

DOCUMENT RESUME

ED 186 469

TM 800 174

AUTHOR Engelhardt, David F.
 TITLE Motivation and Test-Wiseness. Director's Handbook. Topics in Testing, Measurement, and Evaluation, Volume V, Fall 1979.
 INSTITUTION New Jersey State Dept. of Education, Trenton. Div. of Operations, Research, and Evaluation.
 PUB DATE 79
 NOTE 13p.
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Educational Testing: Elementary Secondary Education; *Guessing (Tests): Guidelines: Student Attitudes; *Student Motivation: Teacher Attitudes: Testing Problems: *Test Wiseness
 IDENTIFIERS New Jersey

ABSTRACT

Because motivational factors and test wiseness can contaminate testing used for needs assessments or evaluation purposes, techniques for increasing student and teacher motivation are discussed. Guidelines concerning guessing are also presented. While guessing is encouraged on program evaluation or screening tests, it is not advocated for diagnostic tests; thus, different scoring formulas may be appropriate for different applications of testing. Instruction designed to increase student motivation and test wiseness is described, and is said to be potentially fruitful when several situations are considered: the student's will to win, feelings of individual powerlessness, withdrawal due to previous failure, purposeless testing, and teachers' negative attitudes. Suggestions to combat these problems include feedback to students and parents. Advice is provided concerning several issues in a test wiseness program: when students should guess; strategies for norm-referenced, criterion-referenced, screening, and diagnostic tests; and scoring formulas. It is concluded that students must be taught to pace their testing time, that item construction cues can be taught, and that practice is useful. (GDC)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

David F. Engelhardt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DIRECTOR'S HANDBOOK TOPICS IN TESTING MEASUREMENT AND EVALUATION

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY



MOTIVATION AND TEST-WISENESS

David F. Engelhardt

Introduction

This paper cites instances where two variables, motivation and test-wiseness, can contaminate needs assessment and evaluation studies when using the types of tests most often given by districts in the T&E process. Although tests claim to measure attainment of skills labeled in the item specifications, the scores often reflect variations in motivation or test-wiseness (these variables contribute to both invalidity and unreliability of data).

The author suggests some methods to increase motivation of students by generally increasing the chances of rewarding students and teachers. The teaching of test-wiseness and the practice of test-wise behavior is advocated. A caution against guessing is given for diagnostic tests, while guessing is advocated on tests used for program evaluation or screening (needs assessment). If a test is used for both diagnostic and program evaluation purposes, two separate scoring procedures could be used.

Recommendations are given with the spirit of reducing misclassifications of able students as being in need of basic skill remediation, increasing the validity of measurement, and increasing reliability of test scores. If students or non-random groups are being compared, as is done by scores on norm-referenced tests or is done with New Jersey Educational Assessment tests, the elimination of contaminate measures of test-wiseness is advocated as an important goal. To forget cultural or psychological differences in motivation and test-wiseness may lead to gross inefficiency in our remedial and preventive programs as well as incorrect evaluative conclusions.

Why don't test scores respond more easily to instructional effort? Can our kids really lack so many basic skills? Questions such as these can be asked of test results achieved by some public schools as well as of certain program areas (e.g. grammar) in the curricula of some elite, private schools. The questions can apply to results produced through norm-referenced or standardized criterion-referenced testing. As part of a management team, test coordinators ruminate over their assessments and evaluations with eyes of an examiner, statistician and decision-maker. Having done this myself for several years, I began looking at the test from the examinees' standpoint and concluded, in part, that we:

FALL 1979 Volume V

ED186469

NEW JERSEY STATE
DEPARTMENT OF EDUCATION
DIVISION OF OPERATIONS,
RESEARCH AND EVALUATION
Fred G. Burke
Commissioner of Education

164
ERIC
Full Text Provided by ERIC

(1) *Have failed to motivate or reward many students so they desire to perform to the best of their abilities, and*

(2) *Have often failed to improve the ability of students to demonstrate their skills on tests.*

The validity of our testing operations is affected by the above characteristics of the examinees.

This paper deals with observations of good and poor motivation and the need for test-taking instruction which leads to what some call test-wisness. Suggestions and references for remedial action are given so that a local school district can launch its own student motivation program if desired. Several school districts and at least one state have begun to develop test-wisness in their students, often under the rubric of teaching study-skills which have lifetime benefit. These school districts include Philadelphia; Washington, D.C.; Chicago; Dade County, Florida; and Montgomery County, Maryland. I am aware only of two film strip/cassette programs that address test-taking strategies (Guidance Associates of Pleasantville, N.J. and Lamport et al., 1976). It is recommended that these programs be previewed before purchasing. Books or syllabi recommended for use with the Maryland program or others are: Honig (1973), Hook (1967), Huff (1961), Jongma (1975), Millman and Pauk (1969), and Slakter et. al. (1979). Additionally, Erickson (1972), Ford (1973), and Millman et. al. (1965) outline the framework around which a fruitful test wisness program could be constructed locally.

The Will To Win

Of course, such test-wisness programs have little worth unless students wish to do well. Some score adjustments up to chance level of success, can be made to partially counter poor motivation. Such adjustments discussed later actually simulate test-wise behavior. I have seen test-wise seniors in an excellent/private school fail to perform on a Missouri College English Test, even with the headmaster urging students to do their best. This behavior was exhibited even though students knew the headmaster was desiring to legitimately evaluate the school's new Mechanics of English Program. Problems of obtaining the best efforts of any school's students seem to become more frequent as grade level increases. With proper teacher attitude in the primary grades, students often eagerly await the test -- almost as a game. Evidence of this eagerness is generally lacking in secondary students according to my experience. There is a chance that test-wisness instructional units may prove to interest some rebellious students who are intrigued by the idea of beating a system which has heretofore "turned-them-off." Nonetheless, motivation should and can be addressed outside of the test-

wisness unit.

One might conclude, students invariably must see a reward in the test's output for themselves before mustering all their test-taking energies and wisness. Let us consider a few situations affecting motivation:

Feelings of Powerlessness

Especially prevalent in children of lower socio-economic class is the feeling that no matter what they do, they have no power over what happens to them. Being subject to what appears as a capricious environment, such students, opposed to many middle class children, do not seem to develop the attitude that effort leads to success and eventually to better things. Eisenberg (1967) points out that middle class children find reward within a test, feeling that progress in scores is the path to success. A professor of elementary science education said to me once, "Perhaps the best reason to teach elementary science is to show some children that there is order to our world, and that through their mind and actions they can control part of their environment."

A related and hopefully more rare phenomenon can be encountered when a student feels his/her destiny is predetermined in a favorable sense, often by social bias (not ability). When ability is not a determiner of destiny, the student tests on a "birthright" to reach his/her goal. Such an attitude affects learning, but test taking situations may be even more sensitive. Since this attitude may not affect a long period of learning as strongly as a concentrated, sensitive request for demonstration of skills, the demonstration of skills may be more seriously hampered by the non-competitive attitude. Therefore, the test of skills will not reveal the learning which has taken place despite the nonchalant of the student. Some solutions to this might be to:

- (1) *Make tests more interesting.*
- (2) *Convince the student of a competitive world.*
- (3) *Establish a self-competitiveness, an attitude often adopted by star athletes.*

Withdrawal Due to Previous Failure

Eisenberg (1967) confirms what many educators have seen in older children, if children meet with repeated failure, it is much more rewarding for them not to compete at all. Many children, especially minority students, withdraw from a testing situation to "save face"; not to try and fail without trying seems better than failing despite one's best efforts. Unfortunately, the child does not discerningly choose where he/she might succeed if effort were mustered.

Eisenberg found that lower-class children were likely to give any answer that would end the testing, regardless of whether the answer was right or wrong. The child perceives that failure to exert effort taints with vagueness any conclusions as to ability, thereby saving face by having lowered the confidence of statements about the examinee. Generalizing to our situations, we might say that the child has worked toward minimum loss under avoidance strategies. The program's evaluation therefore suffers through measurements of low confidence.

The solution seems obvious: in areas of basic skills and some special talent areas, give the child a taste of success. Appropriate level (out-of-grade) testing may help, but most important is the prior classroom experience. It may be possible to encourage students to engage in test activities by giving them successful and recent experiences with material similar to the forthcoming test.

Purposeless testing, and teacher attitude:

Who in testing has not at sometime heard the complaints of teachers when examination time approaches? Students are quick to absorb the sense of purposeless test-taking where results are rarely used by the teacher, never shared with students. It is well recognized by many test coordinators that teachers' attitudes and overt concerns regarding test results are major incentive factors in student performance. Of course, the purposes and consequences of the testing should be explained before testing -- not as an after-thought, which would have little effect on test-taking strategy.

Convincing teachers and students of the purpose served by a test is not a small task. Furthermore, the process may backfire in special circumstances where program evaluation or screening is a major purpose: It is possible that a vengeful student may capitalize on such knowledge to attack a teacher, principal, or the system. In another situation, revealing the purpose of Title I testing in a suburban New York area school system lowered test scores because children wanted to qualify for the special summer program which involved field trips, games, and reading skill instruction. Since New Jersey has provided programs for under-achieving gifted or talented students, it wouldn't be surprising if some gifted students might desire to qualify for gifted and talented compensatory education programs by scoring low on achievement tests. Such attempts can be dampened by using other criteria to verify screening test scores.

As long as a system has healthy relations between staff and students, and attractive program alternatives, explaining purposes of testing and making the test purposeful at student and teacher levels will probably increase scores. Some suggestions are:

- (1) **Make sure teachers receive all interpretative manuals and order time-saving reports from computerized scoring services. Don't expect hand tallying.**
- (2) **Involve staff in the selection of tests and provide opportunity for criticism of the tests and report format. With mandated tests, allow teachers to help construct or declare certain items non-relevant as a district. Obviously this only pertains in a criterion-referenced approach.**
- (3) **Teachers must be able to interpret test results. Unfortunately, central office staff rarely have time to do justice to in-service or conferences regarding test results. Unless administrators shift priorities to allotting more in-service training time to test data analysis, we must rely on better teacher training, clearer report formats, and teacher desire to self-educate. Building committees may help educate teachers with more flexibility than district training.**
- (4) **Increase feedback to parents and students with computerized reports and teacher conferences: Purposefulness of any test wanes with delay in the return of results. Feedback should be as current as possible and in time for decisions predicated on test results. Such feedback should be expedited by methods discussed elsewhere in this Handbook. Pretend, as a test coordinator, that you are processing blood samples for diagnostic work in a hospital. Schedule testing so mails can work for you over weekends and holidays. Pre-correction processing might take place on a 20-hour work schedule. Ship results by air. The added cost is minor compared to the effort and cost expended in testing.**

Of course, in-house or cooperative correction, can give the best turn-around time on correction. For certain testing programs, self-scoring sheets or student scoring can be utilized. Calculators and teaching machines can offer immediate feedback with a record for the teacher.

In Long Branch, we have had the experience of students calling our guidance department during the summer to find out results of their Planning Career Goals Test Battery. The Battery contains guidance information matched with basic skills assessment. This group of students has been tested as juniors in the late spring and in spite of a motivationally difficult time to test, the students exhibited interest because of the chance for personal benefit. However, we were not adequately staffed to respond to these summer inquiries. If anything is to be learned by this occurrence, it is to suggest that provision for summer counseling should be built in as a follow-up to spring testing as a stimulus to increase test

performance. Under current regulations, such follow-up is not considered fundable under compensatory education. Yet failure to provide such follow-up by funding a summer guidance counselor probably will increase the compensatory education load since students see less purpose in the test. Relating test scores to career goals (even if these goals are temporary appeals to many students in the upper grades; the appeal is much greater than can be aroused through threatening more homework or a poor program evaluation.

- (5) Although the main purpose of some testing may be for program evaluation, stress some personal use of test data for the student. Durost's (1974) indictment of some Title I test-taking strategies demonstrates that program evaluators must be sensitive to high guessing and to poorly motivated students. As suggested above, using guidance related test or establishing guidance related norms may serve as an effective way to interest students in test results. Currently most commercial norming does not provide interesting norms for the student. A secretarial student may be interested in how well he/she does on a grammar section of a test in relation to other commercial students, not in relation to all students in large cities or in the northeast. On the other hand, the test coordinator would want the more general norm for assessment. Planning Career Goals (CTB-McGraw Hill) allows such dual comparisons. When given the opportunity to take extra sections of the test, 50% of the students in one district used their free time for more testing.

Techniques used to reduce testing time by dividing items among students may lower motivation if students see no individual consequence of taking the test. Such approaches as matrix sampling should be carefully monitored.

- (6) Provide guides for student interpretation to relieve teachers or guidance counselors of some interpretation. The Planning Career Goals Test has exemplary student materials. Most other batteries have limited hand-out materials.
- (7) Provide instructional program response to demonstrated need.
- (8) Include certain results of standardized or departmental criterion-referenced tests as part of a student's grade or as extra credit. Embed certain evaluation or assessment items in normal classroom activities. Most test manual directions have stressed the reduction of anxiety on the part of students. Possibly the pendulum has swung too far, many students lack any hint of anxiety --

some even sleep during assessment tests! Certainly some standardized tests (e.g. MLA foreign language tests, Howell Geometry Test, Missouri College English Test) are valid enough for particular courses to warrant credit be awarded toward a student's grade.

Variety in the testing program:

Providing variety in the testing program, which a student experiences, may have motivational consequences, although I am not aware of any formal studies on this factor. Even if the standardization programs that generate norms for our tests suffer from possible fatigue and boredom (especially "alternate form" norming), it is probably not wise to try to duplicate such negative factors. Test Coordinators can try to avoid examinee boredom by varying the testing approach and series. I have heard counselors remark more than once, that after four or six years of the same test, students just don't try to perform on the test -- even if the questions are different (but of the same style.) It seems that the title page of the test is enough to dissuade some students from trying, even though the various levels in a series do have different questions. In New Jersey, the minimum skills test may provide a break from the yearly administration of a test series. Some life-skill tests may be utilized to good motivational-end in providing variety in a district's test schedule. Measurement of growth on one scale might have to be delayed a year or two, but this may increase the validity of the measurement.

Wade Boykin's research at Cornell University shows cultural differences in reaction to variety in test stimuli. Perhaps unsuccessful students look forward to trying new test situations in which to prove their abilities, whereas successful students look at "variety" as a threatening challenge.

Teacher enthusiasm:

If teachers are confident that they can reach a program improvement goal, their enthusiasm to demonstrate such may stimulate students to perform. Impossible program achievement goals may tend to dampen teacher enthusiasm, which leads to a poor orientation of students toward the test. The use of short-range goals with "front-line workers" is often a better management technique than revealing long-range goals, or, even worse, evaluating the worker on long-range goal standards. Teacher confidence also can be increased by adequate pre-test orientation of teachers, even if only in the area of examiners' instructions. A confused teacher during administration of the test also leads to a poor orientation of students toward the test. A confused teacher can hardly be expected to show much enthusiasm for the test

which contributes to such unwanted insecurity or embarrassment.

Test-Wisness

Assuming that teachers and students desire to perform, the next concern for the test-coordinator is to allow pertinent skills to be evaluated or assessed with reliability and validity. The Concern Is Not To Measure Or Compare Non-Relevant Skills. If we wish to measure communication skills, let's not measure the ability of the students to pace themselves on the test or some other test-taking skill.

Certain criticisms levied against test-wisness training often accuse the school of "teaching to the test." I am not suggesting rehearsal of test questions appearing verbatim on the test; nor am I suggesting cramming for the content of a test. What I am suggesting has been well established in the literature (Eakins et al., 1976; Erickson, 1972; Fenton and Mueller, 1977; Ford, 1973; Maryland Department of Education, 1975; Millman et al. 1965; and Sabers, 1975); the teaching and acquisition of test-wisness -- *the ability to reliably demonstrate the full extent of one's pertinent skills and knowledges through the medium of a valid test, including the demonstration of mastered and partially developed skills.*

Without such training, many students who fail to achieve minimum competency scores may possess satisfactory skills in reading and mathematics. Such failures to prove competency contribute to the overload on remedial services. Special funding is allotted to teach reading when, for some students, it might be best to teach test-wisness. Concern for cost-effectiveness should urge development and implementation of test-wisness units and provision for good testing environments. The quest for valid instruments in the minimum competency movement may be severely confounded due to the variability in test-wisness and testing environment (including teacher attitudes).

Ford (1973) reports that "coaching" (teaching content area of the test and cramming) before tests has not been shown to raise scores as much as test-wisness study which avoided instruction in the subject matter to be tested (Eakins et al., 1976). For instance, Barron's guide How to Prepare for College Entrance Examinations (Brownstein and Weiner, 1969) is more a coaching book as compared to Honig (1973), Hook (1962), Huff (1961) and Millman (1969). Private schools and some public schools might bear Ford's conclusions in mind when constructing preparatory courses and selecting appropriate materials for PSAT and SAT examinations.

With poorly defined domains, it is difficult to ascertain what an item intends to measure and what is

germane to test-wisness; to teach the former is coaching (if done specifically for the many domains just prior to testing) while to teach the latter is what concerns us now. Fenton and Mueller (1977) point out that to teach to the domain of the test is legitimate. If done well in advance of a battery or long evaluation test, such teaching is the essence of the instructional program. No one advocates teaching specific items to be found on the test. Sabers (1975) emphasizes that psychologists, through the American Psychological Association, deemed it essential that the examinee be given the strategy to maximize his/her test score. The following factors are considerations when devising a test-wisness unit.

Should a student guess on examinations?

Perhaps the most significant and controversial facet of test-wisness pertains to guessing on multiple choice tests. The author's opinion is that we should not shy from increasing the pace of students or urging the use of certain techniques so that all answers are completed, even if this may result in some blatant guessing. In fact, some scoring formulas correct for leaving questions unanswered by adding to the number of right answers, the chance score that might have been obtained by answering all unanswered questions. This yields a corrected raw score which is equal to or higher than the number right. If it is not possible to alter present correction formulas (due to economics or inflexibility of either current computer programs or standardized test correction procedures), physical alterations of answer sheets can approach the same end. Such credit for unanswered questions gives each examinee his/her maximum benefit from test-wisness in guessing strategies as if he/she were very test-wise. Of course this method makes it very clear that assigning meaning to raw scores at or below chance level is an erroneous procedure except when measuring the tendency to guess. Such a statement is true even when the "number right" scoring formulas are used.

If we are comparing students or non-random groups, as is done by scores on norm-referenced tests or with item performance results on the New Jersey Educational Assessment Program tests, it would behoove us to eliminate contaminant measures such as test-wise guessing. Giving credit for unanswered questions eliminates the penalty for students who were too cautious, withdrew from competition, were mistakenly cautioned not to guess, or exhibited other unwise test behaviors. Students not needing to guess or having greater partial knowledge will have increased chances of being correct as compared to "pure guessers" and will still obtain higher scores even though the knowledgeable examinees may not have their scores increased as much by the above scoring correction.

To use number-right correction formulas ignores the problem of the extraneous guessing variable to provide the above correction and places this variable at its maximum estimate so that comparative interpretation can be improved. With criterion-referenced tests, this maximum estimate can be considered a floor of performance levels--obtainable with no relevant skill.

It is extremely difficult to prevent guessing (although providing points for unanswered questions retards guessing), but we can reach its maximum limit by encouraging completion of the test or through scoring formulas. The limit varies with the ability of the examinee in one sense; the higher the ability, the less guessing will have occurred. However, on multiple choice tests, a certain score could be achieved without reading the test and blackening the first answer for each question. This score could be considered the maximum limit of the chance score. An individual's maximum guessing limit is the difference between an achieved score reflecting the student's actual ability and a score obtained by adding the remaining possible score points due to chance on any given assessment instrument.

If knowledge accounted for part of the score which might have been attributed to chance, that degree of knowledge is of little practical consequence. Furthermore, only non-standard scoring techniques, not available through most test correction services, would distinguish between some guessing and confident answers.

Ultimately, to advise students not to guess leaves the examiner uncertain if some examinees actually guessed. Therefore to reach the minimum extent of guessing is not as reliable nor realistic as to reach the maximum extent.

Are There Different Guessing Strategies For CRTs and NRTs? Crocker and Benson (1976) distinguish advice based on criterion-referenced tests (CRT) and norm-referenced tests (NRT) based in part on student reaction to the different test types. While it is true that the scoring of criterion-referenced tests can be more easily adapted to measuring confidence of answers or giving credit for indicating wrong answers (coombs-type scoring, Jacobs, 1974), the uses and the anxiety reactions to CRTs, and NRTs do not seem so different to me as to deserve different test-taking strategies. Standards to be set for CRTs might consider equalizing points gained through chance guessing by ensuring completion of all items and then acknowledging that part of the test score could be accomplished by random guessing (millman, 1973). It should be noted that some tests employ a high number of answer options to reduce chance levels on scores. Edits Diagnostic Mathematics In-

ventory (1977), for example, uses seven options.

Aren't There Sophisticated Approaches To Measuring Which Avoid The Guessing Problem? At this juncture, it should be stated that nonconventional approaches to measuring student abilities can eliminate problems of guessing. This paper will not attempt to deal with confidence testing (see Brennan, 1974 and Stegman, 1973) or latent trait theory (see Waller, 1974), neither of which is generally utilized with most standardized tests.

Do We Urge Consistent Strategies With Screening and Diagnostic Tests? Rather than distinguish between CRTs and NRTs as to appropriate answer strategy it does seem fruitful to distinguish between diagnostic and screening tests. When a student takes a diagnostic test, benefit comes from knowing what he/she may need to review or learn. Diagnostic tests usually have more questions on each defined skill. An incorrect or missing response has a chance to be confirmed or negated by other questions.

In The Case of Diagnostic Tests. students should not be encouraged to guess since considerable benefit comes from missing questions and therefore getting needed review. If diagnostic tests are used for summary data comparisons as in evaluation, the proportion of omitted answers that might be answered by chance could be added to the raw score in a separate scoring procedure. Lord (1975) points out that such a formula correlates with traditional "guess correction" formulas which subtract a portion of wrong answers. Of course, adding points for omitted items does raise the score in absolute terms. It may be of interest to note that if we instruct students that on diagnostic tests correction we will give points for omitted answers, guessing will decrease significantly more than if we just instruct students to "not guess." (Diamond and Evans, 1973).

On Screening Tests. not to be generally used as a diagnostic test, the upper limit of a person's true score (partial knowledge included) can be approached by urging completion of all questions unless the test is speeded. The tendency to guess can be considered so variable that comparisons of norm groups would be more accurate by at least completing the entire test. This would partially counter scores poorly motivated students that might not benefit from chance answers. It would minimize test-wisness differences in large cities and suburban districts -- at least in one major factor, risk-taking. Durost and Hodges' study of Title I testing (1974) showed that mean raw scores rose about five points on some of the 1973 Stanford Achievement "Reading" tests; however, sex differences in performance existed. Standard deviations also rose in some cases, although reliability of test scores tended to increase with

guessing in other experiments (Diamond and Evans 1973 and Rowley and Raub, 1977). No clear empirical data of which the author is aware shows that increased guessing would make it harder to find significant gains in evaluation studies.

Rowley and Traub (1977) concluded that examinees having risk-taking personalities benefit from "do not guess" directions, and that most examinees cannot distinguish pure guessing from informed guessing. Numbers of studies have shown that scores are higher when guessing or faster pacing is encouraged. When left to their own test taking strategies, the personality characteristics of students are strong influences over guessing (Cross and Frary, 1976; Diamond and Evans 1973; Durost and Hodges 1974; Ford, 1973; Rowley and Traub, 1977; and, Sherman, 1976). The tendency to guess (or not to respond when any doubt exists) is probably related to socio-economic class and seems to be strongly associated with groups of minority students (Sherman, 1976). It should be noted that Maryland's attempt to teach test-wiseness arose from a suggestion from an advisory committee on minority relations in Montgomery County.

Should We Urge Guessing When Correction Formulas Are Applied? Almost all references agree that even with correction formulas being applied for guessing, a person who can eliminate one wrong answer option from the remaining is legitimately raising his/her score by guessing. When correction formulas are used (as with College Entrance Examination Board Exams, but rarely with batteries used by districts), blind guessing (blackening in the spaces) wastes time -- time which might be better used in reasoning a difficult question. For groups tested with correction formulas for guessing, averages rarely go down even with blind guessing. Scores will obviously not go lower if a number right scoring system is used. Furthermore, Rowley and Traub (1977) present thesis data using ninth graders which demonstrate that 41% of answers claimed to be non-informed guessing were correct guesses, when 25% would have been at the chance level. They conclude that students cannot distinguish between informed and blatantly random guessing. This suggests that misjudgment on the part of a student does no harm when it results in guessing when correction formulas are used. Severe harm is done when a student neglects to guess on a test corrected by the number right formula. One should not urge blatant guessing when guessing correction formulas are applied. However not much is lost if the student "over-generalizes" from the normal type of school exam.

Where Can I Find Additional Discussion on the Guessing Question? Opponents to encouraging guessing and Durost and Hodges (1974), Lord (1975), and

Sherman (1976). The reader may wish to consult these references prior to establishing a test-wiseness program. Those advocating guessing, especially with number-right/correction formulas and particularly with the possibility to eliminate one wrong answer, include Ford (1973) and Rowley and Traub (1977). All advocate test-wiseness instruction.

What Might One Conclude From The Ideas Presented In This Paper? In summary, on assessment and evaluation tests, completion of the test does not seem to harm evaluative, group comparisons and may eliminate some variance due to test-wiseness. If scoring formulas that add "chance points to raw scores" cannot be found, mechanical blackening in of answer sheets can accomplish the same adjustment. Such an adjustment at least puts all students in comparison groups on the same footing in regard to guessing. Differences in scores will then be more heavily dependent on the variable intended to be measured. The standard deviation of group scores might be reduced, whereas the effect on individual scores is controversial. Empirical studies do not seem to confirm theoretical models which disregard "educated guessing."

Durost and Hodges' (1974) research deserves scrutiny beyond this paper since it contains data quite generalizable to needs assessment operations in New Jersey and also to an experimental variable encouraging students to complete the screening instrument. Insightful comments on their research reveal that test-wise behavior was lacking in many New Hampshire youths, with cautions stated about the meaningfulness of much of the test data. The paper's data may need reworking before the reader is willing to accept Durost and Hodges' conclusion that attempts to eliminate guessing would yield more information on such non-diagnostic tests. Their thoughts on criterion-referenced item analysis and the placement of items on a test in order of difficulty should be carefully evaluated. They conclude, as do many test manufacturers, that mathematical correction for guessing has no benefit in reducing guessing or reordering students in performance rank. It does accomplish lower scores in an absolute sense, but may also discriminate against certain personality types.

(1) *Students Must Be Taught To Pace.* There appears to be some evidence that "speed reading" a test is a test-wise approach to demonstrating abilities to be measured. Miller and Weiss (1976) found providing time limits on difficult items on tests did not reduce accuracy, but did increase test taking speed. The Maryland test-wiseness syllabus (1975), trains students in adjusting pace according to the

type of sub-test (also see Hallberg, 1971). For non-fiction reading items, it recommends reading the question and scanning for the answer. So we see that approach can even vary by item type (Maryland State Department of Education, 1975), much as regular reading or text-studying behaviors might be adjusted for the type of book being read.

Students should be urged not to become disturbed if they cannot answer every item with definite confidence (Sabers, 1975). A student willing to abandon certain questions can speed on to more questions. Spatial relations tests are an excellent example of tests on which a person can become bogged down on certain questions. It is not unknown to experience a change of 70 percentile ranks when taking such a test a second time and pacing oneself in a more cursory manner.

Two necessary ingredients for pacing instructions are practice and being able to have an external time check. Classroom practice and yearly examinations can provide test-wiseness to varying degrees in K-12 curricula. Practice on certain forms of items which will be encountered on a test does influence pacing more than counting on generalization. Possibly, however, even SAT scores may eventually begin to rise as students are provided more testing experiences under general accountability pressures.

The second factor in pacing is being able to see some timing device and to calculate time intervals. The procedure of writing times on the chalkboard or announcing time left is thought by some to be too disturbing (e.g. 26). A silent clock in the room, viewable by all, with finish time written on the board (by diagram for youngsters who can't tell time) is a good solution. It should be noted that many children cannot afford wrist watches and their schools may not have operable clocks. Are normative or evaluative comparisons not made invalid by such problems associated with socio-economic conditions?

The recent trend to digital clocks may require that test coordinators supplement time pieces in testing rooms. Depending upon what fraction of a minute is viewed on a digital or impulse clock, an examinee's "minute" check may be over in a few seconds. Furthermore, many digital clock faces are less visible to students than classical, sweep second wall clocks.

(2) *Some Students are Aware of Item Construction Cues.* Certain groups of teachers and students may perpetuate a naivete on item writing cues because no one ever bothered teaching them how to take a test or how to write items. Diamond et. al. (1976) and Millman et. al. (1965) deal with test-wiseness

construction. One may wonder how test constructors can commit item writing errors so blatant that such test-wiseness instruction can be beneficial. Districts should carefully assess the worth of teaching how to spot incorrect or correct answers by such factors as length of option, matched graphemes in stem and answer and use of ungrammatical alternatives. It is true that some test publishers and many teachers do still commit such errors.

(3) *Provide Psychomotor Practice for Some Test Answer Sheets and Students.* With some answer sheets, young students can benefit from practice in recording score (Sabers, 1975). Some test-wiseness programs provide "answer entry" practice for several days prior to testing (Maryland State Department of Education, 1975). Test coordinators should choose answer recording format and correction services with care. Extra money spent in assessing can save unnecessary remedial expenditures.

Some have discovered that the rewriting of horizontal math problems poses a test of small muscle coordination rather than math competency. If conclusions drawn from assessments might be connected with psychomotor preparation, notations during testing or during test-wiseness preparation could be used during item analysis to increase the appropriateness of instruction. Hopefully, test-wiseness instruction could have a psychomotor component to increase the validity of screening tests following such instruction.

(4) *Practice With Format of Items and Allow Students to cultivate Familiarity with Directions.* The student who understands directions well and does not have to refer to directions during a timed test has a definite advantage over the student unfamiliar with item format and the directions. Some types of questions are so involved that weeks of practice are needed. Practice embedded in the normal instructional routine is more efficient than specific units in format practice (Eakins, et al., 1976; Ford, 1973; and, Maryland State Department of Education, 1975).

(5) *Urge Students to Check Answers.* Although unsure guesses may generally be correct on the first try, reasoned answers have proved more correct when corrected or checked once. Continual change of answers is dangerous. Ford (1973) warns against pondering over a question at length.

(6) *Demonstrate That Eliminating at Least One Answer Helps Raise Scores.* Ford (1973) recommends reading all options before deciding on one option. If time is limited, locating at least one wrong option and then choosing what looks like a good answer may prove more efficient. Once again, blank answers or "don't knows" are not useful in the author's opinion.

Reaction To This Paper

Not all existing programs, not all pertinent journal articles, have been reviewed for this paper. It is hoped that readers will reply with added suggestions for further development of issues discussed here. A test-wiseness program developed with suggestions contained in this paper will do no harm to student scores, nor will it confound evaluation practices. We will, however, make an effort to continue dissemination of possible improvements to tests-wiseness programs.

We hope some response is forthcoming from test publishers to concerns expressed in the paper, especially in revising test correction formulas so districts might select appropriate techniques.

Our aim in suggesting test-wiseness courses is to improve the measurement process, not destroy it.

About the Author...

David F. Englehardt is Assistant Superintendent of Schools for the Long Branch Public Schools, Long Branch, New Jersey. He received his B.S. degree from Yale University, an M.A.T. from Duke University and an Ed.D. from Harvard Graduate School of Education.

His areas of specialization are measurement and program evaluation, research utilization, and long-range educational planning. Prior to his present position he has worked with the Massachusetts Department of Education, served as an educational consultant working with clients on the eastern seaboard and has taught biology.

In addition to being a member of numerous professional organizations including AASA, NJASA, Phi Delta Kappa, AERA and NCME,

Small editing corrections
have been made in this copy
by the author. Comments
for the next revision may
be sent to his home:

50 Rosalyn Court
Long Branch, N. J. 07740

References

- Brennan, Robert L. The Evaluation of Mastery Test Items. NCERD Project: Final Report. Educational Resources Information Center (ERIC), ED 092,593, 1974.
- Brownstein, S. C. & Weiner, M. Barron's How to Prepare for College Entrance Examinations. Woodbury, New York: Barron's Educational Series, Inc., 1969.
- Crehan, K. D. et al. 1977. Developmental Aspects of Test-Wisness. Paper presented at the New York meeting of the American Educational Research Association, April, 1977. Educational Resources Information Center (ERIC), ED 137,394, 1977.
- Crocker, L. & Benson, J. 1976. Achievement, Guessing, and Risk-taking Behavior Under Norm Referenced and Criterion Referenced Testing Conditions. American Education Research Journal, Summer 1976, 13, 207-215.
- Cross, L. & Frary, R.B. A Study of Omitted Responses Under the Conventional Correction for Guessing. Paper presented at the San Francisco meeting of the American Educational Research Association, April, 1976. Educational Resources Information Center (ERIC), ED 120, 231, 1976.
- Diamond, J. J. et al. 1976. Are Inner City Children Test-wise? Journal of Educational Measurement, Spring, 1976, 14, 39-45.
- Diamond, J. & Evans, W. The Correction for Guessing. Review of Educational Research, Spring, 1973, 43, 181-191.
- Hallberg, A. Facing the Test: English. Miami, Florida: Dade County Public Schools. Educational Resources Information Center (ERIC), ED 064720, 1971.
- Honig, F. 1973. Taking Tests and Scoring High: How to Study for Tests and Pass Them. New York: ARCO, 219 Park Avenue, 1973.
- Hook, J.N. 1967. Testmanship: Seven Ways to Raise Your Examination Grades. New York: Barnes and Noble, Inc., 1967.
- Huff, D. Score: The Strategy of Taking Tests. New York: Appleton - Century - Crofts, Inc., 1961.
- Jacobs, S.S. Behavior on Objective Tests Under Theoretically Adequate, Inadequate, and Unspecified Scoring Rules. Paper presented at the Chicago meeting of the American Educational Research Association, April, 1974. Educational Resources Information Center (ERIC), ED 096312, 1974.
- Jongsma, E.A. The Effects of Instruction in Test-taking Skills Upon Student Performance on Standardized Achievement Tests: Final Report. Baton Rouge, La.: Louisiana State University Foundation. Educational Resources Information Center (ERIC), ED 114408, 1975.
- Klaas, A.C. A Reassessment of Standard Error of Measurement. Paper presented at the Washington, D.C. meeting of the American Educational Research Association, April, 1975. Educational Resources Information Center (ERIC), ED 104949, 1975.
- Lampert, J., Perle, R. & Reinarch, J. Taking Tests and Relaxing. Film strip and cassette programs. Norwalk, Conn.: New Dimensions in Education, Inc., 83 Keeler Avenue, 06854, 1976.

7
Lord, F.M. Formula Scoring and Number - Right Scoring. Journal of Educational Measurement, Spring, 1975, 12, 7-11.

Maryland State Department of Education. Improving Student Attitudes and Skills for Taking Tests. Occasional Paper Series on Accountability, Author. Educational Resources Information Center (ERIC), ED 128352, 1975.

Miller, T.W. & Weiss, D.J. 1976. Effects of Time Limits on Test Taking Behavior. Office of Naval Research Grant, University of Minnesota: Office of Naval Research Grant. Educational Resources Information Center (ERIC), ED 125123, 1976.

Millman, J. Passing Scores and Test Lengths for Domain-Referenced Measures. Review of Educational Research, Spring, 1973, 43, 205-216.

Millman, J., Bishop, C. & Ebel, R. An Analysis of Test-Wisness. Educational and Psychological Measurement 1965, 25, 707-726.

Millman, J. & Pauk, W. How To Take Tests. New York: McGraw Hill Book Company, 1969.

Moore, J. Schultz, R. & Baker, R. 1966. The Application of a Self-Instructional Technique to Develop a Test-taking Technique. American Educational Research Journal, 1966, 3, 13-17.

New York State Education Department. SPPED Cloze Training Manual, Form C82. Educational Resources Information Center (ERIC), ED 127353, 1976.

Oakland, T. Effects of Test-Wisness Materials on Standardized Test Performance of Pre-school Disadvantaged Children. Journal of School Psychology, 1972, 10, 355-360.

Rowley, G.L. & Traub, R.E. Formula Scoring, Number Right Scoring, and Test-taking Strategy. Journal of Educational Measurement, Spring, 1977, 14, 15-22.

Sabers, D. Test-taking Skills. Tucson, Arizona: Arizona Center for Educational Research and Development, College of Education, University of Arizona. Educational Resources Information Center (ERIC), ED 133341, 1975.

Sherman, S.W. 1976. Multiple Choice Test Bias Uncovered by Use of an "I Don't know" Alternative. Paper presented at the San Francisco meeting of the American Educational Research Association, April, 1976. Educational Resources Information Center (ERIC), ED 121824, 1976.

Slakter, M., Kochler, K., & Hompton, S. 1970. Learning Test-Wisness by programmed texts. Journal of Educational Measurement, 1970, 1, 217-254.

Stegman, C.E. Subjective Probability and the Administration of Objective Tests. Paper presented at the Northeastern Educational Research Association Convocation, 1973. Educational Resources Information Center (ERIC), ED 091427, 1973.

Wahlstrom M. & Boersma, F. The Influence of Test-wisness Upon Achievement. Educational and Psychological Measurement, 1968, 28, 113-120.

Waller, M.I. Removing the Effects of Random Guessing from Latent Trait Ability Estimates. ETS Research Bulletin, 1974, 7, 1-32. Educational Resources Information Center (ERIC), ED 099417, 1974.

PUBLICATIONS IN THIS SERIES INCLUDE:

- Volume I. Purpose of Testing.....*David W. Pearsall*
- Volume II. Test Selection and Evaluation of Tests.....*Malcolm J. Conway*
- Volume III. Organizing and Managing a
District Testing Program.....*Robert L. Jackson, II*
- Volume IV. Some Uses of Standardized
Test Results.....*Eugene J. Travers*
- Volume V. Motivation and Test Wiseness.....*David F. Englehardt*

Published by the New Jersey Department of Education, 225 West State Street, Trenton, New Jersey 08625. The opinions expressed in this publication are those of the invited authors and do not necessarily reflect the position or policy of the New Jersey Department of Education.

ACKNOWLEDGMENTS

Thanks are extended to the Planning Committee for this publication and to the editorial review panel consisting of Dr. Curtis Banks, Princeton University; Dr. Douglas Penfield, Rutgers University; and, Dr. Paul Van R. Miller, Lehigh University.

Appreciation is also expressed to Ms. Vivian Sobke for preparation of the manuscripts.