DOCUMENT RESUME

ED 045 294 RE 003 069

AUTHOR Maxwell, Martha J.
TITLE Evaluation of College Reading and Study Skills Programs.
PUB DATE May 70
NOTE 13p.; Paper presented at the conference of the International Reading Association, Anaheim, Cal., May 6-9, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$0.75
DESCRIPTORS Academic Achievement, *College Programs, College Students, Educational Objectives, *Program Evaluation, *Reading Programs, Standardized Tests, Student Attitudes

ABSTRACT

A number of evaluation techniques for college reading programs are described and discussed. The techniques indicated include (1) determining a clear definition of objectives and specific criterion tasks that are consistent with program objectives, (2) using standardized tests for describing group change, (3) analyzing academic achievement as shown in course grades and grade-point averages, and (4) assessing students' needs and attitudes. It was pointed out that a variety of techniques are needed to diagnose each individual student's needs and evaluate his progress. References are included. (DE)

EVALUATION OF COLLEGE READING AND STUDY SKILLS PROGRAMS

By

Martha J. Maxwell
Director, Reading and Study Skills Service
University of California, Berkeley

The evaluation of a college reading program cannot be separated from the goals, objectives and practices of the program. Evaluation, of necessity, must differ from research studies. Rarely can the typical research paradigms be used nor can the rigid assumptions necessary for rigorous statistical tests be met. Students in a college reading and study skills program are either selected for special help because of their deficiencies or seek help voluntarily. To set up randomly selected groups including non-treatment controls or even to randomize treatments is usually impossible since the goal of the reading program is to offer service to students, not to experiment. Heterogeneous grouping is usually either impossible or undesirable.

Evaluation seeks to answer questions like "How effective is our service in meeting our objectives?" "In what ways are students improving their skills?" as well as the negative questions such as "What students fail to benefit from our program?" and "Are some students harmed through their experiences in the program?"

Evaluation is essential in making decisions as to how the service might be improved, in planning and selecting materials and instructors, in determining whether a service should be expanded or contracted, and in justifying your existence to budget committees.

The first step in evaluation is to determine your goals and specific objectives based, of course, on assessment of student and faculty expressed needs. If objectives can be clearly stated in behavioral terms, then the job of evaluating the program is easier.

Data need to be collected and accurate records kept. Evaluation must be a continuous process beginning when the program is in the planning stages. In order that accurate and adequate data be collected, decisions as to what information is to be kept must be made at the start of the program, not after it is finished. Some kinds of data may be essential for annual reports of the service and budget committees. The usual demographic information collected about students who participate in the program include number of students by sex, year in college, curriculum, prior grade point average (if any), use made of service, type of problem, scholastic aptitude test scores, etc. Knowledge of these background factors is useful for assessment and planning.

Limitations of Standardized Tests in Evaluation

Although standardized tests are used in most programs for screening students in reading and study skills work, they have limited usefulness in evaluating the program. First of all, most standarized tests do not measure the specific goals of most programs and are only tangentially related to the activities that students perform in the program. Traditionally, college reading tests measure reading rate (often on a very limited scale, such as the one-minute timed sample on the Nelson-Denny, a most unrealistic time-sample of the reading of the typical college student), vocabulary and comprehension (usually based on understanding paragraphs.) If the objective of a reading and study skills course is to teach students effective techniques for getting the main ideas and significant details from a textbook chapter, scores on a standardized post-test are not likely to reveal much about their competence in mastering this skill. Nor does the typical test show whether a student is reading flexibly for different purposes, can skim and scan, nor whether he can retain the major concepts from studying a chapter for several days until he has time to review it again.

Standardized tests in reading are useful as predictors of college success, as screening instruments for picking students with exceptionally high or low scores, and for describing groups in general. What the standardized test does not do is reveal specific information about a given student's difficulties, nor does it necessarily reflect the progress he has made in pursuing a reading course.

Also standardized tests, as they are often used, mask individual differences particularly when means are used to describe the results of a course. They do not reveal whether a student whose scores have remained unchanged despite one's best efforts has actually been harmed by taking our program -- for example, we all know students who have inappropriately taken a commercial speed reading course, but because they lack basic vocabulary and comprehension skills and failed to increase their speed were left feeling even more inadequate about their reading skills.

Another problem in using standardized tests to measure changes in reading as a result of a program concerns working with deficient readers. For example, if the students are low in reading skills and sectioned into a reading workshop one would expect that their scores would improve through chance alone (regression toward the mean.) Even studies which have attempted to use control groups have their limitations. Usually in this case, a group of students who are not given the reading program is matched on the basis of sex, college year, curriculum and reading ability with those who take advantage of the reading course. If the reading program is a voluntary one, students who enter it may be more highly motivated than those who have not sought help even though they need it equally as much.

Most tests are confounded by the rate factor (Stroud, 1958). The Nelson-Denny is a good example, since the slow reader is typically unable to complete many of the vocabulary or comprehension items, therefore his scores are low. Showing him how to increase his speed in taking tests may result in improved scores.

Another problem cited by Davis (1961) is that of guessing. This can spuriously raise test scores particularly when administered to low-achieving students who may randomly mark answers even though they have not read the items.

Alternate forms of standardized tests may not be exactly parallel. For example, if raw scores are used in the computation of change on the Nelson-Denny and Form A is administered first, we find that a freshman student reading at 250 words per minute would score at the 50th percentile on the manual norms whereas a student scoring at 250 on Form B would be at the 60th percentile. At the upper levels students scoring at 376 words per minute on Form A would be at the 90th percentile, whereas a student scoring at only 356 words per minute would be at the 90th percentile on Form B. Obviously there are differences throughout the test norms that would affect the results of a pre-post-test comparison. This problem can be handled by converting the raw scores into standard scores so that the student's rank within the group becomes the measure. However, since this involves some statistical operations and reading people are typically averse to computing, this is rarely used. As a result, many of the conclusions reported in the literature that are based on raw score data are spurious. Tracy and Rankin (1967) describe a residual gain statistic based on either of two computational methods -- one derived from a Z-score formula for equating pre- and post-test results and another formula using raw scores. These are attempts to statistically equate the pre-and post-test scores of an individual. The authors stress that it is necessary to compute and graph each individual reading class or group.

Another weakness in using standardized tests is the fact that unless one has developed local norms on his own institution, using manual norms may be deceptive. For example, at the University of Maryland we found that entering freshmen at our school averaged scores comparable to the college senior norms

listed in the manual. Although there is some value in comparing your group with national groups for prestige and status, it is more important to know how an individual student ranks or a class ranks in regard to the specific institution in which they are enrolled. In other words, it is important to know where the student stands in relation to the competition in his own college on reading skills. In summary, standardized tests have limited usefulness in the assessment of a specific college reading program for the following reasons: 1) they rarely measure the objectives of the program that is being taught, 2) alternate forms may give spurious results unless some Z-transformation or standard score is computed to equate the two, 3) if the reading program involves students who are weak in reading skills then regression toward the mean effects will undoubtedly occur and mask any real changes, 4) since standardized tests are by definition both reliable and valid, they are not subject to change readily as a result of a brief instructional program, 5) the use of mean score gains masks the variable that occurs in growth in the typical reading class.

Academic Achievement as a Criterion

Since effective reading and study skills are related to college success, it has been generally accepted as a foregone conclusion that if you provide a program that offers students who are deficient in these skills the opportunity to learn more effective techniques, their college grades will improve. However, in recent years it has been the rare reading program that systematically assesses and reports grade point average improvement. For example, in a recent survey of 17 compensatory education programs in the California community colleges, only one program described the academic success of students in the program with a control group who had not had special reading and study skills help. (Berg and Axtel, 1969)

Ten years ago reading researchers were much more adamant about the necessity of using grades as a criterion. Entwisle and Entwisle (1960) stated that improvement in overall scholastic average as compared with a control group was the only adequate criterion measure of improvement following a college study skills course. In summarizing 14 studies, they concluded that the modal gain in overall G.P.A. was between .4 and .5 of a grade point and further noted that improvement is "almost always maintained when follow-up studies are done."

Wright (1962) in reviewing 31 studies which purported to measure the relationship between reading training and college success, found only 11 with comparable control groups and only 7 of these reported significant improvement in grades for the students taking reading improvement programs. He concluded that the differential results could be attributed to other variables such as the curriculum studied by the student, personality differences between students, nature of the training program, length of course and competence of the instruction.

Wright further describes his study in which students were randomly assigned to control or experimental groups and both groups were retested on reading ability at the end of the academic year. Two grade point averages were computed for each student in the study: one based on grades in English, social studies and humanities courses (Verbal G.P.A.) and one on science and mathematics courses (Quantitative G.P.A.). The experimental subjects who completed the reading course not only showed significantly higher scores on all the Nelson-Denny sub-tests but also had significantly higher verbal G.P.A.'s at the end of the year then the controls did. However, there were no differences between those taking the program and controls on Quantitative G.P.A.

Although the majority of studies reporting effects of reading and study skills programs on improvement in grades show favorable results, there remains the question of the representativeness of the reported studies since editors un-

doubtedly view studies with positive results as more desirable for publication than those with negative results. Furthermore, as Wright (1962) has demonstrated the use of overall G.P.A. may not be realistic in assessing reading and study skills programs, since the majority of college programs stress English and social studies reading and place minimal emphasis on skills in science and mathematics.

A recent study by King, Dolland and Walter (1969) reports that students attending the University of Missouri Reading Improvement Program over a six-year period did not show significantly higher post-grade-point-averages than a control group. However, through analyzing their data by grouping students according to initial reading rates, they found that only the middle group (those whose initial reading speeds were between 200 and 250 words per minute) showed significant improvement in grade point averages. (Students reading slower than 200 words per minute or above 250 words per minute initially did not show grade improvement.) They concluded that the students reading in the 200 to 250 words per minute group initially were at a level where increased reading rate would make a significant difference in their studying while those reading more slowly initially probably had attitudinal problems, were perfectionistic or compulsive readers and hence harder to change. Those reading above 250 words per minute initially, they feel, were probably already reading well enough to keep pace with their college assignments.

This study also illustrates the complexity of the relationships between improvement in reading skills and grades and points up the needs for carefully thought out and well designed studies.

In conclusion, it is important for the college reading administrator to collect data on students' pre- and post-grades, but it is equally important that these be viewed in terms of the specific objectives of the reading program. It

would be folly to expect that a program geared to increasing reading rate on literary materials would transfer to problem-solving in an advanced mathematics course. Similarly, if English course grades were based primarily on the student's ability to write critical essays, then gains in speed reading might not be reflected in improvement in English grades. Ability to complete all the reading assignments might be a more relevant criterion in the latter case.

Assessing voluntary self-help reading and study skills program poses additional problems. If students seek help from the service late in the semester after failing an exam or two, the probability of improving their grades in the course is statistically very remote. Also voluntary programs typically attract a more heterogeneous group of students including some with honors grades as well as those with low achievement.

Straight-A upper division students may be attracted to the program in hopes of maintaining their averages with less effort, while less capable students may need intensive work in basic skills. Attempting to combine such divergent students into one group and examining mean pre-and post- Grade Point Averages would have little meaning. However, examining how long high, average and low-achieving students remain in the program and what they accomplish does have value in developing insights into the characteristics of students who profit from the program and in planning ways it could be improved. For example, at the University of Maryland we found that students with high reading score profiles tended to remain in a voluntary reading program longer than those with average or low profiles. (Maxwell, 1965) This finding prompted us to reexamine our program procedures to determine how we could better help the students with poor skills cope with college demands. Although these students need the service more than the others, they are also handicapped in finding the time to devote to skills improvement when heavy course demands take all of their time.

Grade point averages on students are useful to have but probably individual course grades, particularly those which have relevance to the objectives of the reading and study skills program or the specific area on which a student has worked seem more important to collect and analyze.

Certainly more carefully controlled studies on the effectiveness of college reading and study skills programs for the disadvantaged need to be made. These kinds of studies do raise ethical and political questions, for they require that an equally deficient and equally motivated group of students be deprived of the "benefits" of the program and serve as a control group. These questions mitigate against using traditional experimental methodology and force us to look for meaningful but less direct and different ways of evaluating our programs. However, if clear behavioral objectives are stated at the beginning of the program, data can be systematically collected on the percentage of the group that achieves the criterion by the end of the program. (Such objectives might include the ability to read and answer a general discussion question about a chapter in history in 30 minutes, or to skim an essay to determine the author's main premises for his argument in 3 minutes, etc.) To the extent that these tasks represent "job samples" of the assignments the students are expected to do in courses, then one might legitimately expect that students who learned these skills would attain higher grades in the specific course. If it is determined that performance on these tasks is not related to specific grades, then the reading director should try to determine whether the skills have not been adequately learned or whether they are inappropriate or irrelevant for the course in question.

Student's Attitudes

Ironside (1969) stresses the need for students to be involved in assessing both the goals and their progress in a reading program. He also mentions the over-use of single factor tests in evaluating a program that involves many skills and recommends a grid of 21 skills which can be used to set goals and can also be used to develop criterion tasks for assessing the most common reading skills that could be taught in a course. He stresses that feedback to the student is essential so that objectives can be matched to the instructional program and progress assessed.

Wood (1961) proposed attrition as a criterion for evaluating non-credit reading programs assuming that if students in a voluntary program persisted then this would suggest that they were gaining something from it.

Knafle (1965) studied personality characteristics of students enrolled in a reading and study skills program and instructors' ratings and found that for poor readers, instructors apparently used different criteria to assess improvement. Students with higher scores in dominance who were poor readers were more likely to be assessed by instructors as making greater improvement in reading than students with better reading skills who were equally dominant. On the other hand, students with higher reading scores who had high scores on achievement via independence on the California Psychological Inventory were assessed as making greater improvement than those with low scores on this dimension. Thus there seemed to be interraction effects in terms of teacher expectations and personality patterns of students persisting in a reading program.

Post-questionnaires assessing students' attitudes toward the program are frequently used and can provide valuable information about student reactions and also serve to give the students an opportunity to express their feelings

about a program. Student evaluations, provided they are anonymous, will yield valuable insights. Such questionnaires are subject to the halo effect and should be anonymous to get maximum information.

In summary then, there are a number of techniques that can be used to assess college reading programs. Most important is to clearly define your objectives and set specific criterion tasks that are consistent with the objectives of the program. Standardized tests can be used to describe group changes, but have their limitations if the program's viability is to hinge on the performance of students at the end of a program. Certainly information about grades and grade point averages should be collected since in essence most of our programs do aim to help students improve in their academic work. If the program is to be strengthened then it should be built on the students' needs and without objective knowledge about the kinds of students who do succeed or fail in the program, it is difficult to do long-range planning. If a reading and study skills program is restricted to low-ability or low-achieving students, then the problem of stigma being associated with the service may be a real one. This may affect the students' progress in the course and their attitudes toward the reading specialists who run it. My personal conviction is that college reading programs should meet the needs of all students who want help and this implies using a variety of techniques diagnosing each individual student's needs, and evaluating his progress in these particular skills.

BIBLIOGRAPHY

1. Berg, Ernest H. and Axtell, Dayton. Programs for Disadvantaged Students in the California Community Colleges. Peralta Junior College District, 1968.

2. Davis, Frederick B. "The Assessment of Change," Phases of College and Other Adult Reading Programs: Tenth Yearbook of the National Reading Conference, Emery P. Bliesmer and Albert J. Kingston, Eds. (Milwaukee, Wisconsin: The National Reading Conference, 1961), pp. 86-95.

3 Entwisle, George and Doris R. Entwisle, "Study-skills Courses in Medical Schools." Journal of Medical Education, Vol. 35, No 9, September 1960.

4. Ironside, Roderick A. "Who Assesses Reading Progress?" in Clay A. Ketchum (Ed) Proceedings of the College Reading Association, Vol. 10. p. 64-69. Easton, Pa. 1969.

5. King, Paul T., William D. Dellande and Terry L. Walter, "The Prediction of Change in Grade Point Average from Initial Reading Rates," Journal of Reading. Vol. 13, No. 3. Dec. 1969. pp. 215-218, 245-246.

6. Knafle, June D. "An Investigation of the Relationship Between Personality Characteristics and Social Adjustment and Three Criteria of Reading Effectiveness for a Group of Low Achieving Prospective College Freshmen Enrolled in a Reading Program," Master's Thesis University of Maryland, College Park, Maryland, 1962.

7. Maxwell, Martha J. "Persistence in the Reading and Study Skills Laboratory as a Function of Initial Reading and Study Skills Test Profiles," Reading and Study Skills Brief Reports #1, Counseling Center, University of Maryland, 1965.

8. Stroud, J. B. "Background of Measurement in Reading Improvement," in Starting and Improving College Reading Programs: Eighth Yearbook of the National Reading Conference, Oscar S. Causey and William Eller, eds. (Fort Worth, Texas: Texas Christian University Press, 1958), pp. 77-88.

9. Tracy, Robert J. and Earl F. Frankin, Jr. "Methods of Computing and Evaluating Residual Gain Scores in the Reading Program," Journal of Reading, Vol. 10, No. 6. (March 1967), pp. 363-71.

10. Wood, Roger L. "Attrition As a Criterion for Evaluating Non-Credit College Reading Programs," Journal of Developmental Reading. Vol. 5, No. 1. 1961. pp. 27-35.

11. Wright, Eugene S. "The Effect of Reading Training on College Achievement," in Alton L. Raygor, Ed. College and Adult Reading, First Annual Yearbook of the North Central Reading Association, University of Minnesota, 1962, pp. 32-50.

Reference

Farr, Roger (Editor), Measurement and Evaluation of Reading, Harcourt, Brace & World, Inc., New York, 1970.