests that this study was primarily concerned.

Problems and limitations of the multiple-choice test

Even though the advantages of the multiple-choice test are easily discernable and attested to, it is by no means free of some of the same limitations shared by other forms of testing. The multiple-choice test has not been able to eliminate the following variables that are known or suspected to influence the scores (Anastasi, 1964, p. 48):

1. Motivation
2. Physical and emotional health
3. Reasoning power
4. Degree of preparation for the test, i.e. study methods
5. Anxiety state, established empirically by Resse (1961) to be inversely proportional to the number of correct responses.
6. Use of test taking strategy
7. Quality of the test

It has been noted that knowledge is the strongest factor in obtaining high test scores (Huff, 1961). Perhaps it is unrealistic to expect to eliminate all other variables entirely, but any attempt to minimize the influence of negating factors would hopefully increase the validity of test scores. It is with the last two variables indicated above, i.e. use of test taking strategy and quality of tests that the present study was concerned.

(For detailed and lengthy instructions on how to construct multiple-choice tests, the reader is referred to Lindquist (1963), Thorndike and Hagen (1962), Remmers et al. (1960), Davis (1965), Hawkes et al. (1938), Orleans (1928), Furst (1961), Wrightstone (1956), Weitzman (1949), and Ross and Stanley (1954). There is considerable duplication and overlapping of recommendations in these authors, but Thorndike and Hagen seems to be the most comprehensive of those listed above.)

Cues

One of the criticisms regarding the use of multiple-choice tests is that they are very difficult to construct well (Orleans, 1928). It is not possible in this paper to summarize the skill of test construction, but the inadvertant inclusion of cues in test items is of interest here. A discussion of some of the types of cues identified by various writers and which were utilized in constructing the test for this study follows:

Cue 1--grammatical construction. Pettit (1960) and Ebel (1963) caution against inconsistency between the stem and the options. The stem combined with any one of the options should result in proper sentence structure. If the stem fails to make a perfect sentence when combined with one option, then that option can be eliminated as a possible correct answer. The most common error of this type is when the verb in the stem requires either a singular or plural option, as the case may be. For example, "The chief food of the people of China is _____" requires a singular answer. If three options are plural like "Meat and potatoes", and only one is singular, like "rice"; then, of course, "rice" is identifiable as the answer wanted.

Cue 2--grammatical construction involving the use of "a" or "an" in the stem, as these words suggest an option beginning with a vowel or consonant as the case may be. Thorndike and Hagen (1962) and Davis (1965) caution against this type of error. For example, "One example of a citrus fruit is an _____" indicates "orange" when the other options are plum, peach, and banana.

Cue 3--positive and negative determiners. Thorndike and Hagen (1962), Ebel (1963), Remmers et al. (1960), and Davis (1965) caution against the use of negative determiners such as "all", "none", "certainly", "never", "always". The words are absolutes and are more likely to be false than true. For example, "All bacteria cause disease", is predetermined as false because of the use of the word "all" even when the answer is not actually known to the student.

Wrightstone (1956) cautions against the use of moderately worded statements such as "generally", "usually", "most", and "often", as these are positive determiners. Ross and Stanley (1954) also add "may" and "as a rule" to the list of positive determiners.

A study closely related to the positive determiners mentioned above was made by Stone and Johnson (1959). The results of this investigation indicate that there is considerable conflict regarding the interpretation of relative terms used in test items to indicate frequency. Terms such as "hardly ever", "very seldom", and "very often", etc. proved to have different meanings to the 158 college students questioned.

Cue 4--position. Pettit (1960) and Chauncey and Dobbins (1963) suggest that most teachers tend to favor the "b" and "c" positions over the "a" and "d" positions, particularly where short answers or numbers are used as options. For example, "How many eggs are in a dozen?" would likely be followed by options in this order: a) 5, b) 10, c) 12, d) 15. Test makers are cautioned against making any consistent preference for positioning the correct answers. (Ross and Stanley, 1954).

Cue 5--"all of these". Thorndike and Hagen (1962) caution against using "all of these" as the last option in tests where "all of these" has not been used previously in the test as a wrong answer. The unexpected presence of this phrase makes it highly suspect as a correct answer.

Cue 6--negatives. Thorndike and Hagen (1962) and Weitzman et al. (1949) caution against using negatives in the stem without underlining or calling attention to it in some way, since a student who actually knows the answer may overlook the "not" and select the wrong answer on that basis. They indicate also that the double negative is confusing and should be avoided. This occurs when a negative appears in the stem and also in one or more of the options.

While it is not likely that these types of errors in test construction would appear, in abundance, in any one test, a poorly worded item could render the decision between pass or fail in borderline cases. The fact that the cues can be overlooked by some students and used profitably by others is reason enough for teachers to take steps to eliminate such errors in their test construction.

PROCEDURES

Hypotheses tested

Hypothesis 1. That sixth grade students who had read the subject matter prior to testing would not improve their scores significantly by using cues.

Hypothesis 2. That those sixth grade students who only read unrelated material prior to testing would not improve their scores significantly by using cues.

Hypothesis 3. That since knowledge is considered the strongest factor in producing high test scores, those Ss allowed to read the subject matter on which they were to be tested would score significantly higher than those Ss who read only unrelated material.

Hypothesis 4. That no learning would take place as a direct result of taking the test without cues before taking the test with cues. It was further hypothesized that "set" would not be established as a result of taking the test without cues before taking the test with cues.

Hypothesis 5. That there would be no significant difference between the percentage of improvement scores for those Ss who rated above the median on the California Achievement Test (CAT) and those below the median in CAT scores. (The percentage of improvement scores were computed by calculating the percent of incorrect answers on Test 1 that were changed to correct answers on Test 2.)

Population and sample used

The subjects for this study were drawn from the sixth grade classes at South Junior High School, Ogden, Utah. Sixth grade students were used for two reasons. One was their availability. The other was the fact that their behavior patterns and curriculum were familiar to this experimenter. Although all of the 190 students enrolled participated in the study, only the data obtained on 169 students with California Achievement test ratings were used in the final analysis.

Classification of subjects

All students were ranked from high to low according to their CAT scores which were on file in the school records. The tests had been given to all students by the school counselors in October 1965. The testing was performed as a matter of district policy at that time. Since such scores were to remain confidential, numbers were used instead of names for the subjects of this study.

For purpose of testing the hypotheses of the study, three groups of subjects were formed. In order to equate the groups as much as possible according to their past achievement, the names of the students, by CAT rank, were dealt like cards from top to bottom into Groups A, B, and C.

Explanation of the tests used

Test 1 contained forty multiple-choice items covering the chapter entitled "Sunlight, the Foodmaker" in the sixth grade science book. This test was written for the purpose of measuring knowledge of the subject matter. No cues were included. By means of this test, it was intended to eliminate knowledge as a variable in Test 2.

Test 2 contained forty multiple-choice items covering the identical subject matter as Test 1 (item for item); but this test was constructed purposely to include in every item a cue designed to aid the student in reaching correct answers that he had missed on Test 1. In the administration of these tests, no coaching was given and no hint in regard to the cues. The cues were mixed indiscriminately, so that no pattern would likely be detected by the Ss. It was the purpose of this study to determine the degree of test taking strategy that normally exists in Ss at this age level. There has been no instruction regarding the use of cues in their curriculum.

There were six different types of cues inserted into the items on Test 2, and those cues are described in detail in the "Review of the Literature" section of this paper. However, they are listed here by number in order that they may be identified in the exact test items where they are used.

1. Where the singular or plural option is indicated by the verb form used in the stem.
2. Where the presence of "a" or "an" in the stem indicates the option starting with or without a vowel, as the case may be.
3. Where a positive or negative determiner indicates the option wanted.
4. Where the selection of option in the "b" or "c" position is an advantage.
5. Where "all of these" is used in the "d" position as the answer wanted.
6. Where "not" is used in the stem without underlining.

Since the first three cues were of primary interest in this study, each of these was contained in 10 items.

Cue 1 was written into items 2, 5, 8, 11, 14, 17, 20, 26, 29, and 23.

Cue 2 was written into items 1, 4, 7, 10, 13, 16, 19, 22, 25, and 28.

Cue 3 was written into items 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30.

Cue 4 was written into items 34, 35, 36, 37, 39, and 40.

Cue 5 was written into items 33 and 38.

Cue 6 was written into items 31 and 32.

It was then assumed that those correct responses on Test 2 that had been missed on Test 1 were reached by means of using the cues. The complete tests will be found in the Appendix.

Treatment of the groups

Since the school principal preferred that all students be tested at the same time, he enlisted the aid of his counseling staff to administer the tests to the subjects. To insure uniformity in the experimental procedure for all three groups, detailed written instructions were given to each of the counselors in charge of a group and to the teachers of the subjects. These instructions are presented in the Appendix.

The three groups received the same treatment in the following respects.

1. They received the same explanation of the purpose of the study. The explanation was brief and intended only to satisfy the curiousity of the students, to allay the barrage of questions that may be expected from sixth grade students, and, hopefully, to provide a degree of motivation.

2. They were required to read for thirty minutes.

3. They were allowed one minute for standing and stretching.

4. They took two tests with 25 minutes allowed for each.

5. The instructions for marking the answer sheets and for each of the above steps was the same (see Appendix).

The three groups received different treatment in the following respects.

1. Group A read the chapter in their science books entitled, "Sunlight, the Foodmaker". This was the material on which they were subsequently tested. They took Test 1 first and then Test 2 in immediate succession.

2. Group B read material that had been previously assigned to them by their teachers. The only requirement for this reading material was that it be something other than science. This group took Test 1 first and then Test 2, in immediate succession.

3. Group C read material that had been previously assigned to them by their teachers. This material also was unrelated to science. They took Test 2 without taking Test 1 first.

Analysis of data

For making the comparisons in hypotheses 1, 2, 3, and 4 the "t" test was used to estimate the significance of the differences obtained.

For making the comparison in hypothesis 5 a fourfold contingency table was set up and the Chi Square value was determined by means of the following formula (Garrett, 1964):

$$\text{Chi Square} = \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)}$$

RESULTS AND DISCUSSION

Table 1 presents the descriptive statistics used for obtaining the "t" values of the comparisons in hypotheses 1, 2, 3, and 4. Figure 1 presents diagrams illustrating the results of the comparisons made in these four hypotheses.

Comparison of Group A on Test 1 and Test 2

In testing hypothesis 1 the comparison made between Test 1 and 2 for those Ss who read the subject matter prior to taking the test yielded a "t" value of 1.57, which was not significant. The null hypothesis could not be rejected.

Comparison of Group B on Test 1 and Test 2

In testing hypothesis 2, the comparison made between Test 1 and 2 for those Ss who read only unrelated material before taking the tests yielded a "t" value of 1.38, which was not significant. The null hypothesis could not be rejected.

The above two comparisons were the main points of interest in this study. Because of the nature of the tests, it was assumed that the greatest part of the difference obtained would be due to the use of cues.

Comparisons of Groups A and B on each test

In testing hypothesis 3, two comparisons were made. (1) A comparison on Test 1 for Group A who had read the subject matter and Group B who read only unrelated material yielded a "t" value of 3.23 in favor of Group A. (2) The comparison on Test 2 for Groups A and B

Table 1. Computations made to determine differences in the comparisons by "t" test, used in Hypotheses 1, 2, 3, 4.

	N	M	M^2	ΣfX^2	SD	SE_m
Group A Test 1	56	23.16	536.3856	31,761	5.55	.742
Group A Test 2	56	24.86	618.0196	36,536	5.89	.7874
Group B Test 1	57	19.64	385.7296	24,040	6.00	.7947
Group B Test 2	57	21.10	455.521	27,013	5.36	.7099
Group C Test 2	57	20.81	433.0561	25,652	4.12	.546

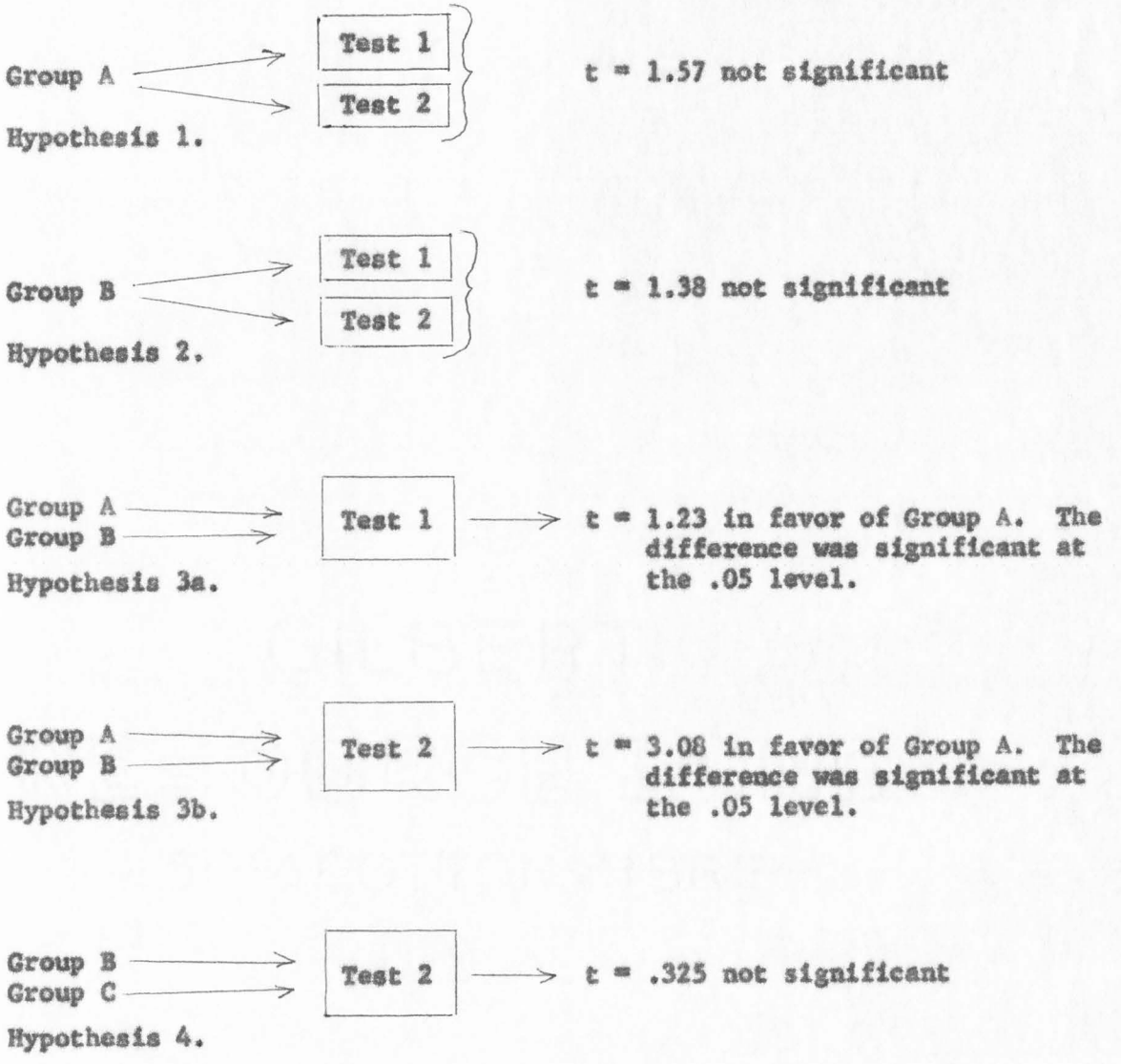

Figure 1. Diagrams of the comparisons made in hypotheses 1, 2, 3, and 4.

yielded a "t" value of 3.08 in favor of Group A. Both of these values were significant at the .05 level.

These comparisons were used as a control of experimental conditions. Since it is an established fact that knowledge is the strongest factor in achieving high test scores (Huff, 1961), compatibility of the findings with this statement indicated the attitude of the Ss; that is, as a whole they did attempt to make correct responses. Group A did score significantly higher than Group B on both tests, as would be expected.

Comparison between Group B and Group C on Test 2

In testing hypothesis 4 a comparison was made on Test 2 between Group B (who had previously taken Test 1) and Group C (who had not taken Test 1). The "t" value obtained for the difference was .325, which was not significant. The null hypothesis could not be rejected. This comparison was made for "set control" in order to determine if the Ss in Group B had acquired a response set while taking Test 1 which would make them inclined to select the same answers for Test 2 that they had selected on Test 1. The findings indicate that this did not happen.

Comparison between California Achievement Test ratings and percentage of improvement scores for Group B

In testing hypothesis 5 the comparison made between the percentage of improvement scores for those Ss above the median in CAT ratings and those below the median in CAT ratings yielded a Chi Square of zero. See Table 2. The Chi Square value indicates that the high scorers on the CAT rating did not benefit from the use of cues in a different way than the low scorers on the CAT rating. We see these results on the group as a whole, which does not rule out the possibility that some

Table 2. Comparison of percentage of improvement and CAT ratings on 56 students.

	Percentage of improvement scores*		
	Improved by 7%	Did not improve by 7%	
Ss who scored above 6 - 5.5 on CAT	A 15	B 13	28
Ss who scored below 6 - 5.5 on CAT	C 15	D 13	28
	30	26	56

$$\text{Chi Square} = \frac{N(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)}$$

$$= \frac{56(15 \times 13 - 13 \times 15)^2}{(15+13)(15+13)(15+15)(13+13)}$$

Chi Square = zero

*Note: The percentage of improvement scores were computed by calculating the percent of incorrect answers on Test 1 that were changed to correct answers on Test 2. The median for percentage of improvement was 7% for Group B.

The median score on the CAT was sixth year - fifth and one half month (6 - 5.5)

This comparison was analyzed for Group B only. The scores for Groups A and B were so similar that it was assumed that the Chi Square value would also be similar.

individuals did benefit from the use of cues. Attention to the following factors contributes plausability to the results found.

Limitations of sixth grade students

This investigation was considered only a small part in studying the extent to which all students make use of cues in tests. Since test taking strategy would be expected to increase with experience and coaching over the years, a significant step in understanding the implications of cues in tests would be to determine whether test taking strategy is manifest at the sixth grade level.

The use of cues involves a cognitive process that goes beyond the obvious problem of, "Do I know the answer to this question?" The thinking process would likely begin with, "I don't know this answer, so I must figure out some way to make the best guess." The degree of complexity of the thinking would vary considerably with each individual. Cues are inconspicuous for the most part, so much so that the teacher does not usually realize she has included them. In order to capitalize on them, a student has to be a relatively "deep thinker". This type of thinking falls into the period of formal operations, when, according to Hunt (1961), the individual's "central processes become sufficiently differentiated and autonomous to permit him to operate with the sum total of possibilities rather than with merely the empirical situation." The period of formal operations begins at the age of eleven to twelve, which would place the sixth grade student in the position of not having fully developed this deductive type of reasoning. Sixth grade students also fall into Piaget's Stage 2 (from 7-8 to 11-12) where "development of propositional logic cannot be handled." (Inhelder and Piaget, 1958)

The findings in this study were compatible with Piaget's experiments.

Limitations of the instruments used

Even though the two tests were revised many times to improve face validity, they are not represented as a perfect measure. The attempt to establish content validity and reliability was considered beyond the scope of this study. No attempt was made toward a complete item analysis. But, in case the tests be considered for further use, it is well to call attention to several items in which a limitation of the item may have influenced the response. On Test 2 item 14 there is a typographical error; the word "protein" should be plural. In items 15 and 18 the negative determiners (never, all, only) are not used in the same context usually found in test questions.

The cues written into the first thirty items appeared more obvious than the ones written into the last ten items. For this reason, a comparison was made to compare the difference in the percentage score between the first thirty and the last ten. Percentages were obtained by the following formula:

$$\frac{\text{Number of correct responses for all Ss for items 1-30}}{\text{Total possible number of correct responses for all Ss for items 1-30}}$$

A similar formula was used for items 31-40. The difference was .02, and the correlation was -.13. The difference was obviously not significant, so analysis was discontinued at that point.

Motivation

Motivation is known to influence test scores in general as proven in studies cited by Anastasi (1964). In the present study it was noted

that 23% of the Ss made higher scores on the "test without cues" than on the "test with cues". This was surprising at first glance, since it was felt that if the S knew the answer on the first test, he should have known the same answer on the second test. The reason for this discrepancy is likely attributed to motivation. The selections made on Test 2 could not have been made with much thought by these particular Ss, and it is possible that the same is true for Test 1. Random markings would be expected to produce results such as these. Therefore, motivation, as a contaminating variable, was apparent here and doubtless affected the over-all results.

SUMMARY

1. This study constituted an attempt to measure the extent to which sixth grade students at South Junior High School in Ogden, Utah benefitted from the use of cues purposely written into a science test.

2. One hundred sixty-nine students were classified into three groups equated by means of individual CAT scores. Groups A and B were used in the experimental procedure, and Group C as a control.

3. The apparatus used was in the form of two tests, identical in content, but different in that one contained cues and one did not. The "test without cues" was expected to identify the correct answers that had been reached primarily by knowledge of the subject matter. When this score was subtracted from the "test with cues", the difference was assumed to be attributable to the use of cues.

4. The findings failed to show a significant difference between test scores for either Group A or Group B. These two groups received the same treatment except that Group A had been required to read the subject matter prior to testing, and Group B had been required to read unrelated material.

5. The support of the null hypothesis in this investigation showed that this sampling of students at the sixth grade age level did not benefit from the use of cues as measured by the instruments used.

6. The comparison of scores on the "test with cues" between Group B and Group C showed no significant difference. These two groups received the same treatment except that Group B took the "test without cues" prior to taking the "test with cues", and Group C did not. These

findings indicated that "set" had not been a contaminating factor in the investigation and also that fatigue had not been a factor.

7. Since the above comparison ruled out "set" and fatigue as independent variables, failure to reject the null hypothesis in this study was tentatively attributed to the stage of development of cognitive functioning in the age group used, possible inaccuracy of the tests used, and lack of motivation.

8. It should be kept in mind, however, that the data reported in this paper deal with the averages of groups, and that they do not show whether or not certain individuals would benefit significantly from the use of cues.

9. If it could be demonstrated empirically that students do benefit from cues, the implications would be threefold: (a) that test scores should not be too heavily weighted in the evaluation process, (b) that teachers make an effort to improve the quality of their tests, and (c) that students develop a test-taking technique as defined in such books as On Your Own in College by Resnick (1963), which is now in use at Weber State College. It would seem important enough to merit further study with subjects at progressively advanced stages of mental development, particularly in view of the fact that the differences reported in this paper, although small, are in the same direction.

10. Another interesting area for investigation along this line is the effect of coaching on the use of cues.

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. New York: The Macmillan Company, 1964.

Beitner, Marvin S. Original not seen; cited in a letter as quoted by Hoffman, Banesh. Tyranny of Testing. New York: Collier Books, 1964, 80.

Buros, Oscar. Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Carroll, John B. Original not seen; cited in Research in Education: Where Do We Stand? Harvard Graduate School of Education Association Bulletin, Winter, 1960 as quoted by Hoffman (1964).

Chauncey and Dobbin. Testing and its Place in Education Today. New York: Harper and Row, 1963.

Davis, F. B. Educational Measurements and Their Interpretation. Belmont, California: Wadsworth Publishing Co., Inc., 1965.

Ebel, Robert L. Writing the Test Item, in Lindquist, E. F., Educational Measurement. Menosha, Wisconsin: George Banta Publishing Co., 1963, 190.

Furst, E. J. Constructing Evaluation Instruments. New York: Green and Co., 1961.

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

Green, John A. Teacher-made Tests. New York, Evanston, and London: Harper and Row, Publishers, 1963.

Hawkes, Herbert E., E. F. Lindquist, and C. R. Mann. New York: Prentice-Hall, Inc., 1938.

Hoffman, Banesh. The Tyranny of Testing. New York: Collier Books, 1964.

Huff, Darrell. The Strategy of Taking Tests. New York: Appleton-Century-Crofts, Inc., 1961.

Inhelder, Barbel and Jean Piaget. The Growth of Logical Thinking from Childhood to Adolescense. New York: Basic Books, Inc. Publishers, 1958.

Lee, J. Murray. A Guide to Measurement in Secondary Schools. New York: D. Appleton-Century Company, 1936.

Miller, Mungo. Original not seen; cited in a letter as quoted by Hoffman, Banesh. Tyranny of Testing. New York: Collier Books, 1964, 79.

Moore, James C., Richard Schutz, and Robert Baker. The Application of Self-Instructional Technique to Develop a Test-Taking Strategy. American Educational Research Journal. Vol. 3, No. 1, January, 1966.

Orleans, J. S. and Glenn A. Sealy. Objective Tests. Chicago: World Book Company, 1928.

Pettit, Lincoln. How to Study and Take Exams. New York: J. R. Rider Publishing Co., Inc., 1960.

Reese, Hayne W. Children's Manifest Anxiety Scale and an Arithmetic Achievement Test. Journal of Educational Psychology. Vol. 52, No. 3, 1961.

Resnick, W. C. and David H. Heller. On Your Own in College. Columbus, Ohio: Charles E. Merrill Books, Inc., 1963.

Remmers, H. H., N. L. Gage, and J. F. Rummel. A Practical Introduction to Measurement and Evaluation. New York: Harper and Brothers, Publishers, 1960.

Ross, C. C., and Julian Stanley. Measurement in Todays Schools. Englewood Clifts, New Jersey: Prentice-Hall, Inc., 1954.

Stone, David.R. and Richard R. Johnson. A Study of Words Indicating Frequency. Journal of Educational Psychology. Vol. 50, No. 5, 1959.

Thorndike, Robert L. and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1962.

Weitzman, Ellis, and Walter J. McNamara. Constructing Classroom Examinations. Chicago: Scientific Research Association, 1949.

Wrightstone, J. Wayne. Joseph Justman, and Irving Robbin. Evaluation in Modern Education. New York: American Book Company, 1956.

* * *

APPENDIX

APPENDIX

Directions given to the teachers whose students participated in the study

1. On Wednesday during your reading class, assign about 30 minutes of reading material to each of your students (all different or all the same whichever you prefer), and tell them that this is what they should read during the observation on Thursday.

2. The observation and testing will begin at 9:00 Thursday.

3. Tell the children they are asked to participate in a study of sixth grade children. They will be asked to read for 30 minutes and then take 2 multiple-choice tests.

4. Since it is a test of the whole group and not individuals, the student shall be known by a number only. They should remember their own number and sign that number on their answer sheets--not their names.

5. Give each student his number; then read off all students in Group A. Instruct each student to stand as his number is called. When they are all stsnding, ask them to take a pencil and science book to rooms 114 - 115.

6. When Group A has left, read the numbers for Groups B & C in the same way and have them take their reading material and a pencil to rooms 120 - 121 for Group B and rooms 118 - 119 for Group C. They will receive further instructions in the testing rooms.

Instructions given to the counselors, who assisted in the testing procedures

1. Call the roll (numbers only) of the students in your group. Be sure you have no one who belongs in a different group. Check to see that each one has reading material and a pencil.

2. Tell or read to the students-

"You are asked to participate in a study of sixth grade students in the anility to take tests. No names will be used in the study. It is important that you do your very best. The results of this observation are expected to show how students may make better test scores in the future and to show teachers how to make tests that are fair to everyone. I can not explain all the details to you at this time, but you can see how your honest efforts are necessary in order to make the results meaningful.

First you will have 30 minutes to read the material your teachers assigned to you. Then you will take 2 tests. Some of

you may not finish reading your assignment. Others may finish before the time is up. In that case, you may read parts of it again to study it more thoroughly, or you may close your books and sit quietly till the time is up. The only thing we ask is that you remain in your seats and do not disturb the others.

You may now begin to read, and I shall stop you at the end of 30 minutes."

3. At the end of the reading period, say - "Close all books and stretch for one minute. Stand beside your chair, but do not talk to your neighbor."

4. Pass out answer sheets. You will note that the answer sheets have "CAT Arithmetic Tests" printed on them. Assure the students that this is not an arithmetic test. Have the students write their numbers on the answer sheets.

5. Explain how they are to select their answers and blacken the appropriate space on the answer sheet. Let them do the two "for practice only" problems. The small block for samples A and B should be crossed out and not used by the students.

6. Pass out the test papers and explain that they are not to mark on these. They must put all answers on the answer sheets. Ask them to look over the papers to see if all the print is clear. There should be no questions asked after they start the test. Tell them, "Even though you may not know some of the answers, try to figure out which is the right answer and make as high a score as you can. There is no penalty for guessing. Answer all the questions that you have time for. Start with #1 on your answer sheets and go through #40. If you finish before the time is called, you may go over your work or just sit quietly."

7. Note the time and allow 25 minutes. While they are taking the test, walk around and see that everyone has his number on the answer sheet.

8. Collect the tests and exchange tests with the other group.

9. Distribute the next test without delay and proceed the same as for the first test. Allow 25 minutes.

10. At the end of the test, collect all tests and answer sheets.

11. Thank the students and have them return to their home rooms.

Test I

1. Sugar is -
   a) a by-product of physical changes
   b) an ingredient the plant needs to make proteins
   c) a mineral product
   d) a low-calorie diet food

---

2. Foods for snacks which are not so likely to spoil our appetites at mealtime are -
   a) bread and jelly sandwiches
   b) chocolate cake and pie
   c) fruits and popcorn
   d) none of these

---

3. Carrots contain -
   a) vitamins A & C
   b) vitamin B

---

4. One item of food that contains the materials for building bones, muscles and blood is -
   a) a grain of corn
   b) seaweed
   c) an egg
   d) lettuce

---

5. Which energy rays from the sun are used by plants in making food?
   a) red and violet
   b) red and green
   c) green and violet
   d) all of these

---

6. What gas does growing grain return to the air?
   a) carbon dioxide
   b) oxygen

---

7. An Alaskan fur seal may eat -
   a) a crab
   b) an octopus
   c) a walrus
   d) a porpoise

---

8. Foods which contain a great amount of starch are -
   a) molasses and honey
   b) corn and potatoes
   c) beef and pork
   d) milk and eggs

---

9. When lake water evaporates into the air -
   a) it can not return to earth
   b) it will fall to earth again some day

---

10. Yeast is -
    a) an amoeba
    b) a cluster of many tiny animals
    c) an animal
    d) a kind of fungi

---

11. The chief food of the people of Mexico is -
    a) rice
    b) cheese
    c) corn
    d) wheat

---

12. A test used by chemists to detect the presence of carbon dioxide is with -
    a) litmus paper
    b) limewater

---

13. A kernel of grain contains a tiny baby plant called -
    a) a seed germ
    b) an embryo
    c) a yolk
    d) wheat

14. The food making machinery of a plant is its -
a) protein
b) thorn
c) blossom
(d) chlorophyll

15. Carbon dioxide is found in human breath mostly during -
(a) exhaling
b) inhaling

16. The lack of a particular vitamin in your diet may cause -
a) an allergy
(b) a deficiency disease
c) an attack of appendicitis
d) any of these

17. The ingredients necessary for plants to make food are -
a) nitrogen and hydrogen
(b) sunlight and carbon dioxide
c) oxygen and iron
d) all of these

18. Fat contained in the cells will produce a waxy "raincoat" on the leaves of -
a) cactus
b) mushrooms
(c) apple trees
d) cactus & mushrooms

19. Because fat has more fuel value than sugar and starch -
a) it forms the largest portion of the diet of people living in tropical climates.
(b) it is an especially useful in cold climates.
c) it is the quickest energy food.

20. Which minerals are especially important for bone growth?
a) iron and nitrogen
(b) calcium and phosphorus
c) calcium and nitrogen
d) potassium and calcium

21. Protein may be found in -
a) dead cells
(b) living cells
c) living cells and dead cells
d) neither of these

22. Almost everywhere in the world, grain is -
a) a product very difficult to raise.
b) a plant that requires heavy rainfall.
c) used to feed chickens, cattle, and hogs more than humans.
(d) an important food crop.

23. Which mineral is needed in our red blood cells to carry oxygen?
a) phosphorus
b) lead
(c) iron
d) potassium

24. Molds are -
(a) useful to man when making penicillin
b) white
c) poisonous to touch
d) sweet smelling

25. A lime is -
(a) an example of citrus fruits
b) a source of vitamin B
c) a small sweet lemon
d) a product grown mostly in cold climates

---

26. Water and carbon dioxide are changed into sugar in the bodies of -
(a) plants
b) animals
c) bacteria
d) all of these

---

27. Chlorophyll may be found in -
a) mushrooms
b) fungi
c) bacteria
(d) oak leaves

---

28. If you get the odor of burning feathers when you burn a piece of food, then that food contains -
a) acid
b) alkali
c) sugar
(d) protein

---

29. The gas contained in the smoke of burning wood and coal is -
a) oxygen
(b) carbon dioxide
c) sulphuric acid
d) hydrogen

---

30. You should eat candy -
a) when you are hungry
b) before meals to increase your appetite
c) during your meals
(d) after meals as a dessert

---

31. Which of these foods are seeds?
a) potatoes and carrots
b) lettuce and tomatoes
(c) peas, beans and nuts
d) asparagus

---

32. Water -
a) undergoes a chemical change when it turns to ice.
b) becomes lost forever as it evaporates.
(c) keeps traveling in an endless cycle.
d) is a pure element found in nature.

---

33. The energy found in the meat we eat -
a) was taken by plants from the minerals of the ground.
b) was absorbed by the animals from the sunlight.
c) actually comes from sugar content of the meat.
(d) was taken from the plants which were eaten by the animal.

---

34. Which vitamin is sometimes called the sunshine vitamin?
a) A
b) B
c) C
(d) D

---

35. Four loaves of bread have about the same fuel value as how many heads of lettuce?
(a) 100
b) 500
c) 710
d) 1000

36. Chlorophyll can be recognized by its -
(a) color
b) taste
c) smell
d) all or none of these

37. The sprouting of a seed while it is still using its stored food is called -
(a) germination
b) pollinization
c) fermentation
d) all or none of these

38. Old leaves and dead wood soften and decay because of the work of -
a) bacteria
b) fungi
c) moisture and warmth
(d) all or none of these

39. Which of these foods is grain?
a) beets
(b) oats
c) carrots
d) all or none of these

40. What part of a plant supplies most of the minerals and water for the plant?
a) leaves
b) blossoms
(c) roots
d) all or none of these

* * *

Test II

1. Sugar is an -
   a) by-product of physical changes
   (b) ingredient the plant needs to make protein
   c) mineral product
   d) low-calorie diet feed

---

2. Foods for snacks which are not so likely to spoil our appetites at mealtime are -
   a) a sandwich
   b) a piece of cake
   (c) fruits and popcorn
   d) none of these

---

3. Carrots usually contain both vitamin A and vitamin C.
   (a) true
   b) false

---

4. One item of food that contains the materials needed for building bones, muscles and blood is an -
   a) grain of corn
   b) seaweed
   (c) egg
   d) lettuce

---

5. Which energy rays from the sun are used by plants in making food?
   (a) red and violet
   b) green
   c) red
   d) violet

---

6. All plants return carbon dioxide to the air.
   a) true
   (b) false

---

7. An Alaskan fur seal may eat an -
   a) crab
   (b) octopus
   c) walrus
   d) porpoise

---

8. Foods which contain a great amount of starch are -
   a) honey
   (b) corn and potatoes
   c) meat
   d) milk

---

9. When lake water evaporates into the air, -
   a) it can never return to earth.
   (b) it sometimes remains in the air for many days.

---

10. Yeast is a -
    a) amoeba
    b) animal cluster
    c) animal
    (d) kind of fungi

---

11. The chief food of the people of Mexico is -
    a) fish and rice
    b) crackers and cheese
    (c) corn
    d) wheat and oats

---

12. A test used by chemists to detect the presence of carbon dioxide is -
    a) always with limewater
    (b) sometimes with limewater

13. A kernel of grain contains a tiny baby plant called an -
a) seed germ
(b) embryo
c) yolk
d) wheat

14. The food making machinery of a plant is -
a) protein
b) thorns
c) blossoms
(d) chlorophyll

15. Carbon dioxide may be contained in human breath -
(a) mostly during exhaling
b) never during inhaling

16. The lack of a particular vitamin in your diet may cause a -
a) allergy
(b) deficiency disease
c) attack of appendicitis
d) any of these

17. The ingredients necessary for plants to make food are -
a) nitrogen
(b) sunlight and carbon dioxide
c) oxygen
d) iron

18. Fat contained in the cells will produce a waxy "Raincoat" on the leaves of -
a) all cacti
b) mushrooms only
(c) apple trees
d) cacti and mushrooms only

19. Because fat has more fuel value than sugar and starch, it is an -
a) forms the largest portion of the diet of people living in tropical climates.
(b) especially useful food in cold climates
c) quick energy food

20. Which minerals are especially important for bone growth?
a) iron
(b) calcium and phosphorus
c) nitrogen
d) potassium

21. Protein may be found in -
a) dead cells only
(b) living cells of both plants and animals
c) plants only
d) animals only

22. Almost everywhere in the world, grain is an -
a) product very difficult to raise.
b) plant that requires heavy rainfall.
c) used to feed chickens, cattle, and hogs more than humans.
(d) important food crop.

23. Which mineral is needed in our red blood cells to carry oxygen?
a) calcium and phosphorus
b) lead and zinc
(c) iron
d) potassium and calcium

24. Molds are -
(a) sometimes useful to man in making penicillin.
b) always white
c) always poisonous to touch
d) all sweet smelling

25. A lime is an -
(a) example of citrus fruits
b) source of vitamin B
c) small sweet lemon
d) product grown in cold climates

26. Water and carbon dioxide are changed into sugar in the bodies of -
(a) plants
b) an animal
c) a bacterium
d) none of these

27. Chlorophyll may be found in -
a) mushrooms only
b) all fungi
c) only bacteria
(d) oak leaves

28. If you get the odor of burning feathers when you burn a piece of food, then that food contains a -
a) acid
b) alkali
c) sugar
(d) protein

29. The gas contained in the smoke of burning wood and coal is -
a) oxygen and nitrogen
(b) carbon dioxide
c) sulphuric acid and oxygen
d) helium and hydrogen

30. You should eat candy -
a) never
b) any time you are hungry
c) only during meals
(d) after meals as a dessert

31. Which of these foods is not a seed?
a) nuts
b) peas
(c) carrot
d) beans

32. Water is not -
a) a liquid
b) found in the atmosphere
(c) a pure element found in nature
d) capable of being reused by plants and animals

33. The energy found in the meat we eat -
a) came first from the sun
b) was taken into plants by their chlorophyll
c) was taken from the plants which were eaten by the animals
(d) all of these

34. Which vitamin is sometimes called the sunshine vitamin?
a) A
(b) D
c) B
d) C

35. Four loaves of bread have about the same fuel value as how many heads of lettuce?
a) 10
(b) 100
c) 500
d) 1000

36. Chlorophyll can be recognized by its -
a) taste
(b) color
c) smell
d) all of these

37. The sprouting of a seed while it is still using its stored food is called -
a) pollinization
b) germination
c) fermentation
d) all of these

---

38. Old leaves and dead wood soften and decay because of the work of -
a) bacteria
b) fungi
c) moisture and warmth
d) all of these

---

39. Which of these foods is not grain?
a) wheat
b) oats
c) beets
d) none of these

---

40. What part of a plant supplies most of the minerals and water for the plant?
a) leaves
b) blossoms
c) roots
d) all of these

* * *