

REVIEW OF: TESTING, TESTING, TESTING

John R. Hills
Director, Testing and Guidance
Regents of the University System of Georgia
Atlanta, Georgia

What is Testing, Testing, Testing?

I have read the pamphlet Testing, Testing, Testing now many times and with great care. I'm going to tell you in a minute what it contains, but I think one must ponder the whole document, indeed the whole endeavor that produced the document, to comprehend what it means-- what it signifies. I think that the pamphlet is a symbol. It symbolizes a wail of anguish on the part of educational administrators at their lack of sophistication in the area of evaluation, one of the critical parts of the educational endeavor. It is a circuitous expression of their unhappiness at having their ignorance found out and at being forced to go back and learn something that they earlier avoided because it might prove difficult. Rather than take that course of action, they harken back to the good old days before testing became so ubiquitous that everyone is involved, everyone knows a little bit, and everyone expects the professional educator to know a lot more!

What is Testing, Testing, Testing about?

Testing, Testing, Testing primarily is a presentation of a single topic. That topic is a criticism of the educational profession from the teacher through the entire school or system. The criticism starts on page 14 and continues through the last page, 32, with only one letup for more than a single page. For example, on page 14 the school administration in one state is criticized for giving five tests when one would do; on page 15 administrators are criticized for abdicating professional responsibility and submitting to community pressures; on page 16 there is a brief description of what must be some sloppy test interpretation and counseling, but the schools are roasted for their belief that tests are infallible, and so on, page after page. Three different times educators are exhorted to do something about their weaknesses in the area of evaluation and measurement.

There are some strange passages in this pamphlet. Some are nonsense, e. g., "Testing programs often cause displacement within the school," or "To evaluate ... means to appraise, not just to measure," or "Though the standardized test seemingly has merit of reducing guesswork and increasing accuracy and objectivity, the process is reductionistic." Some are irrelevant, e. g., the passage stating that, unfortunately, the reward for high performance on tests is placement in one of a limited number of specialized fields of high remuneration. Some passages are innuendo, e. g., the statement that the interests of some parents would be better served by "understanding the effects of external test pressures on the mental and emotional health of their children," or the one indicating that external testing programs may detrimentally impinge upon the relationships between democracy and education. Aside from criticism of educators, and this nonsense and innuendo which give the document a flavor vaguely opposed to testing, there is some discussion of educational measurement. This discussion is not particularly good; in fact, some of the nonsense appears here along with unproven assertions, naive discussions, and erroneous statements. However, the authors do state in bold type, "TO TEACH WITHOUT TESTING IS UNTHINKABLE," and they say that success can to some extent be predicted. They even go so far several times as to say that testing should be more inclusive and more extensive. So, though you may not have realized it on a quick reading, the authors of the pamphlet blasted educators while appearing to talk knowledgeably about someone else!

The Recommendations in Testing, Testing, Testing

There are eighteen or more recommendations following the introductory pages, whose content I have described above. The authors number eighteen recommendations, but several are double-barrelled, and two should probably be combined (and then discarded together). I will take the recommendations in order, but for economy I will paraphrase briefly, trusting you to have read the pamphlet or perhaps to have it in your hand as I proceed.

1. The first recommendation (on page 30) is that each school have a sound policy about testing in the school. Who can disagree with that?

2. The second recommendation is that professional knowledge be used in developing the policy. That is even better. They did leave out

the contribution that might be made by including the professional knowledge of testing experts, but I assume that this was a mere oversight!

3. The school periodically evaluate its testing program. Hurray!

4. Schools should evaluate the national testing programs before recommending their adoption. Right! In fact, they should evaluate anything before recommending its adoption!

5. Scores on a single national test should not be used as the sole measure of the quality of a school's educational program. I would even venture that a single number should rarely be used to evaluate anything about a school that is of any importance.

6. A record should be kept of the man-hours devoted to testing. Now this seems like a silly recommendation to make to any but an incompetent school administrator; it certainly shows a low opinion of their fellows by the authors of the pamphlet. Time and motion study is useful for certain purposes. If a school administrator has a hypothesis about his school which can be illuminated by such study, of course he should use it, whether the topic is basket-weaving, baton twirling, or learning-by-doing. However, if you have followed recommendations 1, 2, 3, and 4, the record keeping mentioned in number 6 is likely to be a useless annoyance.

7. Recommendation 7 is one of the double-barreled ones. It states that external tests for scholarships should be free to the student. This sounds meddlesome to me, unless, of course, the schools are volunteering to pick up the tab. More likely, this is a personal axe being ground and should have been excised from the pamphlet by a judicious editor. At any rate, consider the results of such a policy. First, many more people would probably take the scholarship examinations if they were free. But even now there is much wasted testing because so many people take those examinations who can be told from data already available that they have no chance at the prize scholarships. May I refer you to Dr. Frank Womer's article on this point in the November, 1962, issue of the Personnel and Guidance Journal. The second result, assuming that you ask the sponsor to make the tests free, would be fewer scholarships available. If the schools are meant to pay for the tests, it would mean higher taxes or less school services of some other sort; money has to come from somewhere, and it can only be spent once.

7b. The second part of recommendation 7 is that donors get, and follow, school administrators' advice on finding and aiding worthy students. I feel certain that donors have sought such advice--if not, it was indeed a blunder. For school men to recommend that their advice be followed is, in a way, amusing. Do you suppose that you could ever get all school men to agree on a topic as complicated as this? And if you could get agreement, suppose they agreed on a technique that was wrong! Let's reword 7b to be acceptable, yet probably to contain the worthy kernel of the committee's idea. Donors should seek the best ideas from any source on finding and aiding worthy students. They should try these ideas out, keep the ones that work, discard the others, and strive continuously to improve their procedures.

8. Recommendation 8 is that schools stop coercing or defrauding students into taking scholarship examinations. If three national associations have to tell their colleagues to stop coercing students to take scholarship examinations, education is in a bad way. This recommendation also says that we need effective counseling so that students will realize their probabilities of success in scholarship competition. With that, one must agree.

9. Recommendation 9 is "School teachers, administrators, and other professional personnel become more knowledgeable in the field of measurement and evaluation." I agree emphatically!

10. Here the three associations recommend that they, themselves, inform the public about the potentials of testing programs. I would suggest only that they follow their ninth recommendation, becoming more knowledgeable, before they tackle number 10, informing others.

11. This is another double-barrelled recommendation. First, it is recommended that test scores not be used to compare students, schools, or states. The only fault with this recommendation is that they left out the word "unsound" or "frivolous." If comparisons are sound, careful, and appropriately qualified, then they may be very helpful, and should be made by all means. To the extent that the public deserves to be informed about such comparisons, or to the extent that public knowledge of them aids education, why should the public be kept in the dark? Only three pages earlier in the pamphlet the authors were stressing that the schools belonged to the people!

11b. The second part of number 11 points out that the use of test scores by testers or schoolmen for publicity is unprofessional. It

should embarrass the education profession that this needs to be said about them.

12. Once again we get the double barrel. First, it is recommended that external tests be given outside of class time and at the convenience of the faculty and students. To the extent that this is possible, it does not seem like a bad idea. However, if I were a school administrator, I would think it rather impertinent of three national associations to be recommending details concerning how I ran my school. I would evaluate a test in terms of its value to my program and its cost to that program. If its value was greater than its cost, I would propose to adopt it. In its value I would include consideration of such things as whether it might replace other tests I was using, whether I could use its results, whether my students would be aided by having it available, and so on. In its cost I would consider not only the money but also whether it interfered with anything else in my school's program. The results might often be that, for participation and cooperation, I would want to have a test given during class time. The nature of the test, such as its security, would have to take priority even over my desires. The school administrator, himself, can't always have things the way he wants them just when he wants them.

12b. Schools are told to refuse to participate in tests unless they are worth their cost, which, after what I said, reveals that I'm not entirely out of step, after all. The best comment I can make about this recommendation is perhaps to list several words that could be inserted in place of tests, e.g., athletics, music, baton twirling, contests of all sorts, driver training, fund drives, school lunches, building programs, etc., etc. Refuse to participate in anything unless it is worth the while!

13 and 14. These two should be considered together. They are really one recommendation. They recommend that equivalency tables be developed to permit scores on one test to be translated into scores on other tests, and that these estimated scores be used interchangeably with actual scores for college admissions purposes. It is unfortunate that the committee which wrote Testing, Testing, Testing did not explore whether or not this recommended procedure was a possibility. They can find out how remote the possibility is, by studying Dr. William Angoff's paper presented at the Educational Testing Service's Invitational Conference this year. The problem is that it is not possible to provide a single equivalency table for each pair of tests. Special tables would also be required for each pair of tests for each sex, for

each college, for each purpose for which the tests were to be used, and so on. For precision, myriads of tables would be required. If each of the standardized admission tests measured precisely the same thing, the problem would be simple. However, they don't even have the same number of scores. Five scores are used for admission in Florida; four or five are used in Alabama; and, for the most part, we use two in Georgia. Besides that, many of the tests don't even try to agree in what they measure. There is a rough comparability among them, to be sure, but it is about like the comparability between height and weight, not like the comparability between yards and meters.

One might estimate scores on one test from scores on another through correlation and regression, but it would not be proper to use the estimated scores as though they were actual scores because estimated scores of weak students would be too high and of strong students would be too low; they would be regressed toward the mean. A weak student would get an advantage by taking the undesired test; a potential honors student would be hurt. The problem is exaggerated because high and low scores suffer the most from conversion, and these are the very scores of greatest interest to the admissions officer. I'll not take more time here to discuss this problem. I could do no better than to read Angoff's paper to you. It is indeed disillusioning that the committee responsible for the pamphlet we are discussing did not consult with someone like Angoff before leading you astray with a recommendation which sounds so plausible but whose overwhelming problems could readily have been delineated.

Before I leave this topic I would like to suggest that what may be insurmountable problems for the function of the admissions officer may not be so bad for the purpose of the counselor. The counselor can tolerate more error than the admissions officer. In fact, counseling about academic potential or probability of success in college should probably be looked on as helping the client to obtain a sound but rough idea of his status. The job is to help the brilliant person recognize his talent, to help the slow one to accept his talent and make the most of it; and to help the ones in the middle in the same way. Fairly rough measures of academic potential will serve this purpose. There may be merit in development of equivalency scores of some sort, but for counseling purposes, not for admissions. When it comes to a decision about rejecting a student for admission, or for honors program, he deserves all the accuracy we can reasonably give him.

15. The committee recommends the addition of new areas to national testing programs which seek to identify academic competence.

The committee is recommending even more extensive, time-consuming, costly, testing, testing, testing! Would you have believed it? Unfortunately, their recommendation brings into question their own competence. No professional in the area of evaluation would recommend that areas be added to a test unless they could be shown to contribute to the purposes of the test. It really sounds as though the committee wants to add these areas to influence the school curriculum, but they decry this very tendency only three pages earlier in the pamphlet. If measures of arts, homemaking, health, or even baton-twirling add appreciably to prediction of academic success, they should be added to the tests designed for that purpose. If they do not so contribute, I recommend that you refuse to participate in efforts of the National Association of Secondary School Principals, the Council of Chief State School Officers, and the American Association of School Administrators to have such measures inserted into the tests.

16. The recommendation that scholarships from tax funds include the vocational fields has no relevance to the topic of testing. The editor slipped again.

17. Item 17 recommends that schools inform parents about coaching schools. I would only add that the former schools be sure they know what they are talking about before they inform on the latter schools.

18. The last recommendation is free. The committee recommends that the associations keep their members informed. As with number 10, I suggest only that they first follow recommendation 9.

A Condensed Set of Recommendations

The Joint Committee on Testing of the three national associations spent nearly two years studying the nature, scope, and effects of testing, particularly external testing, in schools throughout the nation. In terms of the analysis which I have just reported to you, their main cogent criticisms were of the schools rather than the tests, and many of their recommendations do not withstand the light of critical scrutiny. What can be salvaged; what should schools and school people do about the problems of testing--external, internal, and eternal? May I suggest that if the following recommendations are followed, most of the problems connected with testing will seem to take care of themselves.

First, I recommend that every school system develop an over-all policy governing all tests administered under the auspices of the school--including the tests the teacher makes for the guidance and evaluation of her instruction. A sub-recommendation under this is that in formulating the policy governing testing, the professional knowledge of all available persons who have sound training in testing as it applies to education be obtained and used. This would include measurement-sophisticated teachers, principals, and superintendents and might well include consultants in the measurement area. A second sub-recommendation would be that each school system periodically evaluate its total testing program, internal and external, of course, to determine its soundness, adequacy, and necessity in all of its parts, and that each school modify its testing program wherever appropriate. A third sub-recommendation would be that available national testing programs be included in the school's total testing program wherever such inclusion is appropriate to the goals of the school, and not otherwise. Unnecessary redundancy of measurement should, of course, be avoided.

A fourth sub-recommendation would be that, wherever possible, coordination between school systems, within states and within larger geographic areas, should be vigorously sought.

My second and last recommendation is that school teachers, administrators, and other professional personnel become more knowledgeable in the field of measurement and evaluation.

If the above recommendations have a familiar ring, you should note that they are essentially recommendations 1, 2, 3, 4, and 9 of the Joint Committee on Testing, with the addition of broader geographic considerations in formulation of a school's testing policy and program. The Joint Committee had the answer all the time; they just buried it in so much fuzz that it had to be ressurected.

* * * * *

The annual meeting of the Florida Educational Research Association will be held at the time of the annual meeting of FEA, April 23-25, 1964, in Miami Beach. Mrs. Karen Addington Strattan, President-Elect of FERA will serve as program chairman. Specific information about program and location will appear in a forthcoming issue of Florida Education.

* * * * *

The Eighth Annual Invitational Conference on Testing, sponsored by the Florida Educational Research Association, will be held at the Holiday Inn, Tallahassee, on January 23, 24, and 25. Mrs. Karen Addington Strattan, President-Elect of FERA, is serving as program chairman and H. A. Curtis, Florida State University, is in charge of local arrangements. Program outlines and registration forms will be sent to all members of FERA.

* * * * *

Deadlines for receipt of Cooperative Research Program research proposals are April 1, September 1, December 1, and March 1. The Cooperative Research Program is administered by the U. S. Department of Health, Education, and Welfare and it is operated under terms of Public Law 531, 83rd Congress, which authorizes the Commission on Education to "enter into contracts or jointly financed cooperative arrangements with universities and colleges and state educational agencies for the conduct of research, surveys, and demonstrations in the field of education." Six kinds of activities are funded: (1) Basic and Applied Research; (2) Demonstration; (3) Curriculum Improvement; (4) Small Contract; (5) Research and Development Centers; and (6) Developmental Activities. Applications and instructions are available from Director, Cooperative Research Program, U. S. Office of Education. Public school personnel who plan research of the kind supported by the Cooperative Research program should contact Dr. Bob Gates, Coordinator, Florida State Department of Education, for information regarding conditions under which public school personnel might participate in the Program. For complete information about the Cooperative Research Program, request from the Commissioner of Education, U. S. Department of Health, Education, and Welfare, Washington, D. C., a copy of OE-12017: Cooperative Research Program, Application Instructions for Research Contracts.

* * * * *

Various titles of the National Defense Education Act contain provisions for funding educational research. Interested researchers might write directly to Director, National Defense Education Act, U. S. Office of Education, or to Dr. Bob Gates, Coordinator, Florida State Department of Education.

* * * * *

The Department of Educational Research and Testing of the Florida State University has graduate research assistantships available at the master's and doctoral levels for the 1964 Summer Session and for the 1964-1965 school year. Address inquiries and requests for application forms to Dr. H. A. Curtis, head of department.

* * * * *

Dr. H. H. McAshan, Director of Research and Projects, Brevard County, has written a book on research methodology which is entitled Elements of Educational Research, copyrighted in 1963, and distributed in soft back by the McGraw-Hill Book Company.

* * * * *

A CURRICULUM STUDY CENTER TO DEVELOP AND TEST
APPROACHES TO SEQUENTIAL CURRICULUM IN
JUNIOR HIGH SCHOOL ENGLISH

A Curriculum Center in Junior High School English, directed by Dr. Dwight L. Burton, has been established at the Florida State University under the auspices of Project English, U. S. Office of Education. The associate directors are Dr. Kellogg W. Hunt, Professor of English at Florida State, and Mr. Paul Jacobs, Consultant in Language Arts, Florida State Department of Education. Running from July, 1963, to December 31, 1967, the Center aims to develop three distinct sequential English programs for grades seven, eight, and nine; to try out these programs in six junior high schools in Florida and to evaluate the programs as fully as possible; and to develop appropriate syllabi, teaching materials, and tests. The six participating schools with their county supervisors include Dade County--Mrs. Elizabeth S. White, Supervisor, Ponce De Leon Junior High School and Brownsville Junior High School; Duval County--Miss Mabel Talmage, Supervisor, Joseph Stillwell Junior High School; Escambia County--Mrs. Ruth Duncan, Supervisor, Warrington Junior High School; Pinellas County--Dr. Lois Arnold, Supervisor, Madeira Beach Junior High School and Tyrone Junior High School.

Each of the experimental curricula will involve the study of literature, language structure and certain other linguistic matters, and written composition. Advising the Center in these matters are four

consultants: in composition, Dr. Albert R. Kitzhaber, University of Oregon; in literature, Dr. Louise M. Rosenblatt, New York University; in linguistics, Dr. Thomas H. Wetmore, Ball State Teachers College; and in general matters, Dr. Francis G. Townsend, Florida State University. Dr. John S. Simmons and Mrs. Helen O'Hara Rosenblum of Florida State University are research associates, and other personnel include Mr. William T. Ojala of Florida State, Mrs. Judith Dodd, secretary of the Center, and several graduate assistants.

The first curriculum is arranged sequentially by facets of subject matter. Thus, language, literature, and composition are planned as separate units of work in each of the three grades. In the seventh grade, for example, language is planned in three units, "Power Over Language," "Form Classes," and "Lexicography." The first of these units is a broad introductory unit on the English language dealing with the functions of language, purposes and abuses of language, operation of the language, and some characteristics of effective and ineffective use of language. The second unit deals with the characteristics and inflections of the four great form classes of words. The third seventh grade language unit discusses various aspects of the dictionary. The eighth grade work on language has two major concerns. The first is with varieties of language and the second with basic syntax and morphology. The ninth year work on language deals with basic phonology, dialects, and basic semantics. Literature and composition are similarly dealt with as subject-matter blocks.

In the second curriculum a theme, concept, or abstract idea is placed in a position of central importance. The theme is first identified and described for the students. It may then be discussed. From the outset, then, all activities, materials and devices for evaluation of student contributions are developed in light of the theme and the proposition it makes concerning human experience. Within the three years there are six major or encompassing themes. For each of these, three sub-themes have been indicated, one for each succeeding grade level, with each sub-theme more complex and subtle than the one previously selected. The accompanying activities and materials will also reflect this growth of complexity and subtlety.

Based on Jerome Bruner's spiral curriculum in The Process of Education, the third experimental program integrates skills of language, literature, and composition first with an emphasis on the process involved in the cluster of skills and later with an emphasis on particular aspects of the discipline of English. In the seventh grade,

for example, there are four phases in the curriculum. The first phase emphasizes structure and small units of meaning in language, literature, and composition. Language and composition activities are complementary here. Literature structure concerns discussion of plot, setting, characterization as well as anticipating the study of Northrop Frye's modes of fiction by introducing common fictional devices such as mistaken identity as part of either romantic adventure or comedy and "excess" as part of comedy or tragedy. Activities, too, follow Bruner's suggestions. Pupils are asked to complete stories which follow the pattern of comedy or romantic adventure or mystery.

The second phase in the seventh grade spiral emphasizes small units of meaning but continues the relationship with structure. The importance of context in the English language is a major area, here, as well as beginning work with symbols. Language, literature, and composition skills are grouped around the study of "meaning." The third phase of the year's program emphasizes organization or structure on a more complex level. The fourth phase emphasizes are on enlarged areas of meaning. The eighth grade progresses in a similar manner with increasingly more complex activities and integration of skills until in the ninth grade the focus of instruction has shifted from English as a "skill subject" to English as a "content subject." The earlier work, concerned as it is with "approaches" to the subject of English, anticipates later effecting these approaches in the more complex phases of the discipline.

The Curriculum Center is actually an experiment comparing groups of students who have gone through three years of study under one of three different English programs. The three experimental groups will be compared with each other and with the control group composed of those students who have been in the school three years but have not been in an experimental program.

The classes for the three programs are to be randomly selected and thus assumed to be heterogeneous in population. In the spring of 1967, a battery of tests will be given to all ninth graders in the six schools. The test data will be analyzed by computer. In addition, all teachers in experimental programs will be asked for subjective evaluations of the program. Objective tests will be given in writing, linguistics, and the reading of literature. In the analysis of test results, students who have not spent three years in the particular school will be eliminated. Dr. Howard Stoker of Florida State University, who heads the state-wide ninth-grade testing program, is advisor on research design for the Center.

The end products of the four-year undertaking include syllabi and teachers' guides for each type of curriculum at each grade level, with some specially-prepared teaching materials; a set of tests in alternate forms; a report on the statistical analysis of test scores, theme ratings, and on the evaluations by teachers and others involved; and, finally, a report on the entire project, with recommendations growing out of it. The pressing need for the development of a sequential and cumulative curriculum has been recognized throughout the English teaching profession. Dozens of articles and reports have cited the disjointed and often chaotic nature of the English curriculum. It is also true that a great deal of curriculum work has been done by local, state, and national groups, and many bulletins and curriculum guides have been published recently. Further, each new series of textbooks proposes some sort of sequence. Still, the lack of a research basis is apparent. The Curriculum Center in English at Florida State University hopes to aid in providing such a basis.

Dwight L. Burton
Florida State University

* * * * *

What PROJECT ENGLISH Grant #1998 is About

The Title: DIFFERENCES IN GRAMMATICAL STRUCTURES WRITTEN AT THREE GRADE LEVELS, THE STRUCTURES TO BE ANALYZED BY TRANSFORMATIONAL METHODS

Director: Kellogg W. Hunt, Professor of English.
Investigator: Dwight L. Burton, Professor and Head,
English Education.
Florida State University, Tallahassee, Florida.

The Problem

The teacher of writing knows a good deal about a student's vocabulary growth. Indeed, at any given time he can measure rather precisely whether a student's vocabulary achievement is ahead of the average schedule or behind it. But the teacher knows very little about a student's growth in syntax, even though writing involves syntax as well as vocabulary. Choosing the right words is not enough. They have to be put in the right order, and the right order is a matter of

sentence structure, of syntax. Any writing teacher senses this, but he knows very little about the order of steps through which a student's written structures develop.

Even without that knowledge, sentence structure is now taught, and grammatical concepts are carefully arranged into what is supposed to be a sequential program. The teacher feels some vague conviction that a student's command of sentence structure must develop throughout his school career. But the existing studies give only the most fragmentary evidence of what changes occur after the student has once learned to speak and understand his native tongue.

The hundreds of studies of language development tell much more about growth in speech than in writing, and about vocabulary than syntax.

The fact that a useful picture of syntactic growth in writing has failed to emerge from hundreds of studies does not prove that no such growth occurs. It proves only that little has been perceived by the instruments used for that perception.

A new instrument might or might not lead to new perceptions.

Analysis Using Transformational Terms

The new method to be used for perception in this study is transformational analysis. Such an article as Lees' "A Multiple Ambiguous Adjectival Construction in English." (Language, 36, 2, 207-221) shows convincingly that transformational analysis is a more incisive method for describing syntactic structures than is either the parsing done in traditional grammar, or the immediate constituent analysis done in structural syntax.

This project, then, will use transformational analysis to describe each sentence produced by the students in this study.

The students will be in grades four, eight, and twelve. The syntactic structures used at each grade level will be tabulated and then studied statistically to see whether at these three grade levels covering a span of eight years some structures emerge later than others.

It may be possible also, using the concept of "degrees of grammaticality" developed by the transformationalists, to specify the

structures that are produced inaccurately at each grade level. If this can be done, then a second criterion of growth will be available. It would then be possible to say: "Such-and-such a structure is not attempted by fourth graders. Such-and-such a structure when undertaken by fourth graders is used unsuccessfully this often."

What Writings Will be Used

The writings will be those produced by "average" students in the fourth, eighth, and twelfth grades. "Average" will here mean with I. Q. between 90 and 110. Equal numbers of boys and girls will be used.

Each child will produce 1,000 words, probably during several class periods. An attempt will be made to get writing from each student that is "a normal sample" for him. At least one writing sample will be produced in response to a stimulus common to all students: all will watch a film that should interest all three grade levels, and then all will write about it.

Kellogg W. Hunt
Florida State University