DOCUMENT RESUME

ED 237 567 TM 830 832

AUTHOR Murray, Stephen L.; Braverman, Marc T.
TITLE Insuring Equity in Student Selection under Chapter 1. Research on Evaluation Program. Paper and Report Series No. 84.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Research on Evaluation Program.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jul 83
CONTRACT 400-80-0105
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Compensatory Education; *Decision Making; Educationally Disadvantaged; Educational Research; Equal Education; Models; *Program Evaluation; Research and Development Centers; Research Needs; *Resource Allocation; *Student Placement
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1; Northwest Regional Educational Laboratory; Technical Assistance Centers

ABSTRACT

This paper describes and illustrates a decision theory model to analyze student selection strategies in Chapter 1 programs. The relationship between the method and the practice of moral evaluation activity is based on viewing justice as a fundamental moral concern in distributing resources, one which is accomplished, in part, through student placement decisions. The decision theory is well suited to promote services for the most disadvantaged if one accepts the premise that academic advantage is best measured by standardized achievement tests and that schools in poverty neighborhoods require additional support to serve those students. The model, its rationale, application, and limitations are discussed: (PN)

paper and report series

No. 84 INSURING EQUITY IN STUDENT SELECTION UNDER CHAPTER 1

STEPHEN L. MURRAY
MARC C. T. BRAVERMAN
Northwest Regional Educational Laboratory

Research on Evaluation Program

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204
Telephone (503) 248-6800

No. 84 INSURING EQUITY IN STUDENT SELECTION UNDER CHAPTER 1

STEPHEN L. MURRAY
MARC T. BRAVERMAN

Northwest Regional Educational Laboratory

July, 1983

Nick L. Smith, Director
Research on Evaluation Program
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue, Portland, Oregon 97204

Published by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work upon which this publication is based was performed pursuant to Contract No. 400-80-0105 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

PREFACE

The Research on Evaluation Program is a Northwest Regional Educational Laboratory project of research, development, testing, and training designed to create new evaluation methodologies for use in education. This document is one of a series of papers and reports produced by program staff, visiting scholars, adjunct scholars, and project collaborators--all members of a cooperative network of colleagues working on the development of new methodologies.

How can we assess the justice and equity of selection procedures used to determine which disadvantaged students receive augmented educational services? This paper presents a decision theory model based on assigning values to various possible correct and incorrect placement decisions as one method of determining the equity of placement decisions in selecting students for Chapter 1 programs. The model, its rationale, application, and limitations are discussed here.

Nick L. Smith, Editor
Paper and Report Series

TABLE OF CONTENTS

INSURING EQUITY IN STUDENT SELECTION UNDER CHAPTER 1[1,2]

When Congress passed Chapter 1 of the Educational Consolidation and Improvement Act (ECIA) of 1981, it signalled its continued commitment to the concept that students residing in poor neighborhoods and demonstrating achievement deficiencies in basic skills were entitled to augmented educational services. Chapter 1 replaced Title I of the Elementary and Secondary Education Act (ESEA) of 1965, and, although its regulatory structure is modified, Chapter 1 is essentially the same program that Title I was. It carries an annual appropriation of about 3 billion dollars, roughly 20 percent of the total federal education budget.

Federal policy, in the form of law and regulations, structures the selection of schools and students for Chapter 1 programs. Chapter 1 (formerly Title I) Evaluation Technical Assistance Centers (TACs) have provided consultation and training since 1976 on evaluation procedures, including issues relating to the selection process. An external evaluation of Title I evaluation practice found notable improvement in student selection in the ten-year period from 1972 to 1982 (Reisner, Alkin, Boruch, Linn & Millman, 1982). Reisner et al. observed considerable variability, however, in the quality of selection data. This may be due to varying distributions of evaluation duties in local school districts. Many Chapter 1 staff who fill evaluator roles in small and medium size districts do so only on a part-time basis. They lack the technical knowledge that trained professional evaluators often take for granted. To these staff, designing a student selection strategy presents significant technical and moral issues.

It is often the technical issues which receive the most attention. These concern the choice of a valid and reliable selection instrument which has a sufficiently high correlation with some predetermined criterion measure of selection soundness, such as the pretest. However, the moral issues must not be

overlooked, and these revolve around justice and equity in selection. From this added perspective, the adequacy of selection and placement strategies is decidedly not reducible solely to the technical soundness of the instruments; one must also take account of the consensual or negotiated institutional values upon which the selection strategies are implicitly based. These institutional values can be operationally defined as the utilities that are assigned to the various possible correct and incorrect placement decisions.

Justice may be situationally defined as fairness and impartiality in the way resources are distributed to those who are entitled to them. Federal budget reductions have heightened concern for justice, since fewer students can be served with fewer dollars. This paper addresses the concepts of justice and equity as they impinge on student placement decisions in Chapter 1. It also describes a methodology that local district evaluators can use to model and appraise alternative student selection strategies, thereby being responsive to concerns regarding how equitably Chapter 1 resources have been distributed.

Approaches to Determining Equity

Federal statute requires that the placement of students into Chapter 1 programs occur through the application of an objective selection process. In practice this generally takes the form of a rank ordering produced through a district testing program, often in combination with some form of teacher judgment. Chapter 1 participants are selected by starting with the lowest-ranked student and continuing until all available program spaces have been filled. Students requiring other special services (e.g., special education) may be excluded. Typical elements of the procedure that local district personnel must decide are which test to use, whether or not to include teacher input, and, if so, what form that input will take.

The standard model for evaluating Chapter 1 programs, the norm-referenced evaluation model, requires that students be selected on the basis of a test administration other than the one that will be used to measure their pretest status. Usually different tests are used for the two purposes. The separation of selection testing and pretesting allows the group's pretest percentile to be used as an estimate of the posttest status that would have resulted in the absence of a Chapter 1 program (Tallmadge, Wood, & Gamel, 1981). The separation is a control that allows for the occurrence of regression to the mean, which would otherwise lead scores of the selected Chapter 1 group to be closer to the mean of the local population on the pretest than they were on the selection test. The pretest is viewed as a more precise indication of pre-program status, since it avoids the problem of a nonnormal distribution of error variance associated with the selection of a extreme-scoring group.

The pretest is sometimes used as a check or a criterion measure to estimate the adequacy of the selection procedure. A technical analysis of the selection problem typically focuses on the content validity and reliability of the two measures, and on the correlation between them. If the pretest score is indeed taken as a criterion of proper placement, the correlation between the selection measure and the pretest can be interpreted as an indicator of predictive validity. The higher the predictive validity, the fewer students there would be who outscore nonselected students on the pretest, should both groups take both tests. The objective is to maximize predictive validity in the local context.

Under the norm-referenced evaluation model, districts must record their program participants' pretest and posttest status in terms of national norms. Let us assume that a district establishes some national percentile level on the pretest as an achievement standard setting off those students in need of supplementary services. If we further assume that Chapter 1 is predicated on an entitlement principle of justice, then nonselected students falling below this standard, as well as

selected students falling above it, create dissonance for local Chapter 1 coordinators. The burden of proof is frequently placed on the local school district when a noticeable proportion of students is found to be misclassified in either of these two ways.

The application of a classical psychometric perspective is useful in identifying possible sources of inadequacy in a selection strategy, but it requires methods of analysis that are not always technically feasible and may be beyond the expertise of small district project staff. When used in practice the analysis problems frequently encountered due to missing data and restricted ability range in the Chapter 1 population can create interpretation difficulties. Moreover, the psychometric perspective emphasizes underlying technical weaknesses without providing a direct basis for evaluating and improving a local strategy from either a technical or moral standpoint. For this reason an assessment of the selection process based on decision theory can be of benefit to local personnel.

Decision Theory

A decision theory perspective provides project staff with simple empirical methods to evaluate alternative selection strategies as they may be expected to work locally. The objective of a sound selection strategy in decision theory terms is to minimize incorrect decisions, or, conversely, to maximize correct decisions.

Students are placed into two groups on the basis of some criterion, such as the pretest. Those below a prespecified pretest standard are classified as academically deficient, while those above this standard are classified as not academically deficient. The same population of students is dichotomized on the basis of the selection information. Some students are identified as low-achieving, and therefore qualified for Chapter 1 services, while others are identified as not qualified. A single score, a composite score or a multistage

selection process may be used to support classifying students as qualified or not qualified. (Note that the term "qualified" in the present discussion refers to low rather than high achievement levels.)

Having dichotomized independently on the two bases of selection information and the criterion, the evaluator prepares two by two contingency tables with the selection information on one dimension and the criterion information on the other (see Figure 1). The proportion of correct identifications (correct decisions) and the proportion of incorrect identifications (decision errors) are readily derived from the tables. For this analysis is to be complete, criterion scores must be present for the nonselected group. This will therefore rule out use of the pretest in those districts pretesting only their Chapter 1 students. In those cases local evaluators wishing to employ a decision theory design can use a districtwide test for criterion information. The selection of the criterion will be discussed later in the paper.

Two types of decision errors are possible. First, a student identified as qualified on the basis of selection information may be subsequently judged as academically proficient on the basis of the criterion information. These students are "mistakenly" placed in the Chapter 1 program. A second type of decision error occurs when a student classified as not qualified is subsequently determined to be academically deficient. These students are "mistakenly" excluded from the Chapter 1 program. Alternative selection strategies can be compared on the basis of the rate of decision errors. The following paragraphs define key concepts in the approach.

The <u>selection decision</u>, typically based on a measure or measures of achievement, is a judgment to select or not select a student for the Chapter 1 program. The measures may consist of a single test score or a combination of measures in which case the judgment may be based on a weighted composite or on a sequential decision strategy. Students selected are those who fall below a designated cutoff score or, in the case of sequential decision

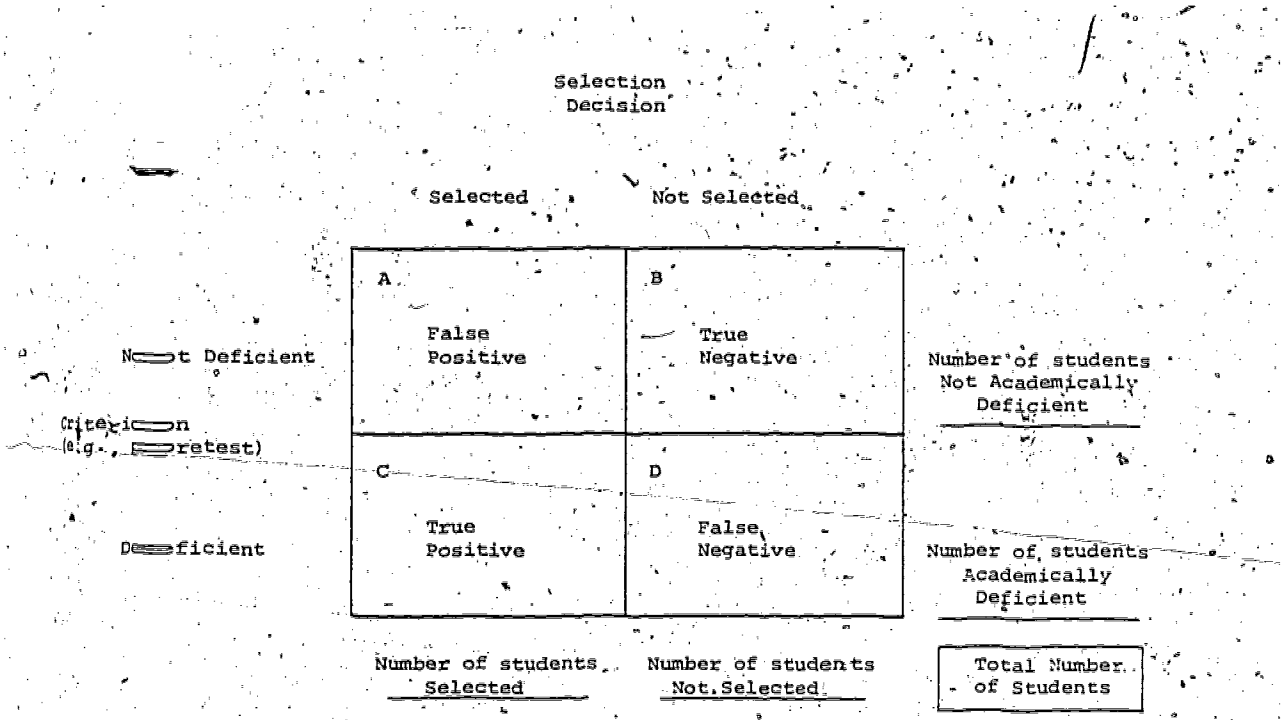

Figure 1: A Model for Evaluating Student Selection Strategies

rules, are characterized by a pattern of scores across multiple measures. Students not selected are those falling above the cutoff score.

Academically deficient students are those judged to be in need of the skills taught in the Chapter 1 program. Falling below a designated standard of skill attainment on the criterion measure serves as the operational definition of this classification.

A correct identification results when a student selected for the Chapter 1 program is also determined to be academically deficient (true positive), or when a student not selected for the Chapter 1 program is also determined not to be academically deficient (true negative).

An incorrect identification results when a student selected for the Chapter 1 program is determined not to be academically deficient (false positive), or when a student not selected for the Chapter 1 program is determined to be academically deficient (false negative).

The correct decision rate is the number of students correctly identified divided by the total number of students in the local population. The purpose of the analysis is to maximize this ratio, given a criterion of satisfactory achievement level and alternative selection decisions.

$$\text{Correct Decision Rate} = \frac{n_B + n_C}{N},$$

where n_B and n_C are the numbers of students in quadrants B and C, respectively, and N is the total number of students considered in the selection process, i.e., $n_A+n_B+n_C+n_D$.

The base rate is the proportion of students judged to be academically deficient, based on the criterion measure.

$$\text{Base Rate} = \frac{n_C + n_D}{N},$$

with symbols being defined as above.

The selection ratio is the proportion of the student population that is selected, or the number of available positions in the Chapter 1 program divided by the total number of students from which the selection is made. The selection ratio is influenced by such factors as district policy, available resources, and the nature of the instructional strategy. The selection ratio, therefore, may differ from the base rate, as would be the case when resources do not allow all academically deficient students to be served.

$$\text{Selection Ratio} = \frac{n_A + n_C}{N},$$

with symbols being defined as above.

Uses of the Model

The decision theory perspective can be a useful addition to the psychometric approach for several reasons. First, consider the different foci of the methods. Psychometric approaches emphasize test characteristics while decision theory emphasizes the selection process itself.

An example is found in the Regression Model for selection decisions (e.g., Cronbach & Gleser, 1965), illustrated in Figure 2. The selection measure and the criterion measure are both considered to be on interval scales, and the least-squares regression equation is calculated to determine a student's predicted criterion performance on the basis of his or her selection measure score. The regression line $y' = a + Bx$ produces this predicted score for each possible selection score. In a bivariate regression the regression coefficient B is equal to the product-moment correlation between the variables.

A cut-off score x^* is chosen on the selection measure which determines entry into the Chapter 1 program. This score is generally the one which sets off the number of students which the program can accommodate. If we set aside for the moment exceptions arising from parental consent issues or policies

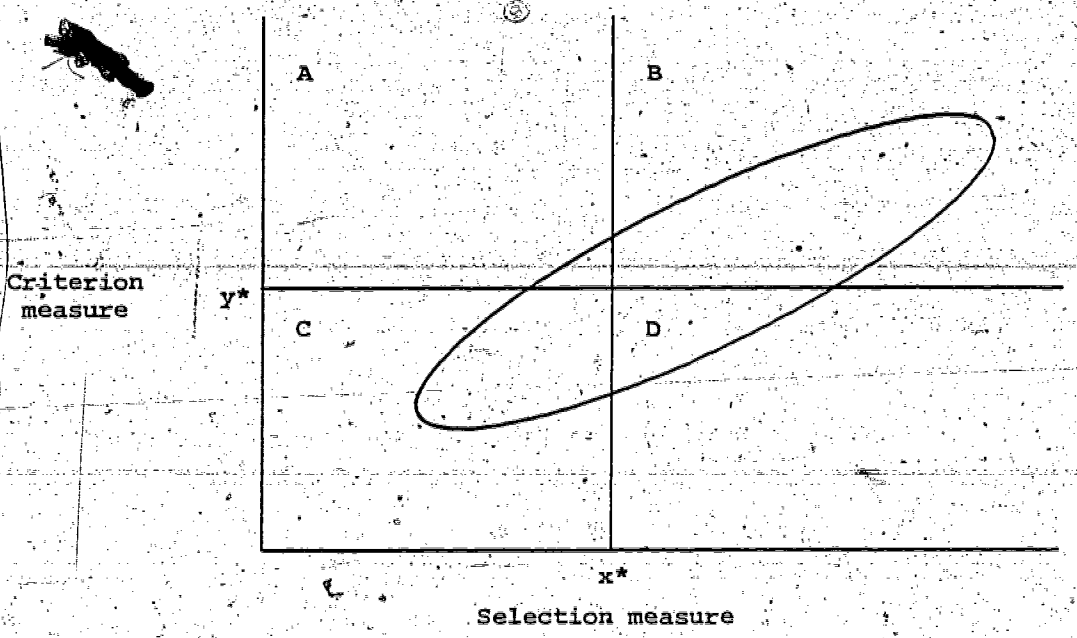

Figure 2: The Regression Model for Student Selection

regarding previous Chapter 1 participation, neither of which affect the analysis, we can state that all students falling below this cut-off go into Chapter 1; all students falling above it do not.

To compare this method with the one based on decision theory, let us add the one line $y = y^*$, where y^* is the cut-off on the criterion measure which identifies the achievement level representing minimum academic proficiency. As Figure 2 illustrates, the two cut-off points divide the student population into the quadrants encountered in Figure 1.

Judging the adequacy of the selection measure on psychometric considerations entails several steps. First the evaluator examines the reliabilities of the two measures and ascertains that they are both sufficiently high. The content validity of the criterion test will indicate its adequacy in determining academic proficiency. The correlation between the two measures, represented by the regression coefficient B, is taken as an indicator of the criterion validity of the selection measure. (For the selection measure, criterion validity rather than content validity is the critical psychometric feature, since the test's express purpose is to duplicate the criterion information. Nevertheless, criterion validity and content validity will be related, and in practice the latter is usually considered an important attribute.)

A selection procedure involves a series of dichotomous placement decisions on individual students. The greatest potential for selection error will occur in the area around the intersection of the two lines formed by the cut-off scores. Yet the correlation coefficient is sensitive to all points in the bivariate distribution. If discrepancies in the rank orderings created by the two measures exist anywhere on the ordinate scale, these will lower the coefficient B and decrease the apparent attractiveness of the selection measure. In a general sense this perception is correct, since, given a bivariate normal distribution, discrepancies will occur in a predictable pattern across the range of scores. However, one must remember that in

the context of a selection process the correlation coefficient is an indirect indicator. Minor discrepancies at the extremes of the ordinate scale will generally not have a bearing on the selection process, since in those cases the dichotomous selection decision will be unaffected. Therefore one of the weaknesses of the psychometric analysis is that the statistical summary it provides does not directly bear on the information needed to judge the selection process. In particular, to the extent that the variables do not follow a bivariate normal distribution, the correlation coefficient is an inexact reflection of the correctness of the selection decision. This is a special consideration in districts with small numbers of students, since in those cases one can expect greater deviation from normalcy in the distributions, with a correspondingly decreased predictive power to identify the tests' relationship in the critical ranges near the cut-off points. In sum, the purpose of a selection procedure is more directly reflected in the decision theory conceptualization: to maximize the number of correct classifications in student placement.

Another advantage of the decision theory perspective is that it accommodates noninterval-scaled classification on the selection measure. Most districts do not in fact use only test information in making selection decisions. Some submit tentative selection lists based on test rankings to students' previous-year teachers for confirmation or disconfirmation; as a result some students will be moved onto the Chapter 1 list and others off of it. Other multistage selection processes are sometimes employed as well. In addition, some districts use composite scores formed by weighting and adding teacher ratings to test scores; it is not clearly justified to assume interval scaling for such meaures. Use of the Regression Model is not recommended in those cases where the selection measure is not interval-scaled. The decision theory model, however, is appropriate in all cases; it requires simply the end results of the classification process.

Utilities. The decision theory approach is well suited to reflect district values regarding justice through the flexibility inherent in the assignment of utilities. Districts can assign these utility values to the various categories of classification. When computing summary values, the total in each cell would be multiplied by the cell's utility value. The four resulting numbers would be summed and divided by N.

The simplest coding scheme, and the one which we have been implicitly assuming thus far, is to assign a value of +1 to correct classifications and a value of 0 to both categories of misclassification (false positive and false negative). Summing the utility-coded totals across the four cells and dividing by N would then be equivalent to calculating the previously noted Correct Decision Rate.

The technique also allows for any other coding scheme deemed appropriate by district evaluators. One alternative strategy is to value the two misclassification types unequally. For instance, utility values might be set at 0 for false positives and -1 for false negatives. From a justice standpoint, a district might adopt this strategy if it decided that withholding Chapter 1 services to an academically deficient child was a greater disservice than providing them to a child who is achieving satisfactorily. The strategy might also be desirable from a practical standpoint if false positives proved easier to rectify than false negatives; for instance, it could be the case that children mistakenly placed in Chapter 1 would be identified within a few weeks and sent back to the regular classroom, whereas children mistakenly excluded from Chapter 1 would have to be retested or await program openings before entering the program.

The choice of unequal utility values would not be particularly beneficial in comparing selection processes that entailed equivalent selection ratios. This can be seen from a glance back at Figure 1. Let us assume that for a particular year the base rate and the selection rate in a district are fixed, though not necessarily equal. With the marginal totals thus specified, determining any of the four cells will determine

the whole matrix. The cells are not free to vary and thus any two selection strategies can be compared on the basis of one entry (e.g., number of correct Chapter 1 classifications) rather than a utility-coded summary value computed across the four cells.

However, the unequal coding scheme would be helpful in comparing selection strategies with differing selection ratios. As an example, one district we have worked with set out to modify its selection process so as to admit fewer students to the Chapter 1 program. Central office staff, believing that too many students who did not need Chapter 1 help were being referred and accepted into the program, planned simply to lower the local cutoff score to decrease the number of program participants. Data from the previous year's evaluation were used to model and evaluate alternative selection strategies. On reviewing the data, the district was compelled to take into account the effect that reducing the cutting score had on increasing the number of students erroneously excluded from the program. If these false negatives (erroneous exclusions) were judged more serious threats to equity than false positives (erroneous admittances), a utility-coding scheme reflecting this value would cast doubt on the new selection strategy to a greater extent than would a scheme assigning equal value to the two forms of error.

Limitations of the Model

The decision theory model is presented as an aid to the assessment of equity in a district's Chapter 1 selection plan, but it is not intended to replace the more usual psychometric considerations. In particular, we must note two shortcomings of the model, both of which concern the possibility of arbitrary judgment relating to the criterion measure.

First is the choice of the criterion measure itself. The model would be misleading if it caused one to infer a lack of error in this dimension. In practice the criterion measure is subject to as much error as the selection measure, as is evident from the case when both are standardized tests. To remedy this

possible interpretive error, we suggest that the user view the model as an index of agreement between two sources rather than an index of the selection procedure's matching an absolute standard of "truth."

We can understand the genesis of this problem if we consider that the decision theory model is used most notably in personnel decisionmaking (e.g., Cronbach & Gleser, 1965). In that field, there is indeed an ultimate standard against which the selection plan is judged, namely subsequent employee performance. Thus measures of employee performance are conceptually, temporally, and operationally distinct from measures used in selection. The selection instrument is merely a more or less successful predictor of the criterion. In education, however, the direct corollary of employee performance, later success in the Chapter 1 program, is not the variable that determines the correctness of the original Chapter 1 selection decision, lest the student be blamed for programmatic shortcomings. Federal statute requires that the admittance decision be based on level of academic proficiency; this classification, then, is the aim of the criterion measure as much as it is the aim of the selection measure, and in neither case is it made without some degree of error.

Given that the criterion should be used to measure agreement with the selection measure, it can be any accepted, objective measure of student performance. Probably the most logical candidate is the test used for pretesting, since substantively and temporally it is closely related to the selection instrument. However, other measures could be used as well, including the previous spring's posttest, teacher ratings, or any other standard that district personnel feel is suitable. As mentioned earlier, in many districts it is not possible to use the pretest as a criterion, because only the selected Chapter 1 students are given this test and therefore complete matrix information is lacking.

The second shortcoming of the model lies in the dichotomy required for judging who is academically deficient. When a

district counts its program participants from the lowest achieving student on up, the choice of a cut-off score on the criterion measure may have little more significance than an afterthought. Evaluators may find it difficult to dichotomize a variable that is most clearly conceptualized as continuous. We feel this is the greatest weakness of the model. To overcome the possible arbitrariness of the cut-off judgment, a district should select a locally meaningful standard, such as a percentile level that has served a previous delimiting function, the percentile equivalent of a targeted grade equivalent score, or a specified number of instructional objectives mastered.

Conclusion

We have described and illustrated a method that can be used to analyze student selection strategies in Chapter 1 programs. The relationship between the method and the practice of moral evaluation activity is based on viewing justice as a fundamental moral concern in distributing resources, one which is accomplished, in part, through student placement decisions. A general conception of justice requires us to give consideration to fairness and impartiality.

The decision theory approach described does not consider the consequences of placing students in Chapter 1 programs. The relative effectiveness of one program over another in terms of achievement gain is not considered. The local Chapter 1 program might be no more effective than the regular school program, or it might even have a hindering effect on its participating students. In principle, the use of decision theory models could be expanded through the designation of achievement gain or treatment effect as the ordinal matrix variable, thus allowing the consequences of placement to be considered. Cronbach and Gleser (1965) provided the mathematical basis for these extended models years ago. However, it might be more realistic to assume that periodic program evaluations will detect neutral or harmful effects should they be present.

The decision theory analysis assumes that the objectives and focus of the Chapter 1 program have been set prior to analysis. The grade levels and subject areas for the program must be established prior to using the decision theory technique. While annual needs assessments are required for Chapter 1 funding, there is considerable variability in the soundness of those assessments (Reisner et al., 1982). More important, there is little reason to expect that making these decisions can be reduced to technical criteria (Cronbach and associates, 1980).

The decision theory technique is well suited to promote services for the most disadvantaged if one accepts the premise that academic disadvantage is best measured by standardized achievement tests and that schools in poverty neighborhoods require additional support to serve those students. Thus, the decision theory technique is compatible with the intent behind Chapter 1.

FOOTNOTES

1 A version of this paper was presented at the annual meeting of the American Educational Research Association, Montreal, April, 1983.

2 A previous report in this series contains an earlier, less procedural discussion of this topic. See S. L. Murray "Utility and Equity in Student Placement." ROEP Paper and Report Series No. 56, Northwest Regional Educational Laboratory, Portland, Oregon, May, 1981.

REFERENCES

Cronbach, L.J., & Gleser, G.C. Psychological tests and personnel decisions (second edition). Urbana: University of Illinois Press, 1965.

Cronbach, L.J., & associates. Toward reform of program evaluation. San Francisco: Jossey-Bass Publishers, 1980.

Reisner, E.R., Alkin, M.C., Boruch, R.F., Linn, R.L., & Millman, J. Assessment of the Title I Evaluation and Reporting System. Washington, DC: U.S. Department of Education, 1982.

Tallmadge, G. K., Wood, C. T., & Gamel, N. N. User's Guide: ESEA Title I Evaluation and Reporting System. Mountain View, CA: RMC Research Corporation, 1981.