DOCUMENT RESUME

ED 273 649 TM 860 497

AUTHOR	Dowell, David A.
TITLE	Evaluation of an Out-of-School Tutoring Program: CROSSROADS.
PUB DATE	Jun 86
NOTE	32p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; *After School Education; *Basic Skills; Criterion Referenced Tests; Educational Assessment; Educational Policy; *Evaluation Problems; Junior High Schools; Mastery Tests; Parent Participation; Pretests Posttests; Reliability; *Skill Development; Standardized Tests; *Tutorial Programs; *Tutoring; Validity
IDENTIFIERS	*Crossroads Program

ABSTRACT

This paper deals with the evaluation of an out-of-school tutorial program, CROSSROADS, which was funded under the U.S. Department of Education's Basic Skills Improvement Program. Like many projects funded under the Basic Skills program, CROSSROADS was terminated after two years even though it had been funded with the expectation of four years for program development and demonstration. Subjects of the evaluation were 166 tutees from four junior high schools in Long Beach, California. Progress rates of the tutees were measured by pre- and posttests, standardized tests, school grades and criterion referenced measures. Evaluation results suggested that tutees learned some basic skills but not enough to impact upon standardized test scores or grades. Results also suggested that out-of-school basic skills programs will have a difficult time in producing substantial gains in basic skills in lower achieving children due to problems with tutee motivation and logistics of running out-of-school programs. Although the evaluation did not substantiate strongly the effectiveness of the CROSSROADS model for basic skill improvement, findings and issues are pertinent to efforts to improve basic skills and to educational policies and evaluation practice. (Author/JAZ)

EVALUATION OF AN OUT-OF SCHOOL TUTORING PROGRAM: CROSSROADS

June, 1986

David A. Dowell, Ph.D.

California State University

Department of Psychology

Long Beach, CA 90840

Running Head: Evaluation of Out-of-School Tutoring

Abstract

This paper deals with the evaluation of an out-of-school tutorial program. CROSSROADS, which was funded under the U.S. Department of Education's Basic Skills Improvement Program. Like many projects funded under the Basic Skills program. CROSSROADS was terminated after two years even though it had been funded with the expectation of four years for program development and demonstration. Evaluation results suggested that tutees learned some basic skills but not enough to impact upon standardized test scores or grades. Results also suggested that out-of-school basic skills programs will have a difficult time in producing substantial gains in basic skills in lower achieving children due to problems with tutee motivation and logistics of running out-of-schools programs. Although the evaluation did not substantiate strongly the effectiveness of the CROSSROADS model for basic skill improvement, findings and issues are pertinent to efforts to improve basic skills and to educational policies and evaluation practice.

In 1980, the U.S. Department of Education launched the Basic Skills Improvement Program. About one hundred and fifty demonstration projects were funded nationwide at a time of national concern about falling achievement test scores. The Department of Education encouraged a broad range of approaches to basic skills improvement in three general categories: in-school projects, out-of-school projects, and parent involvement projects. The Basic Skills Improvement Program was terminated after two years even though many projects had been funded with the expectation of four years of funding for program development and demonstration. This paper deals with the evaluation of an out-of-school tutorial program, CROSSROADS. Although the evaluation did not strongly substantiate the effectiveness of the CROSSROADS model for basic skill improvement, findings and issues are pertinent to efforts to improve basic skills and to educational policies and educational evaluation practice.

That tutoring <u>can</u> be effective seems to be well-established based upon the accumulation of evaluative literature (Fitz-Gibbons, 1979). This extends to the conclusion that tutoring can be more effective than ordinary classroom instruction (Bausell, Moody, and Walzl, 1972). Part of the importance of tutoring lies in the tutor-tutee relationship (Fitz-Gibbons, 1979). If a lower achieving tutee works with a higher achieving tutor, the tutee may identify with the tutor, raising the child's academic self-concept, educational confidence and aspirations. This process might be especially important for families lacking a role model of academic success.

CROSSROADS

Given this knowledge, one reasonable approach to basic skills improvement in 1980 was to develop a supplemental tutoring program for children at the lowest end of the basic skills achievement spectrum. CROSSROADS had a fairly complex impact model, or plan for how program activities would achieve program objectives and goals. This impact model had five general components including tutee variables, environmental conditions, family variables, program services, and proximal and distal outcomes. Each of the program services was based upon one or more aspect of the impact model.

PROGRAM SERVICES

TUTORING. The tutoring model was an individualized, highly structured approach using paid college students as tutors. The program was based upon materials in math and reading developed by the Southwest Regional Educational Laboratories and the SCOPE writing program (Gardner, 1979). Materials were criterion referenced with frequent mastery test. Tutoring took place for one to two hours per week, off-campus at libraries, and recreation centers.

Tutors were recruited primarily from university departments of psychology and education. Tutors received several hours of training from the services coordinator in tutoring and in record keeping. Tutors were supervised for one hour each week in small groups. Tutors was closely monitored to insure that they met with the tutees as scheduled, followed up on tutee absences and kept careful records of tutee achievement. Most of the time of

the full-time services coordinator was spent in monitoring tutor performance. Data on contact hours are given below.

PARENT EDUCATION. Monthly workshops for parents focused upon educationally relevant topics such as careers, how to help children study, and college entrance requirements. Not all parents attended even one of the workshops although some attended regularly. In addition, a newsletter was published quarterly and disseminated to parents, teachers, counselors, and other school personnel about the program. The newsletter included material similar to the parent workshops.

Family involvement is a significant variable in academic attainment (Bear, Hess and Shipman, 1966; Dowling, 1982; Filipczak, Lordeman and Friedman, 1977; Hess, 1967; Scout and Borders, 1979). This construct includes family attitudes toward education, parental or sibling skills at encouraging achievement, and family behaviors which support (or detract from) achievement. Family involvement was expected to be influenced by family acculturation. The program included children from a large proportion of first and second generation immigrant families reside in the Long Beach area, principally from Mexico and from Southeast Asia. Families presented a great variety of attitudes and values toward education.

OUTCOMES. The program services described above were intended to produce both proximal and distal changes in tutees and families which would be conducive to educational achievement. Identification with a higher achieving role model, the tutor, was hoped to impact achievement motivation. Success experiences for the tutee on some of the mastery tests was also hoped to

contribute to increased academic self-concept. It was hoped that family involvement could be increased directly through the family workshops as well as indirectly by the increased enthusiasm on the part of the children.

The program services, interacting with the client/family impacts, were hoped to produce achievement gains in the tutees which would be measurable in several ways. Four achievement measures were used: (1) progress through the tutorial materials, (2) pre-to-post test gains on criterion referenced tests, (3) school-administered standardized achievement tests, and (4) school grades.

EVALUATION QUESTIONS

CROSSROADS evaluation efforts had several purposes including evaluating attainment of process and product objectives. Process objectives were generally met in a very timely and complete manner. Educational materials which resulted from CROSSROADS included a tutoring manual for out-of-school programs. However, attainment of process and product objectives are necessary but insufficient conditions for the success of remedial educational interventions. The key evaluation question is whether educational achievement gains are measured. This report focuses on the the basic skill gains of the tutees.

METHODS

SUBJECTS

Subjects were 166 tutees of CROSSROADS from four junior high schools in Long Beach, California. Schools chosen for CROSSROADS services were four in the city whose students had the lowest standardized test scores. They served "transition" areas whose demographics were characterized by high family mobility (due to a Navy installation), large proportions of first and second generation families from Mexico and Southeast Asia, and a core-city area with characteristics such as deteriorating housing quality and low family incomes. To be eligible for CROSSROADS, children had to have a "C" or less average or score below the 50th percentile in at least one area on the most recent school-administered standardized achievement test.

In response to a drop-out problem, the staff adopted early in the second program year, a criterion of eight (8) tutoring contacts as requisite before a tutee would be considered a "program tutee." Only "program tutees" are considered for the key evaluative analyses. A larger group of approximately 474 children participated to a lesser extent but moved from Long Beach, dropped out of CROSSROADS or were terminated before qualifying under the 8-session criterion. In the second program year, the staff made extra efforts to prevent dropping-out. This was achieved by requiring greater commitment from children and parents before tutoring began. Parents were required to return consent forms and to telephone CROSSROADS. These efforts succeeded at reducing dropping-out to a low level. However, program tutees were clearly a select group of children.

MEASURES

PROGRESS. The tutorial materials were very structured with frequent mastery tests embedded within them. Each level of the materials contained 12 to 20 units or steps, each with a mastery assessment (including parallel forms for multiple tries). Each mastery test covered only one fundamental skill (e.g., blending consonants) and included from six to thirty items which assessed the skill. There was a total of 12 levels in the three areas, reading, writing, and math.

The program used an 80% pass criterion for these mastery tests; tutees who failed to meet this criterion returned for more drill on the specific material and later took a parallel form of the mastery test. This repeated up to three times until the 80% criterion was met by each tutee. This criterion was chosen based upon the following logic. Mastery tests assessed only one fundamental skill. Correctly responding to the substantial majority of the items should reflect understanding of the essential skill and rule out the possibility of chance attainment of the criterion. One hundred percent mastery might be cost-inefficient since children may make errors due to carelessness even when they understand the principle and it would not be desirable to repeat the unit with such children.

Because of these "hurdles" which each tutee had to pass in order to progress through the materials, progress rates were an indication of how much and how rapidly tutees were acquiring basic skills. Progress rate was regarded as the achievement measure most similar to the content of the tutorial materials.

and hence, the most criterion-referenced.

PRE-POST TESTS. Upon entering the program, tutees were given a battery of placement tests keyed to the tutorial materials. In each subject area, tutees were placed for tutoring at the lowest level at which they failed to show mastery, again using an 80% criterion. Upon completing a level of tutorial materials, tutees were tested using a post-test which was a parallel form of the assessment test, also keyed to the tutorial materials. The pre-post gains were also regarded as a criterion referenced achievement measure.

STANDARDIZED TESTS. Data from school administered standardized achievement tests (the Comprehensive Test of Basic Skills) were gathered for several years prior to entry into the program and for one or two years (as available) subsequent to the receipt of tutorial services.

SCHOOL GRADES. School grades were also gathered from schools for several prior years and for one or two years following receipt of services.

DESIGN

CRITERION REFERENCED MEASURES. Rates of progress through the tutoring materials constituted a useful measure of tutee progress and program efficiency. Progress rates were examined by comparing the first to the second year of the program.

Tutees were placed into tutorial levels based upon placement tests; parallel forms were administered upon completion of a level. Therefore the design for analyses of criterion-referenced tests was pretest-posttest.

STANDARDIZED TESTS AND GRADES. Because of the high drop-out

rate and the select nature of the "program tutees," any comparison between CROSSROADS tutees and children not enrolled in the program would be uninterpretable. To cope with this serious design problem, the plan for evaluating standardized test and grade results took advantage of the fact that tutees were divided into three partially overlapping groups based upon tutoring areas (reading, math, and writing). Children tutored in one area were compared to children tutored in other areas. For example, in analyses of math achievement, math tutees were compared to children not tutored in math but tutored in reading and/or writing. For analyses of reading and language scores, children tutored in either reading or writing were compared to children tutored in math only because of the potential overlap of reading and writing effects. Children tutored in reading were examined for possible gains in both reading and language test scores as were children tutored in writing. Since all treatment and comparison subjects were selected into CROSSROADS using the same criteria, the effect was to block on the variables causing selection into the program, presumably motivational variables. Validity threats are discussed below.

RESULTS

Results associated with four different types of dependent variables are reported here. Each dependent variable is a measure of academic achievement.

TUTEE PROGRESS. Tutees received an average of 22.3 sessions of tutoring representing a period of contact between the program and tutees ranging from approximately 11 to 23 weeks. Tutees averaged completing 1.45 levels of criterion-referenced tutoring materials.

The average rate of progress over the two years of the program was one level of materials per 15.4 sessions of tutoring. There was a substantial increase in the progress rate from the first to the second year of the program. In the first year, the average tutee required 20 sessions to complete one level. For 27 tutees who entered in the second program year, the rate was one level for each 12.8 sessions. Therefore, program efficiency apparently increased from the first to the second year of the program.

CRITERION-REFERENCED TEST GAINS. The mean pretest score for tutees was 62.3 percent (standard deviation=10.8 percent). This figure includes only levels in which program tutees eventually received tutoring. Drop-outs and tutees who "tested-out" of a level by scoring above the 80 percent criterion were excluded. The mean posttest score on the criterion tests was 84 percent (standard deviation=10.1) for program tutees. This represents a statistically significant gain ($p < .01$, one-tailed).

STANDARDIZED ACHIEVEMENT TESTS. Table 1 contains results from the standardized achievement tests expressed as expanded

standard scores. This table shows contrasts between background test scores of children tutored and untutored in each tutorial area. Groups tutored in reading and writing were contrasted with the children tutored in math only in terms of reading and language test scores. In addition, two contrasts combine children tutored in reading with those tutored in writing for analyses of reading and language scores.

Background contrasts suggest that the criterion referenced placement tests for reading and writing were sensitive to reading achievement histories of the children. The contrasts for language scores do not indicate a similar sensitivity to language achievement histories.

Table 2 shows the parallel results for followup scores. The pattern is similar with the tutored group having generally lower achievement than the untutored group. For math, the gap has widened, to a statistically significant level. For children tutored in writing, the gap in reading scores has narrowed to a nonsignificant level.

Table 3 compares background and followup scores expressed as expanded standard scores. Changes for all groups, tutored and untutored are statistically significant. This pattern of findings creates doubt that the program is responsible for gains over time observed. To further examine this, additional analyses were conducted. First, homogeneity of the within group regressions was examined. Table 4 presents the results of regressing followup expanded standard test scores on background scores for tutored and untutored groups separately. Tutees were

grouped in several ways for these analyses based upon whether they were tutored in math, reading, or writing. Unstandardized betas and y-intercepts differ for comparisons involving language scores. Betas differ for one comparison involving reading scores and the comparison of the y-intercepts almost reaches significance at the five percent level for this comparison as well. For these significant contrasts, the correlation coefficients and unstandardized betas are not consistently lower in the tutored groups as would be expected if tutoring reduced the strength of the association between background and followup test scores.

To further examine these findings, regression analyses were conducted to test for the magnitude and direction of the effects of tutoring (Table 5). Only one result approaches significance. that for math. This is in the direction opposite that predicted by the hypothesis that the program was effective.

Because the tests for homogeneity of within group regressions suggested differences in three sets of equations, additional tests were conducted. Untutored groups' regression equations were used to predict posttest scores and residuals were examined (Tables 6 and 7). These indicate no significant differences in residual scores.

SCHOOL GRADES

Changes from background to follow-up were examined for school grades with no evidence of statistically significant changes in the grades of tutored or untutored children. Tutees averaged just below a "C".

SUMMARY AND DISCUSSION

The criterion referenced results from progress and pre-posttest measures show evidence that the tutees learned some basic skills. Criterion referenced materials were keyed to grade levels. Hence, for an average child to proceed through almost one and one half levels of materials in twenty-two sessions is evidence that some skill acquisition is occurring. Similarly, the pre-posttest gains indicate this.

However, there are two serious problems interpreting the criterion-referenced results. One is that the simple pre-post design used with these data does not permit us to infer whether the gains were due to program services; some or all of the learning might have occurred in school. Second and most seriously, the criterion referenced measured have an uncertain generalizability to the domain of basic skills. The "criterion" to which the tutorial mastery tests were keyed was the content of the tutorial materials, not a more meaningful criterion. This is the weakness of criterion referenced measures: the extent to which they reflect socially important amounts of learning is unknown. Given these drawbacks of the criterion referenced data, we must turn to the test and grade data for evidence about the social significance of tutee gains and evidence that the program is responsible for these gains.

The school test results suggest that tutees improved from background to follow-up in all basic skills areas. The results of the regression and residual analyses provide, however, no support for the hypothesis that the program was responsible for the changes in these measures. Finally, grade data show no

evidence of improvement.

VALIDITY ISSUES

There are several questions about the validity of this evaluation of CROSSROADS. Since this design involves nonrandom assignment to treatment and comparison groups, regression effects must be considered. Procedures for selection into the program and placement in a tutorial area are key factors in evaluating the likelihood of regression effects. Children were eligible for the program if they had scored below the 50th percentile on the most recent standardized achievement test administration by the school district. However, once referred to the program, children were placed into a tutorial area based upon their performance on placement tests, not based upon their standardized test score below the 50th percentile. Children were also eligible if they had grades below a "C", in a tutorial area; again, children were grouped on the basis of pretests, not grades. Therefore, for analyses involving test scores and grades, children were not grouped on the basis a measure used as a covariate. Regression effects due to testing error are therefore unlikely.

Another potential validity threat is motivational differences between groups. Had the comparisons been between CROSSROADS tutees and untutored groups, this would have been a very serious problem. However, because both treatment and comparison children had to demonstrate similar motivation to enter the program, the design succeeds in blocking on this variable.

More troublesome is the possibility of maturational effects

arising because skill development occurs at different rates in different children. If the tutees had been compared to a group of children with higher achieving backgrounds, this would have been especially problematic. Since procedures for enrolling tutees created groups with similar overall achievement backgrounds, there is little reason to expect that one group would show a sudden acceleration of overall maturation. However, there is the possibility of skill-specific maturation. Tutees were assigned to tutorial areas based upon placement tests which reflected a history of development of a specific skill area. Skill-specific maturation was examined with the tests for homogeneity of within-group regressions. For the one pair of groups which differed on the within-group regressions, skill specific maturation was examined with a test for the difference in residuals. These results suggest that skill-specific maturation is not concealing significant program effects.

Another consideration is the possibility that effects of the program are general rather than skill-specific by improving tutees study habits, academic self-concept or motivation, rather than by improving skills in a specific area. This would be consistent with the results showing improvements in test scores for both children tutored and those not tutored in each area. However, a competing explanation for the overall improvements in test scores is that the tutees who did not drop out represent a subset of tutees who are especially motivated to improve.

Another methodological issue is connected with the use of standardized tests. A controversy exists about their usefulness as measures of the effectiveness of remedial educational

programs. There are two major problems with the use of standardized tests for evaluation purposes. One is a validity issue and the other is a reliability issue.

Floden, Porter, Schmidt, and Freeman (1980) analyzed several standardized educational tests and found surprisingly little overlap in content. This suggests that no one test adequately covers the domain of skill purported to be assessed. Popham (1976) has pointed out that the procedures for standardization tend to remove items which are the core of educational curricula because these items are have low discriminability. Since they are the core of curricula, virtually all children learn them. Thus standardization results in test items peripheral to the curriculum.

The reliability issue is also troublesome. For children scoring at very low percentiles, the number of items useful for discriminating children is small because almost all children from these groups fail a large number of items. Guessing introduces a proportionately larger error component for lower scoring groups than for average scoring groups (Cole and Nitko, 1981). Fitz-Gibbons (1979) asserted that standardized tests are insensitive to the content of program curricula and usually understate gains achieved (cf. Popham, 1976). Weber (1978), however, in a critical review of "uses and abuses of standardized tests" indicated that such measures are useful for program evaluation.

Standardized achievement measures may be poor indicators of the impact of remedial programs. However, alternative indicators, criterion-referenced measures, are also inadequate, because the lack of standardization makes their educational

significance uncertain. This problem is well exemplified by the evaluation of CROSSROADS. The criterion-referenced results indicate that some learning took place but the standardized test results suggest that the magnitude of learning was not of educational significance. We might conclude that the program was ineffective except that questions about the sensitivity of the measures have to be admitted to the discussion. The evaluator is left in the position of being unable to estimate with confidence educational impact at remedial levels. The strongest conclusion which is warranted is that the program was not effective enough to measurably impact the standardized tests. The residual uncertainty is unfair to programs which carry the burden of proof to show effectiveness and puts the evaluator in a questionable ethical position.

Although a major undertaking, it is possible to construct a measure of basic educational skills with the advantages of criterion-referenced and standardized tests. Such a measure would be keyed to instructional materials used in classrooms; it would be standardized with one important change of procedure: item discriminability would not be the principal basis for item inclusion. Items would be retained in proportions which reflect their prevalence and importance in curricula. The test would have one disadvantage: compared to existing standardized measures, it would be lengthy. However, for evaluation purposes and perhaps for individual assessment, such a measure would be superior to any existing. Development of such a test should be a priority for the educational evaluation community.

To summarize the methodological issues, we may note that the

design succeeded in providing control for the most important competing hypotheses, regression, maturation, motivation, and selection. Results are therefore to be regarded as interpretable. However, because the design was quasi-experimental rather than experimental, competing hypotheses cannot be ruled out with absolute certainty. The principle caveat is that the potential insensitivity of the standardized tests creates the potential that the program had a modest impact which was not detectable with these measures.

POLICY ISSUES

CROSSROADS was originally intended to be a four year program but was terminated after two because of policy changes under President Reagan. A question arises about the appropriateness of a summative evaluation on the program since since it was terminated after less than its planned lifetime (see Benassi and Murray, Note 1).

It is certainly conceivable that, given two more years, the program could have improved. Therefore, this is a report on the program as implemented, not the program as conceived at the outset or as potentially developed. This is characteristic of evaluation; most programs are altered in the transition from the drawing board to implementation. The case of CROSSROADS is probably not an extreme example since slightly more than two years of service delivery took place, thanks primarily to diligence on the part of a competent staff who began providing services within a few weeks of funding. It is incumbent upon an evaluator not to ignore meaningful evidence even when programs

are terminated early.

A distinct question is whether CROSSROADS existed long enough to deliver an adequate "dose" of intervention. The average tutee was in contact with the program for a period of less than 22 weeks, about one-fourth of the life of the program. In the second year, 27 tutees entered who eventually passed the 8 session criterion for inclusion among "program tutees". These figures suggest that "program tutees" received an amount of tutoring which would have been typical of the larger group of tutees CROSSROADS might have served over four years.

If we accept that it is appropriate to interpret the results from CROSSROADS, the important question remaining to be discussed then is why the program was of apparently limited effectiveness. The answer may lie in the out-of-school model. The out-of-school model is characterized by several problems which seem inherent.

The most significant problem is motivational. A much larger number of children were involved at a minimal level than were involved at a level which might reasonably have had educational benefits. This indicates that the out-of-school approach has the formidable problem of recruitment and continuing the involvement of the tutees. This problem is much more significant for the out-of-school model than for in-school approaches. Related factors which were more specific to CROSSROADS included tutee and, to some extent, tutor absences from scheduled tutoring sessions. In addition, the paid, individualized tutoring model is labor-intensive. These inefficiencies did not derive from lack of core staff competence. Staff were diligent, worked long hours, were well organized, and met process objectives in a

timely and tho ough manner. Rather, the inefficiencies seem to be inherent in the out-of-school approach.

A volunteer tutoring approach would reduce cost but not eliminate the problems of drop-outs and absences. Creating a reward system for the tutees might reduce the problems of drop-outs and absences but might be costly. The Department of Education apparently believed that out-of school tutoring programs held promise. However, the experience of CROSSROADS seems to indicate that such approaches may find it very difficult to compete on cost and efficiency grounds with in-school models.

Although no single study can provide definitive conclusions about out-of-school tutorial programs, this evaluation attempts to provide a contribution to the methodological issues in the area of evaluation of out-of-school tutoring and to discussion of the policy issues associated with the national Basic Skills Improvement Program.

NOTE

1. Benassi, V. and Murray, P. Comment on Dowell's "Evaluation of Out-or-School Tutoring". Manuscript submitted to ERIC.

REFERENCES

Bausell, R.B., Moody, W.B., & Walzal, P.N. (1972) A factorial study of tutoring versus classroom instruction. American Educational Research Journal, 10, 591-598.

Bear, R. M., Hess, R. D., and Shipman, V. C. Social class differences in maternal attitudes toward the teacher and school. Paper presented at the 75th convention of the American Psychological Association, New York, 1966 (ERIC ED 018475).

Cole, N.S. and Nitko, A.J. (1981) Measuring program effects. In Berk, R. A. (Ed.) Educational Evaluation Methodology. Baltimore, MD: John Hopkins University Press.

Clarke, D.E. (1973) Measures of achievement motivation. Review of Educational Research, 38, 231-244.

Dowling, C. Parent Attitudes and Involvement in Supplemental Education Programs (Thesis). California State University, Long Beach, 1983

Fitz-Gibbons, C. An Examination of the Literature on Tutoring. Los Angeles: Center for the Study of Evaluation, University of California Graduate School of Education, 1978.

Filipczak,J., Lordeman, A., and Friedman, R. M. (1977) Parental involvement in the schools: Toward what end? New York: Institute for Behavioral Research, Inc., ERIC (ED 143104).

Floden, R.E., Porter, A.C., Schmidt, W.H., & Freeman, D.J. (1980) Don't they all measure the same thing? Consequences of standardized test selection. In Baker, E.C. & Quellmalz, E.S. (Eds.), Educational Testing and Evaluation: Design, Analysis, and Policy, Beverly Hills: Sage.

Gardner, L. (Ed.) (1979) Scope English Program. New York: Scholastic Book Services.

Gourash, N. (1978) Help seeking behavior: A review of the literature. American Journal of Community Psychology, 6, 413-423.

Hess, R. D. (1967) The cognitive environment of urban pre-school children. Manuscript, Stanford University, School of Education.

Popham, J.W. (1975) Classroom Implications of Criterion-Rereferenced Tests: Curriculum-Instruction-Evaluation. Los Angeles: Instructional Objectives Exchange, P.O. Box 24095, 90024.

Scout, T. H., and Borders, S. G. (1971) Parent involvement and confidence in the schools. High School Journal, 63, 23-29.

Tallmadge, G.K. (1978) Cautions to evaluators. In Wargo, M.J. & Green, D.R. (Eds.) Achievement Testing of Disadvantaged and Minority Students for Educational Evaluation. New York: CTB/McGraw Hill.

Weber, G. (1978) Uses and Abuses of Standardized Testing in the Schools. Washington, D.C.: Council for Basic Education, 725 Fifteenth Street, N.W., 20005.

Weinstein, M.S. (1969) Achievement motivation and risk preference. Journal of Personality and Social Psychology, 13, 153-172.

TABLE 1

CONTRASTS BETWEEN TUTORED AND UNTUTORED GROUPS ON BACKGROUND TEST MEASURES (EXPANDED STANDARD SCORES)

TUTORIAL AREA	GROUP	CRITERION VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION
READING	UNTUTORED	READBKG*	19	462.9947	78.467
	TUTORED		82	416.3598	62.464
READING	UNTUTORED	LANGBKG	19	456.2263	63.909
	TUTORED		82	431.1378	81.466
WRITING	UNTUTORED	READBKG**	19	462.9947	78.467
	TUTORED		73	450.6247	54.632
WRITING	UNTUTORED	LANGBK	19	456.2263	63.909
	TUTORED		74	460.2203	66.669
READING OR WRITING	UNTUTORED	READBK*	19	462.9947	78.467
	TUTORED		118	430.7754	62.085
READING OR WRITING	UNTUTORED	LANGBK	19	519.5737	67.875
	TUTORED		127	490.8000	63.468
MATH	UNTUTORED	MATHBK	28	450.9929	65.074
	TUTORED		108	440.5907	78.479

*P<.05; **P<.01

NOTE: The groups designated as "untutored" in reading and in writing respectively included children who were tutored in neither reading or language; that is, the untutored children for these contrasts included children tutored only in math. The reason for this is that the writing and reading placement tests were correlated.

TABLE 2

CONTRASTS BETWEEN TUTORED AND UNTUTORED GROUPS ON FOLLOWUP TEST MEASURES (EXPANDED STANDARD SCORES)

TUTORIAL AREA	GROUP	CRITERION VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION
READING	UNTUTORED	READING*	19	519.5737	67.875
	TUTORED		92	479.8457	57.660
READING	UNTUTORED	LANGUAGE	19	526.2211	65.482
	TUTORED		92	504.3402	61.433
WRITING	UNTUTORED	READING	19	519.5737	67.875
	TUTORED		82	506.4524	62.311
WRITING	UNTUTORED	LANGUAGE	19	526.2211	65.482
	TUTORED		82	532.0293	52.892
READING OR WRITING	UNTUTORED	READING	19	456.2263	63.909
	TUTORED		118	442.0407	77.601
READING OR WRITING	UNTUTORED	LANGUAGE	19	526.2211	65.482
	TUTORED		127	515.2622	61.389
MATH	UNTUTORED	MATH**	32	556.1406	75.452
	TUTORED		114	518.1263	64.986

*P<.05; **P<.01

NOTE: The groups designated as "untutored" in reading and in writing respectively included children who were tutored in neither reading nor language; that is, the untutored children for these contrasts included children tutored only in math. The reason for this is that the writing and reading placement tests were correlated.

TABLE 3

BACKGROUND TO FOLLOWUP TEST SCORE COMPARISONS
FOR TUTORED AND UNTUTORED GROUPS

TUTORIAL GROUP	CRITERION		NUMBER OF CASES	MEAN	STANDARD DEVIATION
READING					
TUTORED**	READING	BACKGROUND	79	418.0215	62.065
		FOLLOWUP	79	477.9215	57.511
TUTORED**	LANGUAGE	BACKGROUND	80	432.9075	80.594
		FOLLOWUP	80	502.6250	59.055
UNTUTORED**	READING	BACKGROUND	52	461.6519	59.879
		FOLLOWUP	52	515.2962	64.248
UNTUTORED**	LANGUAGE	BACKGROUND	52	460.8923	62.692
		FOLLOWUP	52	535.2019	55.235
WRITING					
TUTORED**	READING	BACKGROUND	69	449.0188	54.862
		FOLLOWUP	69	505.5188	58.023
TUTORED**	LANGUAGE	BACKGROUND	70	458.6057	67.465
		FOLLOWUP	70	530.7200	49.192
UNTUTORED**	READING	BACKGROUND	62	420.1177	71.411
		FOLLOWUP	62	478.5548	65.240
UNTUTORED**	LANGUAGE	BACKGROUND	62	427.3645	80.176
		FOLLOWUP	62	498.2274	65.646
READING OR WRITING					
TUTORED**	READING	BACKGROUND	112	430.6491	61.159
		FOLLOWUP	112	488.2080	61.027
TUTORED**	LANGUAGE	BACKGROUND	113	441.8646	76.846
		FOLLOWUP	113	513.6487	58.615
UNTUTORED**	READING	BACKGROUND	19	462.9947	78.461
		FOLLOWUP	19	519.5737	67.875
UNTUTORED**	LANGUAGE	BACKGROUND	19	456.2263	63.909
		FOLLOWUP	19	526.2211	65.482
MATH					
TUTORED**	MATH	BACKGROUND	103	439.8262	75.073
		FOLLOWUP	103	516.6155	64.825
UNTUTORED**	MATH	BACKGROUND	28	450.9929	65.074
		FOLLOWUP	28	543.2393	69.442

TABLE 4

SUMMARY OF THE REGRESSION DISCONTINUITY ANALYSES

TUTORIAL AREA	DEPENDENT VARIABLE	CORR. COEFF.	UNSTANDARDIZED COEFFICIENTS BETA	STD.ERROR OF BETA	CONSTANT (Y-INT)	STD.ERROR OF CONSTANT
READING						
TUTORED	LANGUAGE	.6758	.495*	.061	288.262*	26.924
UNTUTORED	LANGUAGE	.6361	.652*	.192	228.888*	88.294
READING						
TUTORED	READING	.5672	.526*	.087	258.235	36.750
UNTUTORED	READING	.8451	.731*	.112	181.121	52.633
MATH						
TUTORED	MATH	.6094	.526	.068	285.171	30.394
UNTUTORED	MATH	.5228	.558	.178	291.627	81.261
WRITING						
TUTORED	READING	.6917	.732	.093	177.035	42.206
UNTUTORED	READING	.8451	.731	.112	181.121	52.633
WRITING						
TUTORED	LANGUAGE	.7267	.530*	.061	287.713*	28.151
UNTUTORED	LANGUAGE	.6361	.652*	.192	228.888*	88.294
READING OR WRITING						
TUTORED	READING	.6659	.664	.071	202.065	30.871
UNTUTORED	READING	.8451	.731	.112	181.121	52.633
READING OR WRITING						
TUTORED	LANGUAGE	.6930	.529*	.052	280.083*	23.406
UNTUTORED	LANGUAGE	.6361	.652*	.192	228.888*	88.294

NOTE: The groups designated as "untutored" in reading and in writing respectively included children who were tutored in neither reading nor language; that is, the untutored children for these contrasts included children tutored only in math. The reason for this is that the writing and reading placement tests were correlated.

TABLE 5

REGRESSION ANALYSES USING EXPANDED STANDARD SCORES

DEPENDENT VARIABLE	PREDICTOR	UNSTD. BETA	STD. BETA	F-VALUE	SIG.
MATH	MATH BACKGROUND	.531	.583	68.421	.000
	MATH TUTORING	-20.689	-.128	3.299	.072
READING	READING BACKGROUND	.581	.637	65.233	.0001
	READING TUTORING	15.527	.100	1.614	.207
READING	READING BACKGROUND	.731	.735	100.240	.0001
	WRITING TUTORING	3.833	.026	.129	.721
LANGUAGE	LANGUAGE BACKGROUND	.515	.661	75.674	.0001
	READING TUTORING	11.592	.076	.990	.322
LANGUAGE	LANGUAGE BACKGROUND	.533	.696	81.133	.0001
	WRITING TUTORING	-3.183	-.024	.104	.748

TABLE 6

TEST OF THE DIFFERENCE OF READING RESIDUALS USING UNTUTORED GROUP PREDICTION EQUATION AND SEPARATE VARIANCE ESTIMATES

GROUP	NUMBER OF CASES	MEAN RESIDUAL	STANDARD DEVIATION	STANDARD ERROR
UNTUTORED READING	52	-.1569	35.103	4.868
TUTORED READING	79	1.6922	52.734	5.933

T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
-.24	128.97	.810

TABLE 7

TEST OF THE DIFFERENCE OF LANGUAGE RESIDUALS USING UNTUTORED GROUP PREDICTION EQUATION AND SEPARATE VARIANCE ESTIMATES

GROUP	NUMBER OF CASES	MEAN RESIDUAL	STANDARD DEVIATION	STANDARD ERROR
UNTUTORED READING	80	-8.5187	45.328	5.068
TUTORED READING	19	-.1265	50.528	11.592

T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
-.66	25.33	.513

GROUP	NUMBER OF CASES	MEAN RESIDUAL	STANDARD DEVIATION	STANDARD ERROR
UNTUTORED READING OR WRITING	113	-3.3350	43.309	4.074
TUTORED READING OR WRITING	19	-.1265	50.528	11.592

T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
-.28	22.67	.796