1976

The effect of test orientation training on children's achievement test scores.

Sarah. Powell
University of Massachusetts Amherst

THE EFFECT OF TEST ORIENTATION TRAINING ON CHILDREN'S ACHIEVEMENT TEST SCORES

A Thesis Presented

by

SARAH POWELL

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

December, 1976

Psychology

THE EFFECT OF TEST ORIENTATION TRAINING ON CHILDREN'S ACHIEVEMENT TEST SCORES

A Thesis Presented

by

SARAH POWELL

Approved as to style and content by:

Ronald K. Hambleton, Chairperson of Committee

James M. Royer, Member

J. Gregory Olley, Member

James R. Averill, Assoc. Chairperson
Department of Psychology

ACKNOWLEDGEMENTS

I would like to express my appreciation to my chairperson, Ron Hambleton, for his help, encouragement and patience. And to my committee members, Mike Royer and Greg Olley, go my thanks for their supportiveness and thoughtful criticism.

My very special thanks go to Mr. Charles Burack, principal of Elm Park Community School and to Miss Eleanor Looney, principal of Belmont Street Community School, for their most gracious cooperation, without which this study would not have been possible. And to the teachers in those schools who conducted the training program, I am deeply indebted for their enthusiasm, their friendliness, and their time and effort. It was a pleasure to work with them.

I would also like to express my gratitude to Chris, who gave me emotional support throughout this project and who, even more than Ron Hambleton, bears the brunt of my neurotic foibles.

Finally, I acknowledge all my friends of the Pioneer Valley Song and Dance Society, without whose constant distraction I might have finished this thesis a long time ago.

ABSTRACT

A training program was developed for teaching test orientation (TO) skills to disadvantaged children. The skills selected for training were: (a) use of machine-scorable answer sheets; (b) following oral and written test instructions; (c) pacing and careful use of testing time; and (d) using partial knowledge to guess wisely. In addition, training was designed to raise test-taking motivation. It was argued that TO skills such as these are logically independent of academic achievement, that the exercise of such skills affects performance on standardized academic achievement tests, and therefore that individual differences in TO skills contribute to invalidity of achievement test scores. It was hypothesized that a TO training program would bring students to a more uniform skill level and thus: (a) raise test scores overall, and (b) increase the validity of scores. The training program was administered to 93 fourth- and fifth-grade students in two inner-city schools in Worcester, Massachusetts. Another 96 students served as controls. Classroom teachers conducted the training sessions. All subjects took the Vocabulary subtest of the California Achievement Tests pre- and post-training.

Training had no effect on mean Vocabulary score, either overall or for selected subgroups. Nor was there any effect on reliability or predictive validity (against the criterion of class rank in reading achievement) of test scores. Two TO variables, number of answer-sheet

marking errors and number of items omitted, were also unaffected. The failure of training to produce the expected results was interpreted as due to the inadequacy of the treatment rather than as refuting the hypothesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Background

Almost anyone who has been administered an objective test would agree that one could develop skills for taking such tests, skills that are unrelated to one's knowledge of the test content. Writers in the testing field have long been concerned that unequal test sophistication among examinees may be a source of invalid test variance. If, indeed, this is the case, then it would seem to present an important problem for standardized testing in education. Goslin (1963; also Goslin, Epstein, & Hallock, 1965) has surveyed testing programs in elementary schools and reports a growing tendency to use standardized tests in the diagnosis of individual differences and placement of children into particular classes at an early age. Since early diagnosis and placement must profoundly influence the child's educational opportunities and growth, the effect of test sophistication on the validity of standardized tests is a question of some practical significance.

There have been two main foci of research concerned with test sophistication. One body of work is concerned with the effects of rather broadly defined coaching procedures on test performance; the other examines the influence of training in specific elements of test wiseness (TW), as defined by Millman, Bishop, and Ebel (1965). In this chapter, the literature on coaching and TW is summarized briefly. The

issue of TW and test validity is discussed, and the existing research is criticized on the grounds that it has not adequately dealt with the validity question. A rationale is then developed for a new approach to remedying differences in test sophistication in children. This approach will be called a test orientation procedure. Finally, the purposes of the present investigation are delineated.

Coaching

Research on TW, and on the related topic of coaching, has focused mainly on whether test scores can be increased by training. The work on TW of the last decade is antedated by the coaching research of the 1920's and 1950's, the concern of which seems to have been primarily practical: Can the scores of some students on standardized intelligence tests be unfairly boosted by practice or coaching on tests or test-like materials? It was thought that something like "test sophistication" might result from such training. The term "coaching" as it is used in the literature apparently subsumes a variety of treatments designed to improve test performance. Coaching can mean drill on the actual test items, drill on parallel items, familiarization with a test-like situation, discussion of the principles underlying item types, and/or training in the concepts or material to be tested. Those coaching procedures not involving the actual test content were somewhat similar to procedures used later to teach TW skills.

Findings of the early coaching work are somewhat inconsistent and,

because the training procedures are incompletely documented, rather difficult to interpret. Perhaps because the early investigators were concerned simply with whether it was possible to raise scores, rather than with the specific factors responsible for gain, their descriptions of training procedures tend to be too gross to permit identification of some important dimensions (for example, it is usually unclear whether instruction concerned item format or item content, nor is it clear whether students were given feedback on the correctness of their responses during training, nor whether there were any reinforcement contingent on correctness of responses). It has generally been found, however, that coaching produces significant mean gain on both individually-administered intelligence tests (Casey, 1928; K. Greene, 1928; Holloway, 1954; Kinnie & Sternlof, 1971) and a group test (Dempster, 1954; Wiseman, 1954). At least one study though (Davidson, 1928) did not find consistent superiority for the coached group. The studies just mentioned used school children at different age levels ranging from preschool to about eleven years. Because different age levels, tests, and procedures were used in these studies, it is difficult to generalize about the size of the coaching effect, though the mean increase is probably fairly small. Results on the duration of the coaching effect are very inconclusive. It is fairly clear that practice on test-like materials is an important element in coaching. Wiseman (1954) compared (a) combined coaching and practice, (b) coaching alone, and (c) practice alone, for 10-year-old children.

Coaching without practice resulted in only a slight mean improvement over controls, while practice without coaching produced a larger gain and the largest gain resulted with the combined treatments. In addition, greater coaching gains seem to result on non-verbal types of tests (Kinnie & Sternlof, 1971; Vernon, 1954). Kinnie and Sternlof found the significantly greater increase of their treatment groups' WPPSI scores was due entirely to improvement on the performance scales. Another fairly strong indication is that the size of the treatment effect is related to the similarity of the coached materials to the target test. Holloway (1954) and Kinnie and Sternlof (1971) found greater gains when coaching was on test items employing the same format (though different content) as the target test. Greene (1928) found greater gains when coaching was on the actual test items (i.e., identical content) than when it was on items with similar format but different content. Finally, there is some indication, though weak, that the size of the increase due to coaching may be negatively related to the age of the child (Casey, 1928), and that within age levels gain is unrelated to initial ability level (Holloway, 1954; Wiseman, 1954).

Test Wiseness

Millman, Bishop and Ebel (1965) have defined test-wiseness for objective tests as "...a subject's capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score. Test-wiseness is logically independent of the examinee's

knowledge of the subject matter for which the items are supposedly measures." Since the Millman et al. analysis, a number of investigations have focused on specific elements of TW. The usual approach has been to examine the effects of training in TW skills.

Millman et al. (1965) identified two major classes of test-taking skills, the first independent of, the second dependent on the test constructor and/or purpose. Skills in the first class are essentially technical skills such as strategies for pacing oneself and for avoiding careless errors, while skills in the second class include deductive strategies (e.g., choose neither or both of two alternatives each of which implies the correctness of the other) and cue-using strategies (e.g., use information contained elsewhere in the test to infer the correct answer to an item). The use of the second type of skill then is to some extent dependent on poor test construction, and thus less applicable to standardized tests.

Two studies using older students and providing training in the second class of skills outlined by Millman et al., have employed direct measures of TW to assess training effectiveness (rather than looking at the effect on scores on some target test). Measures of TW are typically constructed by writing multiple-choice items which deliberately incorporate cues to the keyed response. The item content is fictitious so that there is no truly correct answer; the keyed option can be selected only by applying a particular TW skill. Gibb (1964) trained college students to recognize seven types of secondary cues in

multiple-choice items (e.g., resemblance between the stem and the correct option, absurd options, and the use of specific determiners such as "always"). The trained students scored significantly higher than controls on a 70-item history test incorporating the cues. Slakter, Koehler and Hampton (1970) trained 12th graders in four TW skills; the trained group scored significantly higher than controls on three of the four skills.

Moore, Schutz and Baker (1966) used a different method to directly measure TW. They trained 8th graders in pacing and guessing strategies appropriate for different testing conditions (speed versus power tests, penalty for guessing versus no penalty). They used as a dependent measure the number of items attempted on a standardized test given under four different combinations of speed and penalty instructions. The trained group did vary its strategy to fit the test conditions, while controls did not. Similarly, Slakter et al. (1970) trained a group of 12th graders to "always respond". This group responded to a significantly greater proportion of fictitious items than did controls.

It seems fairly clear, then, that older students at least, can be taught to better utilize cues deliberately included in multiple-choice tests. A study by Wahlstrom and Boersma (1968), however, suggests that such training is not particularly helpful for taking normal tests. Ninth graders were taught test-taking strategies from Millman's outline. Trained students performed significantly better than controls on a test

incorporating common multiple-choice-item faults, but on the same test without the faults there was no treatment group difference. Since standardized tests typically contain very few of these faults or cues, the importance of the second category of TW skills in standardized testing is questionable.

A somewhat different approach to TW training has been employed with younger children. Three investigators have tried to improve the standardized test performance of younger students by training them in TW skills from the first category of the Millman et al. outline (following instructions, pacing, etc.), and by providing them with practice on test-like materials. TW skills of the first category should be more appropriate and useful for standardized tests, while practice was expected to be helpful to children having limited experience with the formats of standardized tests. Tinney (1968) trained high- and low-SES fifth graders on materials adapted from the target test, the New Developmental Reading Test. As expected, high-SES children scored higher. However, there was no main effect of the training. Oakland and Weilert (1971) trained disadvantaged preschool children, pre- and posttesting them on the Metropolitan Reading Readiness Test. Trained children showed only very slightly greater improvement, and only on certain subtests. Callenbach (1973) gave second graders training based on the format of the Stanford Reading Test. The experimental group did gain significantly more than controls, but the difference was actually quite small. Unfortunately, none of these studies

directly assessed the abilities they sought to train. It cannot be determined, therefore, whether training failed to raise scores substantially because it was ineffective in teaching the desired skills, or for some other reasons.

The Problem

More important, of course, than the question of whether the mean score of a group can be raised, is whether individual differences in TW lower the validity of test scores. It may be that only certain individuals will benefit in terms of score gain from TW training. Other individuals who are not deficient in TW may not benefit, yet the validity of the test will still be increased. What this suggests is the need to identify those individuals lacking in TW.

The preceding paragraph contains the implicit assumption that deficiencies in TW do in fact make a person's test score less valid. Whether this is a correct assumption is still an open question. The question may be approached logically in terms of construct validity, and empirically in terms of criterion validity. Of achievement tests, it will ordinarily be true that technical skills such as correctly using an answer sheet, or knowing when it is expedient to guess, are irrelevant to the construct to be measured (i.e., mastery of a particular content or curricular area). To the extent that the lack of such skills prevents any person from expressing his/her true mastery of the content tested, the construct validity of the test will be lowered.

Similarly, if an achievement test affords opportunities to the test-wise student to infer correct answers through deductive and cue-using strategies, construct validity will probably suffer. In practice, this second type of TW is unlikely to be a threat to the construct validity of achievement tests, since careful test construction will usually eliminate secondary cues to the correct responses. The most practical approach to the validity problem posed by technical TW skills would seem to be to ensure that all persons to be tested have the requisite skills, providing skill training where it is needed.[1]

How TW affects criterion validity is, of course, an empirical question. To this writer's knowledge, the question has been addressed directly in only two studies. The correlation of ability test scores with teachers' ratings of ability was higher after coaching on the test for retarded children (Schuchman, 1960) and for immigrant children (Ortar, 1960).

For criterion validity to be increased by TW training, the scores of different individuals must be increased by different amounts, since

---

[1]The question of the effect of TW on the construct validity of ability tests is less straightforward, depending on the theoretical formulation of the ability to be measured. Difficulty introduced by novel item formats, for example, may or may not be relevant to the construct of interest.

if all the scores are increased by the same amount, the correlation with the criterion will remain unchanged. Some evidence, however, suggests that the effect of training is an additive one; in two studies, the correlations between pretest and posttest scores were reduced only very slightly by interposed training (Callenbach, 1973; Dempster, 1954). However, there are two other possible explanations for the apparent additive treatment effect. The students in each of these studies may have been unusually homogeneous in TW before training; each group was, in fact, quite homogeneous with respect to educational background. Alternatively, the training provided may not have been optimally effective for those most deficient in TW, resulting in smaller gains for them than potentially could be achieved.

The foregoing suggests that if training in TW is to increase test score validity, those individuals most deficient in TW must be identified. It might then be possible to determine reasons for the deficiency and, most important, to tailor the training procedure so as to make it optimally effective for that group. A major shortcoming of TW research to date is that it has not attempted to do this.

In the absence of any data, the obvious suggestion presents itself that the least test-wise individuals are those who have had little experience taking standardized tests (primarily young children). This possibility bears investigation, though some findings on practice effects suggest that lack of experience may not be a critical factor. Practice gains on verbal tests are typically fairly small (for example: Frankel, 1960; Kinnie & Sternlof, 1971; Levine & Angoff, 1958), while

gains on non-verbal tests tend to be greater but are probably related to factors other than developing TW (E. Greene, 1937).

Another population which might plausibly be expected to lack TW is the educationally disadvantaged (primarily lower socio-economic status persons). It is known that SES and standardized test performance are positively related; perhaps the poorer performance of low SES students is partly due to their being less test-wise. Some support for this notion is found in Jensen's (1970) citation of evidence that low-SES children perform relatively more poorly on group intelligence tests as compared to individual tests, than do middle-SES children. This suggests that some factors in the group test administration, possibly related to TW, interfere more with low-SES performance. If this were true then TW training programs should increase the scores of low-SES children more than those of middle- and high-SES children. In fact, though, neither Kinnie and Sternlof (1971) nor Holloway (1954) found any SES-level difference in gain after training. Tinney (1968) actually found that low-SES children gained less. However, it may be that low-SES children <u>are</u> the least test-wise, but that training programs used to date have not met special needs of this group and thus have not been effective for them. Tinney, in fact, noted his subjective impression that the low-SES subjects responded to training with much less apparent motivation or interest.

Tinney's remarks and several findings on low-SES test performance suggest that motivation may be very important to the success of a

training program. Eells et al. (1951) analyzed errors made on an IQ test. Low-SES children appeared to have done more "blind" guessing, and middle-SES children more "educated" guessing. Anastasi (1968) has noted that observation of children taking tests suggests that lower-class children work more rapidly and carelessly. One conclusion that may be drawn is that low-SES children are less motivated to do well on tests (although it is also possible that their behavior is the result of simply knowing less of the tested material). A working hypothesis of the present study was that test motivation, in addition to directly facilitating success on tests, is a necessary element for the development of TW. Any training program, then, that is intended to teach TW, must also increase test motivation if it is to succeed with disadvantaged children. Zigler and Butterfield (1968) have advanced the notion that culturally deprived children are less motivated to be correct for its own sake, more motivated to seek praise and attention from adults, and more motivated to seek tangible rewards. This is in accord with other research showing that the test performance of retarded and minority group children can be raised significantly by offering money or tokens for correct responses (Ayllon & Kelly, 1972; Klugman, 1944). These findings suggest ways in which past efforts to equalize TW have not been designed so as to be effective for those individuals who most lack it. Unless this can be done, it will probably not be possible to eliminate the invalidating influence of differential test sophistication.

Purposes of the Investigation

The purposes of the present study were to develop a program for training children in test orientation, and to study the effects of this training on standardized achievement test performance and validity. Test orientation (TO) is conceived here as consisting both of technical elements of test-wiseness (facility with tests and answer sheets), and elements of workstyle and motivation (attending carefully to all instructions; guessing wisely; persisting in the face of difficult problems; working steadily). As discussed in the previous section, the latter elements are thought to be deficient particularly in lower-SES children. The training program was intended primarily for use with lower-SES children, rather than for the general population of school children. In this respect, the present study differs from past test-wiseness training studies.

Further, the program was designed so that, if successful, it could be practically put to use by classroom teachers working with groups of students.

It was intended that the skills taught be specific to test-taking and that training not have any generalized effect on academic performance. While it is probable that some aspects of TO are general characteristics which also facilitate academic achievement (e.g., use of efficient problem-solving strategies, responsiveness to achievement cues, attentiveness), the purpose here was to treat those aspects of TO which are most likely to affect test validity, that is, those skills

specific to test-taking.

This study examined fairly gross effects of training. It was not possible to assess the contributions of isolated components of training, but simply to determine whether test scores and test validity could be increased by giving training in all the skills expected to be most beneficial. One reason for taking this approach was that it was desirable to produce a training program of immediate practical value. Another reason was that the present study was preliminary; if the treatment package were successful in raising scores, its component parts could be evaluated separately in some future research project.

In addition to examining training effects on the test scores, preliminary analyses were made of effects on several other variables. These variables, including number of items omitted and number of marking (i.e., non-content) errors, were examined in an exploratory fashion. If training were successful in raising scores, any observed differences between trained and untrained students on these variables might help illuminate the nature of test orientation.

CHAPTER II

METHOD

Sample

The subjects were the fourth- and fifth-grade students (approximately 250) in two inner-city elementary schools in Worcester, Massachusetts. Most of the students in these schools are from lower-SES families. Seventy percent of the subjects have families whose self-reported gross income makes them eligible for the Federal free lunch program (e.g., less than $8810 for a family of eight).

The training program was designed for fourth- and fifth-graders because most standardized tests are first administered with separate answer sheets at that level. Since there is some evidence that children have difficulty using separate answer sheets, training in their use at the earliest age was expected to be beneficial.

For those students who had been in the Worcester Schools since kindergarten, previous testing experience included the Metropolitan Reading Readiness Test (kindergarten), and the complete Stanford Achievement Battery and the Otis Quick-Scoring Mental Abilities Test (in grade 3).

After random assignment to the two treatment conditions had been made, it was found that some of the students assigned to the training condition would be unable to participate due to one or more of the following reasons:(a) inability to read at the level required for the

practice tests;(b) inability to understand spoken English at the level required for the instructional portion of training;(c) emotional problems. Teachers were therefore requested to identify these students as well as similar individuals in the control group. All such identified students were then dropped from the study. Eleven training-group and six control-group subjects were dropped.

Instruments

Standardized test. Two subtests of the California Achievement Tests (CAT), Level 3 (grades 4-6), 1970 edition, were administered to all subjects pre- and post-training. The CAT was chosen as being representative of achievement batteries commonly used in elementary schools. The Reading Vocabulary and Arithmetic Concepts subtests were chosen as representing two important subdivisions of the battery, while requiring only a small amount of testing time (10 and 7 minutes respectively). Form A was given as the pretest and Form B as the posttest. Due to prohibitive cost, a Xeroxed fascimile was used in lieu of the publisher's answer sheet. The fascimile was designed to resemble a machine-scorable answer sheet as much as possible (a copy is presented in Appendix A).

The completed answer sheets (without identification as to treatment group) were hand scored by the experimenter and an assistant. The raw score was the number of items correct, with no correction applied for guessing. However, items judged by the scorers to be improperly

marked (e.g., with more than one option marked, or a mark barely dark enough to be seen) were considered incorrect. The goal of the procedure was to imitate the scoring of an optical-scanning device. It was assumed that if poor answer-sheet-marking skills hamper test performance, it is at least partly because correctly answered but improperly marked items are discounted by the scoring machine. It was therefore necessary in order to assess the effect of TO training on this particular problem, to try to make machine-like scoring decisions. This meant that some items were scored wrong even though the apparent intended option was correct. For example, if two options were marked and one crossed out, the item was scored wrong even if the remaining marked option was the correct one.

Criterion measure. Each teacher rank-ordered his or her students on the basis of achievement in two areas: Reading/Language and Arithmetic. (A copy of the rating form used by the teachers is included as Appendix B). Class rank in Reading/Language and class rank in Arithmetic were obtained to use as criterion measures to assess the criterion validity of the Vocabulary and Arithmetic Concepts subtest scores respectively.

While a criterion-referenced measure of achievement would have been best for our purposes, the difficulty of asking teachers to make valid judgements of achievement on an absolute scale was considered too great. In asking for such judgements, the danger would be that the resulting ratings would be an unspecified mixture of criterion- and

norm-referenced ratings, where neither the criterion nor the norm group were known. It was decided, therefore, to ask for a normative judgement based on a clearly specified (and highly familiar) norm group, and to treat the obtained measures as such.

Socio-Economic Status measure. Seventy percent of the students in the study get their school lunches free through a federal poverty program. Participation versus non-participation in the free lunch program was used as a dichotomous measure of SES. Eligibility for the program is based on self-reported gross annual income. Some representative cut-offs are $5640 (for a family of four), $7310 (family of six), $8810 (family of eight), $10,190 (family of ten). Since participation was voluntary, and based on self-report, it was probably not entirely valid as an SES indicator. Another limitation of this measure was that the SES of non-participants was not known. In these cases, the assumption made was that the individuals were not in the low SES group.

Procedure

A pretest-posttest control group design was employed, with half of the students in each class randomly assigned to each of the two treatments.

Pretesting was conducted during the last two weeks of March. Training was conducted during the first, second, and fourth weeks of April (the third week was the spring vacation). Posttesting and col-

lection of achievement ranks was done during the third week of May.

The CAT subtests were administered to all students in each class by the regular teacher. Scheduling of the test within the one or two week period was left to the individual teacher.

The training materials were packaged as two sessions, but each session had a natural breaking point such that training could be conducted in two, three or four sessions. The teachers were allowed to schedule the sessions at their convenience over a three-week period, but were asked to try to space the sessions as evenly as possible and to make each session approximately 40 minutes in length. All of the training teachers covered the material in either two or three sessions, and total training time varied from about 2 1/2 to 3 hours.

In the first school, classes containing fourth, fifth and sixth graders are grouped into two multi-age family units, with four classes per unit. One teacher from each unit was asked by the principal to conduct the training of the experimental-group fourth- and fifth-graders in that unit. Half the students in each class were randomly selected for training. The two training groups consisted of 24 and 21 students. Training sessions were conducted in the school's science room, while control students went about their normal routine under the supervision of the remaining unit teachers.

The second school has two units containing fourth- and fifth-grade classes. One unit consists of three fourth grades and one mixed fourth-fifth grade. Two of the four teachers volunteered to conduct training

for the unit. Half of each class was randomly assigned to receive training. Each trainer trained her own students and those of one other teacher. These two training groups had 19 and 20 students.

The second unit has one very large fifth grade and a large mixed fourth-fifth-sixth grade, each group having two teachers. One teacher from each group volunteered to train the half of her or his students assigned to training; the training groups consisted of 29 and 16 students. In both units at this school, training was conducted during the early morning reading period. The training session was held in one area of the unit's space, while the control group students had their normal reading period in another part of the space.

Treatment

Copies of the actual training materials are presented in Appendix C. The training teachers used the written materials as the basis for oral instructions over each topic. Teachers were permitted some flexibility in adapting the training material for their own classes and personal teaching style, but were asked to make notes of any modifications made.

After the oral instruction on each segment, a short practice test was given. Each test was scored immediately and discussed, with the discussion focusing on the elements just covered in the instruction.

Elements of TO included in training are listed below:

1. The purpose of testing. This section involved a brief

explanation intended to help raise test motivation. No practice test was given with this section.

2. Correct marking of answer sheets. Instruction covered the most common marking errors as well as the reason for having to mark carefully.
3. Paying attention to and following instructions. The objective of this section was to counter the observed tendency of lower-SES children to mark items carelessly, before processing all the information needed to answer correctly.
4. Pacing and careful use of time. The section was also designed to counter too-rapid, careless work styles.
5. Guessing strategy. On the practice test for this section, the child received a peanut or several small candies for each correct answer. This was the tangible reward intended to foster test motivation.

The rationale for the choice of elements included in training has been discussed above and in the introductory chapter. The rationale underlying the instructional method is primarily practical, but also has some basis in the literature. The practical considerations are those of time, cost, and acceptability to school personnel. These require that training be suitable for groups rather than require individual administration. Freedom to modify the instruction was built-in to make the procedure more acceptable to teachers, even though

some control over the procedure was thus sacrificed. It was decided that a procedure that is to be implemented by teachers must be potent enough that its effects are not swamped by individual trainer effects. This, rather than rigorous experimental control, was the goal in designing the procedure.

One of the primary objectives of training was to get the children to use a careful, thorough, reflective work style. Some of the research on modification of conceptual tempo was drawn upon in developing training for this objective. Heider (1971) has given a definition of conceptual tempo as the extent to which a subject tends to pause before answering in problem-solving tasks. It is not logically identical with the application of careful problem-solving strategies, however it does tend to correlate highly with errors on a number of experimental tasks. One approach to modifying impulsive conceptual tempo (Kagan, Pearson, & Welch, 1966) has been to train children by enforcing a period of delay before responding to the problem is permitted. Typically, this approach is effective in increasing response latency, but not in decreasing errors. Therefore, this method was not considered for the present study. Meichenbaum (1971) found that training children to covertly verbalize instructions was successful both in increasing response latency and in reducing errors. Covert verbalization was incorporated in the present study. Finally, Heider (1971) compared the effectiveness of three training procedures in modifying the impulsive tempo of middle- and lower-class nine-year-old boys. For lower-class

boys, the procedure most effective in lengthening latency and reducing errors (on a sentence construction task) was task strategy instruction. The instructed strategy was "not to say his sentence aloud until he had ... been able to say it to himself." This was more effective than forced-delay training or offering tangible rewards. For the middle-class boys, on the other hand, all procedures were about equally effective. In the present study, an attempt was made to provide instruction on strategies for approaching tests and test items, although of course the problems and strategies are considerably more complex in this case, and it cannot be assumed that the same procedure will necessarily be effective for complex tasks.

The rationale for the stress on practice is based on Wiseman's (1954) finding that coaching without practice resulted in only slightly greater improvement than the control groups', while coaching combined with practice resulted in significant gain.

The content of the four practice tests was chosen with the goal of producing quite easy items. Language and content selection were guided by examination of second, third and fourth grade textbooks.

Practice Test 1 consisted of very straightforward items and was intended simply as a vehicle for practicing marking answer sheets.

Practice Test 2 was designed to illustrate the necessity of paying attention to and following instructions. It consisted of a variety of items with unusual or complicated instructions. After each set of two or three items, the correct answers were given and the instructions for

the set reviewed.

Practice Test 3 had two objectives. First, in order to encourage skipping over hard items and returning to them later, it was attempted (in a rational manner) to arrange items so that difficult and easy items alternated with each other. The second objective was to encourage use of the full amount of time allowed for a test, and to this end the test was given with what was hoped to be a liberal time limit. Virtually all the children did finish this practice test; however it is not known whether they utilized all the allowed time to do so.

Practice Test 4 had the same objectives as Practice Test 3, and the additional one of encouraging guessing with partial knowledge. However, it was not considered desirable to encourage guessing by writing items with "absurd options," since Wahlstrom and Boersma (1968) showed that learning to guess on tests containing absurd options did not improve performance on well-constructed tests. Instead, it was attempted to encourage guessing by using fairly difficult items. Again, item difficulty was rationally judged rather than empirically determined.

CHAPTER III

RESULTS

Attrition

Of 114 and 117 subjects assigned to the training and control groups respectively, 106 and 114 were given the pretest. This was the most complete set of scores obtained and was assumed to be an unbiased indicant of overall group performance. The pretest raw score means of those subjects with complete data (i.e., having both pretest and posttest scores) were identical to those of the total group (with the one exception that the Vocabulary Test scores of the control subjects differed by 0.1 raw score points on a 40-point test). It was considered desirable to have pre- and posttest scores from the same subjects, and since there appeared to have been no selective attrition, only those subjects with complete data were included in subsequent analyses. There were 96 such subjects in the training group and 93 in the control group.

Non-Equivalence of Treatment Groups

In spite of random assignment within classes to treatment condition, the training group pretest raw score means were significantly higher than the control group means (2.9 points on the Vocabulary subtest and 1.8 points on the Arithmetic subtest). For this reason, analysis of covariance with pretest score as the covariate was used to

assess training effectiveness.

Examination of treatment group composition (see Table 2) revealed that, relative to the training group, the control group had slightly more fourth-graders, boys, and low-SES children. This imbalance might have been expected to contribute to the differences in group means. Inspection of the pretest grade-equivalent score means in Table 2, however, shows that the initial superiority of the training group persists within grades and SES categories.

Achievement Test Score Gains

After administration of the posttest, it was discovered that the pretest and posttest forms of the CAT (Forms A and B, respectively) were sufficiently non-parallel to make the comparison of raw scores undesirable (the means in the norms tables differed by nearly five points). Therefore, raw scores were converted to grade equivalency (GE) scores. This resulted in the loss of the Arithmetic Concepts score as a dependent measure, since a separate GE conversion table was not available for that subtest.

Table 1 gives pre- and posttest Vocabulary GE means and mean GE gain for the two treatment groups. Mean control group GE gain was almost significantly greater than training group gain (2-tailed $p=.053$). Since both groups were below the mean of a hypothetical comparable sample from the standardization population (a representative sample consisting of one-half fourth and one-half fifth graders would be

Table 1

Mean CAT Vocabulary Grade Equivalent Scores

Test Administration	Treatment Group	
	Training[a]	Control[b]
Pretest	4.44 (1.84)[c]	3.84 (1.75)
Posttest	4.76 (1.75)	4.45 (1.59)
Gain (posttest minus pretest)	+ .32 (1.07)	+ .60 (.94)

[a] n=96

[b] n=93

[c] Values in parentheses are standard deviations.

expected to have a mean GE of 5.2 in March and 5.4 in May), some gain due to regression toward the mean was expected in both groups. Both groups did in fact gain more than the 0.2 GE points which would be predicted due to growth alone over a two-month period; the excess gain probably represents the expected regression.

The results of analysis of covariance were interpreted as suggesting that the greater mean gain of the control group was primarily a regression artifact as well. The effect of training on posttest score, after removing the variance associated with pretest score, was not significant, $F(1,186)=1.25$, $p>.10$. (An initial regression solution showed that the Treatment x Pretest interaction was non-significant, $F(1,185)=.049$; hence, in the subsequent solution interaction variance was pooled with error variance.)

One of the concerns of this study was tailoring a training program to the capacities of those students most deficient in test-taking skills. It was hypothesized specifically that lower-SES children would benefit most from the training program. Therefore, CAT GE score gain was examined separately in two SES categories, as well as in extreme high- and low-scoring groups on the pretest, in special and regular education students, and in several other subject categories (these data are presented in Table 2). Although the overall analyses suggested there was no training effect, these subgroup comparisons were made to determine whether there was some training effect within subgroups that was not apparent in the analysis of the overall effect.

Examination of Table 2 reveals few results which are not most

Table 2

Mean CAT Vocabulary Grade Equivalent Scores by

Treatment Group and Subject Classification Variable

Classification Variable	Level of Classification Variable		Treatment Group	
			Training	Control
Pretest Score	Highest			
		Pretest	7.24 (13)[a]	6.72 (13)
		Posttest	7.04	6.58
		Gain	- .20	- .14
	Lowest			
		Pretest	1.50 (13)	1.39 (13)
		Posttest	2.66	2.48
		Gain	+1.16	+1.09
Socio-Economic Status	Free Lunch Recipients			
		Pretest	4.03 (66)	3.50 (67)
		Posttest	4.55	4.06
		Gain	+ .52	+ .56
	Non-Recipients			
		Pretest	5.34 (30)	4.74 (26)
		Posttest	5.23	5.45
		Gain	- .11	+ .71

Table 2 (continued)

Classification Variable	Level of Classification Variable	Treatment Group	
		Training	Control
Special Education Status	Special Education		
	Pretest	3.05 (8)	3.07 (8)
	Posttest	3.05	3.79
	Gain	0	+ .72
	Regular Education		
	Pretest	4.57 (88)	3.92 (85)
	Posttest	4.92	4.51
	Gain	+ .35	+ .59
Grade	Fourth		
	Pretest	3.84 (38)	3.21 (42)
	Posttest	4.27	3.79
	Gain	+ .43	+ .58
	Fifth		
	Pretest	4.84 (58)	4.37 (51)
	Posttest	5.08	4.99
	Gain	+ .24	+ .62

Table 2 (continued)

Classification Variable	Level of Classification Variable	Treatment Group	
		Training	Control
School	School 1		
	Pretest	4.53 (39)	4.69 (30)
	Posttest	4.53	4.98
	Gain	0	+ .29
	School 2		
	Pretest	4.37 (57)	3.44 (63)
	Posttest	4.92	4.20
	Gain	+ .55	+ .76

[a]Values in parentheses are subgroup ns.

easily interpreted as regression effects; and with only one exception, those results that are inconsistent with the regression explanation suggest a detrimental effect of training. Negative results were obtained for training-group subjects in: (a) the non-recipient SES category (i.e., not lower class); (b) the special education category (but note the very small ns); and (c) School 1. In these subgroups the expected upward regression did not occur. The one positive result (initially low-scoring trained students gained .05 GE points more than initially-low control students, in spite of having a slightly higher initial mean score) is far too small to be of practical significance.

Pretest-Posttest Correlation

The Pearson product-moment correlation coefficients of pretest with posttest GE scores were computed for each treatment group. The values of r obtained were .82 and .84 in the training and control groups respectively. A substantially lower r in the training group would have signalled the possible presence of a Subject x Treatment interaction. The virtually identical coefficients, together with the failure to find either overall or subgroup training effects, suggest there was no effect of training for any group or individual.

Validity

In light of the absence of training effects, either additive or non-additive, no pretest to posttest change in the criterion validity of the Vocabulary test was expected. Validity was assessed as

the correlation between Vocabulary GE score and class rank in reading/language (provided by each classroom teacher).

As there was marked variation in ability level (as reflected in CAT scores) among classes, rank in class could not be considered comparable across classes. Therefore, correlation coefficients were computed separately for the training-group and control-group students within each class. In addition, when a single class contained both fourth and fifth graders, ranks were obtained and correlations computed within grades within classes. This was done to avoid spurious inflation of the correlation coefficient by expansion of the range of either variable. It should be noted, however, that when one of the variables is rank, the Pearson coefficient is strongly influenced by the size of the group (since the variance of a set of ranks is determined by n). The classes in this study were of different sizes, but since the object was to compare correlations of the training and control groups, and since an equal number of students in each class was assigned to each group, it was assumed that the degree of spuriousness would be equal for the training and control portions of each class. Thus, the coefficients obtained within each class may be validly compared, but should not be given an absolute interpretation.

The class validity coefficients were averaged using Fisher's $\bar{z}$ transformation (Table 3). The standard errors given in Table 3 are only approximate, being based on a sampling distribution for large samples (Olkin, 1967). The sampling distribution of $\bar{r}$ (average) for

Table 3

Class Validity Coefficients

(Correlation of CAT Vocabulary Scores with Class Ranks in Reading/Language)

Treatment Group		Average $\underline{r}$	$SE_{\underline{r}}$	Lowest $\underline{r}$	Highest $\underline{r}$
Training[a]					
	Pretest	.617	.107	-.72	.96
	Posttest	.520	.116	-.68	.95
Control[b]					
	Pretest	.564	.106	-.70	.89
	Posttest	.618	.101	-.27	.89

[a]Thirteen class groups with $\underline{n}$s between 4 and 20.

[b]Twelve class groups with $\underline{n}$s between 5 and 22.

small samples is unknown. However, the wide range of class rs obtained, together with the approximated standard errors, suggest that the treatment-group differences in pretest and posttest average validities are unlikely to be significant.

There is apparently no available statistical test for the difference between pre- and posttest validity coefficients within treatment groups. Some statistical texts and papers (e.g., Hendrickson, Stanley, & Hills, 1970) have proposed tests for null hypotheses of the form $\rho_{12}=\rho_{13}$ (in one sample), but these are not appropriate when the estimates of ρ are averaged rs. Even if statistically reliable, however, the obtained differences would be too small to be of practical significance.

Reliability

The effect of training on the reliability of CAT scores was evaluated by comparing pretest and posttest reliabilities in each treatment group. Reliability, estimated as the correlation between odd- and even-half scores corrected by the Spearman-Brown formula, was virtually unchanged in either group. Pretest and posttest reliabilities, in that order, were .93 and .89 for the training group, and .89 and .91 for the control group.

A split-half reliability estimate is not strictly appropriate here, since the tests are to some extent speeded. This is less a problem for the posttest than the pretest, since the posttest seems to

have been about equally speeded for both groups (mean numbers of items omitted were 4.14 and 4.83, out of 40 items, for the training and control groups respectively). So although the rs might be somewhat inflated they should still be comparable across groups. On the pretest, however, the mean numbers of items omitted were 3.76 (training) and 6.00 (control). Thus, the pretest reliability coefficient obtained in the control group may be spuriously inflated to a greater degree than in the training group.

Test Orientation Variables

Table 4 shows pretest and posttest means of two variables examined in order to further evaluate the effect of training. Answer sheet marking errors include: incomplete erasures, more than one option per item marked, and marks too light to be picked up by an optical scanning machine (as judged by the experimenter). Since a portion of training focused directly on correct answer sheet marking, this measure was expected to serve as a check on the effectiveness of this training as well as to help show whether answer-sheet-marking skills are related to test performance and/or validity. Similarly, part of the training focused on working steadily throughout the allotted time; thus the number of items omitted was expected to reflect the effectiveness of this portion of training, as well as the relation of this skill to test performance.

Inspection of Table 4 suggests that training was not effective in

Table 4

Means of Test Orientation Variables at Pretest and Posttest

Variable		Treatment Group	
		Training	Control
Number of Answer Sheet Marking Errors			
	Pretest	.104 (.747)[a]	.204 (.543)
	Posttest	.292 (1.065)	.409 (.875)
Number of Items Omitted			
	Pretest	3.76 (7.07)	6.00 (8.73)
	Posttest	4.14 (7.07)	4.83 (6.92)

[a]Values in parentheses are standard deviations.

either of these areas. Both treatment groups made more marking errors on the posttest than on the pretest; and while the controls omitted fewer items on the posttest, the trained group omitted more.

CHAPTER IV

DISCUSSION

Summary of Major Results

Contrary to expectation, no effect of training was found. Every effort was made to explore the data for possible relationships. Mean achievement score gain was examined in the total sample as well as in special subgroups such as high and low scorers on the pretest, middle and low SES, fourth and fifth graders, and regular and special education students. In addition, analysis of covariance was done on post-test scores using pretest score as a covariate. Both the reliability and validity of achievement test scores were considered as dependent variables. Finally, TO variables such as number of marking errors and number of items marked were considered. In each analysis, there was essentially no difference between the trained and untrained students.

Conclusions

There are several possible interpretations of these negative results. The first, obviously, is that TO deficiency is not a source of invalidity for standardized achievement tests for this population. The plausibility of this interpretation will be most easily discussed in conjunction with two other possibilities.

The second interpretation is that TO deficiency <u>is</u> a source of test invalidity, but that the present study did not focus training efforts on the most important aspects of TO. In particular, this study ignored fear and test anxiety. The scope of the training program

was limited by practicality, and test anxiety was not expected to be high in a population that was expected to be lacking in test motivation. However, anxiety may actually be an important aspect of TO. Motivation is another aspect of TO which may not have been treated adequately in this study. The present study treated test motivation in a restricted manner, by trying to (a) provide incentives for correct answers in the form of tangible rewards, and (b) foster an appreciation of the importance of doing well on tests. Motivation might also be construed much more broadly, however, as for example, a general approach to situations (including tests) as problem-solving situations. Some researchers (e.g., Cohen, 1971) have suggested that lower-SES children do not have an analytic cognitive style, which means that they do not abstract information from stimulus input in an analytic fashion. Such a style might be very important to successful performance on a typical standardized achievement test. Moreover, test motivation could be very broadly conceptualized as embracing such a cognitive style or approach. This casts the problem of teaching test motivation as a problem of teaching a very generalized problem-solving approach, rather than the simple incentive-providing problem treated in this study.

The TO measures included in the study provide a way of looking at whether the trained elements of TO were important to test validity, or whether other elements might have been more important. Given the obtained results, if TO as reflected in these two measures had increased

due to training, it would have suggested that answer-sheet marking and number of items marked are unimportant to test validity. However, TO as measured did not improve, so this conclusion cannot be drawn.

Instead, a third interpretation is suggested: the important aspects of TO may or may not have been treated here; the implementation of training was not adequate to teach the skills it was intended to teach, and thus no conclusions may be drawn about their importance. Other support for the "inadequate treatment" interpretation comes from the comments of teachers who carried out the training. They stated that students had much difficulty concentrating on the material and paying attention to instructions; they also felt that the sessions were too long and covered too much material. One teacher felt that the students had rote-learned the content of the sessions but were not able to apply it (this seems to suggest that more practice tests would be valuable). The same teacher stated that the students, even after training, worked on standardized tests in a manner seemingly reflective of low motivation; that is, they wasted large amounts of testing time doing things other than working on the test. This last is also in accord with the failure of training to result in more items completed. Further, the limitation imposed by cost on the type of tangible rewards that could be given was, from the outset, expected to restrict the possible effectiveness of that component of motivation training. Finally, the present study was a pilot study in the sense that the training materials were untested prior to this time. Although

training vocabulary and practice-test questions were intended to be appropriate for the third-grade level, several of the training teachers said the material was too difficult for their students.

In conclusion, then, in light of the inadequacies of the training program as implemented, the present study has not demonstrated by its negative results that TO is not a source of test invalidity. The most plausible interpretation seems to be a combination of inadequate design and implementation of training, and inadequate scope of the treatment of test motivation.

Suggestions for Future Research

Future TO training studies might be more successful in teaching the skills treated by this study if more and shorter sessions were given involving more practice tests, and offering tangible rewards for performance in an extended series of tests in which rewards would eventually be "faded out." It must be noted, however, that any study which purports to develop a practical training program will necessarily be constrained in the expenditure of time, money and materials, and that it may not be possible to teach TO skills within those constraints (even though given optimal conditions it might be quite possible).

Another future study should attempt to broaden the conceptualization of test motivation, as suggested in this chapter, and to find other methods for increasing motivation. One approach that could be tried is structured group discussion in which the discussion leader tries to

facilitate the development of group standards for good test performance.

A weakness of the present study which should be corrected in any future study was not using the publisher's answer sheet. Using a fascimile of the answer sheet could be argued to limit the generalizability of any obtained result.

Another desirable change would be to ensure that the difficulty level of training materials was appropriate by preparing them in collaboration with the teacher of the target students. Selection of vocabulary and practice-test content could be guided by examination of school materials actually used by the students.

A final suggestion is that more TO measures need to be developed. If time and resources permitted, valuable data might be obtained by: (a)interviewing children, especially about whether they feel it is important to do well on tests, whether their friends, parents, teachers feel it is important, how they feel and what they think about while taking tests, etc.; and (b)observational study of children's test-taking behavior. The present study shared with previous studies the weakness of trying to teach test orientation before adequately exploring what it is.

REFERENCES

Allyon, T., & Kelly, K. Effects of reinforcement on standardized test performance. Journal of Applied Behavioral Analysis, 1972, 5, 477-484.

Anastasi, A. Psychological testing, (2nd edition). New York: Macmillan, 1968.

Callenbach, C. The effects of instruction and practice on content-independent test-taking techniques upon the standardized reading test scores of selected second-grade students. Journal of Educational Measurement, 1973, 10, 25-30.

Casey, M. Three studies on the effect of training on similar and identical material on the Stanford-Binet test scores. Twenty-Seventh Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1928.

Davidson, H. Three studies on the effect of training on similar and identical material on the Stanford-Binet test scores. Twenty-Seventh Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1928.

Dempster, J. Symposium on the effects of coaching and practice in intelligence tests. III. Southampton investigation and procedure. British Journal of Educational Psychology, 1954, 24, 1-4.

Eells, K., Davis, A., Havighurst, R., Herrick, V., & Tyler, R. Intelligence and cultural differences. Chicago: University of Chicago Press, 1951.

Frankel, E. Effects of growth, practice and coaching on Scholastic Aptitude Test scores. Personnel and Guidance Journal, 1960, 38, 713-719.

Gibb, B. Test-wiseness as secondary cue response. Unpublished doctoral dissertation, Stanford University, 1964.

Goslin, D. The search for ability: Standardized testing in social perspective. New York: Russell Sage Foundation, 1963.

Goslin, D., Epstein, R., & Hallock, B. The use of standardized tests in elementary schools. Technical Report No. 2 on the Social Consequences of Testing. New York: Russell Sage Foundation, 1965.

Greene, E. Practice effects on various types of standardized tests. American Journal of Psychology, 1937, 49, 67-75.

Greene, K. The influence of specialized training on tests of generalized intelligence. Twenty-Seventh Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Company, 1928.

Heider, E. Information processing and the modification of an impulsive conceptual tempo. Child Development, 1971, 42, 1276-1281.

Hendrickson, G., Stanley, J., & Hills, J. Olkin's new formula for the significance of r_{13} vs. r_{23} compared with Hotelling's method. American Educational Research Journal, 1970, 7, 189-195.

Holloway, H. Effects of training on the SRA Primary Mental Abilities and WISC. Child Development, 1954, 25, 253-263.

Jensen, A. A theory of primary and secondary familial mental retardation. In N. R. Ellis (Ed.), International review of research in mental retardation, Vol. 4. New York: Academic Press, 1970. Cited in J. Sattler, Assessment of children's intelligence. Philadelphia: W. B. Saunders, 1974.

Kagan, J., Pearson, L., & Welch, L. Modifiability of an impulsive tempo. Journal of Educational Psychology, 1966, 57, 359-365.

Kinnie, E., & Sternlof, R. The influence of non-intellective factors on the IQ scores of middle- and lower-class children. Child Development, 1971, 42, 1989-1995.

Klugman, S. The effect of money incentive versus praise upon the reliability and obtained scores of the revised Stanford-Binet Test. Journal of General Psychology, 1944, 30, 225-269.

Levine, R., & Angoff, W. The effects of practice and growth on scores on the Scholastic Aptitude Test. Research Bulletin, Educational Testing Service, February, 1958.

Meichenbaum, D. The nature of the modification of impulse in children: Training impulsive children to talk to themselves. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, April, 1971.

Millman, J., Bishop, C., & Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 3, 707-726.

Moore, J., Schutz, R., & Baker, R. The application of a self-instructional technique to develop a test-taking strategy. American Educational Research Journal, 1966, 3, 13-17.

Oakland, T., & Weilert, F. The effects of test-wiseness materials on standardized test performance of disadvantaged preschool children. Paper presented at AERA, February, 1971, New York, New York (ED 047 050).

Olkin, I. Correlations revisited. In J. Stanley (Ed.), Improving experimental design and statistical analysis. Chicago: Rand McNally, 1967.

Ortar, G. Improving test validity by coaching. Educational Research, 1960, 2, 137-142.

Schuchman, H. Evaluating the educability of the severely mentally retarded. Psychological Monographs, 1960, 74, No. 14.

Slakter, M., Koehler, R., & Hampton, S. Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970, 7, 247-254.

Tinney, R. The effects of training in test-taking skills on the reading test scores of fifth-grade children of high and low socio-economic status. Unpublished doctoral dissertation, University of Minnesota, 1968.

Vernon, P. Symposium on the effects of coaching and practice in intelligence tests. V. Conclusions. British Journal of Educational Psychology, 1954, 24, 57-63.

Wahlstrom, M., & Boersma, P. The influence of test-wiseness upon achievement. Educational and Psychological Measurement, 1968, 28, 413-420.

Wiseman, S. Symposium on the effects of coaching and practice in intelligence tests. IV. The Manchester experiment. British Journal of Educational Psychology, 1954, 24, 5-8.

Zigler, E., & Butterfield, E. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.

APPENDIX A

Fascimile of Machine-Scorable

Answer Sheet for Pretest and Posttest

CALIFORNIA ACHIEVEMENT TEST LEVEL 3 GRADES 4-5-6	NAME Last First Middle SCHOOL			
DATE OF TEST Year Month Day	PUPIL'S AGE Years Months	BOY GIRL (Circle One)	GRADE	TEACHER OR CLASS

VOCABULARY

Sample Item A. 1 2 3 4 — (0) 0 0 0

Sample Item B. 6 7 8 9 — 0 0 0 0

1. 1 2 3 4 — 0 0 0 0	7. 1 2 3 4 — 0 0 0 0	13. 1 2 3 4 — 0 0 0 0	20. 6 7 8 9 — 0 0 0 0
2. 6 7 8 9 — 0 0 0 0	8. 6 7 8 9 — 0 0 0 0	14. 6 7 8 9 — 0 0 0 0	21. 1 2 3 4 — 0 0 0 0
3. 1 2 3 4 — 0 0 0 0	9. 1 2 3 4 — 0 0 0 0	15. 1 2 3 4 — 0 0 0 0	22. 6 7 8 9 — 0 0 0 0
4. 6 7 8 9 — 0 0 0 0	10. 6 7 8 9 — 0 0 0 0	16. 6 7 8 9 — 0 0 0 0	23. 1 2 3 4 — 0 0 0 0
5. 1 2 3 4 — 0 0 0 0	11. 1 2 3 4 — 0 0 0 0	17. 1 2 3 4 — 0 0 0 0	24. 6 7 8 9 — 0 0 0 0
6. 6 7 8 9 — 0 0 0 0	12. 6 7 8 9 — 0 0 0 0	18. 6 7 8 9 — 0 0 0 0	25. 1 2 3 4 — 0 0 0 0
		19. 1 2 3 4 — 0 0 0 0	26. 6 7 8 9 — 0 0 0 0

27. 1 2 3 4 — 0 0 0 0	34. 6 7 8 9 — 0 0 0 0	40. 6 7 8 9 — 0 0 0 0	
28. 6 7 8 9 — 0 0 0 0	35. 1 2 3 4 — 0 0 0 0		
29. 1 2 3 4 — 0 0 0 0	36. 6 7 8 9 — 0 0 0 0		
30. 6 7 8 9 — 0 0 0 0	37. 1 2 3 4 — 0 0 0 0		
31. 1 2 3 4 — 0 0 0 0	38. 6 7 8 9 — 0 0 0 0		
32. 6 7 8 9 — 0 0 0 0	39. 1 2 3 4 — 0 0 0 0		
33. 1 2 3 4 — 0 0 0 0			

CONCEPTS

Sample Items A. A B C D E — 0 0 (0) 0 0

Sample Items B. F G H J K — 0 0 0 0 0

1. 1 2 3 4 5 — 0 0 0 0 0	4. A B C D E — 0 0 0 0 0	11. 1 2 3 4 5 — 0 0 0 0 0	19. A B C D E — 0 0 0 0 0
2. A B C D E — 0 0 0 0 0	5. F G H J K — 0 0 0 0 0	12. A B C D E — 0 0 0 0 0	20. 1 2 3 4 5 — 0 0 0 0 0
3. F G H J K — 0 0 0 0 0	6. A B C D E — 0 0 0 0 0	13. 1 2 3 4 5 — 0 0 0 0 0	21. A B C D E — 0 0 0 0 0
	7. F G H J K — 0 0 0 0 0	14. A B C D E — 0 0 0 0 0	22. F G H J K — 0 0 0 0 0
	8. A B C D E — 0 0 0 0 0	15. 1 2 3 4 5 — 0 0 0 0 0	23. A B C D E — 0 0 0 0 0
	9. 1 2 3 4 5 — 0 0 0 0 0	16. A B C D E — 0 0 0 0 0	24. F G H J K — 0 0 0 0 0
	10. A B C D E — 0 0 0 0 0	17. F G H J K — 0 0 0 0 0	25. A B C D E — 0 0 0 0 0
		18. 1 2 3 4 5 — 0 0 0 0 0	

APPENDIX B

Student Achievement Rating Form

Student Achievement Rating Form

Instructions

As part of the Test Orientation Training study, we would like you to provide us with an assessment of the class standing of each of your students in the areas of Reading& Language, and Arithmetic. This assessment should be based upon your observation and interaction with the students. Your assessments represent an important element in evaluating the success of the training program, so we hope you will complete this (rather tedious) task as carefully as possible!

Specifically, we would like you to rank order your students from lowest to highest in terms of their mastery of material for their own grade level. If you have both fourth and fifth graders in your class, ranks should be determined separately within each grade and should be based upon mastery of the material for that grade level. (Sixth graders should not be included in the ranking procedure.)

The lowest student in each grade should be given a rank of 1, with higher ranks reflecting higher class standing. If it is impossible to distinguish between the achievement levels of two or more students, they may be given the same rank. Please record students' ranks in the appropriate column on the attached class roster. When you have completed the ranking in the area of Reading & Language, please repeat the procedure for the area of Arithmetic.

Thank you very much for your very valuable help.

APPENDIX C

Training Materials:

1. Teacher's instruction manual, Session 1
2. Answer sheet for Practice Tests 1 and 2
3. Practice Test 1
4. Practice Test 2
5. Teacher's instruction manual, Session 2
6. Answer sheet for Practice Tests 3 and 4
7. Practice Test 3
8. Practice Test 4

TEST ORIENTATION TRAINING MANUAL

Session 1.

Approximate time: 30 minutes.

Topics: 1) Purpose of testing
2) Using answer sheets
3) Importance of following instructions

General outline of session: Introduction.
Topic 1. Brief comments by teacher; answer questions.
Topic 2. Instruction on general principles of using machine-readable answer sheets; blackboard demonstrations of marking; practice test.
Topic 3. Instruction; practice exercises.

General instructions for teachers:

Most of the instruction in this session has been written out verbatim in this manual (all verbatim material is indented). It is not necessary that you stick to the printed text word-for-word if this is not comfortable for you. The text is intended to indicate the substance of the instruction. If you do not use my wording, please try to cover the points I have made.

At all times during this session, except while a practice item is actually being done, you are encouraged to answer questions, give special procedural help, and, most important, to continually check students' answer sheets and give feedback on the correctness of their marking. Please feel free to expand the instruction or discussion at any point where students have difficulty.

Finally, if you have any comments or suggested changes for the training materials, please write any comments you wish to make in this booklet.

Specific instructions and procedure:

I. Introductory statement

Let me explain to you what we will be doing in here today and on the other two days that we will meet. All of you have taken achievement tests before, right? Those of you who were at _(school)_ for the third grade took the Stanford Achievement Tests then.

What you will be doing in here is learning some ways that you can do better on achievement tests. The first thing we'll do is talk about why you are given those

tests and what they are used for. Then I'm going to tell you about some things you can do to get the best score you can. You'll get a chance to practice the things I tell you about on some short practice tests, which should be fun to take.

We won't be studying the subjects that are on the tests, such as vocabulary and arithmetic. It's up to you to learn those things in school. The things we will talk about have to do with knowing how to take a test so that you can show everything you do know about your school subjects.

For example, do you think you know how to guess on a test? Did you ever think about there being a right way to guess? You 'll be learning about guessing.

Another thing we'll talk about is using answer sheets. These separate answer sheets, which are made for computers to read, can be rather confusing for people. We'll talk about how to avoid making mistakes when marking them, and you'll get to practice marking a number of them.

We'll also talk about getting nervous when you're taking a test. Not everyone gets nervous when they take a test, but many people do, and being nervous or scared can keep you from doing as well as you could. So we'll talk about why we get nervous about tests, and try to find some ways to help you not to be nervous .

These are some of the things you will learn about in our three meetings. What I hope will happen by the end is that you will know how to take tests well, and that you will feel good about tests and enjoy taking them. Then you will be able to do your very best.

(Go on to the next page.)

II. Purpose of testing

(Rather than write out verbatim what you should say to the students, I have simply listed the points I would like you to cover in a very brief discussion. This portion of the session should not take more than five minutes. The purpose of this discussion is convince the students that tests are important, even though they will probably not see their scores.)

Points of discussion:

1) Tests show how much you have learned. They show your teachers whether you need special help, or should be in a different class. Test scores help your school to give you the kind of teaching that is best for you.

2) Tests also show how good a job your school is doing. People in our state and city governments are interested in your test scores to see how well this school is teaching students, as compared to other schools in Worcester and in Massachusetts. (They don't look at anybody's actual test or know how any particular student has done, but they see all the scores of the school reported as a group.)

3) Knowing how good a job it is doing helps your school to know whether it should change anything in the way you are being taught. For example, if all the students in a school scored very high in vocabulary and rather low in arithmetic, the school might decide to spend less time teaching vocabulary and more time on arithmetic.

4) Students usually don't get their test papers back or find out their test scores on achievement tests. For this reason it is sometimes easy for them to think the tests don't count for anything. They should be aware though, that their test scores are used in making important decisions about their education. In order for these decisions to be made wisely, the test must show the very best work of each student.

III. Answer Sheets

(Read to students:)

> One of the tricks of being a good test taker is knowing how to mark answer sheets in the correct way. There are many different types of answer sheets; you have probably already seen one or two different ones. Even though these types of answer sheets are different in some ways, they are basically alike in the way they are used.

(Hold up one of the pink-printed answer sheets for demonstration.)

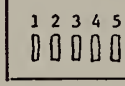
> This answer sheet is made to be read by a machine. So are almost all the answer sheets you will use when you take achievement tests in school from now on. When you have marked your answers on this

sheet, your test can be corrected by a machine without a person ever having to look at it. This saves a great deal of time. The achievement tests you take in school are sent to a company whose business is testing students. Your filled-out answer sheets are scored by machines which can correct the papers much faster and more accurately than any person could do it. The machine is much faster than a person, but unfortunately, unlike a person, the machine cannot look at your paper to try to figure out which answer you <u>really</u> meant to mark if you made a marking error.

All machine-readable answer sheets, though they look different, are basically the same. The answer to a question is marked by blackening in a space with your pencil. The spaces to be filled in look somewhat different on different answer sheets, but the basic idea is always the same.

The answer sheet that I am holding, and which you will practice with in a few minutes, has answer spaces that look like this:

(Draw on blackboard:)

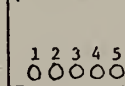


On this sheet you mark an answer by completely filling in one of the little rectangles.

(Demonstrate.)

Some answer sheets have answer spaces like these:

(Draw on board:)

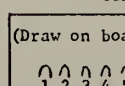


Here the answer choices are circles, and you answer by filling in one of the circles.

(Demonstrate.)

Here is another type of answer space you sometimes see. Can you tell how you would mark your answer here?

(Draw on board and demonstrate marking procedure.)



Sometimes the answer choices will go up and down instead of across:

(Draw on board and demonstrate marking:)

[1]
[2]
[3]
[4]
[5]

Sometimes instead of numbers the choices will have letters:

(Draw and demonstrate:)

A B C D E
○○○○○

Sometimes there will be more or less than five choices. But, as you see, the basic idea of blackening in a space is always the same.

Now, because these marks are being read by a machine and not by a person, you must be very careful to make your marks correctly.

This is what I mean by marking correctly:
1) Mark only one answer space.
2) Whatever the shape of the answer space, fill in that space completely.
3) Make your mark dark.
4) Don't make any marks outside of the answer space.
5) Use only a pencil. Usually the correct kind of pencil will be given to you. If you write in pen, the machine will not be able to read your marks.

The reason you must be so careful to mark your answers exactly right is that the machine cannot tell the difference between a wrong answer, and one that is right but has a mistake in marking. Any answer that has a marking mistake (like accidentally marking two spaces) is counted wrong.

(Pass out one pink answer sheet to each child.)

Let's get some practice marking a real answer sheet. We'll start by filling in the identification boxes, because they use answer spaces too. This is how the scoring machine reads your name, since it cannot read printing.

(Show how to turn the answer sheet so the name grid is at the top. Have them locate the places for last and first names; have them print this information. Have them blacken in the correct spaces under the letters of their names. Demonstrate this on the board if neccessary. As they fill in the name spaces, walk around checking for accuracy of marking. Take plenty of time and insist on accuracy.)

(When they are well along, say:)

Did anyone make a mistake and fill in a wrong letter space? When you make a mistake, or want to change your answer, be sure you erase your old answer completely. If there is any pencil left on the old answer, the machine may read it and think you have marked two answers, and will count your answer wrong.

Now find the place below the name box where it says GRADE.

(Instruct them on filling in the spaces for GRADE, BIRTHDATE, and SEX. Again, walk around and make sure marking is accurate.)

Now turn your answer sheet so that the name box is at the top.

Whenever you are given an answer sheet to use, look over the whole sheet carefully before you begin the test. See how the answer spaces are arranged on the page. Notice the numbering. This will help you to keep your place on the page when you begin taking the test.

Look at this answer sheet. The first thing you may notice is the type of answer space . Each question has five little rectangles, one of which you will blacken in for your answer. Notice also that the rectangles are numbered 1-5.

Next you will notice that there are places for 160 answers on this page, and that they are divided into four sections. Each section is labeled with a Roman numeral on the left side of the page.

(Give them time to notice the things you point out.)

Find Section I. and follow along with me. Put your finger on item #1. Starting with #1, the numbers go down until you come #5. Then you go to the top of the next column to find #6. Read down to #10, then go back to the top of the next column to find #11. If you read through the numbers on the first section, you'll find that Section I has places for 40 answers. Then Section II begins back on the left side of the page with #41.

You need to pay attention to which way the numbers on your answer sheet are going. When you are taking the test, make sure the number of the item you mark on the answer sheet is the same as the number of the question you're answering on the test.

Now I'm going to give you a little test so that you can practice using the answer sheet. The test is very easy; everyone will know all the answers. The reason for taking this test is to get practice using an answer sheet. The more familiar you are with answer sheets, and the better you are at marking them correctly, the easier it will be for you when you take a real achievement test. If you are already good at figuring out answer sheets and marking them, then you won't have to

waste time thinking about it while you're taking a test. Instead you'll be able to put all your attention to figuring out the answers to the questions.

Let's review the important points about marking answers.

(Try to get them to volunteer the major points:

1) only one answer per item
2) fill space completely
3) dark marks
4) no marks outside answer space
5) if changing answer, erase old one completely
6) mark sure number on test corresponds to number on answer sheet)

(Make sure each student has a pencil with an eraser.
Pass out Practice Test 1 face down.)

This test has 12 questions. When I say "Begin", turn over the test and begin working. Mark your answers on the pink answer sheet. You will have 2 minutes.

Begin.

(After 2 minutes say:)

Stop. Put down your pencils.

(Have them trade papers with each other to check for correctness of marking. Read the correct answer numbers to them.

Key:	1. 2	7. 5
	2. 3	8. 4
	3. 2	9. 3
	4. 5	10. 3
	5. 1	11. 2
	6. 2	12. 5

As they check each item, also have them check for correctness of answer space marking. Tell them not to count an item as correct unless the correct answer is marked and all the criteria for proper answer sheet use are met. This will undoubtedly require a lot of guidance and checking on your part. Walk around the room helping them decide whether marking has been done correctly. Be quite stringent. Encourage questions and try to generate some discussion of right and wrong ways of marking.

Have them circle the item numbers (on the answer sheet) of missed items. Have them count the number of items correct and write it in the upper-left corner of the answer sheet.

Collect the tests (but not the answer sheets).)

IV. Following Instructions

In order to be a good test-taker it's extremely important to pay very careful attention to any directions that are given. As soon as you hear the announcement that a test is going to be given, you should be listening for any directions that may follow.

This is important because students often think they know what they are supposed to do on a test, and they go ahead without listening to instructions, only to find out they have done it wrong. Even if you know the right answer to a test question, you can get it wrong just because you don't follow instructions on how to answer. For example, the directions might tell you to mark your answers in a certain part of the answer sheet; if you didn't hear the directions and marked your answers in a different place, all your answers would be counted wrong.

Some instructions will be read to you, others you will read yourself. Some instructions will come before the test and will apply to the whole test. Sometimes a particular question or group of questions will have its own instructions. So you must always be alert in case some new instructions are given.

I'm going to give some practice questions, which you will answer on your answer sheets. The purpose of these exercises is to give you practice paying careful attention to instructions. We will go through the questions one or two at a time; after each set of questions we will correct and discuss it. To get these questions right you are going to have to read and follow the instructions. Some of the things you are instructed to do will be silly. But remember this is not a real test, this is an exercise to learn to follow instructions, so do exactly as you are directed.

(Pass out Practice Test 2 face down.)

Now turn your test face up. Notice that the first question is numbered 81. Find #81 in the section of your answer sheet marked Roman numeral III. This is where you'll begin marking your answers.

(Make sure everyone has the right place.)

Go ahead and do the two questions in the first box, #81 and #82. Take your time, answer carefully, and when you have finished, stop and wait.

(Wait until all have finished.)

For #81, did you mark answer 3 ? Sad is the <u>opposite</u> of happy. The directions tell you to mark the word that means the opposite. But #82 says to mark the word that means the <u>same</u>, so you should have marked answer 4, fast. If you missed either of these

questions, circle the number of the missed question on your answer sheet.

(As they correct each item in this series, walk around the room checking that items are marked in the correct location, marks are made correctly, etc. This will take some time and it needs to be done carefully. Encourage questions and discussion about the correct way of marking answers. Remind them, if neccessary, of the previously discussed points concerning marking.)

Now go ahead and do questions 83, 84 and 85. When you finish them, stop and wait.

(When all have finished, have them check answers as before.
Key: 83. 1 (True) 84. 1 (True) 85. 2 (False).
Have them circle the numbers of missed items.)

Now find item #96 on your answer sheet. The next question on your test is #96, so you will answer it in space #96.

(If anyone has trouble locating particular items on the answer sheet, tell them to run their finger down the item numbers, going from column to column, until they locate the one they want.)

Now go ahead and do the next two questions on your test.

(Wait until all are finished.)

Did you pick choice 2 to question 96? If you did you were wrong, because the directions said to mark an answer that was <u>not</u> correct. So you could have marked 1, 3, 4 or 5, but <u>not</u> 2.

Of course you would not see a direction like this on a real test. This is just an exercise to get you into the habit of paying close attention to test instructions.

Did everyone skip 97 on the answer sheet and go to 98, as the directions said to do ? The answer to 98 is 5, furious. Circle the number of any items that you missed.

Now do the questions in the next box.

(When all have finished, have them check answers as before.
Key: 99. 4 100. 4 101. 5 102. 5.)

Now do the question in the next box, #103.

(The option numbers for this item are reversed, i.e. 5-1 rather than 1-5. If any questions are asked about this while students are doing the problem try to answer them privately, saying that they should use the option numbers <u>as they are</u> on the test.)

The answer is #2. Did you notice that the numbers of the choices were backwards? If you didn't notice it you probably marked #4, since the correct answer was in the fourth position.

Again, you won't ever see a question like this on a real test. This was just an exercise to teach you to stay on your toes.

Now do the last problem.

(Wait a few seconds.)

Check your answer to be sure that you completely erased choice 1. There should be no pencil marking left on choice 1.

(Have them trade answer sheets with each other and check each other's papers for proper marking (not for correctness of answers). Tell them to mark incorrect (by circling the item number) any item with incomplete erasures, more than one answer marked, incomplete blackening of option, very light marking, or marking outside the space. Encourage them to be critical, but monitor their judgements.
Have them count the number of correct answers on Practice Test 2 (Section III on the answer sheet) and write that number in the upper-right corner.)

(Collect tests and answer sheets.)

(Remind students they will have two more sessions on test taking, and then dismiss them.)

PRACTICE TESTS

NAME Last First Middle

SCHOOL

DATE OF TEST Year Month Day	PUPIL'S AGE Years Months	BOY GIRL (Circle One)	GRADE	TEACHER OR CLASS

PART I

Sample Item A. 1 2 3 4 — C

Sample Item B. 6 7 8 9 — 0

1. 1 2 3 4 — 0
2. 6 7 8 9 — 0
3. 1 2 3 4 — 0
4. 6 7 8 9 — 0
5. 1 2 3 4 — 0
6. 6 7 8 9 — 0
7. 1 2 3 4 — 0
8. 6 7 8 9 — 0
9. 1 2 3 4 — 0
10. 6 7 8 9 — 0
11. 1 2 3 4 — 0
12. 6 7 8 9 — 0
13. 1 2 3 4 — 0
14. 6 7 8 9 — 0
15. 1 2 3 4 — 0
16. 6 7 8 9 — 0
17. 1 2 3 4 — 0
18. 6 7 8 9 — 0
19. 1 2 3 4 — 0
20. 6 7 8 9 — 0
21. 1 2 3 4 — 0
22. 6 7 8 9 — 0
23. 1 2 3 4 — 0
24. 6 7 8 9 — 0
25. 1 2 3 4 — 0
26. 6 7 8 9 — 0
27. 1 2 3 4 — 0
28. 6 7 8 9 — 0
29. 1 2 3 4 — 0
30. 6 7 8 9 — 0
31. 1 2 3 4 — 0
32. 6 7 8 9 — 0
33. 1 2 3 4 — 0
34. 6 7 8 9 — 0
35. 1 2 3 4 — C
36. 6 7 8 9 — 0
37. 1 2 3 4 — 0
38. 6 7 8 9 — 0
39. 1 2 3 4 — 0
40. 6 7 8 9 — 0

PART II

Sample Items A. A B C D E — 0 0 C 0 0

Sample Items B. F G H J K — 0 0 0 0 0

1. 1 2 3 4 5 — 0 0 0 0 0
2. A B C D E — 0 0 0 0 0
3. F G H J K — 0 0 0 0 0
4. A B C D E — 0 0 0 0 0
5. F G H J K — 0 0 0 0 0
6. A B C D E — 0 0 0 0 0
7. F G H J K — 0 0 0 0 0
8. A B C D E — 0 0 0 0 0
9. 1 2 3 4 5 — 0 0 0 0 0
10. A B C D E — 0 0 0 0 0
11. 1 2 3 4 5 — 0 0 0 0 0
12. A B C D E — 0 0 0 0 0
13. 1 2 3 4 5 — 0 0 0 0 0
14. A B C D E — 0 0 0 0 0
15. 1 2 3 4 5 — 0 0 0 0 0
16. A B C D E — 0 0 0 0 0
17. F G H J K — 0 0 0 0 0
18. 1 2 3 4 5 — 0 0 0 0 0
19. A B C D E — 0 0 0 0 0
20. 1 2 3 4 5 — 0 0 0 0 0
21. A B C D E — 0 0 0 0 0
22. F G H J K — 0 0 0 0 0
23. A B C D E — 0 0 0 0 0
24. F G H J K — 0 0 0 0 0
25. A B C D E — 0 0 0 0 0

PRACTICE TEST 1
Do Not Mark On This Test

1. The town we live in is
   1 Boston
   2 Worcester
   3 New York
   4 Springfield
   5 Philadelphia

2. This year is
   1 1776
   2 1876
   3 1976
   4 1971
   5 1973

3. Which of the following is not a mammal?
   1 raccoon
   2 frog
   3 bear
   4 horse
   5 dog

4. Where would you go to check out a book?
   1 grocery store
   2 gas station
   3 drug store
   4 church
   5 library

5. The first president of the United States was
   1 Washington
   2 Lincoln
   3 Jefferson
   4 Ford
   5 Franklin

6. The season which follows summer is
   1 winter
   2 fall
   3 spring
   4 summer
   5 Saturday

7. 2+2=
   1 0
   2 1
   3 2
   4 3
   5 4

8. The capital of the U.S.is
   1 Boston
   2 NewYork
   3 Worcester
   4 Washington D.C.
   5 Los Angeles

9. 4-2=
   1 0
   2 1
   3 2
   4 3
   5 none of these

10. A chicken is a
   1 fish
   2 reptile
   3 bird
   4 mammal
   5 none of these

11. Which of these words is spelled incorrectly?
   1 heavy
   2 howse
   3 clock
   4 comb
   5 none of these

12. A good source of protein is
   1 lettuce
   2 candy
   3 bread
   4 fruit
   5 meat

PRACTICE TEST 2
Do Not Mark On This Test

Mark the number of the word that means the opposite.

81. happy

1 smart
2 joyful
3 sad
4 pleasant
5 none of these

Mark the number of the word that means the same.

82. quick

1 energetic
2 slow
3 happy
4 fast
5 none of these

Stop.

For the next three questions, mark 1 if the statement is true, 2 if the statement is false.

83. An apostrophe is a punctuation mark.

84. Light i a form of energy.

85. Cold air tends to rise.

Stop.

Mark one answer that is not correct.

96. 28 ÷ 7 =

1 2
2 4
3 6
4 8
5 10

Skip 97 on your answer sheet. The next question is 98. Go on to the next column.

Mark the number of the word that most nearly means the same.

98. angry

1 jealous
2 defeated
3 sad
4 amazed
5 furious

Stop.

For questions 99-102, decide if each statement is a complete sentence. If it is a complete sentence, mark answer space 4; if it is not a complete sentence, mark answer space 5.

99. It is spring.

100. The weather is getting warmer.

101. As the days become longer.

102. Playing outside with our friends.

Stop.

---

103. 5 x 3 =

5 10
4 5
3 17
2 15
1 12

Stop.

---

104. Mark choice 1 on your answer sheet. Then change your answer to choice 4.

Stop.

Session 2.

Approximate time: 45 minutes.

Topics: I. Review of Session 1.
II. Pacing and careful use of time.
III. Good guessing strategy.

General instructions to teacher:

Rather than write out the instructional material verbatim, as I did for Session 1, I have given it here in outline form. However, certain instructions to be read to the students are indented, while instructions to you are left-justified.

It is most important to go through the points slowly and thoroughly, checking constantly to be sure the students are understanding and correctly executing the practice exercises.

Again, I encourage you to write your comments in this booklet.

I. Review

This should be a short discussion in which you remind the students that they had a previous session on test-taking, give them a hint as to what it was about, and then prompt them to remember the major points of the session. These were:

Purpose of Testing

1) Test scores are supposed to show how much you know.
2) Teachers use them to help decide how to teach you best.
3) People in the state and city governments look at the scores made by students in our school to see how good a job this school is doing, and to tell whether the school needs to change its teaching to help students better.
4) It's important for everyone to try hard and get as many questions right as they can.

Answer Sheets

1) When you're given an answer sheet, look it over first to see how it is arranged, how items are numbered, what type of answer spaces it has, etc. This is to help you keep your place on the sheet while you are taking the test.
2) Always use a pencil to mark answers.
3) Mark only one space per item.
4) If you change an answer, erase the old one completely.
5) When marking an answer, fill in the answer space completely, but don't go outside the space. Make a dark mark.
6) Keep checking to be sure the number of the item you mark on the answer sheet is the same as the number of the question on the test.

Following Instructions

1) When you are told that a test is about to be given, start listening for any directions that may be given.
2) Don't go ahead and mark anything until you know exactly what the directions are for marking it. If you don't follow the exact instructions you may do it wrong.
3) Carefully read everything in each test question before you answer. Watch out for tricky instructions.

(Pass out one answer sheet to each student and have them fill in all the identification spaces.)

Outline: Pacing and Guessing

II. Pacing

A. Instruct students on the following points, in the order given:

1) Use all the time allowed for the test. Never stop working from the time the teacher says "Begin" till she or he says "Stop".

2) Take the time to read each question completely. Then read all the answer choices carefully. Think over each choice before choosing and marking your answer. Do not answer too quickly. (You might remind them again of the earlier discussion of the importance of thoroughly reading instructions.) Before beginning each new question, remind yourself by saying silently to yourself "READ AND THINK". Say this before each question, and then do it.

(Write the example items below on the blackboard, one at a time:

Example 1 Mark the number of the word that means most nearly the same.

enormous

1 small
2 large
3 unhappy
4 rich
5 huge

Tell them to say "READ AND THINK" to themselves before answering the question in their heads. Point out that if one answered quickly, without considering all the choices, he or she might pick large, when huge is really a better answer and the correct one.

Example 2 Pat has four cookies and Shelly has two brownies. If they share all the food equally, each of them will have

A four cookies and two brownies.
B one cookie and two brownies.
C two cookies and two brownies.
D two cookies and one brownie.
E none of these

Again, have them say "READ AND THINK" befor answering the question to themselves. Then point out that the answer choices look very much alike, so that one must read all of them carefully in orier to select the right one. It's also neccessary to read each choice all the way through, since each one has two parts to consider (a "cookie" part and a "brownie" part).)

3) Remind yourself to keep working by saying to yourself "DON'T LOOK UP" . If you catch yourself looking around the room, quickly tell yourself "DONT LOOK UP" and look back at your test. Make it a contest with yourself to try not to look at anything but your test until time is called.

4) Work through the questions in order. When you come to a question you can't do (after having read and thought about it carefully), skip that question and go to the next one. Be very sure to skip that number on your answer sheet, too.

5) If you get to the end of the test before the time is up, go back to the beginning and try again to do the questions you skipped. (A trick that will help you find them quickly is to put a tiny pencil mark next to each question you skip when you go through the test the first time. Erase the marks after you answer the questions.)

6) If after you have tried all the skipped questions a second time, and either answered them or decided that you just can't do them, there is still time left, go back to the beginning again. Start checking your answers. That is, read each question again and see if you still think the answer you marked if right. If not, change it. Check answers until time is called.

B. Summarize the main thrust of the preceding points, which is that one should concentrate complete attention on the test for as much time as is given. Especially critical is the point about taking time to carefully read each question and all the answer choices before deciding on an answer.

C. Give PRACTICE TEST 3, which is to be answered in Part I of the white answer sheet. Three minutes are allowed for the test, which should provide more than ample time for checking answers. Make sure each child has a pencil with an eraser and an answer sheet with the identification blanks completed. Review the points discussed above:

1) Don't stop working; use all the time.
2) Skip hard questions. After finishing the ones you can do easily, come back to the hard ones and try again.
3) See if you can keep your eyes on your test until time is called. Remind yourself by saying "DON'T LOOK UP".
4) If time remains after you've done all the questions you can, check your answers.
5) Before doing each question, remind yourself "READ AND THINK". Read the whole question and all the answers, then think carefully before marking your choice.

After passing out a test to each child (tell them to keep PRACTICE TEST 3 facing up), read these instructions:

> This is a practice test. You will have 3 minutes to work on it. Find Part I near the top of your answer sheet. Mark your answers in Part I, beginning with #1. Ready, begin.

While they work on the test, watch for drifting attention. Remind them when necessary: "DON'T LOOK UP". If anyone stops working before the time is up, remind them to check their answers again.

After 3 minutes, say:

Stop. Put down your pencils.

Have them trade papers and correct the tests. Tell them not to count an item correct unless a) the answer is correct, and b) the answer sheet is properly marked (help them make this judgement). Circle the number on the answer sheet of any missed item. Write the number correct at the top-left corner of the answer sheet.

Key:	1. 2	5. 4
	2. 6	6. 8
	3. 3	7. 1
	4. 8	

Next, have the papers returned to their owners. Try to get them to volunteer whether they:

a) kept working the whole time,
b) said "READ AND THINK" and "DON'T LOOK UP"
c) skipped hard questions and came back to them (you might ask about particular questions; prompt them to describe which items they skipped, etc.)
d) checked their answers.

III. Guessing

A. Instruct students on the following ideas about guessing. When should one guess at the answer to a question? If you have tried a question at least twice, and are sure you don't know the answer, you should make a guess only if you are pretty sure of at least one of the choices that is a wrong answer. (Show them what you mean using the question below as an example. Write the item on the blackboard. Discuss how, if one didn't know the answer to this question, he or she might be able to eliminate at least one choice (most likely A) and thus could profitably guess from among the remaining alternatives.

Example President Ford became president of this country in:

A 1960
B 1973
C 1974
D 1975

If most of them do not know the answer, you might have them all guess and see how many guess right.

When one has no idea at all about any of the choices, one should not guess. Leave the question unanswered. The reason for this is that one is too likely to guess wrong, and there may be a penalty for marking wrong answers.

B. Give PRACTICE TEST 4, which is on the reverse side of PRACTICE TEST 3. They should still have the page turned so that PRACTICE TEST 3 is facing up. Each child will need a piece of scratch paper. This test will be answered in Part II of the white answer sheet. Seven minutes are allowed. Peanuts will be given for correct answers. Announce the practice test and tell them they will win one peanut for each correct answer. Then before beginning, review all the major points once more:

1) Pay careful attention to instructions, both spoken and written;

2) Mark answer sheet correctly (if mistakes are made in marking, the question will be counted wrong) ;

3) Use all the time; don't look up;

4) Read each question and all its answer choices completely;

5) Skip over hard questions and come back to them if you have time;

6) Guess only if you know at least one choice to be incorrect, otherwise do not guess.

Read these test instructions:

> Do not turn over your test paper until I say begin. Find Part II of your answer sheet. This is where you will mark your answers. Do not make any marks on the test. Use your scratch paper to do any figuring. There are 14 questions; for every one you answer correctly, you will get a peanut. You will have 7 minutes to work on this test.
> Ready, turn over your test and begin.

After 7 minutes say:

> Stop. Put down your pencils.

Have them trade and correct papers (correcting for right answers only), circling the numbers of missed items.

Key:

1. 4	6. A	11. 3
2. B	7. J	12. A
3. G	8. D	13. 5
4. B	9. 3	14. B
5. J	10. D	

Have each student bring his/her own sheet to you to be checked for proper answer sheet marking. Mark any item wrong that has a marking errer. Write the number correct in the upper-right corner and give that number of peanuts to the child. Collect answer sheets and tests.

If any time remains, try to generate group discussion of pacing and guessing strategies employed (especially, which items were hard and were skipped over, and which were guessed at and on what basis), other techniques employed (e.g., did they tell themselves "DON'T LOOK UP" ?), and any problems encountered.

This is the end of Session 2.

NAME GRID: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z (repeated for each column)

TEACHER ONLY: STUDENT ABSENT FOR PART: I II III IV

STUDENT NUMBER: 0 1 2 3 4 5 6 7 8 9 (each digit column)

SEMESTER: FALL / SPRING; A / B

SEX

BIRTH DATE: MO (JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC) / YEAR 0 1 2 3 4 5 6 7 8 9

GRADE: 1 2 3 4 5 6

FORM OF THIS TEST IS: A B C D / 1 2 3 4

TEST ______ GRADE ______ INSTRUCTOR ______

Each item below has answer spaces 1 2 3 4 5.

I

1	6	11	16	21	26	31	36
2	7	12	17	22	27	32	37
3	8	13	18	23	28	33	38
4	9	14	19	24	29	34	39
5	10	15	20	25	30	35	40

II

41	46	51	56	61	66	71	76
42	47	52	57	62	67	72	77
43	48	53	58	63	68	73	78
44	49	54	59	64	69	74	79
45	50	55	60	65	70	75	80

III

81	86	91	96	101	106	111	116
82	87	92	97	102	107	112	117
83	88	93	98	103	108	113	118
84	89	94	99	104	109	114	119
85	90	95	100	105	110	115	120

IV

121	126	131	136	141	146	151	156
122	127	132	137	142	147	152	157
123	128	133	138	143	148	153	158
124	129	134	139	144	149	154	159
125	130	135	140	145	150	155	160

PRACTICE TEST #3
Do Not Mark On This Test

1. Which one of the following is not part of a plant?

   1 root
   2 soil
   3 stem
   4 fruit

2. Which sentence contains a mistake?

   6 There is many kinds of parks.
   7 Some are small, and some are large.
   8 Some are near where you live.
   9 Others are far away.

3. Will had nine apples. He gave three to Kent. Then Chris gave Will five more. How many apples does Will have now?

   1 6
   2 9
   3 11
   4 17

4. 9 x 4 =

   6 27
   7 32
   8 36
   9 none of these

5. Which sentence contains a mistake?

   1 What do you like to eat?
   2 Food is used by your body to help you grow.
   3 In the morning, you have breakfast.
   4 How many times a day do you eat.

6. All but one of these things are alike in a certain way. Which one does not belong with the rest?

   6 ruler
   7 scale
   8 telescope
   9 thermometer

7. 11 - 8 =

   1 3
   2 4
   3 5
   4 none of these

Stop.

PRACTICE TEST 4
Do Not Mark On This Test

1. Ron had 6 crayons. He found 2 crayons. How many crayons did he then have?

   1 2
   2 4
   3 6
   4 8
   5 10

2. Which of the following is an insect?

   A spider
   B ant
   C crab
   D worm
   E none

3. Solve: 325
   -57

   F 168
   G 268
   H 278
   J 372
   K 382

4. Emily needs 7 pieces of string, each 3 feet long. How many feet of string does she need altogether?

   To solve this problem, you must

   A subtract
   B multiply
   C divide
   D subtract, then multiply
   E none of these

5. Which word is spelled wrong?

   F while
   G which
   H whisper
   J whoose
   K none of these

6. What part of the carrot plant do we eat?

   A root
   B stem
   C leaf
   D fruit
   E seed

7. Mark the letter of the part of this sentence which contains a mistake.

   F G H J K
   Linda lives on Oak st. in Boston.

8. Mark the letter of the word which has a different vowel sound than the others.

   A line
   B right
   C find
   D sing
   E bike

9. Janice made 14 cupcakes. She took 6 of them to a friend. How many did she keep?

   How would this problem be written in the language of arithmetic?

   1 6 + 14 =
   2 6 - 14 =
   3 14 - 6 =
   4 14 x 6 =
   5 14 ÷ 6 =

10. Doug left his house at 3:20. He rode his bike directly to Dan's house. It took him 15 minutes. At what time did he arrive at Dan's house?

    A 3:05
    B 3:15
    C 3:25
    D 3:35
    E none of these

11. Which two words both have long vowels?

    1 and, than
    2 back, ball
    3 same, made
    4 saw, was
    5 bad, dad

12. To weigh something, you need a

    A scale
    B ruler
    C clock
    D gauge
    E thermometer

13. A table is 1 yard and 1 foot long. How long is it in inches?

    1 18
    2 24
    3 30
    4 40
    5 none of these

14. Which month comes after March?

    A February
    B April
    C June
    D August
    E October