DOCUMENT RESUME

ED 096 346 95 TM 003 952

AUTHOR Page, Ellis B.; Breen, Thomas F., III
TITLE Seeking a Measure of General Educational Advancement: The Bentee. Final Report.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO BR-2-A-014
PUB DATE Dec 73
GRANT OEG-1-72-0004
NOTE 100p.

EDRS PRICE MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; *Educational Assessment; Educational Development; *Educational Needs; Educational Planning; High School Curriculum; Management Systems; Objectives; Operations Research; *Social Values
IDENTIFIERS *Bentee

ABSTRACT

Education suffers from the lack of any common measure of overall educational advancement, related in some clear way to the objectives of society. The present study proposes such a measure in the "bentee" (standing for benefit T-score) and provides evidence about its utility and form. Creation of a general "bentee" permits the design of a tree of values, leading by clearly defined steps from the highest values of society to the lowest, most directed, and definable objectives and test items. The mathematics of such a tree is analyzed, permitting the tracing of any subarea of performance in terms of its overall value. Two methods of assigning values, at a given level of the tree, are compared experimentally, using 101 sample judges from two regions of the U. S. It was found that the values had clear and consistent rank orders when group means were considered; that the less elaborate token methods served as well as the latest technique; that the relative values for boy high-school seniors were only slightly different from those judged for girls; and that educators did not differ, as a group, from noneducators in their assigned values. Individuals, however, differed enormously from one another in their appraisal of educational values. (Author)

Final Report

Project No. 2-A-014
Grant No. OEG 1-72-0004

Ellis B. Page and Thomas F. Breen III
Department of Educational Psychology
The University of Connecticut
Storrs, Connecticut 06268

SEEKING A MEASURE OF GENERAL EDUCATIONAL ADVANCEMENT: THE BENTEE

December 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

Final Report

Project No. 2-A-014
Grant No. OEG 1-72-0004

SEEKING A MEASURE OF GENERAL EDUCATIONAL ADVANCEMENT: THE BENTEE

Ellis B. Page and Thomas F. Breen III

The University of Connecticut

Storrs

December 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development
(Regional Research Program)

SEEKING A MEASURE OF GENERAL EDUCATIONAL ADVANCEMENT:

THE BENTEE

Ellis B. Page and Thomas F. Breen III

(Abstract)

Education suffers from the lack of any common measure of overall educational advancement, related in some clear way to the objectives of society. The present study proposes such a measure in the "bentee" (standing for benefit T-score) and provides evidence about its utility and form.

Creation of a general "bentee" permits the design of a tree of values, leading by clearly defined steps from the highest values of society to the lowest, most direct and definable objectives and test items. The mathematics of such a tree is analyzed, permitting the tracing of any sub-area of performance in terms of its overall value.

Two methods of assigning values, at a given level of the tree, are compared experimentally, using 101 sample judges from two regions of the U.S. It was found that the values had clear and consistent rank orders when group means were considered; that the less elaborate <u>token</u> method served as well as the <u>latent</u> technique; that the relative values for boy high-school seniors were only slightly different from those judged for girls; and that educators did not differ, as a group, from non-educators in their assigned values. Individuals, however, differed enormously from one another, in their appraisal of educational values.

Preface

The idea of the "bentee" (or of some similar system of combining information into an overall rating of educational accomplishment) probably has a long ancestry. But for the first author, the beginning notions first took form during the Spring of 1971, no doubt influenced by a good many experiences, but notably by these: work with the beta coefficients in multiple regression; work in Venezuela for a year on the problem of national goals; the chairmanship of an Applied Research Planning Session for the U.S. Office of Education; an earlier concern with educational philosophy as related to behavioral science; and most importantly, the experience of frustration in attempting application of management science to education, this difficulty resulting from the absence of any general measure of benefit beyond a single test score.

The debts are many to the persons who suggested approaches and criticized the early thinking: I am indebted to Sanford Temkin for conversations in 1967, when he spoke of searching for a unit he called a "bentile," which was conceptually to be constructed in a very different way; also for guiding me to certain writings such as those by C. West Churchman, and to his own work within Research for Better Schools, Philadelphia. For other valued comments, I am indebted to Robert Gable, Arthur Jensen, Robert Linn, Richard McCann, Richard Otte, Richard Page, Dieter Paulus, Harry Silberman, and Julian Stanley. Later valuable comments were recieved from Carl Bereiter, Bruce Rogers, and Skip Livingston. Peter I. Tillett made some good technical recommendations, and Harry Posten solved a substantial statistical problem for us. Important later comments came from William E. Coffman, and from Harold Gulliksen.

The general ideas of the bentee project have been encouraged by editors and by program chairmen: Robert Linn published the beginning idea in the Journal of Educational Measurement (1972). Scott Gehman accepted a further theoretical and mathematical analysis for Educational and Psychological Measurement (1974). Gilbert Austin arranged an invited address for the American Educational Research Association (1972); and Lewis C. Solmon, of the National Research Council, encouraged its presentation to a panel at Woods Hole concerned with higher education (these proceedings have now been published). And the findings and conclusions were provided a very useful hearing by William E. Coffman, at the University of Iowa Dedication of the Lindquist Center for Educational Measurement, in 1973 (these proceedings have also been published).

The writers are greatly indebted to the officials of the U.S. Office of Education, Boston, who provided a small but vital financial support to the research. And we thank Wayne Martin, who collected data in the area of Cleveland, Ohio. We are grateful to Richard Bloomer, who helped evaluate the proposal and the data, as part of the Dissertation Committee for the second author. And we acknowledge the help provided by The University of Connecticut Computer Center, with support from the National Science Foundation.

-- Ellis B. Page (Storrs, Conn.) and Thomas F. Breen (São Paulo, Brazil)

CONTENTS

LIST OF TABLES

(continued)

List of Tables (continued)

CHAPTER 1

DEFINITION, BACKGROUND, HYPOTHESES

The purpose of this research is to propose and evaluate a measure of overall educational accomplishment. This measure is referred to as the "bentee" (defined and described later), and may be considered roughly analogous to a "dollar" in economic values. This chapter will accomplish the following: Section A will define the problem of overall value. Section B will examine the related literature. And Section C will describe the hypotheses governing this research.

A. Definition of the Problem

Educational research and development have repeatedly run aground in the fog of undefined goals. Long-range human goals, such as "happiness," "adjustment," or "equality," seem too remote from curriculum to be useful in educational planning. Yet without reference to the largest purposes, the sub-goals selected have at best an arbitrary quality: too often they seem to self-serve particular educators or sub-groups of society, and to be lacking in any obvious legitimacy. Educational planners and administrators are constantly encountering competitive goals which are literally "incomparable," since no technique exists for comparing them with each other. To make clear this difficulty, let us consider three cases requiring administrative decisions. They are common problems, and thus exhibit our fundamental confusion:

> Case #1. Recently, in a well-published study, a researcher gave low performers remedial work in mathematics. The students improved, somewhat reducing their lag in that one subject. Left unreported, however, was what happened to them in other subjects, from which the extra time was taken. Let us assume they suffered slightly in these others (as would surely be true in a well-run school). Then how would one decide whether the combination gain-and-loss is desirable?

> Case #2. A school board is considering possible allocations of teaching effort. One option is to give heavy emphasis to programs for the academically weak; another is to spread available resources more evenly across the board of talent; a third is to emphasize the gifted. How may we establish cost-effectiveness ratios for the three options, in order to compare them? If we define "effectiveness" for the weak students, the measure will not work for the gifted. How can the benefit to the total youth of the district be measured? Surely, the choice should not be resolved purely at the level of politics or public relations.

> Case #3. A high school is swept up in a fashionable movement for student "freedom" in course selection. As a result, a large number choose no math, no hard science, no foreign language. One body of teachers argues that the students are developing "responsibility" and "creativity," and these values are more urgent than the purely academic considerations. Is there any responsible technical procedure for balancing these claims, and for interrupting the confusing circle of warring opinion?

Yet this fundamental confusion exists in the last third of the twentieth century, after five thousand years of educational experience, three hundred of astonishing scientific development, and more than eighty years of a rapidly improving behavioral science. In the words of one writer (Churchman, 1961, p. 1): "Probably the most startling feature of twentieth-century culture is the fact that we have developed such elaborate ways of doing things and at the same time have developed no way of justifying any of the things we do."

Objective Functions. Yet by comparison with education, the practitioners of certain other disciplines have overall goals which serve as, at least, an approximate measure of achievement, so that they may select among competing alternatives: The engineer can use efficiency of manufacture, measured in cost or number of operations. The industrialist has corporate profit. The physician, client recovery. The businessman, personal income after taxes. And for certain purposes, the presence of such objective functions becomes a great convenience.

None of these objectives (steps of manufacture, client recovery, income after taxes) is a completely adequate representation of the presumed total objective. Yet they are very useful as approximations and enter fruitfully into various formal "games," techniques, and strategies, such as linear programming, queueing analysis, and decision theory (Churchman et al, 1957; Hare, 1967; Banghart, 1969; Van Dusseldorp et al, 1971). Wagner (1969) writes that for a problem to be susceptible to operations research, it must have four characteristics: "a primary focus on decision-making ... measurable values [such as economic values] that unequivocally reflect the future well-being of the organization ... reliance on a formal mathematical model ... [and] dependence on an electronic computer" (pp. 50 6). All four of these criteria clearly turn around the second: these "measurable values" which will make possible the mathematical models, will justify the complexity requiring the computer, and will validate rational decision-making. Lack of an overall effectiveness criterion, therefore, makes it very difficult to apply management-science techniques in education.

Attributes of a Solution. Suppose that a suitable measure of educational benefit were at hand. Then what would be its characteristics? Some would be: a universality of application; an interval scale of measurement; a non-arbitrary definition; a democratic basis for establishment; provision for expert management; responsiveness to appropriate change; and a recursiveness of function, so that it might be applied at a number of different levels. These are stated cryptically here but may become clearer later in the presentation. In any case, a model of these qualities is the dollar (or other monetary unit), a brilliant social invention established in a marketplace of goods and services.

Let us therefore postulate a "marketplace of educational values." What would be the features of such a marketplace? The principal feature is that many different "cases" should be traded in some useful way, so that the appropriate values may be established on the basis of one common scale. Another is that the judgments made should have a potential for translation, eventually, from value-space to test-space.

Over the last several decades, psychologists and educators have become expert in the measurement of traits, and only somewhat less so in the analysis of trait "profiles" (multi-trait vectors) of students or groups of students (Cronbach, 1970, chs. 10, 11). We know how to study relationships among traits; yet the separate traits still remain separate: we do not have accepted techniques for trading off one ability against another, or for deciding easily whether one student (or group of students) is better educated or prepared than another. Yet historically, progress has been made by overriding the difficulties and ambiguities in such judgments, and by functioning within some clear reductionist rules, even though recognizing their limited correspondence with the ideal. This proposal investigates an attractive possibility for reducing profiles to single scores.

Means and Ends. Since World War II a new management science has emerged which has developed techniques for the analysis of systems. Its main interest is the improvement of the operations of man-machine systems through the use of the scientific method of inquiry. Scientific inquiry includes logical, systematic, and empirical analysis of all constituent parts of a system, their functions, and their value in terms of the overall objective.

Logical analysis begins with the notion of the purposive act itself. "There are three classes of elements, entailed in any purposive act. These are (1) the decision-maker, (2) a set of alternative actions, and (3) a set of goals" (Churchman, 1961, p. 137). The decision-maker specifies one goal from the set of possible goals and then attempts to optimize the accomplishment of that goal by selecting from the set of alternative actions those which have the greatest probability of bringing him to the goal. If the goal has not been specified exactly enough, it will be impossible to choose among alternative programs to optimize it.

We can speak of good and bad actions in two senses: a) in an ethical sense, in terms of morally good and bad--a distinction which is outside the present realm of discussion--and b) in a practical, prudential sense, in terms of which actions have better or worse chances of accomplishing some specified end. A good action is an action which is better suited to attaining an end. In ethics the designation of a morally good act is predicated on an analysis of ultimate ends and goals and of the relationship of human acts to those ultimate goals. In practical, prudential judgments, actions are designated as good or bad after an analysis of their chances of accomplishing their specified purpose. The selection of educational programs may involve some ethical considerations, such as the possible waste of time, energy, resources, and people, but of more importance to this discussion are practical considerations of how well various investments of people and programs accomplish the ultimate end of education. Without a definition of the end of education there is no way to evaluate the means.

The Schools as Means. Our schools are the approved and accepted method of socializing our young. Through schooling the accouterments of society and culture are passed from one generation to the next. It is the official means adopted by American and Western European cultures to prepare younger members to take their full place in the culture as adequately functioning adults. "In its very essence, the school is an institution established by society for the purpose of preparing the young to participate in that society" (Stanley et al., 1956, p. 2).

This educational system has developed over many generations into its now rather complex structure. And its ultimate purpose may have been lost in that development. Our schools get bigger, our curricula more varied; we employ more advanced technology and machines, and consume more money and resources in the process. "Admittedly, numerous substantial gains have been realized from some of these developments. However, the question must ultimately be raised: . . .by what standard shall the value. . .of any educational activity be measured?" (Thut, 1957, p. 382).

The schools have a tendency, as does any organization, to trap people in a "mindlessness," where they fulfill their jobs without considering why they are doing them, without questioning their jobs' purpose. As Silberman (1970) states:

> . . .by and large, teachers, principals, and superintendents are decent, intelligent, and caring people who try to do their best by their lights. It they make a botch of it, and an uncomfortably large number do, it is because it simply never occurs to more than a handful to ask why they are doing--to think seriously or deeply about the purposes or consequences of education (p. 11).

He quotes Henry A. Kissinger further on this problem of mindlessness:

> . . .an increasing amount of energy has to be devoted to keeping the existing machine going, and in the nature of things there isn't enough time to inquire into the purpose of these activities. The temptation is great to define success by whether one fulfills certain programs, however accidentally these programs may have been arrived at. The question is whether it is possible in the modern bureaucratic state to develop a sense of long-range purpose and to inquire into the meaning of the activity (p. 20).

Without some adequate definition of a goal for the educational system, it is nearly impossible to attempt to justify any of the means educators use in schools. It is impossible for the educational decision-maker to make a rational choice between alternative actions and programs without a thorough definition of the ultimate purpose for which they are designed. The evaluation of alternative programs depends on the relative probability that each program furthers the ultimate goal of the entire system, the program with the highest probability of success, with the highest measure of accomplishment, being considered best.

It is striking, then, to realize that education lacks even any coherent effort to define some measurable objective. Our schools share some of the poor symptoms of other twentieth century institutions. To repeat the words of Churchman (1961, p. 1): "Probably the most startling feature of twentieth-century culture is the fact that we have developed such elaborate ways of doing things and at the same time have developed no way of justifying any of the things we do."

A rational choice from alternative programs demands a scientific analysis of the educational system, of its procedures and decision-making processes, and of the relationship of all operating parts to the final goal.

Operations Research. There is a science, operations research or systems analysis, which contains the tools needed for such an analysis. It is an adaptation of the scientific method to the operation of human enterprise.

> The scientific method has been man's outstanding asset in his technological development. Its success has been manifest not only in his greater understanding of the natural world and his ability to predict its behavior, but also in material advances which have lengthened his life, increased his leisure time, and enabled him to pursue a far greater variety of activities.
>
> In the field of applications man has pragmatically accepted these advances as sufficient evidence of the usefulness of the scientific method. Yet, perhaps for historical reasons, there remained a stubborn resistance to admitting the scientific method into the more personal and less materialistic affairs of man himself.
>
> . . . .Early in World War II, the barriers were lifted, and the scientific method, after centuries of confinement to technical problems, was at least called forth to struggle with the operations involved in human enterprises (Saaty, 1959, p. 1).

Although Banghart (1969) admits that the terms "operations research" and "systems analysis" are used interchangeably, he does distinguish between them:

> Confusion in the literature regarding systems analysis and operations research has been over whether or not the two phrases are synonymous. In general, the phrase systems analysis is more comprehensive and implies a wider look at the problem. The phrase operations research denotes one of a set of specific mathematical techniques for problem analysis (p. 37).

Definitions. As has been indicated above, systems analysis or operations research is the application of the scientific method to human enterprises. Most of the definitions of this science are couched in general terms. Saaty (1959, p. 3) gives as the "usual" definition of operations research "an aid for the executive in making his decisions by providing him with the needed quantitative information based on the scientific method of analysis."

According to Banghart (1969), systems analysis is

> . . .the set of quantitative-scientific techniques that assists the. . .administrator in the decision-making process. . . .Although the concept of systems analysis implies an investigation of a total system, the term denotes a set of quantitative-scientific tools available for analytical purposes (p. 3).

And although Churchman et al. (1957) think that operations research "is perhaps too young to be defined in any authoritative way," the authors do provide a "tentative working definition:"

> Operations research is the application of scientific methods, techniques, and tools to problems involving the operations of a system so as to provide those in control of the system with optimum solutions to the problems (p. 18).

Because of its great size and complexity, there is a necessity for applying operations research and systems analysis to the educational system. Operations research is a development aimed at exercising some control over the complexities involved in modern systems. Essentially, "a system is an integrated assembly of interacting elements, designed to carry out cooperatively a predetermined function" (Flagle et al., 1960, p. 58). Chorafas (1965) shows the combination of the two concepts:

> A system is a group of interdependent elements acting together to accomplish a predetermined purpose. Systems analysis is an attempt to define the most feasible, suitable, and acceptable means for accomplishing a given purpose (p. 2).

Whatever scientific and quantitative tools are used for research into, and analysis of, systems are subsumed under the heading of operations research.

The School System. We generally speak of the "school system" or the "educational system," and it can be shown that the schools do form a system in the technical sense. They are complex purposeful organizations. Feyereisen et al (1970) express this clearly:

> A system is a set of components organized in such a way to constrain action toward the accomplishment of the purposes for which system exists. Applied to a school organization this means that a school system is organized to receive uneducated children as its input and through a process called education produce educated young adults as its output (p. 55).

The size and complexity of the system demand new, more rigorous methods for the development and administration of programs. As Banghart says:

> . . .the problem has as its focal point the increased size and complexity of educational organizations. As a result of the increased size and complexity, new techniques had to be developed to help the administrator with his administrative decisions (p. 7).

However, education has tended to be one of the "more personal and less materialistic affairs of man" which have often shown "a stubborn resistance to admitting the scientific method" (Saaty, 1959, p. 1). But the scientist must continue to try to make available to the educator "the best tools to predict the results of any course of action he may propose and to analyse dispassionately his aims and objectives" (p. 1).

The historical or psychological reasons why educators may have resisted opening their affairs to the dispassionate scrutiny of the scientific method are not to be argued here. What is being pointed out is that, although the schools constitute a system, they have not been operated according to systems concepts.

> . . .Schools exist as systems having interdependent parts or subsystems. And, as is typical of systems, they have inputs, processes, and outputs. They have a purpose, a mission, a goal. While schools have been called systems in the literature of the past, they have not been organized, nor have they operated, according to systems concepts (Feyereisen et al., 1970, p. 7).
>
> Systems concepts are used to design a self-regulating and adaptive management system for a school which is capable of compensating for the stresses and tensions found in the system and redesigning itself to ensure that the objectives of the system are accomplished (p. 55).

The reason schools have not operated according to systems concepts is that one of the intrinsic characteristics of a purposeful organization has been missing. That essential element is the control function. According to Huysman (1970, p. 4), "a control function, which compares the achieved with the desired outcomes," is an integral part of a purposeful system. This function cannot be implemented in education until the desired outcomes are strictly specified. As Page (1972, p. 34) says, "Lack of an overall effectiveness criterion, therefore, makes it very difficult to apply management-science techniques in education."

The Purpose of Education. Therefore, the purpose of our educational system must be specified. The goal in need of specification is the overall purpose of the schools to educate. For Silberman (1970, p. 6) "education, to be education must be purposeful."

> Education is defined. . .as the deliberate or purposeful creation, evocation, or transmission of knowledge, abilities, skills, and values. To emphasize the deliberate and purposeful is not to deny that non-deliberate influences may be more powerful; it is to assert that man cannot depend upon a casual process of learning. Unless men are forced to rediscover all knowledge for themselves, they must be educated, which is to say that education to be education, must be purposeful (p. 6).

There has to be some specification of the purposes or mission of the schools before operations research and systems analysis techniques can be applied to other than trivial problems. Banghart (1969) is very particular about pointing this out:

> The most difficult part of the entire systems study involves establishment of very specific objectives to be accomplished. Because of the quantitative nature of systems analysis, it is necessary to be extremely specific in determining the objectives to be achieved. To state objectives in general terms such as the improvement of curriculum, enhancement of the educational system, or development of a better building program is totally inadequate (p. 39).
>
> An evaluation of performance is done in terms of effectiveness. The design of the problem solution determines how effective the new system will be in fulfilling the organization's mission. It is necessary to devise standards for measuring effectiveness (p. 42).
>
> Inherent in establishing the performance criteria is the concept of validity. Validity is used here in the statistical sense as in psychological testing. That is, the procedure or performance is valid if it does what it is supposed to do. Therefore, the problem solution is valid if the solution accomplished what it is supposed to accomplish (p. 42).

If operations research methods are effectively applied the specification of objectives will of necessity be of the overall and most general objectives of the system. What is needed is some measure of what it would mean to be educated. For, as Churchman et al. (1957) say, "operations research tries to find the best decisions relative to as large a portion of a total organization as possible" (p. 6). It is "not effectively used if it is restricted to one-shot projects" (p. 8). They continue:

> The concern of operations research with finding an optimum decision, policy, or design is one of its essential characteristics. It does not seek merely to find a better solution to a problem than the one in use; it seeks the best solution. It may not always find it because of limitations imposed by the present state of science or by lack of time, funds, or opportunity. But operation research's efforts are continually directed to getting to the optimum or as close to it as possible (p. 8).

The present limitation in the schools is the lack of a specified measurable overall objective for education. Van Dusseldorf et al. (1971) show some applications of operations research to subproblems within the educational system, but an optimal solution, to the entire system, demands a specified overall objective.

A Possible Solution. While it may be very difficult to formulate this objective, Feyereisen et al. (1970) do not view construction of specific objectives as impossible.

> The goals of an organization express its reason for being as well as its aspirations. They project expectations and define products of the enterprise. Educational organizations define their functions in terms of educational outcomes and in language which connotes the development of persons in the acquisition of such skills and abilities, concepts and perceptions, attitudes and values, and knowledge and methods of inquiry as are necessary to personal fulfillment and societal improvement. Four principal sources of objectives are generally identified as (1) the learners--their needs, interests, and purposes; (2) society--local, state, national, and world communities; (3) human activities and social processes; and (4) the organized bodies of knowledge or disciplines embodying the accumulated experiences of mankind. For each of these referents it is possible to articulate specific objectives in terms of the outcomes of learning at any level of educational effort (pp. 117-118).

Once specific objectives have been articulated, what would remain would be to combine them, through appropriate weighting techniques, into the overall measure that is needed to develop optimum decisions about programs and practices. Such a measure might have the following attributes which Page (1972) and Page and Breen (1972) consider essential to a suitable measure of educational benefit. They would be:

> a universality of application; an interval scale of measurement; a non-arbitrary definition; a democratic basis for establishment; provision for expert management; responsiveness to appropriate change; and a recursiveness of function, so that it might be applied at a number of different levels (Page, 1972, p. 34).

With the development of such a measure, it would be possible to take a more systematic and scientific approach to program planning and development. It would become possible to assess more rigorously the overall effects of curriculum changes and to calculate returns for various investments in people, money, time, and resources. Such scientific analysis is impossible in the present state of undefined educational goals.

Section B will examine some of the goals and purposes stated for education. Not one, however, is found that fulfills all the needs of the effectiveness criterion, that is necessary for the introduction of systems methodology into curriculum planning. That criterion must be developed.

B. Related Literature

The above section tried to establish the need for an overall measure of educational advancement to serve as an effectiveness criterion for educational programs and planning. Advancements in management science and systems analysis, which have been so effectively applied in other complex operational systems, will find very limited application in education without some overall effectiveness criterion. A search of the literature will determine whether such measures now exist, or whether a methodology need be developed to produce it.

Already expressed is a desire for some measure which would have many of the qualities of the dollar, such as interval scale of measurement and a value established in a marketplace of goods and services. A theoretical solution to this problem would be to turn out different types of school graduates into the social marketplace to determine what price will be paid for each one. From this theoretical background there has emerged one widely used measure of educational benefit and, in terms of its usefulness for research, it has been highly successful. It is a consideration of schooling's impact on a student's wage-earning ability.

Human Capital. As it turns out, the researchers who follow this reasoning have used a measure which not only has many qualities of the dollar, but is, in fact, the dollar itself. The investigators, many of whom are economists, have restricted their research to the economic effects of education and have adopted personal income as the criterion variable. This is the basic idea behind the concept of "human capital," a notion that society rewards its better-educated and better-skilled members with increased economic benefits." Becker (1964) points to the

> tremendous amount of circumstantial evidence testifying to the economic importance of human capital, especially in education. Probably the most impressive place of evidence is that more highly educated and skilled persons almost always tend to earn more than others (p. 2).

Within the human capital perspective, people invest in further education and skill improvement to increase their economic welfare. According to Becker and Chiswick (1966),

> Each person is assumed in effect to maximize his economic welfare by investing an appropriate amount in human capital, and the distribution of earnings is determined by the distribution of investments and the rates of return (p. 368).

This is the underlying notion of an earnings function, that earnings are dependent on schooling and education. A basic statement of the function can be found in Hanoch (1967):

> The flow of net earnings (y) that an individual expects to receive at a given time is assumed to be a function of his age (t), his schooling level (s), and various additional factors, lumped together. . . in a vector of variables (z) (p. 310).

At first, schooling level was measured solely in terms of years of schooling since those were the data most easily available and since they did show a significant relationship to earnings.

> Limitations of the data available have reduced the scope of the empirical analysis to investment in formal education as measurable by years of schooling. Evidence from states and regions within the United States and from several countries indicates that schooling usually explains a not negligible part of the inequality in earnings within a geographical area and a much larger part of the differences in inequality between areas. These and other findings are generally quite consistent with the implications of the theory (Becker and Chiswick, 1966, p. 368).

However, further research has refined some of the measures of schooling level so that it became "important to know how much people learned while in school" (Hansen <u>et al.</u>, 1970, p. 409). These authors say that it is

> well known that the level of educational attainment (LEA) is positively correlated with level of earnings. . .It is erroneous to attribute all of the observed differentials in earnings associated with

> education to differences in years of schooling (p. 409).
>
> The significance of years of schooling per se diminishes with the introduction of a measure of learning. . . .A more fundamental approach. . .would examine the factors affecting learning. . ., i.e., ability, motivation, home environment and quality of schooling, to determine their individual and collective roles (p. 417).

All of this combines to help form the notion that schools should be the instrument to relieve the economic inequalities in the nation. This idea has been exported to other countries also, where the building of a school is seen as another step toward economic development and social equality. In this country the idea has been tested in a number of major studies, culminating in the highly publicized and controversial work on Inequality by Jencks et al. (1972).

In the popular mind this work has labeled schools as failures:

> The public schools are failing dismally in what has always been regarded as one of their primary tasks--in Horace Mann's phrase, to be "the great equalizer of the conditions of men" facilitating the movement of the poor and disadvantaged into the mainstream of American economic and social life" (Silberman, 1970, p. 53).

Limitations of Human Capital. The human capital line of reasoning has proved very successful in methodological terms. The criterion has been an objectively measured, universally meaningful variable--dollars and cents. Criticisms about methodology have been leveled at many studies. Levin (1972), for example, strongly faults Jencks et al. (1972) for their computation of variables and dismissal of certain effects as "negligible."

Yet more important than the computation of variables are the philosophic limitations of the method itself. There is, first of all, the possibility that the economic effects of education are only part of its total aim, and possibly only a minor part. Becker (1964) had some restrained but harsh words to say about such critics:

> Passions are easily aroused on this subject and even people who are generally in favor of education, medical care, and the like often dislike the phrase "human capital" and still more any emphasis on its economic effects. They are often the people who launch the most bitter attacks on research on human capital, partly because they fear that emphasis of the 'material' effects of human capital detracts from its 'cultural' effects, which to them are more important (p. 2).

Thus, there are probably many people who would not want to see courses in the fine arts offered only to budding artists or physical education courses only to prospective professional athletes. The effect of the human capital view is to reduce cultural values to a question of economic profit.

Another limitation, pointed out by Schultz (1967), is that there is "all too little evidence on the relationship between social and private rates of return" (p. 308). When human resources are considered in economic terms only, they are not measured in terms of their value to society in general but only in terms of their value to individuals. It is, in fact, impossible to place all members of society in the highest paying positions.

<u>Alternative Goals</u>. Not all educators are ready to accept the economic analysis of human capital as the prime measure of the benefits to be derived from education. Although easily measured, it is not as universally accepted as an overall measure of educational benefit should be. Jencks himself admits in an interview (Robinson, 1972) that there are other important outcomes of schooling beside income.

> . . .moral development. . .seems to me a rather important issue, to say the least. And we don't know anything about whether schools have any effect on this. . . .There's a long list of things we haven't measured. So we can't say what kind of impact schooling makes in many areas of life (p. 257).

The famous Coleman Report (Coleman <u>et al</u>., 1966; Mosteller and Moynihan, 1972a) assessed educational opportunity in terms of academic achievement, the results of tests of academic achievement. This, again, is only one kind of impact schooling makes. As Mosteller and Moynihan (1972b) point out:

> academic achievement is but one output, and. . . schooling is expected to produce many others. Retention rates, proportion going to college, income and occupation of graduates, even happiness, are a few of many outputs that might be measured (p. 6).
>
> The three R's do not end the list of things we want schools to give our children. (Surely, for example, we want children to emerge from their school years with a sense of social dignity and place, and with a commitment to their community. And we want them to learn how to work and live with others) (p. 27).

At this point a synopsis will be presented of some of the purposes of education expressed by educational philosophers, thinkers, and critics. One cannot guarantee that all voices will be presented. It is also impossible to point out all the deficiencies in each case which make it impractical to apply operations research techniques to education. It should be sufficient, however, to see that there is general disagreement among the many purposes. Some of these are disagreements of emphasis, but many are fundamental differences about the nature of education itself.

For what an individual values in education may depend to a great extent on both his experiences with education and his present relationship to it. For example, Feyereisen et al. (1970) feel that "a statement of objectives prepared by academicians will usually be subject-matter oriented" (p. 138). It may be possible to see such relationships in the objectives presented here.

Specified Objectives. Since our discussion is particularly about the ends of education in the United States, which may be distinct from other cultures, this survey begins with the great American philosopher, John Dewey, who emphasized in Moral Principles of Education, that "the moral purpose is universal and dominant in all instruction--whatsoever the topic" (quoted in Silberman, 1970, p. 9).

Not all people would agree. For Ebel (1972), moral education deserves high priority, "but it does not deserve the highest priority. That spot must be reserved for the cultivation of cognitive competence" (p. 5). He feels that education must move away from trying to solve the problems of young people to imparting useful knowledge.

> Schools have been far too willing to accept responsibility for solving all the problems of young people, for meeting all of their immediate needs. That schools have failed to discharge these obligations successfully is clearly evident.
>
> Schools are for learning. They should bend most of their efforts to the facilitation of learning.
>
> The kind of learning on which schools should concentrate most of their efforts is cognitive competence, the command of useful knowledge (p. 7).

Rogers (1969) is in general agreement with the term "facilitation of learning," but makes a definite qualification:

> Teaching as imparting of knowledge makes sense in an unchanging environment. We face a new situation. The goal of education is the facilitation of change and learning. The educated man is he who has learned how to learn; how to adapt (p. 104).

Whitehead (1929) protests against dead knowledge and inert ideas. "Except at rare intervals of intellectual ferment, education in the past has been radically infected with inert ideas" (p. 15). Education should not be the imparting of knowledge as such, but of the ability to use it. "Education is the acquisition of the art of the utilization of knowledge" (p. 18, quoted in Stone and Schneider, 1965, p. 30).

In this regard he is in some agreement with individuals and organizations who think that education should have a great deal to do with the ability to think. According to the Educational Policies Commission of the National Education Association (1961), "the purpose which runs through and strengthens all other educational purposes--the common thread of education--is the development of the ability to think" (p. 12). And Ennis (1962, p. 82) writes that "the teaching of critical thinking. . . is thought to be one of the educator's main jobs."

But for others, such as Pritzkau (1970), education is a form of general self-realization and transformation. He speaks of education for the "authentic," calling for the development of authentic people through some encounter, facilitated by the teacher, with the authentic meanings of people and things.

And for Fréire (1970), from his experience in northeastern Brazil, its poverty and social ferment, education should be an instrument for the transformation of society, for social and cultural revolution. Education should be for the liberation and humanization of people, for training their consciences to see the evils and oppression in society and to want to do something about it.

Finally, education is seen as an instrument for teaching people how to live, which may be a return to Dewey's ideas. According to Howe (1963):

> The purpose of education is not simply the transmission of information but the bringing into being of persons of responsibility and integrity who, in a world of persons and things, can be instruments of love which, in the words of Paul Tillich, 'moves everything toward everything else that is.' The educator's responsibility is to recognize that each single unique person is the bearer of a special task of being which can be fulfilled through him and him alone (p. 70).

The above selection of views on the purpose of education shows that the unity of thought which would be desirous on a topic of such vital concern to educators is certainly lacking. A further deficiency is the immeasurability of many of the goals expressed. And thirdly, some statements lack specificity of the purpose, i.e., what types of activities are expected to achieve the goal.

There have been some efforts to develop a systematic analysis of education and its purposes. One of them originates from the Commission on the Reorganization of Secondary Education in 1918 which decided on seven cardinal objectives: health; command of fundamental processes; worthy home membership; vocational competence; effective citizenship; worthy use of leisure time; and ethical character (Stone and Schneider, 1965, pp. 27-28). Another list of purposes was developed in 1939 by the Educational Policies Commission of the NEA, mentioned above. It summarizes objectives under four main headings: the objectives of self-realization; the objectives of human relationships; the objectives of economic efficiency; and the objectives of civic responsibility (EPA, 1961, pp. 31-33).

The Question of Value. Yet even when one has systematized and specified all the objectives of education, there remains the question of how valuable each objective is in itself and in relation to other objectives. This brings us to the study of values, known as axiology.

> As a rule, a study of values centers around three main issues: (1) whether values are objectives or subjective, that is, impersonal or personal; (2) whether they are changing or constant; and (3) whether there are hierarchies of value.
>
> Objective values exist regardless of man's personal feelings and desires. . . .Subjective values, on the other hand, are relative to personal desire, which confers value on the object considered. . . . Absolute values are constant. . .and are a reflection of reality itself. . . .Changing values, on the contrary, are likely to be responses to man's immediate needs (Kneller, 1964, pp. 13-14).

Brubacher (1962) makes this comment about values:

> We must distinguish what in fact is held valuable from what ought to be held valuable. While we know pretty well how to settle disagreements about facts of value--for they do not differ greatly from other kinds of facts--we find it much more difficult to settle disputes about what ought to be the fact or norm of values (p. 110).

The myriad goals proposed for education give a hint of the difficulty of settling disputes in the context of what ought to be the norm of values. It makes us despair of finding, through the exercise of intellect alone, some "correct" objective function which would illumine the intrinsic merits of all educational goals. It may not be desirable or even necessary to attempt that.

We can call education "the process whereby persons intentionally guide the development of persons" (Phenix, 1958, p. 13). The general purpose of this development is to prepare adults who can function adequately in the society. According to McCleary and Hencley (1965), this preparation, to exist as formal education, may convey

> at least a modicum of each of the following elements: (1) a development of the individual--personal, emotional, and aesthetic aspects of life; (2) modes of thought and techniques of inquiry--disciplined intellect; (3) transmission of the cultural heritage--civic and moral values; and (4) promotion of vital societal needs contributing to economic, social, and political well-being--the realm of technics (p. 195).

Yet the relative value of each of the elements, and of their constituent parts, will depend very much on the nature of the society into which one is to be assimilated.

> The ultimate aim of education, therefore, takes its form from whatever the social stereotype may be. Naturally this stereotype will vary depending on whether society is organized along democratic, fascistic, or communistic lines (Brubacher, 1962, p. 115).

"Therefore," according to Feyereisen et al. (1970, p. 138), "we must start be defining the needs of the individual, the nature and needs of society, and a system of values from which we can derive the objectives of the curriculum."

Most importantly and finally, it is from the needs of society and its system of values that any justifiable measurable objective should be developed.

> In its very essence, the school is an institution established by society for the purpose of preparing the young to participate in that society. Like the family, the church, or the government, the school is a social institution whose fundamental character is determined by the society it serves (Stanley, et al., 1957, p. 2).

Although the instrumentation for accomplishing this governance are not developed for maximum efficiency, society does rule the schools.

> When the error between what the parents want their children to learn in school and what the children actually receive becomes too great, the school board feels this force of public opinion and changes the school policies. In such a manner, the schools are regulated by the mores of the community (Smith, 1958, p. 3).

Solution. A solution would lie in developing objectives out of a study of the values of society as a whole. Jensen (1971) investigated educational values. However, his study dealt with beliefs about the process and methods of education, and not about the products or objectives. More needed is a measure of the judged value of educational objectives.

Each objective of the curriculum should be measured against society's values. Some overall scale should be developed which incorporates each aspect with its individual relative worth. Then this scale can be used to measure the accomplishment of every schooled member of the society. It can be calculated periodically to adjust appropriately to any changes in the society. It can be used to measure the effectiveness of the school by its ability to have its people measure up to the standard. And the procedure of comparing objectives to values can be applied at all levels of the curriculum, if it is found to be organized hierarchically.

These are, in fact, the attributes which Page (1972) and Page and Breen (1972) propose for the acceptable solution.

Churchman et al. (1957) have suggested a method of assigning relative weights to alternative objectives, which is to present interested people with possible outcomes, or sets of possible outcomes, and have them select the most desirable one from each set. The most important outcome is arbitrarily assigned a value of 1.00 and other outcomes are scaled relative to it. Combinations of the outcomes are then compared, on the assumption that values are additive, and the relative weights refined. With this method it would be possible to scale all the objectives along one continuum.

In attempting to measure any value-system it might be more reliable to record people's responses in life-like situations than to elicit from them self-conscious statements of their values. As Raths et al. (1966) writes:

> . . .because values are a part of living, they operate in very complex circumstances and usually involve more than simple extremes of right and wrong, good and bad, true or false. The conditions under which behavior is guided, in which values work, typically involve conflicting demands, a weighing and a balancing, and finally, an action that reflects a multitude of forces. Thus values seldom function in a pure and abstract form. Complicated judgments are involved and what is really valued is reflected in the outcome of life as it is finally lived (p. 27).

In the context of the importance of measuring values in lifelike situations, Page (1972) suggests that a measurable overall educational objective be developed by an empirical strategy. He originally proposed presenting to a random sample of people from the general population, or from any other chosen set of "judges," a number of hypothetical students. The attributes of these students would be presented in a profile format as socres on various "academic" traits. These traits can be structurally linked to all the various aspects of the high school curriculum. Page proposed calling this measure the "bentee," for benefit T-score, a normalized measure of the benefit derived from education.

Method. Accordingly, one method of this study was to abstract from the lifelike situation those elements of possible student-graduates, outcomes of the system, which are felt to relate to a schooling experience. A group of "judges," will be asked to evaluate these outcomes. The solution to the problem of defining and weighting educational objectives will be left to an analysis of the judgments of the evaluators about the hypothetical students. Possible limitations to this solution lie in the quality of the abstraction of student characteristics and in the quality of the selection of judges.

Possible outcomes of the school system, which will be students with various trait profiles, will be presented to a sample of randomly selected people. These people, called evaluators, are asked to rate the students on a one-dimensional continuum according to how "well-educated" or how "well-prepared" the students seem to be. The value or weight assigned by the evaluator to each trait is the correlation between the trait scores and his overall evaluations.

The study is concerned with any systematic relationships between the trait scores and the evaluators' judgments, whether these relationships differ for different identifiable groups, and whether the relative importance of traits are related to judge variables.

A more direct value measure is also obtained from the judges--a weighting of the traits by direct assignment of tokens. This second weighting of the traits may be compared with the trait-value measures derived from the analysis of the profile evaluations.

C. Hypotheses

Hypotheses. The major hypotheses investigated, as expressed in null form, are the following:

1) that there are no differences in the relative importance assigned to the traits as indicated in the evaluators' ratings of the profiles, i.e., that the correlations between trait scores and the ratings are not significantly different from zero:

$$r_{1r} = r_{2r} = \ldots = r_{7r} = 0.$$

2) that there are no differences in the numbers of tokens assigned to each trait by the evaluators;

$$nt_1 = nt_2 = \ldots = nt_7 = 50/7.$$

3) that there is no relation between the weightings obtained from the correlation study of the profiles and the assignment of the tokens, that the rating of profiles is not a reliable measure of educational values;

(for each judge and group of judges)

$$r_{tr} = 0.$$

4) that the value profiles obtained for different strata and different identifiable groups are not significantly different;

$$r_{pj} = r_{lj} \qquad p = \text{professionals}, \quad l = \text{lay}$$

$$nt_{pj} = nt_{lj} \qquad j = 1 \text{ to } n \text{ of traits (7), and}$$

5) that there is no significant relationship between any personal variables and the relative weightings assigned to each trait.

It was predicted that the first three null hypotheses would be rejected at a suitable level of significance ($p < .05$). The last two hypotheses may be accepted in most instances. Appropriate statistics, correlation-procedures and analyses of variance, were applied to the data to test these hypotheses.

Finally, a factor analysis is performed on the weights derived from both the direct (token) method and latent (profile-rating) method. This analysis attempts to ascertain whether in the view of the "judges" the traits seem to fall into groups, whether the "judges" tended to treat traits in subgroups and evaluate "students" on these subgroups rather than on the individual traits.

Summary. A number of proposed objectives and traditional criteria was examined. None was found acceptable as a suitable, comprehensive objective for operations research and systems analysis. The objectives are too narrow in scope, too devoid of specification, or too difficult to measure.

The concept of "human capital" was discarded as a valid overall criterion for educational effects, because it too narrowly limits the work of the school. The same is true of one-subject academic achievement tests.

Many proposed goals for education in the literature are found to disagree with one another, to be not measurable, and to consist of diversified objectives whose relationships are not defined, or which reflect only personal and individualistic views.

The lack of a suitable objective emphasizes the need for measuring the cultural values of society toward the traits of possible outputs of the school system, since it is to society and its values that schools must justify their work. The measurement of values should occur in a lifelike situation so that the actual values which govern society's everyday actions will emerge.

A method is outlined defining an overall criterion for education, through the collection of social values. Hypotheses have been developed about the results of testing this methodology. Subsequent chapters will outline the characteristics of a "tree" of values, will specify procedures of data collection, will analyze the results of such data, and will suggest further desirable research.

CHAPTER 2

CHARACTERISTICS OF THE BENTEE

The first chapter described the problem of educational objectives in terms of philosophical analysis, and proposed the "bentee" as an overall measure of effectiveness. The present chapter addresses itself to the psychometric history of such overall measures, and to some mathematical properties of the proposed tree of values.

A. Historical Perspective of the Bentee

This is not the first time, of course, that someone has suggested some sort of "overall" index to summarize a profile of scores (for an early scale of "attainment" see Greene et al., c. 1933). An "EQ" was created long ago by analogy with the IQ: by dividing the "educational age," however, computed, by the chronological age. And the "AQ" was invented as an "accomplishment quotient," computed by dividing the EQ by the IQ, and thus, in theory, usable to identify over- and under-achievers (Hilden and Skeels, 1935). The AQ, however, suffered from serious measurement shortcomings, quite apart from the scaling feasibility of the EQ (Cureton, 1937; Flanagan, 1951; Angoff, 1971).

In more recent decades, there have been corporate efforts to resolve achievement profiles into some acceptable summary scores, in testing programs of major companies. The National Merit Scholarship Qualifying Test, in order to make possible a selection among students, resolves a profile into a "total" score. The Metropolitan Achievement Tests summarize submeasurements of usage, punctuation, and capitalization, into a "total" score for language. The Secondary School Admission Tests assemble, from part-scores, an "ability total" and a "reading total." The National Educational Development Tests summarize five sub-scores into a "total." The California Achievement Tests advertise a reading total, an arithmetic total, a language total, and just a "total." And the American College Testing Program uses five content fields to generate a "composite." The weighting rules by which such composite summaries are developed, however, are quite capricious and obscure.

Seen analytically, all test scores are somewhat mysterious or insecure in their weighting rules. The most professional and expert test-development techniques now operating, may be resolved into quite arbitrary individual decisions concerning item-inclusion and hence, the implicit weighting of sub-traits into larger traits. The standard quality control techniques in which the profession puts such faith (such as indices of difficulty, discrimination, predictive validity) still do not remove the arbitrary micro-trait weighting built into achievement test scores at almost every level. For they do not explicitly come to grips with the values of society.

Such weighting as has been done depends ordinarily on the implicit judgments made by a small number of involved professionals, on the availability of items, on estimates of predictive validity, on the problems of pleasing prospective customers, etc. A distinguishing feature of the bentee is that weighting is performed by defined, analyzable, and selected populations of experts and/or laymen whose values may be considered by definition appropriate within a given society.

The more technical efforts to assess values have usually concentrated on individuals, and the maximizing of individual satisfactions measured in "utiles" or similar scales (for ways of using values in formal decision-analysis, see Raiffa, 1968; and for theoretical discussion of the individual/group problem see Churchman, 1961, ch. 12). Yet in Western education, we usually recognize that the larger pedagogical values or objectives, whether verbal or behavioral, whether norm-referenced or criterion-referenced, should originate in the larger society, or should at the very least operate with that society's consent. But the ways of translating societal values into educaional actions have not been technically defined.[1] The bentee therefore represents an effort toward such technical definition and translation.

B. The Bentee and Behavioral Objectives

The movement in favor of "behavioral objectives" seems virtually to have monopolized the scientific study of educational punpose. Yet beginning with such detailed objectives may lead to overwhelming difficulties; such a "bottom-up" strategy appears extremely cumbersome to apply, generating huge unmanageable lists of behaviors, with no clearly understood techniques for weighting them or eliminating them from the lists. On the other hand, practicing educators have usually established arbitrary, large, required blocks of study (such as "three years of high school math, one hour each day"), which seem of doubtful authority and genealogy. Also, such blocks do not lead to any overall measure of educational benefit, nor contain any clear technique for evaluation of outcome.

Consequently educational research, development, and evaluation have repeatedly run aground in this "fundamental confusion" of undefined goals, explored in the first chapter. In the present confusion, there is no accepted technical procedure for trading off, for example, a gain of math against a loss in English; of balancing the claims of one group against the needs of the majority; of evaluating new trends setters who claim one dimension of growth to be more important than others. This fundamental confusion, furthermore, makes education inaccessible to management-science techniques, which require "measurable values . . . that unequivocally reflect the future well-being of the organization" (Wagner, 1969, p. 5). For such a "measurable value" would require that any benefits be appraised along the same dimension, with the same scale, the same unit. In economics,

---

[1]The convergence of group opinions toward a statistical "consensus" has been studied in earlier experiments in social conformity, and more recently through the "Delphi" technique (e.g., Dalkey, 1969). Such exploration of belief shift under group influence may eventually have relevance for understanding the bentee or similar applied techniques.

we have such a unit, in the dollar, franc, or similar monitary measure. In education, however, as we have seen, no such measure seems to exist.

For these reasons, as already outlined, such a measure of "general educational advancement" is proposed, to be called the "bentee": a coined term referring to a "benefit T-score" for education, that is, a normalized, equal-interval scale, adjusted to a norm of some comparison group (such as high-school seniors, or any other reasonable population). It would have a mean of 50, and a standard deviation of 10, and therefore a usual range from around 20 to around 80. A bentee score, then, is an ideal, overall appraisal of an individual, or of some sub-group, compared with that norm population. A person who received a bentee of 70 would be considered "very well-educated" or "-prepared". Or a group of students with a bentee score of 40 would be, on the average, one standard deviation below the larger norm population in "education" or in "preparation".

C. The Tree of Values

Before we consider ways to establish such a measure, let us consider the general shape of a top-down value tree. Figure 1 contains a hypothetical tree, extending from the highest value-space almost to the specific

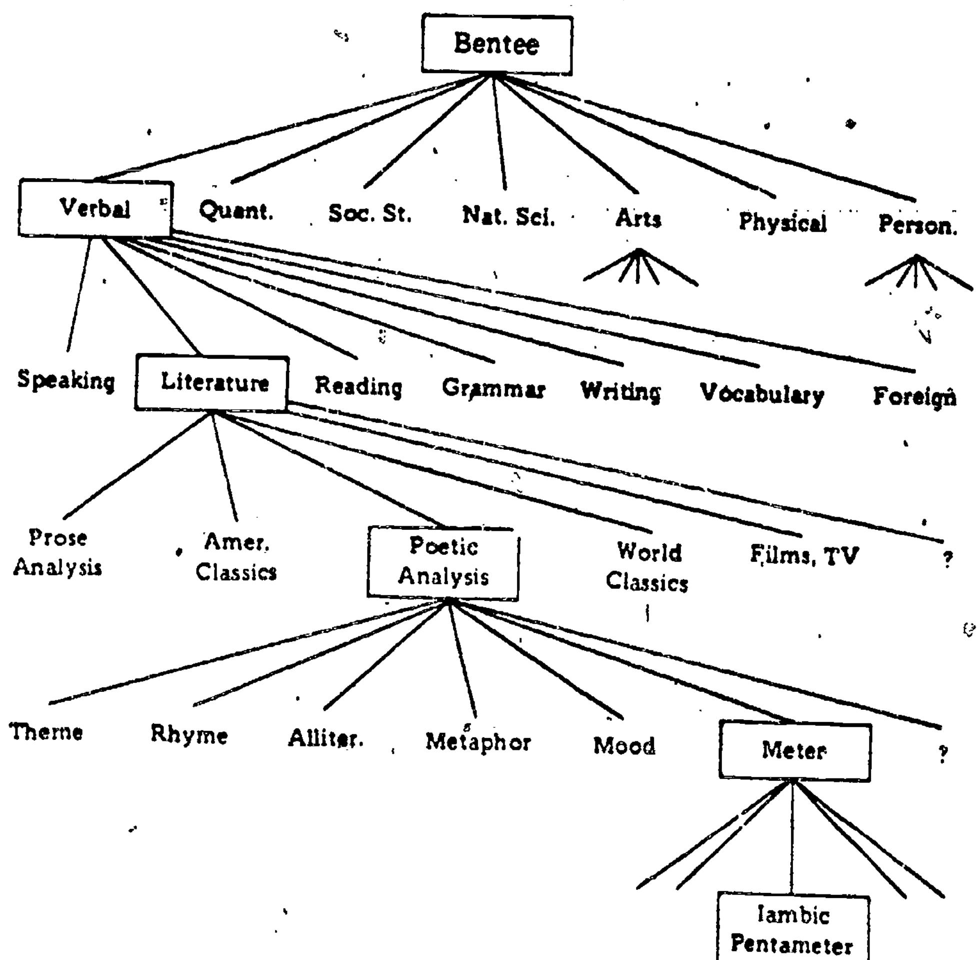


Figure 1. The Possible Recursive Feature of the Bentee Technique. (As analysis moves from the general to the specific, a shift is made from societal to expert opinion, and from value-space to test-space.)

item-space. We begin by arbitrarily declaring the bentee as the overall unit of value, worth the total unity of value, or 1.00. Next, we declare that seven "traits" may be defined such that they include all important educational values. In this illustration, these seven traits are declared to be VERBAL, QUANTITATIVE, SOCIAL STUDIES, NATURAL SCIENCES, ARTS, PHYSICAL and PERSONALITY. (See descriptions of traits in Appendix). Each of these terms becomes a node in the tree, and each node has its own branches, leading to other nodes. These nodes are here presented arbitrarily, but they might be matters for articulation of expert panels, regarding the number of branches beneath each node, the terms which make up the subdivision into branches, and the definition for each node.

In Figure 1, we observe one possible division of the BENTEE, then a possible subdivision of VERBAL into a number of plausible branches, then a possible further subidivision of LITERATURE, then of POETIC ANALYSIS, then of METER, and finally to IAMBIC PENTAMETER, where we are obviously close to the content of a specific test item. Thus we observe that not many steps are required (though the steps here may be altered with exploration) to move from the highest value-space of the judges (who represent "society") to the lowest and most detailed item-space or behavior-space. It is apparent, too, that this tree is "recursive": that is, the same strategy may be applied at each sub-node, without any major philosophical shift, in moving from such value-space to test-space. Furthermore, there is provision for the functioning of different sorts of "judges". At the top level, the judges may be citizens, students, parents, board members, or elected officials. At the lower levels, the judges may be (and indeed would almost have to be) psychological or subject-matter experts who thoroughly understood the details and their relative importance within the topics considered. Thus the bentee provides for a reasonable movement from democratic principles to technical expertise.[2]

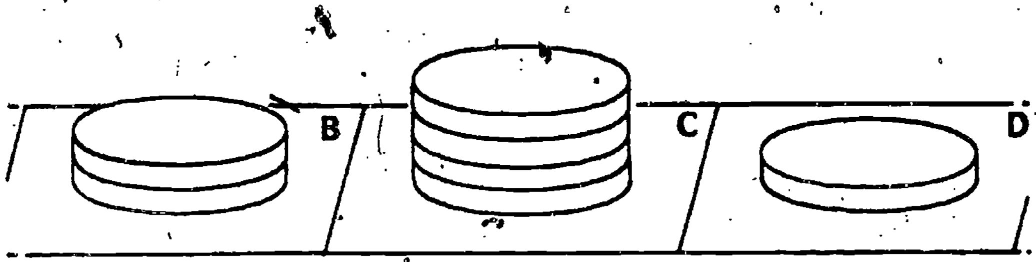
D. Two Techniques for Rendering Judgment

There may be numerous ways of establishing proper weightings for the various nodes. One may be called the "token strategy", and the other the "correlational strategy." In both strategies, one begins by identifying a correct selection of "judges", and by presenting to these judges a description of the sub-traits which make up the branches from a particular node of interest. For example, in Figure 1, we might have brief descriptions of the sub-fields SPEAKING, LITERATURE, READING, GRAMMAR, etc. Then we may ask our judges (acting independently of one another) to weight these branches by one of these two techniques:

Token strategy. Figure 2 suggests what the "token" strategy amounts to: A judge is given a stack of tokens, or poker chips, all of equal value.

[2] It is interesting that use of the tree does not depend on "democratic principles": the top-most allocation (or any subdivision) could be performed by any authorized person or agency. The bentee appears, rather, a technical device for gaining control over the process of education and like most technical procedures, is in itself quite value-free.

Each judge is given a stack of tokens, and asked to "spend" them according to the importance or value of each topic. In this example, judge J has awarded 2 to B, 4 to C, and 1 to D.



For each judge the "value" of each topic is the number of tokens given, divided by the total N of tokens.

Figure 2. The direct, or token, strategy for weighting sub-areas in importance.

Perhaps 100 such chips might be used. Then the judge is asked to "spend" the chips according to his appraised values of the subtraits. If 100 chips are so spent, then the number of chips for each branch indicates the proportion of importance given to that branch (within the node-family). Thus, if 20 chips are placed on LITERATURE, this would indicate that .20 of the value of VERBAL belongs (in the judge's opinion) to literature study. And within each level of the tree, each node has branches which must sum to 1.00. The scale of such weightings is thus considered to be a ratio scale; that is, 30 chips would indicate "twice as much" importance as 15 chips. Zero chips would indicate no importance at all for a particular branch.

Correlational strategy. As Figure 2 suggested the judgmental situation for the token strategy, so Figure 3 suggests the stimulus situation for the correlational strategy. The instructions accompanying such bar graphs may be as follows:

> When we say a person is "well educated" or "well prepared," what do we mean" There are many wasy of making such a judgment, depending on what we regard as important. This present study is an effort to define what is meant by such terms. You are asked to help decide, using as a guide your own preferences and beliefs.
>
> You are asked to judge 25 boys and 25 girls, to rate them for how "well prepared" or "well educated" they are. All you will know about them is contained in a "profile," such as the one here.
>
> You are given this profile of the boy. You are asked to make your own overall evaluation of his preparation and education. These traits are in a chance order (different for every judge like yourself). It is up to you to decide which ones are more important for you. (Page, 1972, pp. 41-42. For experimental instructions, see Appendix E.)

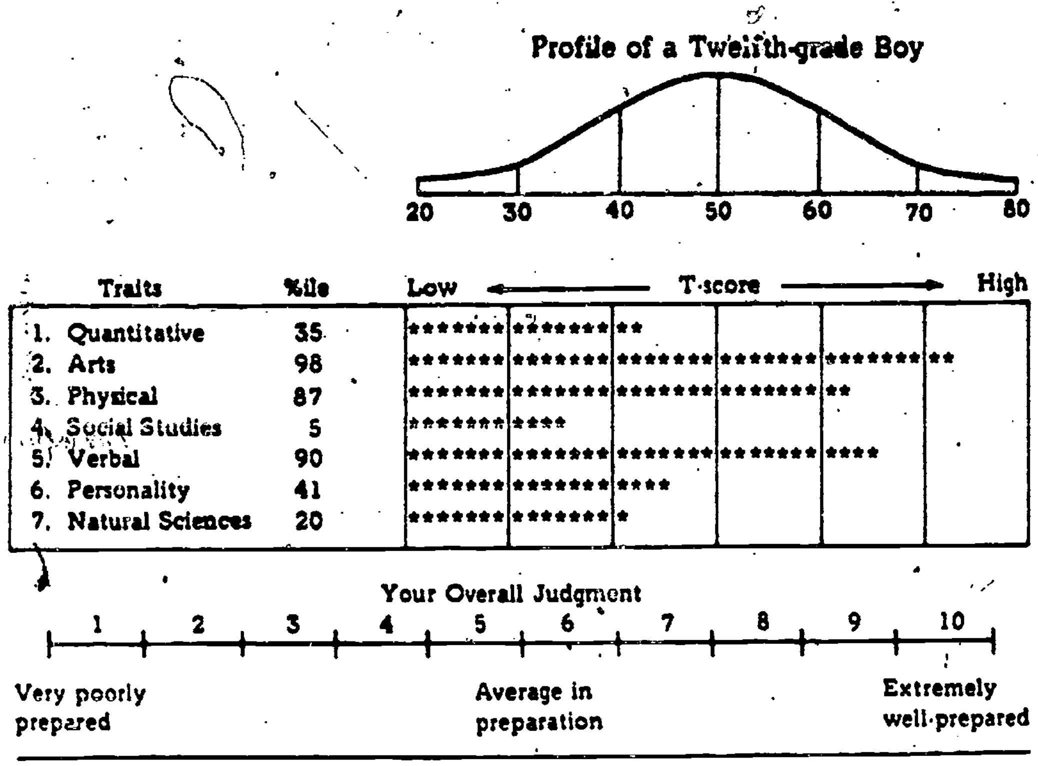


Figure 3. Profile of a Twelfth-Grade "Boy" which the Judge Evaluates by Checking a Spot along the Line of "Overall Judgment".

And these judges, like the token judges, are given descriptions of the traits, as a basis of making their evaluation.

The principal difference between the strategies is that, in the correlational method, the weighting of the traits is performed indirectly, through the individual weighting of individual cases. That is, the weighting may be inferred through discovering the correlations between each trait and the "overall judgments" made by a person or a group. For the present, the research generates these trait-profiles randomly by computer (the program written by Thomas Breen), and no inter-trait correlations are built into the profiles. The resulting matrix of intercorrelations would be similar to that of Figure 4.

An interesting and unusual property of such a matrix is that the correlation coefficients in the last column are also the beta coefficients for the multiple regression R, for the best prediction of the "overall judgment" for a profile of traits. That is,

$$r_{ic} = \beta_i \qquad (1)$$

where i represents any trait, and c represents the overall criterion or judgment, and where β_i represents the ith beta weight for optimally predicting c.

Traits	1	2	3	4	5	6	7	Judged Overall
1. Verbal		0	0	0	0	0	0	.30
2. Quantitative			0	0	0	0	0	.25
3. Social Studies				0	0	0	0	.22
4. Natural Sciences					0	0	0	.20
5. Arts						0	0	.15
6. Physical							0	.10
7. Personality								.10

Figure 4. A Possible Matrix Generated by the Bentee Judgments of One or More Persons, Using the Correlational Strategy. (The trait intercorrelations would be approximately zero, reflecting the way they were generated. The correlations between traits and overall ratings are hypothetical, for a possible individual or group.)

(This is true only when the trait intercorrelations = 0.)[3] By the same reasoning, the multiple regression R for predicting c is

$$R = \sqrt{\sum_{i=1}^{n} \beta_i} = \sqrt{\sum_{i=1}^{n} r_{ic}} \qquad (2)$$

given n traits.

Since these correlations are also beta weights, they may be assumed of ratio scale, so far as judged "importance" is concerned. That is, when one has a correlation of .30, this correlation, like a beta weight, may validly be considered as "twice as important" as a correlation of .15.[4] In this sense, such weightings by correlation are similar to those of the token strategy. In order to make the two strategies still more comparable, a column of correlations may be normalized or "proportionalized" to sum to 1.00. That is, each weighting may be redefined as

[3] This equivalence will become intuitively clear when it is remembered that each beta coefficient represents the correlation of the trait with the criterion minus various products of the intercorrelations. Since the intercorrelations are assumed zero, all of these products are also zero; and the correlation is left unqualified.

[4] Some may be troubled by the fact that the amount of variance accounted for is not linearly represented by the correlation coefficient. Nevertheless, multiple-regression theory shows the beta weightings to be proportional to their magnitude of the coefficients.

$$V_i = \frac{r_{ic}}{\sum_{j=1}^{n} r_{jc}} \qquad (3)$$

In this way, each V_i is the proportion of value for the ith trait, within a given node-family.

E. Technology of the Value Tree

Of course, to establish the values of society, or of a profession, single judgments would not suffice; thus, one would average such proportions across a sample of appropriate judges. And for each node in the value-tree, one would establish proportions of value for the group which would also sum to 1.00 within each node-family. How would one thus appraise a student (or a group of students), given V_i and given a profile of his traits? At the highest level, the appraisal of the jth student would be worth:

$$\text{Bentee}_j = T\left(\sum_{i=1}^{n} v_i m_{ij}\right), \qquad (4)$$

where V_i is the societal value of the ith trait, where m_{ij} is the measure (in standard scores) of the ith trait for the jth student, and where T represents the transformation of the obtained sum of weighted measures to the T-distribution (the calculation of which would depend on the parameters of the norm-group to which the student belongs). And just the same techniques could be applied to each node in the value-tree.

The appearance of this tree, where the values of the branches sum to 1.00 within each node-family, suggests a fruitful analogy: the probability tree for independent events. Thus, in the value tree, it is seen that the value of each node is the product of all the lineal values above it. That is, if LITERATURE is considered worth .20 of VERBAL, and VERBAL is worth .30 of the overall BENTEE, then LITERATURE is worth (.2 x .3 = .06), or just 6% of the overall BENTEE. More generally,

$$V_t = \prod_{k=1}^{L} v_k \qquad (5)$$

where V_t is the overall value of the tth trait, where V represents the value of each direct parent-node between t and the BENTEE (or other selected parent-node below the highest level), and L is the number of such levels or generations between t and the BENTEE. Thus a program is indicated for evaluating any node at any depth in the tree.

Overlapping branches. Yet reflection upon the tree of Figure 1 will lead us to another consideration: the overlap among the various branches will, eventually, become substantial.[5] For example, if ARTS is defined as covering esthetic training, then we may expect that POETRY ANALYSIS may be reached through ARTS, as well as through VERBAL. The weightings, of course, may be justifiably different within one value-family, as compared with another. And whatever the value of behavior b turns out to be in one branching, it may be added to the values of that same behavior b wherever it occurs in the value tree. Thus,

$$V_b = \sum_{h=1}^{H} v_{bh} , \qquad (6)$$

where V_b represents the overall value of behavior b (such as recognition of iambic pentameter); where h represents a specific hierarchy within the value tree, wherein behavior b is found; where H represents the total frequency of occurrences of behavior b; and where v_{bh} represents the judged value of behavior b in the branching h. Thus if the recognition of various poetic meters is worth .0002 in one setting, and .0001 in another, and those two occurrences are the only ones, its total value would be .0003.

Dimensions of Concern. Still another procedural rule of thumb may be suggested by independent probabilities. Suppose we have two dimensions of concern, such as shown in Figure 5, suggesting four branches of social

Taxonomy Levels		Subject-Matter Content Areas 1	2	3	4	
Fact	1	.30 X .10 = .03				.30
Concept	2					.30
Application	3					.40
		.10	.30	.40	.20	1.00

Figure 5. Arbitrary Assignment of Cell Value in Two-Dimensional Framework. $V_{ij} = v_{i}.v_{.j}$. (Where a cell's value is already estimated, that estimate may still be used, and the others assigned by the product method.)

[5] And the tree itself will of course grow to great size. If we consider each node to have 7 branches, and guess that perhaps 8 generations will be adequate to cover the evaluation tree, then we have 7^8, or approximately 4.5 million specific items or cells. But large numbers of these will be duplications, reducing greatly the number of unique items.

studies and three levels of taxonomy, for example, fact, concept, application. Then, we may weigh each cell under the assumption of row-and-column independence, by making the value of each cell the product of the row-value times the column-value. (This strategy was suggested by Dr. José Roig Ibáñez, Germán Espinosa M., and other Spanish educators at a seminar in Valencia in June, 1972.) Where there is a judged interaction of the two dimensions, the already estimated cells may be assigned their estimated weights in advance, and the remaining cells assigned by a system of row-column multiplication.

F. Management of the Value Tree

An earlier criticism was made of the unmanageable nature of the "bottom-up" strategy, which begins near the "behavioral objectives" at the bottom of the tree. Is the top-down strategy of the bentee equally susceptible to such criticism? Apparently not, for it offers procedures for diminishing whole branches of behavior very early in the analysis, and reducing them to trivial size. Such pruning of the tree is extremely important, for it lets the educational system concentrate on those branches where substantial values reside.

Also, as more is learned about the intercorrelations of sub-goals, a large amount of pruning may take place in the measurable behaviors required. This suggests that the bentee weighting may function smoothly within a "matrix-sampling" environment, wherein different students are measured with different items.

For example, suppose an educational system is beginning to construct a new evaluation program from its beginning, and wishes to allocate the efforts of scholars, psychologists, and item writers to obtain the most valuable information possible at minimum cost. Then the bentee seems to offer great utility, regardless of who the judges are, or how the system is administratively organized. That is, if the educational system should happen to be a very hierarchical one, with highly centralized authority, then perhaps the Minister of Education (or some similar official) could, through his staff, allocate values at the top levels, and draw upon experts or officials within various disciplines to allocate values within their own lower subdivisions. In other words, though the bentee recursion permits democratic sampling, it does not require it, and appears politically neutral. But it can balance claims of pressure groups of whatever persuasions.

One difficulty of any top-down strategy is this: that the values of level 2, for example, cannot be precisely appraised until one has sufficiently explored levels 3, 4, etc. For example, the fact that PERSONALITY contains morality, at some lower level, may not be sufficiently realized by the judge at the top level. If he knew this, he might modify his evaluation of PERSONALITY at the top. Such a problem suggests that analogy of the artist painting a picture: he usually allocates the major areas of his canvas, and often brushes in an approximate color wash, before he paints in the detail within the areas. In this way he maintains artistic control of his canvas. Yet the subsequent detail forces the artist to make minor alterations in his various areas, as he goes along. Similarly, the first BENTEE evaluation may not remain the permanent allocation of top-level value; modifications may be made as the lower levels are further defined, and as their values seem to increase or diminish. Yet the top-level allocation permits the system to get underway.

Another objection sometimes raised is that a top-down system may lead to a rigid Procrustean bed of evaluation, which would override differing individual talents, and differing curricular or group goals. This objection seems unwarranted. Indeed, value trees may be created for specific needs of subgroups, sub-disciplines, or sub-types within the system. A "bantee" value tree for the university engineering student would expectably differ from that for the general high-school student. Even for the individual, it is conceivable that one sort of useful bentee, for purposes of self-direction and personal counseling, might be the sum pf products of his trait measurements times his personal allocations of value. In this way, the traditional counseling goal of "self-actualization" might become a respectable, technical possibility.

Properly understood, the bentee tree of values appears very close to what practicing educators have been doing through their present practices, which involve legislative rulings, curriculum committees, department understandings, and teacher interpretations. What the bentee may do is to provide a technical rationalization and formal operating procedure for educational values. Thus it might transform the current top-down system, which has always seemed a natural one to educators and to society, from an intuitive and uninformed strategy to one which is explicit, and quantifiable, and related in clear ways to behavioral objectives. Thus education may capitalize on the great efficiencies which such quantification of purpose has provided for other disciplines.

Subsequent chapters will describe an investigation into the practicality of discovering such social values for education, across differing social and professional groups.

CHAPTER 3

EXPERIMENTAL PROCEDURES

In Chapter 2, the experimental hypotheses were outlined. The present chapter describes the sampling and procedures used for the test of the bensee methods. The goal was to explore the values of society, especially as reflected in two groups: educators and non-educators. Thus, it was thought desirable to stratify the sampling of the judges. In order to measure their reactions, it was necessary to write computer programs producing the stimuli and analyzing the data. This chapter will explain the problems of sampling encountered, the nature of the experimental treatments, the computer programs for producing the stimuli, and the techniques of data preparation.

Sampling. The original sample design called for 100 subjects of whom 50 were to be professional educators engaged in secondary education and 50 lay people who were not professionally connected with secondary education. The actual number of subjects recruited was 101 of whom 52 came from the professional and 49 from the lay group. Table 3 contains a list of the areas in which sample teachers worked.

Half of the sample (51 subjects) was selected from Connecticut, mainly from eastern Connecticut, from areas convenient to the University; while the other half (50 subjects) was from the greater Cleveland, Ohio area, where another researcher assisted with the data collection. Research of this kind often finds no significant effect due to region of the country. Major effects are due, rather, to social class-type variables. Sampling from two regions allowed us, however, to test this assumption, although it may be confounded with any possible effect due to researchers.

The sampling process was designed to take about a half-hour of each subject's time. It took from 20 to 40 minutes. From the lay group it took place in their homes or places of employment. For the professional group it took place in the schools where they worked, although several came to the researchers' offices for the session. Information was collected about the subjects' age, sex, race, education, and location (Cleveland or Connecticut). Table 1 contains a breakdown of the sample by occupation, sex, and location. All subjects were reimbursed for their time from funds awarded for that purpose by the United States Office of Education.

The high school teachers and administrators were contacted through the schools where they worked. Although schools were selected randomly, the selection of teachers was most often made by the principal's office and thus cannot be considered a random selection of teachers within the school. They are probably friends of the principal or teachers who answered a request for volunteers made at a staff meeting. The specialization of the teachers in the sample are listed in Table 2. The three teacher-educators were selected from among doctoral candidates completing their degrees at The University of Connecticut.

An attempt was made to recruit members of the lay sample by calling randomly selected telephone numbers. A standard format was designed for

TABLE 1

OCCUPATION, SEX, AND LOCATION OF SUBJECTS

			Cleve.	Conn.
Professional	Teachers	Male	13	15
		Female	8	5
	Administrators		5	8
	Teacher-Educators		-	3
Lay	Women		11	10
	Blue-Collar Men		6	7
	White-Collar Men		7	8

TABLE 2

SPECIALIZATIONS OF SAMPLED TEACHERS

Social Studies	9	Art	1
English	6	Bookkeeping	1
Mathematics	4	English & History	1
Biology	4	Clothing & Tailoring	1
Business Education	4	English & Spanish	1
History	3	French & Spanish	1
Physical Education	1	French	1
Science	1	Physics	1

this telephone conversation which, it was hoped, would insure the greatest amount of cooperation. However, this method proved to be a failure. Appendix A contains a copy of the ill-fated format for the telephone conversations.

The refusal rate for the first 14 calls was over 90 per cent (13 out of 14). There was a number of reasons for this. The first contact spoke no English. A number simply said they were not interested. One woman could not discuss participating in the study unless her husband was present. One prospective subject wanted to know exactly why he had been chosen as a subject. He did not believe the professed intent of the research and seemed to feel the researcher might be an agent of some investigative agency. At a couple of numbers there was no answer. When the researcher did visit the home of the one person who had agreed over the telephone to take part in the study, he found that the man had recruited his college-educated son to stand in for him because the son could do a "better job." The researcher in Cleveland had similar experiences.

Thus, the telephone approach was dropped in favor of one that could recruit more people. Contacts were made with some individuals through professional, social, and religious affiliations, or through their place of employment. People who were willing were recruited for the sample and were visited by the researcher at their homes, or at times, where they worked. Contacts were also made through some of these subjects with their friends, neighbors, and co-workers. In this way the sample was completed.

Since the sample was not selected randomly, it may be possible that the sampling process has in some way influenced the results of the study. If the study had been an attempt to construct the final measurable objective for education, the exact social values of the traits would have been of prime importance and the lack of a random sample most certainly damaging. But since the purpose of the study was more to test and develop a methodology, an appropriate sample was one that included an adequate number of people from professional and lay populations to extract reliable weightings of the traits. With these it would be possible to begin investigating differences in their educational values. This type of investigation is not invalidated by the selection of a non-representative sample, although the generalizability of the value measures may well be affected.

Profiles. Each of the subjects was asked to evaluate 50 hypothetical twelfth-grade students, 25 "boys" and 25 "girls," according to how "well-educated" or how "well-prepared" each one appeared to be. The definitions of the terms "well-educated" and "well-prepared," which were used synonymously, were not specified. Subjects were encouraged to rate students according to their own values. These hypothetical students were presented to subjects on profiles produced by a computed program. The format of the standard profile appears in Appendix B.

No attempt was made to produce realistic profiles. Although some techniques for generating multivariate data with desired realistic interrelationships have been suggested (Capra and Elster, 1971), it was considered more informative to generate student profiles with no correlations between the traits. If there were systematic relationships between the profile traits, it would be extremely difficult to estimate the unique importance of each trait in the formation of the subject's overall judgments.

Some evaluators, generally from among the more sophisticated professionals, did remark about the unexpected lack of relationship between the traits. They were told that for purposes of the study no attempt was made to produce realistic profiles.

In order that extraneous details of the profiles not systematically affect subjects' ratings, the trait titles and scores were randomly ordered for each subject's set of profiles. Also, the order of the sexes within the profile sets was randomized.

Program PROFILE. A computer program, called PROFILE, was developed to generate these random student profiles. It was written for the IBM 360/65 in The University of Connecticut Computer Center in the IBM computer language PL/I, level F (IBM, 1969, and Weinberg, 1966). It is a program which generates large numbers of profiles of the form required for this study. The program allows the user to specify up to ten traits for each profile. It also produces punched card output, one card for each profile, containing the profile identification data, trait order on the profile, and trait scores.

The PROFILE Program was written because it was the only feasible way to mass-produce five thousand unique student profiles, half boys and half girls, with a random trait-order for each set, and with randomly generated trait scores. At the same time it could automatically produce, with great savings of time and manpower, the punched cards needed for later analysis.

PL/I was chosen as the language for the program because of the ease with which it allows for the manipulation of strings of characters. This was very important since the profile format contains a diagram of a normal curve, multiple printing on the same line, and variable length bar-graphs for the T-scores. These specifications were easier to program in a list-producing language like PL/I than in a mathematical formula language such as FORTRAN.

Because of the input-output restrictions of the University computer, the program had to be designed to produce the profiles in small batches. Input and output devices, especially the printer, are extremely slow compared with the operation of the central processing unit. The operation of the system would be greatly impaired if the input/output devices were taxed to capacity. Output is stored temporarily on a magnetic disk and is printed or punched when the program is completed. The storage capacity of the magnetic disk is far less than what would be needed for the 5,000 (100 evaluators x 50) profiles that were generated, so only a specified number was produced on each run.

Since each profile and its data card have a unique identification number, the first number in the current sequence has to be included as input to the program. Also included in the input parameters are the number of sets of profiles to be generated, the number of profiles for each set, the number of traits to appear on the profiles, and a list of the trait titles.

The program is not specific to the requirements of the present study. It was designed to be readily usable by other researchers, one of whom is currently using similar profiles for a study of his own.

Appendix C contains a flow chart of the program. And a listing of Program PROFILE is available. What follows is a detailed description of the program.

Program Flow. The program first allocates sufficient memory space for storing all necessary arrays and the profile printout format; then it reads the input parameters and the trait titles. After these preliminaries it initiates the random-number sequence by reading the computer clock and converting the time, which appears in hours, minutes, seconds, and milliseconds, to an odd number. This large number becomes the seed number for the subroutine which produces random numbers for the rest of the program.

Next, a counter begins which will check the number of sets of profiles produced by the program against the number requested on the parameter card. This counter serves as the evaluator number. It begins with the evaluator identification number specified in the input register and increments until the correct number of profile sets has been generated.

For each of these sets two batches of profiles will be produced, one to be labeled "twelfth-grade boy" and the other "twelfth-grade girl." The number in each batch will be the same and is specified by the user on the input cards. The order of these two halves is random. It is decided by selecting a random number between zero and one. If this number is greater than one-half (.5), the first half of the profiles is called boys; if it is less than one-half, the first half is girls.

Also, the order in which the trait titles are printed is determined randomly before a profile set is produced. Random numbers are computed between one and the number of traits. After each number is computed it is assigned to the next trait in the original trait input order. As numbers are assigned to traits, the program checks back through the traits for any duplications. If a sequence number duplicates a previous number, the trait being sequenced receives the next random number. This process is terminated when each trait has a unique sequence number.

At this point a second index begins to count the number of profiles in each half-set, until the specified number should be reached. Then the calculation of the percentile scores for the traits begins. The program generates as many random percentiles, between one and 99, as there are traits. Each percentile is generated from a random number between zero and one. The corresponding T-score for the profiles is calculated by a set routine from the same random number, since the T-score is a function of the area under the normal curve, of which the random number can be considered a measure. The percentiles and corresponding T-scores are assigned to the traits in order of original input sequence.

The completed profile is then printed out. A routine prints the standard format when all information is complete. The appropriate sex is printed at the head of the page, and the trait titles, percentiles, and bar-graphs for the T-scores are printed in the randomized order derived at the beginning of the sequence. An identification number, composed of the evaluator number, the sex identification number, and the profile sequence number, is printed at the bottom of the page. (Reference may be made to Appendix B for an example of the completed profiles.)

Next a data card is punched. This card contains a one-digit code, which can be used to identify the set of traits used on the profile, the same composite identification number as the profile, the trait order (the sequence number of each trait on the profile), the percentiles, and the T-scores. This information is produced for the traits in the original trait input order, not the order in which it appears on the printed profile.

After punching the card, the program checks the profile index to see if all profiles for the current sex group have been produced. If they have not, the index is incremented and another profile produced. When the number is complete, the program checks to see if profiles have been generated for both sexes. If not, it will change the sex and begin another set after resetting the profile index. If both sexes have been printed, it will check to see if it has produced the number of profile sets requested by the user. If this check is false, it increments the set counter, i.e., the evaluator number, and begins the operation again by calculating new first sex and new trait sequence.

When all the profiles requested have been produced, the program will attempt to read a new set of input parameters. More parameters and trait titles can be read into the compuer to produce profiles with a different set of evaluator numbers, or profiles with different traits. A new input set will cause the program to begin at the initiation of the set counter. Otherwise, the program terminates after printing a message that no more input cards were encountered. This informs the user why the program has terminated.

Program PROFILE was used to print all the profiles needed for this project. However, since it has been written as a general program, it can be used by anyone attempting to do similar kinds of research, or by anyone wishing to extend the scope of the current project.

The Traits. By nature, the information-processing capacity of the human subject is limited. The average person may be able to process simultaneously about seven pieces of information without undue confusion (Miller, 1956). For this reason, the number of traits was limited to seven. Seven headings were chosen which could include all aspects of the curriculum so that the amount of information could be readily manageable.

The seven traits chosen are arbitrary. They are designed to include in a logical fasion all aspects of the curriculum. They are a redesigned version of the seven proposed for investigation by Page (1972) and thus represent only one possible logical analysis of the educational program. The logical breakdown had been proposed to various educators and researchers prior to its adoption in the study, and no reservations or criticisms were expressed.

The trait titles selected are the following: Quantitative, numbers and related studies; Arts, esthetic and practical studies; Physical, health and hygiene; Social Studies, civics and social sciences; Verbal, communication skills and verbal reasoning; Personality, social behavior and character development; and Natural Sciences, the physical and biological sciences. The complete trait descriptions which were given to the evaluators appear in Appendix D.

As yet, there has been no comprehensive attempt to develop the complete structure of the bentee under these headings. But when, and if, completed, this structure would extend from the value-space down to particular operations and behavioral criteria. The importance of each operation and criterion in terms of the over-arching goals and values of education would be computable from the product of the weights of each operation, sub-trait, and trait of which it is a part. These characteristics were noted in Chapter 2.

One reservation must be made regarding the traits. Although there were no complaints expressed during the research about their theoretical completeness, either by professional researchers or by the subjects, it can be argued that another category should have been included: Vocational education. It was felt that vocational courses could be included under one or more of the other traits without overly stretching their definitions. That the absence of an additional trait went unnoticed may indicate general acceptance of the logical scheme.

<u>Data Collection</u>. The data were collected in a personal interview with each subject. Each session lasted approximately 30 minutes (20 to 40 minutes). The length of the session depended on how quickly the subject understood the instructions and how quickly he was able to work through the materials.

At the beginning of the session, subjects were asked to fill out identification sheets which asked for their name, address, phone number, age, sex, education, and occupation. Each information form also contained space for the subject's identification number, which was filled in by the researcher. The researcher also made note of the subject's ethnic grouping.

After the form had been completed, the researcher explained the general purpose of the study and gave a brief explanation of the material, the profiles. The standard format for the interview instructions appears in Appendix E. These instructions could be modified by the researcher at his discretion to facilitate the subject's mastery of the task.

In conjunction with the description and explanation of the profiles, the subject was given, on index cards, brief descriptions of each trait. Since the arrangement of the trait titles on the profiles was random for each subject, to protect against the possibility that subjects' evaluations could be influenced by trait order, the descriptions were printed on cards so they could be rearranged at each interview session into the same order as the tratis appeared on the computer printout.

After the subject felt he was sufficiently familiar with the trait descriptions, he might have been asked to assign weights to the tratis by sorting 50 pennies according to how important he felt each trait to be. The subject was asked to <u>use all the pennies</u>, but was free to place them where he wished.

Subjects were asked to make independent assignments for each sex. The order of the sexes for this evaluation was the same as the random order in which the profile packet had been arranged by the computer. Half the subjects, by random selection, performed this task before the profiles were explained to them. The other half performed it after they had completed their profile evaluations.

After a subject had read the trait descriptions and assigned the pennies, if that were called for, the researcher explained to him the profile format. He explained what percentiles and T-scores were, pointed out that half the profiles were "boys" and half "girls," indicated the scale under the profile for recording his overall judgment, answered any questions, and offered any further explanations until the subject understood the task.

Subjects were then taken through a sample profile and shown how to record their judgment, what kind of mark to make, and where to put it on the computer printout. After the subject knew what to do and understood the profiles and trait descriptions, he was given the profile packet and left undisturbed to make his evaluations. The trait descriptions were left with him for any necessary reference.

When the subject had completed the 50 profiles, the researcher collected them. If the subject had not weighted the traits prior to evaluating the profiles, he made his penny assignments at this point, and his weightings were recorded.

The researcher then paid the subject his stipend and, if the subject were interested, discussed the study with them. Any significant comments or suggestions were recorded.

Data Preparation. First, the profiles were scored by hand. Each evaluation mark made along the judgment continuum under the profile was scored as a number from 1 to 81. There was no intrinsic or generalizable meaning to this numbering system. It was arbitrarily adopted because the judgment continuum line was composed of 81 segments. Scoring by line segment number was the least taxing and least fallible method which allowed making more than a ten-point discrimination in rating. The rating scores were then punched onto the data cards which had been produced when the profiles were generated.

The information from the identification form was also coded and punched onto another data card. This card contained subject number, a code for location (Connecticut or Cleveland), age in years, sex, race, number of years of education, stratum number, and codes indicating whether the subject performed the token assignment before or after evaluating the profiles, and whether his profile set contained "boys" or girls" first.

This data card also contained the token-weightings assigned to each trait by the valuator. This card and the set of profile cards contained all the data used in the analysis. From these data all variables used in the analysis were constructed. The next chapter describes the computation of these variables and the results of the statistical analyses performed.

CHAPTER 4

EXPERIMENTAL RESULTS USING THE BENTEE

The previous chapters have summarized the theory concerning the bentee, and have outlined the procedures employed, and the sample of judges drawn for the experimental work. Still, it has not been established that human judges may reasonably serve such a function. Is there any reliable consensus of people? Do different groups of judges differ much in their evaluations? Do bentee evaluations vary according to student sex? These and similar considerations occupy the present chapter.

The contents of this chapter will be as follows: A) Results of the token method. B) Results of the profile or correlation method. C) Comparisons of the token and profile methods. D) Values related to personal variables of the subject-judges. E) Factor analysis of the value matrix. And a summary will conclude the chapter.

A. Results of the Token Method

As already noted, each judge was asked to evaluate the importance of seven tratis for the education of boys, and the same seven tratis for the education of girls. Then for each judge, an average may be calculated, combining the weights given for the two sexes, and this is the judge's total evaluation. Thus there are 21 token-assignment variables for the set of judges, and a summary of these appears in Table 3.

The order in which traits were presented to subjects was random; in the analysis they are presented in all tables in their order of importance, as computed by the combined weights received in the token assignment method.

Each set of seven variables forms an ipsative measure. An ipsative measure is one in which the sum of all the values is equal to a constant. Since subjects were given 50 pennies to assign differentially to the seven traits, the sum of all the token assignments for each subject and the sum of means across groups of subjects must always equal 50. It is obvious, then, that the number of tokens assigned to a trait cannot be considered a measure of its absolute value. It is thus impossible to make comparisons between subjects; i.e., it is impossible to say whether Verbal is more important to Subject A than to Subject B. The meaning of the number of tokens assigned to a trait is relative to the other numbers within the same ipsative set (Clemens, 1965).

As was expected, the means for these judgments about the relative values of the traits are not equal. Also, the values of the traits for both sexes are not the same. The judged order of the traits in importance is not the same for the two sexes. And even where the traits appear in relatively the same order, there are differences in the relative weights of the traits.

TABLE 3

TRAIT WEIGHTINGS BY TOKEN METHOD

Trait	Mean	S.D.	Maximum	Minimum
Boys				
Verbal	10.21	4.87	44	0
Personality	8.19	4.37	22	0
Quantitative	7.71	2.71	15	1
Social Studies	7.10	3.12	20	0
Natural Science	6.28	2.33	11	0
Physical	5.98	3.22	19	0
Arts	4.41	2.44	10	0
Girls				
Verbal	10.44	4.91	44	2
Personality	8.94	4.54	25	0
Quantitative	6.56	2.74	13	0
Social Studies	6.89	2.88	20	0
Natural Science	5.57	2.45	11	0
Physical	5.52	3.23	19	0
Arts	5.98	2.62	15	0
Combined				
Verbal	10.33	4.83	44	2
Personality	8.56	4.32	20	0
Quantitative	7.14	2.51	13	1
Social Studies	7.00	2.78	20	0
Natural Science	5.93	2.26	11	0
Physical	5.75	3.12	19	0
Arts	5.19	2.22	10	0

Means are computed over the 101 subjects.

There seem to be some slight biases in subjects' judgments about the relative importance of the various traits for boys and girls. Although Verbal and Personality were considered most important for both sexes, they appear to be judged relatively more important, compared to the other traits, for girls. Quantitative was judged third in importance for boys, but Social Studies was considered more important than it was for girls. And Arts was judged least important for boys, but more important than either Natural Science or Physical for girls.

As can be seen from Table 3, there were great differences of opinion among subjects about all the traits, indicated by the large standard deviations and the extreme ranges of maximum and minimum values. For all the sex-trait combinations, except Quantitative for boys and Personality for girls, there was at least one subject who rated the trait of no, or of absolutely minimal, importance.

For any one trait the differences can be quite drastic, with the trait being judged of least importance by one subject and of greatest importance by another. The maximum values do fall, however, into approximately the same order as traits were judged in importance by the sample. This relationship and a similar one for the standard deviations may indicate that the bulk of the mean differences in judged trait values may be due in largest part to some subjects who had strong opinions about the traits.

A good number of the subjects tended to rate the traits of almost equal importance, making only slight distinctions between them. Many of them gave all the traits just seven tokens except for one judged slightly more important to which they gave eight. Such subjects many times would simply change the location of the one extra token when they made their judgments for the opposite sex. Many gave indications that they would have been more satisfied had it been possible to weight all traits equally. The differences in means for the traits is certainly more a reflection of the judgments of other subjects with stronger opinions, than of these meek subjects.

Such differences in the opinion and rating behavior of the subjects may have been related to the group to which a subject belonged. Subjects who were professionally engaged in some aspect of secondary education may have valued the traits differently from those who were not engaged in education. So the sample was broken down to the two strata, and mean weightings were calculated for each trait. Table 4 contains this breakdown.

TABLE 4

TRAIT TOKEN WEIGHTS
BY TRAIT, SEX, AND STRATUM

	Professionals (N = 52)		Non-Professionals (N = 49)	
Trait	Boy	Girl	Boy	Girl
Verbal	9.94	10.19	10.51	10.69
Personality	8.08	8.77	8.31	9.12
Quantitative	7.63	6.35	7.80	6.80
Social Studies	7.15	6.98	7.04	6.80
Natural Sciences	6.37	5.77	6.18	5.37
Physical	6.04	5.51	5.92	5.53
Arts	4.77	6.44	4.02	5.49

Differences between strata do not appear very large. Both groups of subjects seem to exhibit the possible sex bias mentioned above. The non-professional subjects seem to rate Verbal higher than do the professionals and Arts correspondingly lower. Other differences appear negligible.

An analysis of variance was performed to test whether these observed differences between the trait weightings were statistically significant. A model was chosen which would test the possible effect of sex of hypothetical student on judged trait value and also test for differences in rating behavior between professional and lay subjects.

The design chosen for this analysis of variance is a three-factor repeated measures design adapted from Winer (1962). The three factors are a) stratum, with two levels, professionals and non-professionals; b) sex of the hypothetical students, also with two levels, boys and girls; and c) trait, the repeated measure, with seven levels.

Some reservations must be made about using analysis of variance with the token assignments. As has been mentioned, the numbers of tokens assigned to each trait are not independent. They constitute an ipsative measure, since the sum of the seven traits weighting is equal to 50. No precedent has been found for this type of analysis, and Clemans (1965) makes no mention about analysis of variance with ipsative measures.

A repeated measures design was used for the analysis since it was one way to attempt to control for the fact that the ipsative measures were not independent of each other. The repeated measures design applies in situations where the same subject responds to a variety of stimuli, here the trait descriptions, and his scores may not be independent. It seemed to be the design best suited to test the differences observed in the token weightings.

Since the data were ipsative, finding any significant effects for either of the first two factors, stratum or sex, would be impossible. Since each subject was given 50 tokens to assign to the traits for each sex, his total score would always be 100 tokens. The score for each stratum would be 100 times the number of subjects, and the score for each sex would be 50 times the total number of subjects.

The meaningful tests are for differences between trait value for the interaction of hypothetical student sex and trait value, and for the interaction of stratum and trait value. In other words, the analysis tested whether there were significant differences in the weightings derived for the traits, whether the relative differences in the weightings were similar for boys and girls, and whether the weightings of the traits differed for professional and non-professional subjects.

The summary table for the analysis of variance of the token weights appears in Table 5. As has been mentioned, it is based on the model which appears in Winer (1962, p. 328). Since there are unequal numbers of subjects in the two strata, appropriate changes were made in some denominators and degrees of freedom in Winer's model.

TABLE 5

ANALYSIS OF VARIANCE--TOKEN WEIGHTS

Source	Sum of Squares	df	Mean Square	F
Between Subjects	32.7	100		
Stratum	.34	1	.34	1.04
Subject within group	32.36	99	.33	
Within Subject	20,687.2	1313		
Sex	0.0	1	0.0	0.0
Stratum x Sex	0.0	1	0.0	0.0
Sex x Subject within group	2.4	99	.02	
Trait	3,917.5	6	652.92	25.75***
Stratum x Trait	65.07	6	10.84	.43
Trait x Subject within group	15,064.13	594	25.36	
Sex x Trait	260.4	6	43.40	18.75***
Stratum x Sex x Trait	2.84	6	.47	.20
Sex x Trait x Subject within group	1,374.86	594	2.31	

*** $p < .001$

The null hypothesis of no differences between trait values was rejected at well beyond the .001 level of significance. There definitely were differences in the relative weightings assigned to these seven traits by the direct token method. The null hypothesis for the sex-by-trait interaction was also rejected at well beyond the .001 level of significance. So subjects felt that some traits were more important for students of one sex than for those of the other.

There were no other significant effects or interactions. Stratum had no effect at all. There seem to be no differences between the sample of professional educators and of lay perople in the way they respond to direct evaluations of the traits.

The variance between subject and between sex was determined to be zero, due to the ipsative nature of the measure. That some variance appears in these two factors is due to a slip in arithmetic. For a couple of subjects the numbers of tokens recorded for the seven traits did not sum to 50. These subjects either were given more than 50 tokens or their responses were recorded incorrectly.

The results indicate that by a direct assessment method, in which judges must assign tokens in an explicit way, statistically significant weightings can be obtained for a set of academic traits. Subjects' trait values do not seem to be a function of their professional relationship to education. The profile of trait values for professional subject did not differ significantly from that for non-professionals.

There was definitely a sex bias operating in the assignment of the tokens. Personality and Arts were rated more important for girls, and Quantitative and Natural Science rated less so. The bias may represent what different training a girl should have, to be able to overcome the stereotypes. It does show that when asked directly, subjects do not see students as the women's liberation advocates would want, i.e., with no distinctions on sex.

B. Results of the Profile Evaluation

Profile Evaluation Method. The trait values have so far been presented in terms of a direct assessment of trait by assigning tokens. In such a direct fashion, subjects might well be influenced by expectancy of what others, including the researcher, might think. Or they might respond to what they believed their values would be, although in practice, in making judgments about students, those values might turn out to be quite otherwise. Thus a second method was used to measure the values of the same group of subjects in a somewhat more lifelike situation. This was done by having subjects evaluate a group of student profiles.

Chapter 3 described the computer generation of random student profiles. These contained random percentile scores with their T-score equivalents in a random trait sequence to avoid any possible sequence effect. Subjects evaluated 50 profiles, 25 "boys" and 25 "girls." The 14 scores appearing on each profile were compared with subjects' overall judgments by computing Pearson-product-moment correlation coefficients. Forty-two correlations were calculated.

There are three sets of these correlation coefficients. Each set contains 14 correlations, seven computed using the trait percentile scores and seven using the trait T-scores. The first set contains the correlations calculated on the set of "boys" profiles, the second the correlations from the "girls" profiles, and the third the correlations calculated for all the profiles. The four sets are contained in Tables 7, 8, and 9.

As was stated in the section on procedures, subjects were given 52 profiles to evaluate, 26 "boys," and 26 "girls." However, not all the correlations are computed over 26 or 52 profiles. The correlations are for all profiles for which valid judgments had been recorded. This number is usually 26 or 52, but for some subjects there were fewer valid judgments.

There are a number of reasons for these omissions. First, on some occasions during the explanation of the task, so much use was made of an example profile that it was discarded lest its evaluation have been biased by the discussion. Second, there were some profiles which were garbled

TABLE 7

TRAIT WEIGHTINGS BY CORRELATION METHOD
"BOY" PROFILES

Trait	Mean	S.D.	Maximum	Minimum
Verbal T	.396	.235	.921	-.342
Verbal %	.400	.234	.933	-.330
Person T	.319	.282	.935	-.365
Person %	.324	.284	.948	-.359
Quant T	.286	.254	.800	-.375
Quant %	.290	.259	.825	-.348
SocStu T	.232	.251	.788	-.267
SocStu %	.236	.250	.790	-.274
NatSci T	.197	.219	.607	-.382
NatSci %	.204	.221	.616	-.387
Phys T	.125	.255	.741	-.459
Phys %	.126	.258	.729	-.450
Arts T	.137	.207	.711	-.386
Arts %	.136	.213	.708	-.418

Each mean is the arithmetic mean for 101 subjects.

TABLE 8

TRAIT WEIGHTINGS BY CORRELATION METHOD
"GIRL" PROFILES

Trait	Mean	S.D.	Maximum	Minimum
Verbal T	.435	.218	.908	-.120
Verbal %	.439	.218	.929	-.101
Person T	.325	.270	.925	-.237
Person %	.327	.268	.947	-.227
Quant T	.260	.217	.785	-.274
Quant %	.262	.219	.791	-.208
SocStu T	.229	.241	.737	-.408
SocStu %	.231	.242	.758	-.414
NatSci T	.195	.220	.724	-.389
NatSci %	.195	.222	.723	-.406
Phys T	.142	.240	.735	-.411
Phys %	.147	.240	.725	-.417
Arts T	.173	.244	.724	-.299
Arts %	.169	.246	.732	-.263

TABLE 9

TRAIT WEIGHTINGS BY CORRELATION METHOD, ALL PROFILES COMBINED

Trait	Mean	S.D.	Maximum	Minimum
Verbal T	.414	.189	.922	-.114
Verbal %	.417	.190	.937	-.086
Person T	.324	.235	.924	-.175
Person %	.325	.234	.945	-.176
Quant T	.274	.198	.729	-.183
Quant %	.274	.202	.730	-.203
SocStu T	.230	.196	.707	-.137
SocStu %	.231	.197	.701	-.157
NatSci T	.195	.181	.573	-.294
NatSci %	.197	.184	.593	-.305
Phys T	.133	.187	.691	-.365
Phys %	.136	.187	.712	-.363
Arts T	.157	.165	.633	-.296
Arts %	.154	.166	.636	-.281

in the printing process and had to be discarded. Such profiles were not discovered until they had been given to subjects. And third, occasionally one or more profiles were inadvertently skipped by a subject and were not noticed until they were scored. These occurrences meant a reduction in the number of profiles which could be used to calculate the correlations.

Correlations were calculated from both sets of trait scores on the profiles. There are minor differences between the two sets of correlations. The correlations computed from the percentile scores were used in subsequent analyses. This was done not only because absolute values were slightly higher but also because most subjects seemed to make their evaluations by using the percentiles. These scores appeared on the profiles as numbers rather than as bar graphs like the T-scores. Observations of subjects' behavior and markings left on the profile printouts indicated evaluations based on the percentiles rather than the T-scores. The concept of a percentile seemed easier for subjects to understand than that of a deviation-based T-score.

The means for the correlations as they appear in the tables are arithmetic means and are not computed by the z-score transformation. There may be arguments as to whether they represent unbiased measures of the population correlations. The z-transformation would give a slightly higher estimate of these population values.

As there was with the token weights, so here there is a great deal of disagreement among subjects about how important each of the traits is. For almost every trait-sex combination there is at least one correlation which has a strongly negative minimum value. And the maximum positive values indicate that at least one subject seemed to make overall judgments based almost entirely on the variation of one trait.

Although the correlation measures do not, strictly speaking, form an ipsative measure, there are some limits on their values. If the scores for the traits were uncorrelated, then the squared multiple correlation coefficient between the trait scores and subjects' overall judgment would be equal to the sum of the squares of the seven simple correlations between overall judgment and the trait scores (Darlington, 1968, p. 162). Since the squared multiple correlation must be less than or equal to 1.0, it was expected that the sum of the squares of the seven correlations would also approach 1.0 as a limit.

For many subjects, however, the sum of the squared correlations was either greater than 1.0 or far less than 1.0. The first case would indicate that the randomly generated trait scores were not free of random intercorrelations. The second would show that there were many subjects who were "bad" judges, who either did not make their judgments according to a consistent scheme or made a great number of errors in their calculations, so that as a result the relationship of their judgments to trait scores is only slightly higher than would be expected by chance alone.

Many of the same relationships which appeared in the token weightings, appear with the latent measures as well. But in the correlations there must be a lot of measurement error in the results, since over the whole sample the variance in subjects' evaluations explained by variance in trait scores is less than .56. This is calculated by summing the squares of the seven correlation coefficients.

In spite of the error there are observable differences in the means for each trait. These differences may be due to a number of "good" judges, judges who made a set of evaluations strongly related to one or two traits. A subject who tried to make only minor distinctions in trait value or considered the traits of about equal value would tend to have low correlations for all the traits and be called a "bad" subject.

The traits do seem to receive different values from the subjects despite great differences in opinion between subjects. A summary table prepared by stratum (Table 10) shows some possible differences in trait value for sex.

An analysis of variance was performed on the correlation method weightings, those computed from percentile scores, to test for significance in the differences observed. Prior to the analysis the correlations were transformed to z-scores. The summary correlations presented in Table 10 are computed from these z-score transformations. The same three-factor repeated measures design from Winer (1962) was used for this analysis.

TABLE 10

CORRELATION TOKEN WEIGHTS
BY TRAIT, SEX, AND STRATUM

	Professionals (N = 52)		Non-Professionals (N = 49)	
Trait	Boy	Girl	Boy	Girl
Verbal	.43	.51	.43	.43
Personality	.37	.35	.35	.37
Quantitative	.29	.27	.33	.29
Social Studies	.27	.28	.24	.20
Natural Science	.18	.21	.25	.21
Physical	.15	.15	.13	.15
Arts	.12	.15	.17	.22

The results of the analysis of variance are presented in Table 11. The only significant effect is for differences between the traits, which here were again significant at well beyond the .001 level of significance. So in spite of all the measurement error involved in the instrument the differences in the values assigned by subjects to the traits were significant. However, the sex by-trait-interaction which appeared on the token method was not present here.

There were a number of other effects possible here which were not possible in the analysis of the token method. A significant effect for stratum would have meant that either professionals or lay perople were able to make "better" judgments on the profiles, judgments which resulted in a larger sum of the correlation coefficients. There were some differences for strata, but not significant ones. If there were "better" judges, they did not belong to one group more than the other. In the same manner, it would have been possible for subjects to make "better" judgments on profiles of one sex rather than the other. But any such differences were not significant.

The only significant differences which did appear were between the trait values. Subjects did not seem to distinguish between the sexes when they made their evaluations, only between the traits. It could be that in a more lifelike situation the directly measured sex bias does not appear. It could be also that any such bias was swamped by the errors in judgment which were made on the profiles. Or it could be that subjects were not made as aware of hypothetical student sex as they might have been. As can be seen from Appendix B, sex was indicated near the top of the page while the bulk of a subject's attention was focused on the scores and the judgment line much farther down the page. So it is possible that subjects did not pay much attention and thus did not notice when the sex of the profile changed.

TABLE 11

ANALYSIS OF VARIANCE--CORRELATION WEIGHTS

Source	Sum of Squares	df	Mean Square	F
Between Subjects	6.729	100		
Stratum	.0387	1	.0387	.573
Subject within group	6.6903	99	.0675	
Within Subject	127.343	1313		
Sex	.059	1	.059	1.599
Stratum x Sex	.0214	1	.0214	.580
Sex x Subject within group	3.6616	99	.0369	
Trait	17.774	6	2.9623	24.727**
Stratum x Trait	.5593	6	.0932	.778
Trait x Subject within group	71.1867	594	.1198	
Sex x Trait	.258	6	.043	.761
Strantum x Sex x Trait	.2114	6	.0352	.623
Sex x Trait x Subject within group	33.6116	594	.0565	

** $p < .001$

However, subjects did make distinctions between the importance of the traits by this method. They also made significant distinctions on the direct method. The question is whether the latent weightings derived from the profile evaluations are comparable to the deliberate weightings derived from the token assignments.

C. Comparison of Methods

It was hoped that the methods would be two aspects of the same general behavior and that the judgments made about trait value on one method would be comparable to the judgments made on the other. Therefore, the expectation was for general agreement between the two methods, within the limits of sampling and instrument error.

The first attempt to measure this relationship was to correlate the number of tokens a subject gave to a particular sex-trait combination with the correlation coefficient between his judgments and the percentile scores for the same trait and sex combination. Table 12 contains a summary of the results of this first analysis. Pearson product-moment correlation coefficients were computed between subjects' token weightings and their

TABLE 12

CORRELATIONS BETWEEN DIRECT AND LATENT WEIGHTS FOR 101 JUDGES

Trait		Boys	Girls
Verbal		.365	.488
Personality		.578	.474
Quantitative		.452	.343
Social Studies		.405	.210
Natural Science		.358	.155
Physical		.384	.088
Arts		.169	.331
	Mean =	.387	.298

correlation weightings for the 14 trait-sex combinations. In reference to previous tables these correlations are between the first 14 variables in Table 3 and the percentiles correlations from Table 7 and 8.

As the table shows, the relationships here, although different from zero, are not very high. They are a little higher for boys than for girls; the judgments of the subjects seem more stable across methods for boys. However, the variance in either set of weights accounts for less than 20 per cent of the variance in the other set.

There could be a number of reasons for this low reliability. There was certainly some measurement error in both methods. It is possible that subjects did not evaluate students according to their manifest values. Perhaps subjects made different kinds of judgments on the two methods. Or it is possible that they were not able or willing to use all the information available, seven trait scores and a sex designation, to make their evaluations.

The indication from these correlations is that the less important traits are more subject to error, relative to their manifest judged importance, when profiles are evaluated. The higher correlation coefficients are associated with the less important traits. Rank-order correlations were computed between the trait token means and the reliability coefficients according to the ranks appearing in Table 13.

These rank-order correlations suggest that subjects based their judgments on only a few important variables rather than on all available information. For the greater the importance of the trait, the higher the reliability was across measures. The judgments made on the profiles are not so concise about less important traits as about the more important ones.

A second comparison of methods was made by computing Spearman rank-order correlation coefficients and Pearson product-moment correlation

TABLE 13

RANK-ORDERED TRAITS FOR
RELIABILITY AND TOKEN-WEIGHTS

Trait	"Boys" Rel.	"Boys" Token	"Girls" Rel.	"Girls" Token
Verbal	5	1	1	1
Personality	1	2	2	2
Quantitative	2	3	3	4
Social Studies	3	4	5	3
Natural Science	6	5	6	6
Physical	4	6	7	7
Arts	7	7	4	5
	r = .572		r = .893	

coefficients between each subject's vector of token weights and his vector of correlation weights. These coefficients were computed for the three sets of weights: for "boys", for "girls," and for the two combined. The means for these coefficients for the 101 subjects appear in Table 14. They were computed according to the z-score transformation.

TABLE 14

MEANS OF RELIABILITY COEFFICIENTS (N = 101)

	Spearman Rho	Variance of z-score	Stan. Dev. of z-score
"Boys"	.554	.264	.514
"Girls"	.453	.363	.603
Combined	.607	.255	.505
	Pearson R	Variance of z-score	Stan. Dev. of z-score
"Boys"	.606	.295	.544
"Girls"	.498	.374	.612
Combined	.679	.304	.551

This second analysis also indicated more agreement between methods on trait importance for "boys" than for "girls." The combination of the two measures shows more agreement than for either sex individually. This may indicate that sex differences were not so important as subjects thought when they assigned token weights, that subjects were not aware of sex differences in their profile evaluations, or simply that more reliable measures are associated with increases in number of measurements.

Two correlations were computed, but the rank order correlations may be psychologically more meaningful. Subjects may be attempting to assign priorities to the traits rather than attempting to assign ratio measures. Although these data indicate somewhat more reliability and consistency across the two methods than the previous analysis, the variance in the weights assigned to traits by one method accounts for less than 40 per cent of the variance in the weights from the other. For some reason the average subject did not transfer value schemes exactly from one method to the other.

The bentee, however, was formulated not as a measure of individual values but of group and societal values. A scientific analysis of an educational system would depend on system-wide values and goals, cultural and not individual stereotypes. So the relationship across methods of population and subgroup value profiles is very important.

The traits were ranked according to their judged values for the whole sample. The rank-ordered traits appear in Table 15. The rank-order coefficients are calculated for "boys," "girls," and both groups combined.

TABLE 15

RANK-ORDERED TRAITS
FOR TWO METHODS (101 JUDGES)

	"Boys"		"Girls"		Combined	
	Token	Corr.	Token	Corr.	Token	Corr.
Verbal	1	1	1	1	1	1
Personality	2	2	2	2	2	2
Quantitative	3	3	4	3	3	3
Social Studies	4	4	3	4	4	4
Natural Sciences	5	5	6	5	5	5
Physical	6	7	7	7	6	7
Arts	7	6	5	6	7	6
	r .964		r = .929		r = .964	

For these group values profiles, there is a great deal of consistency across methods. The coefficients are all greater than .90 and indicate that for group responses the two methods result in quite comparable measures.

Profiles of subgroups were also compared across the two methods. Correlations are reported in Table 16. Relationships are still very high although slightly less than for the whole sample. There is more agreement, nonetheless, across methods about "boys" than about "girls." Here, however, the combined measures are slightly less consistent than those for boys, probably because there is extremely high consistency in group judgments about "boys."

TABLE 16

CORRELATIONS BETWEEN RANK-ORDERED TRAITS FOR IDENTIFIABLE SUBGROUPS

	Professional (N = 52)	Lay (N = 49)	Male (N = 67)	Female (N = 34)
"Boys"	1.000	.929	.964	.964
"Girls"	.786	.838	.893	.964
Combined	.964	.929	.929	.964

For groups and the whole sample, the agreement was much higher than for the average subject. There were many "bad" subjects, subjects whose profile evaluations showed no relationship to their token weightings and in some instances showed even inverse relationships. In group analyses these subjects tend to cancel each other out, although they do reduce the reliability of the measure and introduce error which clouds other possible relationships.

An analysis of many individual profile rating results discovered that the correlations between less important traits and overall evaluation did not exhibit more than chance variation from zero. There is, then, more evidence that subjects made their profile evaluations on some subset of all the information available, perhaps just the more important traits.

Yet group reliability analyses seem to indicate that there is concordance of opinion about the seven traits, especially for identifiable groups. It is imperative in developing a standard of educational benefit to be able to measure group values reliably and it would seem that either method can be suitably employed to obtain them. In fact, the direct method may prove to be the only one necessary, and it is certainly easier to administer. It showed an expected sex bias which did not appear in the correlation method. Perhaps refinements might be possible in the profiles to reduce some of the error which seems to be present.

D. Personal Variables

Information was collected on some variables about the subjects and the interview schedule which might explain some of the variance which has been considered error of measurement. At the beginning of each interview session, subjects completed identification forms with information on residence, age, sex, education, and occupation. The interviewer also made note of the subject's ethnic grouping. From these data, six background variables were constructed.

The first variable is Location. This variable also serves as an identification code for the interviewer since, as has been mentioned, data were also collected in the Cleveland and Connecticut areas by two different researchers. Thus interviewer and location effects are inextricably confounded. Subjects in the study recruited from Connecticut were coded 1, and those from the Greater Cleveland, Ohio, area were coded 2. There were 51 subjects from Connecticut and 50 from Cleveland.

The variable Sex is a coding of the sex indicated on the identification forms. Male subjects were coded 1 and female subjects 2. Approximately two-thirds of the sample were males.

The variable Stratum is a nominal variable, a categorization of the occupations of the subjects. Categories one through three contain the professional subjects; the first being high school teachers, the second high school administrators, and the third teacher-educators. Categories four through six contain the non-professional or lay subjects; the fourth comprising all women regardless of occupation, the fifth, blue collar males, and the sixth, white-collar males. Table 1 (in Chapter 3) contains a complete breakdown of the sample by Location, Sex and Stratum.

The variable Age is simply the age in years as reported on the identification form. One female subject did not respond to this question and was assigned the mean age for the sample. Although the sample exhibits a wide distribution on age, the mean indicates that more of the subjects come from the younger end of the curve. This does not appear to be a random occurrence but seems indicative of the fact that older people were more afraid of participating in the study.

The variable Education is the completed grade level as reported by the subject. A high school diploma was coded 12 and a college degree, 16. A master's degree was considered one year's schooling beyond a college degree and coded 17, and a doctorate was considered two years' more work and coded 19. Teachers and administrators who indicated having received a sixth-year professional certificate were scored 18 on this variable. Table 17 contains a description of variables Age and Education.

It may be noted that the mean education is equal to almost a college degree even though post-graduate degrees are scored rather conservatively. Years of schooling for the sample would be expected to be a little higher than for the general population since all the professional subjects would have earned at least a college degree. But the non-professional group

TABLE 17

AGE AND EDUCATION OF SUBJECTS

	Mean	Stan. Dev.	Maximum	Minimum
Age	36.08	10.98	67	18
Education	15.64	2.12	19	8

also tended to be somewhat well-educated, probably a characteristic of the type of people willing to take part in the study. Less well-educated subjects were afraid to make judgments about educational values, pleading lack of proper credentials or suitable intelligence. There were a couple of potential subjects who even listened to an explanation of the task, and then abstained from being subjects, claiming that they were not qualified to make such evaluations.

The variable Race contains the ethnic background of the subjects as recorded by the interviewer. Since the sample was predominently white, only two categories were constructed; whites (coded 1), and non-whites (coded 2). Only six of the 101 subjects were not white, of whom five were Black, and one Puerto Rican.

Two variables recorded the order in which the materials were presented to the subjects. A tabulation of these two variables appears in Table 18.

TABLE 18

INTERVIEW MANIPULATIONS

	"Boys" First	"Girls" First	Total
Tokens Before	32	14	46
Tokens After	22	30	52
Unrecorded	3	--	3
TOTAL	57	44	101

Subjects were asked to make their token assignments before or after they had evaluated their set of profiles. This manipulation was designated prior to each interview session. If a number of interviews were conducted consecutively one subject would assign tokens before evaluating profiles, the next the reverse. However, since no record of the order was kept from one interview session to another, the odd-even split was not systematic throughout the entire schedule of interviews. The unsystematic assignment did result in an almost 50/50 split although a few more subjects assigned tokens after they had rated profiles than before.

Subjects' profile packets could contain either "boys" or "girls" first. This random order was generated by the computer program when it produced the profiles. The token assignment was performed in this same computer-generated order, assignments being made for "boys" or "girls" first according to the order of the profiles. More packets contained "boy" profiles first.

These eight variables--six personal, and two experimental--were correlated with the 14 token weights and 21 correlation coefficients, between subject rating and trait T-score for the "boys" profiles, for the "girls" profiles, and for the combined set. Table 19 contains all significant intercorrelations. Since there were no significant correlations for the interview manipulations they are not included in the tables.

TABLE 19

CORRELATION MATRIX FOR PERSONAL VARIABLES

	Location	Age	Sex	Race	Education	Stratum
Location	--					
Age		--				
Sex			--	.26		
Race			.26	--	-.19	
Education				-.19	--	-.45
Stratum					-.45	--

The intercorrelations between the personal variables all seem to be idiosyncracies of the sample or of the definition of the variables. More minority subjects were women. The one Puerto Rican subject, who was the only subject who had not attended high school, is partly responsible for the race-education correlation since there were only six minority group members. The very strong correlation between stratum and education shows that the professional educators were better educated than the non-professional educators.

Significant correlations between the personal variables and token weights appear in Table 20. According to correlations, subjects from the Cleveland area gave more tokens to Quantitative and Natural Science than subjects from Connecticut, who gave more tokens to Personality and Physical. This relation may be due to an interviewer effect, as has been pointed out.

TABLE 20

CORRELATIONS BETWEEN PERSONAL VARIABLES AND TOKEN WEIGHTS

	Location	Age	Sex	Race	Education	Stratum
Quantitative "Boys"	.32					
Quantitative "Girls"	.27					
Arts "Boys"					.27	
Arts "Girls"						
Physical "Boys"	-.28				-.23	
Physical "Girls"	-.30				-.21	
Personality "Boys"	-.32					
Personality "Girls"	-.27					
Natural Science "Boys"	.21					
Natural Science "Girls"	.21				.31	

The only other significant relationships occur for Education. Better educated subjects thought Arts more important for boys, and Natural Science more important for girls. They considered Physical less important overall than did less well-educated subjects.

Finally, the significant relationships between the personal variable and the correlation weights appear in Table 21.

Subjects from Cleveland still rated Quantitative more important for all students and Personality less important for girls than subjects from Connecticut did. But their biases against Physical and for Natural Science measured on the token method, did not carry over to the latent measurement. The correlation between Location and Personality for the "boys" was not quite significant.

The relationship of Education also changed. On the correlation method, better-educated subjects gave higher weights to Verbal, a different relationship from those which occurred on the token method. Three other relationships occurred. Older subjects rated Natural Science higher than younger subjects; possibly the younger generation is not committed to a traditional curriculum and thinks the hard academic subjects are not so important. And men rated Arts more important for boys and Personality more important for girls; the higher rating for Personality may indicate some

TABLE 21

CORRELATIONS BETWEEN PERSONAL VARIABLES AND PROFILE WEIGHTS

	Location	Age	Sex	Race	Education	Stratum
Quantitative "Boys"	.27					
Quantitative "Girls"	.28					
Quantitative Combined	.31					
Arts "Boys"			-.24			
Arts "Girls"						
Social Studies "Girls"						
Verbal "Girls"					.27	-.22
Verbal Combined					.26	
Personality "Girls"	-.20		-.27			
Personality Combined	-.21		-.20			
Nat. Sci. "Boys"		.22				
Nat. Sci. Combined		.24				

latent "chauvinist" feelings in the male subjects. The correlation between Stratum and Verbal may indicate that educators thought it somewhat more important than non-educators did.

There was one other relationship which was discovered. Correlations were computed between the Sex of the profiles and the overall evaluations. A description of this variable is in Table 22. It appears to be randomly

TABLE 22

PROFILE SEX WITH RATINGS

	Mean	S.D.	Maximum	Minimum
Sex	-0.023	0.189	0.484	-0.434

distributed around a mean of zero. However, this variable received a correlation of .22 with the order of presentation of sex within the profile set. This value indicates that in general the sex which came first was given higher evaluations, i.e. that subjects tended to give higher evaluation to the first half of their set of profiles.

E. Factor Analysis

One remaining question has to do with whether subjects responded to, or valued, traits individually or whether they responded to groups of traits. Since both sets of value variables have many of the attributes of ipsative measures, by giving more weight to a particular trait, a subject must take some weight or emphasis away from some other trait.

In the analysis of the relationship of personal variables, to token trait values, for example, the subjects from Cleveland were shown to value Quantitative and Natural Science higher and Physical and Personality lower than subjects from Connecticut. If such patterns occurred in most subjects, it might be more meaningful to talk about factors rather than individual traits.

A principal-components factor analysis was used to find out how the data might be meaningfully summarized. An analysis was first performed on the weights derived from the direct method of token assignments. The weights used were combined weights for both sexes. The intercorrelations of these weights appear in Table 23. The rather large proportion of negative correlations is due to the fact that the weights form an ipsative measure (Clemans, 1965).

TABLE 23

CORRELATION MATRIX FOR
FOR TOKEN WEIGHTS (N = 101)

	Ver.	Per.	Quan.	SocStu.	Nat.Sci.	Phys.	Arts
Verbal		-.20	-.03	-.40	-.25	-.37	-.45
Personality			-.40	-.23	-.54	.04	-.20
Quantitative				.03	.34	-.30	-.27
Social Studies					.33	-.23	.12
Natural Science						-.26	.17
Physical							.27
Arts							

Three factors emerged from the analysis. A standard varimax rotation of the three factors resulted in the factor loading matrix which appears in Table 24. An oblique rotation produced substantially the same factor loading matrix. The factors were also shown by the oblique rotation to be almost completely orthogonal. The largest intercorrelation between factors was -.14.

TABLE 24

FACTOR LOADINGS FOR TOKEN WEIGHTS

	Factors		
	1	2	3
Verbal	.10	-.55	-.69
Personality	-.91	-.12	.07
Quantitative	.64	-.34	.01
Social Studies	.15	-.12	.87
Natural Science	.74	.04	.43
Physical	-.20	.81	-.29
Arts	.13	.77	.26

The same principal-components factor analysis was also applied to the trait weightings derived from the correlation method. The weights used for the analysis were the correlations computed over subject's entire set of profiles, the combined weights for both sexes. The intercorrelation matrix for these weights appears in Table 25. Since these weights have some properties of ipsative measures, a high proportion of negative correlations appear in the table.

Again, three factors emerged from the analysis. A varimax rotation of these three factors produced the factor loading matrix which appears in Table 26. Further oblique rotation did not change the loadings substantially and showed that the factors were substantially orthogonal. Again, the intercorrelationons of the factors were not much different from zero, the largest being -.10.

Clemans (1965), in his examination of ipsative variables, says that "it will be extremely difficult, if not impossible, to obtain psychologically meaningful results from the factor analysis of a set of ipsative intercorrelations" (p. 51). However, in this analysis at least two meaningful factors did appear. The third factor was not easily identified and

TABLE 25

CORRELATION MATRIX
FOR CORRELATION WEIGHTS (N = 101)

	Ver.	Per.	Quan.	Soc.Stu.	Nat.Sci.	Phys.	Arts
Verbal		-.16	-.04	-.17	.04	-.07	-.11
Personality			-.52	-.10	-.24	.04	.02
Quantitative				.03	.34	-.01	-.08
Social Studies					.15	-.06	.06
Natural Science						.06	.14
Physical							.09
Arts							

TABLE 26

FACTOR LOADINGS FOR CORRELATION WEIGHTS

	Factors		
	1	2	3
Verbal	.15	-.69	-.20
Personality	-.80	.10	.14
Quantitative	.82	.00	-.05
Social Studies	.17	.77	-.15
Natural Science	.65	.14	.30
Physical	.00	-.19	.76
Arts	.00	.27	.64

These three factors account for 59 per cent of the variance in trait weightings.

named. It may be an artifact of the way the data were organized, i.e., as an ipsative measure.

As may be noted from the factor loadings matrices, the three factors which emerge from each method are practically identical. The first factor is definitely a "scientific" factor. Quantitative and Natural Science load very highly on this factor, and Personality has a very high negative loading. This factor makes intuitive sense. Natural Science and Quantitative are seen as cold, unemotional, and technical calculation, devoid of the warm richness and humanity of Personality. Subjects seem to see Personality as not related to the factual subjects of the curriculum. This factor accounts for 31 per cent of the variance of the token measure, and 25 per cent of the variance on the correlation measure.

The second factor for the token weights and third for the correlation weights has been nick-named the "ballet" factor, since it receives high positive loadings for Physical and Arts. For the token weights analysis, this factor may be considered simply a "non-textbook" factor, since it also has some negative loadings for Quantitative and Natural Science.

The third factor does not make as much intuitive sense as the other two. As has been noted, it may be just an artifact of the situation. One would expect Social Studies and Verbal to be strongly related. It might be considered a "social concern" factor, associated with some idea that it is better to be socially aware than to read Shakespears. There might be just such a factor operating in the correlation weights, but the situation is more complicated for the tokens. The positive loading for Natural Science does not make sense unless the factor becomes something of a "practical" arts factor, negatively related to ivory-tower intellectualism. But the two loadings for Arts and Physical do not make any sense to that breakdown. So this factor remains unnamed, and difficult to explain.

The factor analysis indicates that subjects probably respond to groups of traits rather than to traits individually. The "scientific" and the "ballet" factors make intuitive sense to most traditional breakdowns of the curriculum. Science and math programs are collected together, and the stereotype of the engineer is that of an individual with no personality. Correspondingly, the physical education and aesthetic fields are often disclaimed by traditional academicians. These two are factors that one might expect to find in subjects' view of the academic curriculum.

The third factor is not a science factor; there is also no relationship between it and the first factor in the oblique transformation. Perhaps, therefore, it would best fit in the social concern or social awareness category. Its exact nature is hard to discern.

A way to clarify these factors might be to compute factor scores for subjects and investigate the relationships of these factors to personal characteristics of the subjects.

Summary and Conclusions. Subjects were selected from two strata, people professionally engaged in secondary education and lay people from outside the educational system, and from two locations, in Ohio and in Connecticut. There were 101 subjects in the sample, 52 professionals and 49 non-professionals.

Subjects' values toward the seven traits were sampled by two methods: a) directly, by having them assign tokens to the traits; and b) indirectly, by having them evaluate random profiles of hypothetical students. Subjects' evaluations of profiles were correlated with the profile trait scores to produce the indirect measures of trait value.

Subjects' value judgments were analyzed by an analysis of variance. For the direct method there were significant differences between the empirically derived values. There was also a significant sex-by-trait interaction, the values of the traits being slightly but significantly different for boys and girls. For the indirect method there were significant differences between the trait values, but there was no significant sex-by-trait interaction. For neither method was there any significant difference between educator and lay groups.

For the total sample, and for identifiable sub-groups, the measured order of trait values was consistent across methods. For individual subjects the overall order among trait values was not so consistent across methods. And the actual measured values of the traits, the values given by individuals on each method, were even less consistent. The average reliability coefficient was about .35.

Some significant relationships were discovered between personal variables and trait values, but these relationships were not necessarily consistent for both methods. They also were not found to relate to the groupings in the sampling design.

A principal-components factor analysis indicated that subjects' trait values, fall into patterns which emerge as three factors of trait. The first factor was a "scientific" factor, with high positive loadings for Quantitative and Natural Science and a high negative loading for Personality. The second factor was a "non-textbook" factor, with high positive loadings for Physical and Arts. The third factor was difficult to name. It had a high positive loading for Social Studies, a positive loading for Social Studies, a positive loading for Natural Science, and a high negative loading for Verbal. Very comparable factors emerged from both methods.

The study found that it was possible to discover differential weightings between subject areas, i.e., that measurable systematic differences do occur in the judged importance of academic traits. These differences exhibited themselves whether subjects were asked to rate areas directly by a token assignment, or where asked to evaluate student profiles with scores on these areas. In both cases the order of importance of the traits was almost exactly the same for the total sample and for identifiable sub-groups, with only slight differences among some of the traits.

It would thus seem possible to construct a measure like the bentee, a general measure of educational advancement. With vigorous sampling techniques one could, with relative certainty, measure the educational values of the concerned community. Analysis has shown that the direct measurement of values produces as good an estimation of this population value profile as the indirect method does, with far greater ease of administration.

CHAPTER 5

SUMMARY AND CONCLUSIONS

Introduction. The tools of management science and operations research have not yet been applied to general educational problems. They have been utilized in specific areas where specific goals can be optimized, but they cannot be applied to optimum solutions until the overall goals of education are specified. Prompted by the notable lack of a well-defined measurable goal for educational programs and research, we have proposed that such a measure be developed by empirical methods.

Educators have long labored in a fog of underfined goals. For want of a better measure, the effects of schooling and education have usually been measured in dollar terms, by income and social-class effects, or by standard achievement tests. Although educators are in general agreement that the goal of education is not purely to raise income levels or particular achievement levels, their published statements show little agreement about what that goal might be or how one might measure progress toward it.

The research tested a methodology for empirically defining the goals of education and measuring their relative importance. The success of this methodology would allow the development of a universally applicable measure of educational benefit. This measure, derived empirically from social and cultural values, could be used to evaluate students and programs and could be defined in response to social and cultural changes. It could serve as an effectiveness criterion for rigorous scientific systems analysis.

Summary. The high school curriculum was categorized arbitrarily into seven traits: Quantitative, Arts, Physical, Social Studies, Verbal, Personality, and Natural Sciences. These titles are summative of concise definitions which were developed for each trait. Random student profiles were generated with scores for these traits.

Two samples were selected, the first of 52 people professionally engaged in some aspect of secondary education, and the other of 49 people not associated directly with education. Value judgments were elicited by two methods. Subjects were asked to weight the seven tratis by assigning them tokens. They were also given a set of random student profiles and asked to rate each of these hypothetical students on a continuum.

Hypotheses were generated about the results of these two rating techniques. It was expected that some of the traits, possibly Quantitative and Verbal, which are analogous to the main components of many achievement tests, would be rated significantly higher than others. Other hypotheses were concerned with possible systematic differences in judged trait importance for boys and girls, with differences in trait-importance judgments between professionals and non-professionals, and with the relationship of individual differences in judged trait importance and the personal variables which were measured. The two methods were expected to be somewhat equivalent and highly correlated. Finally, subjects were expected possibly to value trait clusters rather than individual traits so that trait factors might emerge.

The data were prepared and variables constructed for the analysis. Various correlational analyses, analysis of variance, and factor analysis, were used to test the hypotheses.

When judged by the total sample, the order of the traits in importance was practically the same for both methods. For the boys this judged order for the token method was Verbal, Personality, Quantitative, Social Studies, Natural Sciences, Physical, and Arts, with the order of the last two traits reversed for the profile-rating method. For the girls the order by the tokens was Verbal, Personality, Social Studies, Quantitative, Arts, Natural Science, and Physical, with Social Studies and Quantitative reversed, and Arts and Natural Sciences reversed, in the profile ratings.

Two analyses of variance were performed, the first using the values from the token method and the second using the trait values from the profile-rating correlation method. Both analyses showed significant ($p < .001$) differences in trait importance. There was no significant difference between the behavior of the professionals and the non-professionals on either method. Differences within these groups were greater than any identifiable group difference. For the token method there was a significant ($p < .01$) sex-by-trait interaction--some traits were judged more important for boys than for girls and vice-versa--but this interaction did not occur in the profile evaluation method.

The study seems to have produced reliable measures across methods, when values were averaged for groups of subjects. The major difference (between the order derived from the token method and that from the correlation method) is that the significant sex difference in trait importance which occurred for the token-assignment method did not appear in the profile-rating strategy.

There were some significant correlations between personal variables and judged trait importance. With such a large number of correlations, a few significant ones would be expected by chance alone. One of the relationships that did occur was for Location; subjects from Cleveland tended to give more importance to Quantitative and Natural Science and less to Physical and Personality. This "regional" effect could be attributable to some experimenter effect, however, since the two samples were collected by different researchers.

There were also significant relationships for education and age. More educated people judged Verbal and Arts to be more important; older people gave greater importance to Natural Science. Finally, women tended to rate Personality and Arts of less importance for boys than did men.

Conclusions. This study has shown that, even with the measurement error inherent in a newly developed and unfamiliar technique, it is possible empirically to weigh various curricular areas in accordance with their importance for an overall measure of educational advancement and benefit. Thus the application of operations research to the educational system is more than just a hypothetical possibility.

Seven "academic" traits were developed from a traditional breakdown of the school curriculum by logical analysis. By both latent and direct measurement schemes, these seven traits were weighted by 101 subjects and received weights which were systematically different from each other. These trait weightings from each method, were comparable when averaged over groups of subjects.

The direct method of weighting by tokens evidenced some expected differences in trait importance by sex. These differences did not appear when values were sampled latently through judgments elicited on student profiles. There was a great deal of error associated with these judgments about hypothetical students, just as there well might be in unguided human judgment about actual people.

The measurement error in the profile ratings may have been due to subjects' unfamiliarity with the instrument or to difficulties in making such complex judgments with so many bits of data. Since the weightings derived from both methods were so comparable for groups, and the latent method involved so much error, it may be unnecessary to make more than a direct assessment of values. Such a method is easier to administer, and seems to supply at least as much information as a latent value assessment.

Whatever the method, the value profiles, the vectors of trait weightings, were not statistically different for comparisons involving a group of professional educators and one of the non-professionals. Since the groups were not selected randomly, it cannot be considered that no differences occur for income and occupational groups, but in this study such differences were swamped by differences between subjects within groups.

A measure as important, potentially, as the bentee, should clearly be a group measure, and care must be taken that it represent the social group. For if students and schools are to be widely compared on this measure to facilitate scientific evaluation and development of programs, that measure must be derived from the society at large.

Suggestions for Further Research. Some refinements are possible in the instrumentation, and in the selection and definition of the curriculum areas or traits. The present study employed a traditional breakdown of the high school curriculum. Further effort may be applied to a development of the logical or latent structure of the curriculum. Some method, such as latent partition analysis (Wiley, 1967), may be profitably used and recursively applied until a latent structure of the curriculum emerges. The techniques of this study may be applied in each level of such a structure until the point is approached where no meaningful differential weightings can be obtained by empirical methods.

Of greater importance, though, is the investigation of the interpersonal differences which occur in this study. The sample for this study tended to be homogeneous, in some respects, consisting generally of those who were comfortable enough with their educational system. It might be labeled a middle class, middle-income, white, youthful and well-educated group; even blue-collar subjects tended to fall into the same classification.

What could be implemented is a more complex sampling design, one that might sample across class, income, and occupational groups with a more effective recruitment method, i.e., one that would influence more subjects to take part in the study. Such a design would allow greater comparison of different groups and might discover some of the factors which contributed to the large within-group differences occurring in the judgments. That there were no differences between professionals and non-professionals may be due to the similar social class backgrounds of the two samples.

Finally, further study might be applied to comparing direct with latent methods for trait weightings. This study found a very strong relationship between them. With fewer traits the error involved in the profile weighting may be reduced.

However, as is indicated by the correlation analyses, there is low individual reliability for the two measures, at least when there are only 50 profiles. On the average, the variance in the number of tokens assigned to a trait accounts for barely ten per cent of the variance in the measured relationship between profile rating and trait score. Since the relationship is weaker for less important traits, it may be partly a function of the number of trait scores a subject could comfortably consider when making his judgments. It might also suggest, however, that subjects made different judgments in their token assignments and profile ratings, which somehow were cancelled out by the groups as a whole.

Nevertheless, rank order correlation coefficients for the trait importance order for each subject on the two methods has a mean greater than .50. This may indicate that differences were due more to measurement error, or to subjects' inability to weigh as many as seven important scores in making an assessment of a student.

A principal-components factor analysis performed on the seven overall trait measures, weightings combined for both sexes, produced equivalent sets of three factors. The first was a "scientific" factor, with high loadings for Quantitative and Natural Science and a negative loading for Personality. The second has been named the "ballet" factor, with high positive loadings for Physical and Arts. And the third, which is difficult to name, has a high positive loading for Social Studies and high negative loading for Verbal. These are all the loadings for the factors derived from the correlation methods. For the token method, Verbal loads negatively on the "ballet" factor and Natural Science loads positively on the third factor. Subjects' judgments do then, as supposed, fall into response sets.

This study has been but a first step toward estimating an empirically verifiable measure of educational values. The search must continue for a commonly held measure of educational advancement, a standard which can be used to measure the benefits of education for each member of the society, and for the society as a whole.

REFERENCES

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971. Pp. 508-600.

Banghart, Frank W. Educational Systems Analysis. Toronto: Collier-Macmillan, 1969.

Becker, Gary S. Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education. New York: Columbia University Press, 1964.

Becker, Gary S., and Barry R. Chiswick. "The Economics of Education: Education and the Distribution of Earnings," The American Economic Review, 56 (2), May, 1966, 358-369.

Brubacher, John S. Modern Philosophies of Education. New York: McGraw-Hill Book Co., 1962.

Capra, J. R. and R. S. Elster. "A Note on Generating Multivariate Data with Desired Means, Variances, and Covariances," Educational and Psychological Measurement, 31, 1971, 749-752.

Chorafas, Dimitris N. Systems and Simulation. New York: Academic Press, 1965.

Churchman, C. West. Prediction and Optimal Decision: Philosophical Issues of a Science of Values. Englewood Cliffs: Prentice-Hall, Inc., 1961.

Churchman, C. West, Russell L. Ackoff, and E. Leonard Arnoff. Introduction to Operations Research. New YOrk: John Wiley & Sons, Inc., 1957.

Clemans, William V. An Analytical and Empirical Examination of Some Properties of Ipsative Measures. Psychometric Monograph #14. Princeton, N.J.: Psychometric Corporation, 1965.

Coleman, James, Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. Equality of Educational Opportunity. Washington, D. C.: U.S. Government Printing Office, 1966.

Cronbach, Lee J. Essentials of psychological testing (3rd ed.) New York: Harper and Row, 1970.

Cureton, E. E. The accomplishment quotient technic. Journal of Experimental Education. March, 1937. Pp. 315-326.

Dalkey, N. Analyses from a group opinion study. Futures, December 1969, 1, 541-551.

Darlington, Richard B. "Multiple Regression in Psychological Research and Practice." Psychological Bulletin, 69 (3), 1968, 161-182.

Ebel, Robert L. "What Are Schools For?" Phi Delta Kappan, 14, September, 1972, 3-7.

Educational Policies Commission. The Central Purpose of American Education. Washington, D.C.: National Education Association of the United States, 1961.

Ennis, Robert H. "A Concept of Critical Thinking." Harvard Educational Review, 32, 1962, 81-111.

Feyereisen, Kathryn V., A. John Fiorino, and Arlene T. Nowak. Supervision and Curriculum Renewal: A Systems Approach. New York: Appleton-Century-Crofts, 1970.

Flagle, Charles D., William H. Huggins, and Robert H. Roy. Opeations Research and Systems Engineering. Baltimore: The Johns Hopkins Press, 1960.

Flanagan, J. C. Units, scores, and norms. In E. F. Lindquist (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1951. Pp. 695-763.

Freire, Paulo. Pedagogy of the Oppressed. New York: Herder & Herder, 1970.

Gable, Robert K. "The Measurement of Work Values: A Normative Framework." Paper presented at the 2nd Annual Convention of the Northeastern Educational Research Association, Grossinger, New York, November, 1971.

Greene, H. A., V. A. C. Hendon, and M. J. Wagener, et al. Unit scales of attainment. Minneapolis, Minn.: Educational Test Bureau [c. 1933].

Hanoch, Giora. "An Economic Analysis of Earnings and Schooling." Journal of Human Resources, 2 (3), 1967, 310-329.

Hansen, W. Lee, Burton A. Weisbrod, and William J. Scanlon. "Schooling and Earnings of Low Achievers." The American Economic Review, 60 (3), June, 1970, 409-418.

Hare, Vancourt, Jr. Systems analysis: A diagnostic approach. New York: Harcourt, Brace and World, 1967.

Hilden, A. H. and H. M. Skeels. A comparison of the Stanford-Binet scale, the Kuhlmann-Anderson group test, the Arthur point scale of performance tests, and the Unit scales of attainment. Journal of Experimental Education, December, 1935, 4 (2), 214-230.

Howe, Reuel L. Miracle of Dialogue. Greenwich, Conn.: Seabury Press, 1963.

Huysman, Jan H. B. M. The Implementation of Operations Research: Publications in Operations Research. New York: Wiley-Interscience, 1970.

International Business Machine Corporation. PL/I Language Specifications. White Plains, N.Y.: IBM Corp., 1969.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. Inequality: A Reassessment of the Effect of Family and Schooling in America.

Jensen, Leonard Lloyd. Structure of Educational Beliefs: Some Correlates with the Staff of a Suburban School System. Unpublished Ph.D. Dissertation, Wayne State University, 1971.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont California: Brooks/Cole Publishing Co., 1968.

Kneller, George P. Introduction to the Philosophy of Education. New York: John Wiley & Sons, Inc., 1964.

Levin, Henry M. "Schooling and Inequality: The Social Science Objectivity Gap." Saturday Review/Education, 55 (46), December, 1972, 49-51.

McCleary, Lloyd Everald, and Stephen P. Hencley. Secondary School Administration: Theoretical Bases of Professional Practice. New York: Dodd, Mead, 1965.

Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." The Psychological Review, 63, 1956, 81-97.

Mosteller, Frederick, and Daniel P. Moynihan (eds.). On Equality of Educational Opportunity. New York: Random House, 1972a.

________. "A Pathbreaking Report." Frederick Mosteller and Daniel P. Moynihan (eds.). On Equality of Educational Opportunity. New York: Random House, 1972b, 3-66.

Page, Ellis Batten. "Seeking a Measure of General Educational Advancement: The Bentee." Journal of Educational Measurement, 9 (1), Spring, 1972, 33-43.

Page, Ellis Batten, and Thomas F. Breen III. "Two 'Top-Down' Strategies for Educational Evaluation." Address: American Educational Research Association Annual Convention, Chicago, Illinois, April 4, 1972.

Phenix, Philip H. Philosophy of Education. New York: Henry Holt and Company, 1958.

Pritzkau, Philo T. On Education for the Authentic. Scranton, Pa.: International Textbook Company, 1970.

Raiffa, Howard. Decision analysis: Introductory lectures on choice under uncertainty. Reading, Mass.: Addison-Wesley, 1968.

Raths, Louis E., Merrill Harmin, and Sidney B. Simon. Values and Teaching: Working with Values in the Classroom. Columbus, Ohio: Charles E. Merrill Books, Inc., 1966.

Robinson, Donald W. "An Interview with Christopher Jencks." Phi Delta Kappan, 54 (4), December, 1972, 255-257.

Rogers, Carl. Freedom to Learn. Columbus, Ohio: Charles E. Merrill Publishing Co., 1969.

Saaty, Thomas L. Mathematical Methods of Operations Research. New York: McGraw-Hill Book Company, Inc., 1959.

Schultz, Theodore W. "The Rate of Return in Allocating Investment Resources to Education." Journal of Human Resources, 2 (3), 1967, 293-309.

Silberman, Charles E. Crisis in the Classroom: The Remaking of American Education. New York: Vintage Books, 1970.

Slagle, James R. Artificial intelligence: The heuristic programming approach. New York: McGraw-Hill, 1971.

Smith, O. J. M. Feedback Control Systems. New York: McGraw-Hill Book Company, 1958.

Stanley, William O., B. Othanel Smith, Kenneth D. Benne, and Archibald W. Anderson. Social Foundations of Education. New York: The Dryden Press, Inc., 1956.

Stone, James C. and Frederick W. Schneider. Foundations of Education: Commitment to Teaching, Volume I. New York: Thomas Y. Crowell Company, 1965.

Suojanen, Waino W. The Dynamics of Management. New York: Holt, Rinehart and Winston, 1966.

Thut, I. N. The Story of Education. New York: McGraw-Hill Book Company, 1957.

Tucker, L. R. and S. Messick. An individual differences model for multidimensional scaling. Psychometrika, 1963, 28 (4) 333-367.

Van Dusseldorf, Ralph A., Duane E. Richardson, and Walter J. Foley. Educational Decision-Making Through Operations Research. Boston: Allyn and Bacon, Inc., 1971.

Wagner, Harvey M. Principles of Operations Research: With Applications to Managerial Decisions. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969.

Weinberg, Gerald M. PL/I Programming Primer. New York: McGraw-Hill Book Company, 1966.

Whitehead, Alfred N. The Aims of Education. New York: Macmillan, 1929.

Wiley, David E. "Latent Partition Analysis." Psychometrika, 32 (2), 1967, 183-193.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

APPENDIX A

Initial Telephone Contact

______ Hello. (May I speak to your mother or your father?)
______ Hello, is this Mrs. ______________? My name is ______________.
I'm sorry to trouble you. I am a research associate at The University of Connecticut where we are conducting a public-opinion survey to determine what people believe makes someone a "well-educated" high school graduate. You are one of the people I have been contacting in this area of (Connecticut or Ohio) to see if they would be willing to participate in this survey. Would you be willing to help us with your opinion?

(No) I am sorry to have interrupted you. Thank you for your time.

(Yes) Thank you. We have a number of response categories in which we place subjects. To determine which of the response categories your household might fall into, could you tell me what your husband does for a living? (White collar or blue collar occupation?) (At this point, determine which response category this household will be put into - Housewife, Blue Collar Worker, or White Collar Worker.)

A. (IF HOUSEWIFE) We have some forms which are profiles of high school students on various activities, which we would like you to look at and evaluate. This should take about 30 minutes of your time. We have been authorized to pay you $6.00. When would be a convenient time for me to come to your home to show you the material?

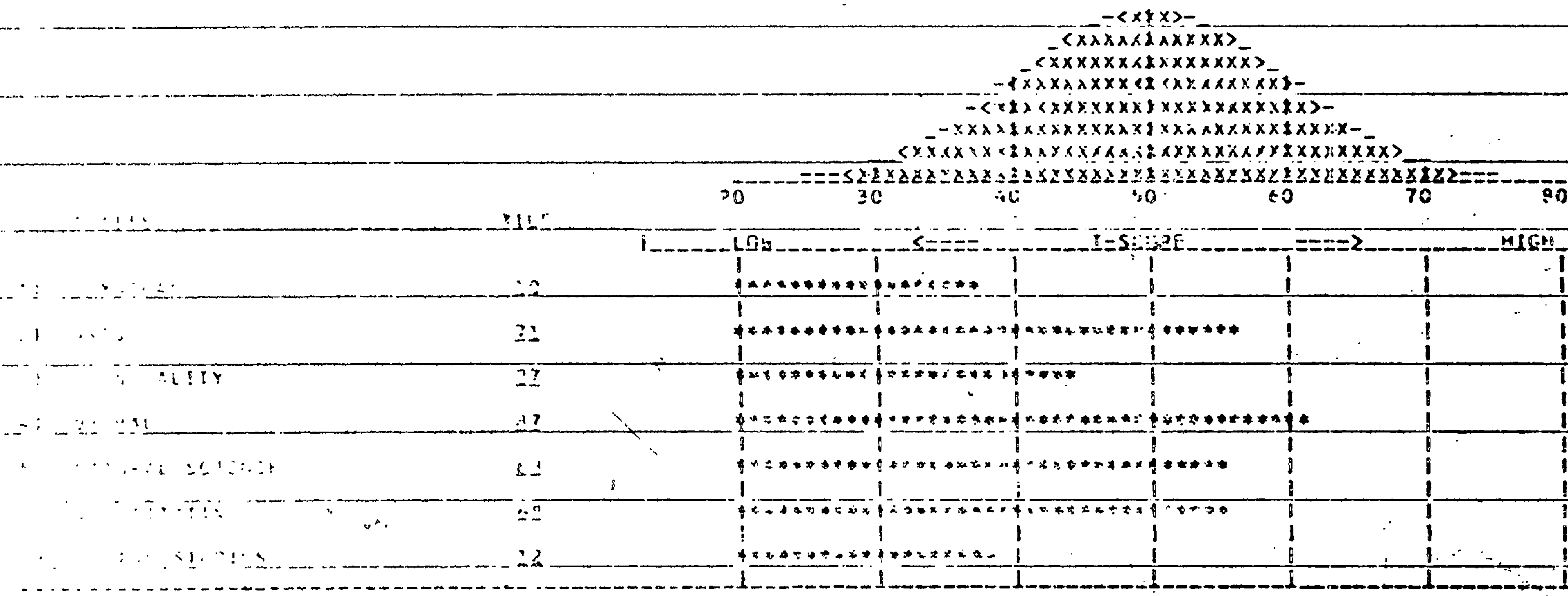
(Make definite appointment)

Thank you for your cooperation. I will see you (at appointment time).

B. (IF BLUE OR WHITE COLLAR WORKER) Do you think your husband would be willing to participate in the survey? We have some forms which are profiles of high school students on various activities, which we would like him to look at and evaluate. It should take about 30 minutes of his time. We have been authorized to pay him $6.00. (IF WIFE AGREES) When would be a convenient time for me to come to your home to show your husband the materials? (Make definite appointment). Thank you for your cooperation. I will see you and your husband (at appointment time).

APPENDIX B

PROFILE OF TWELFTH-GRADE BOY



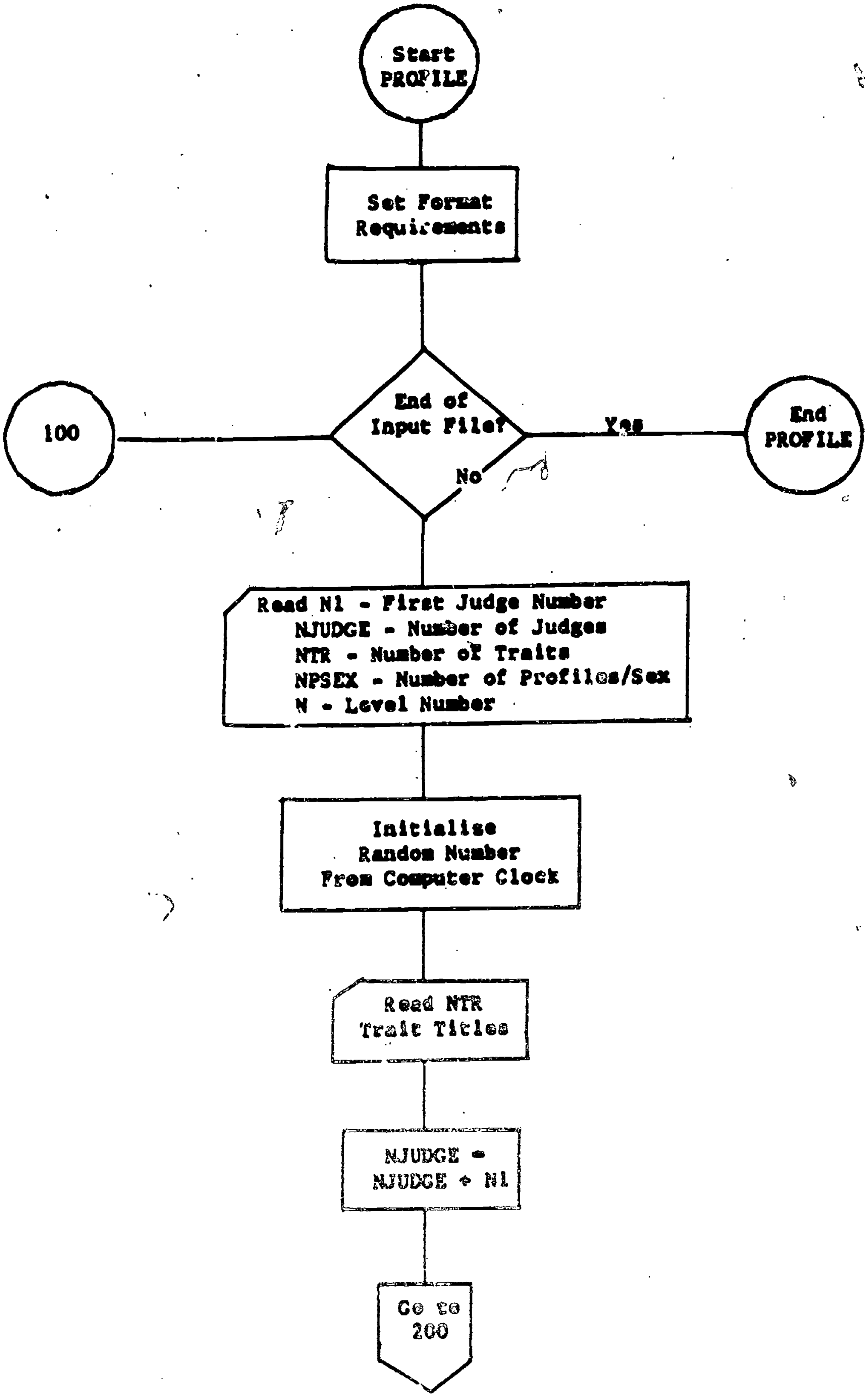
YOUR OVERALL READINESS

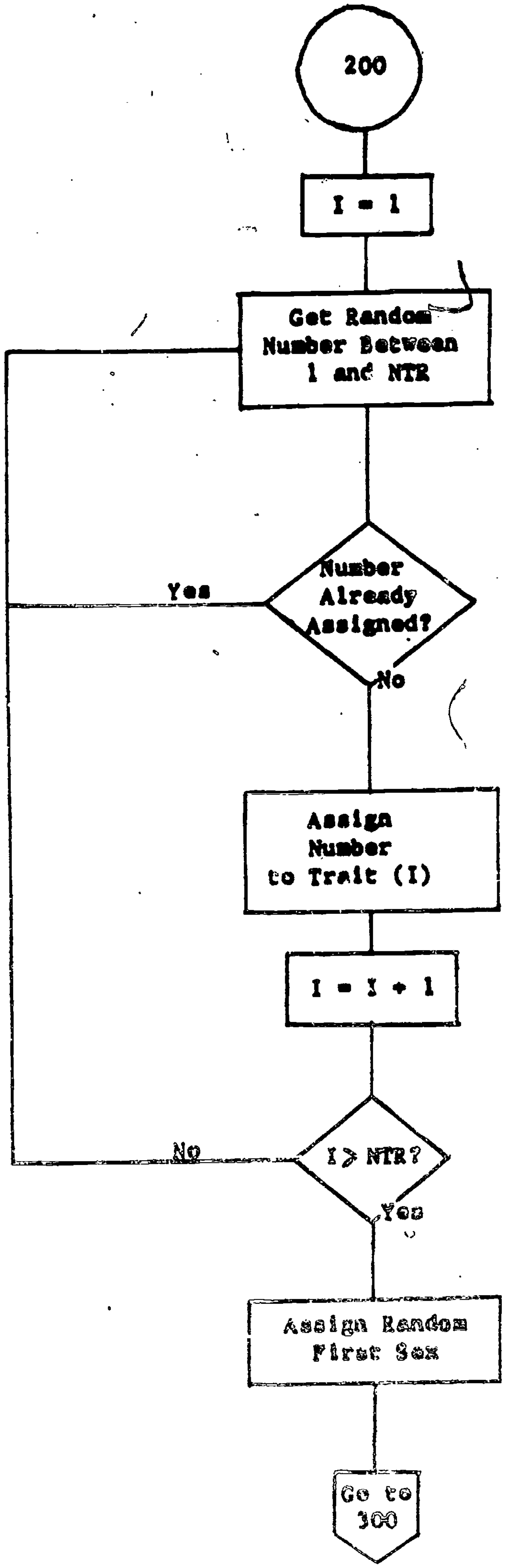
1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10

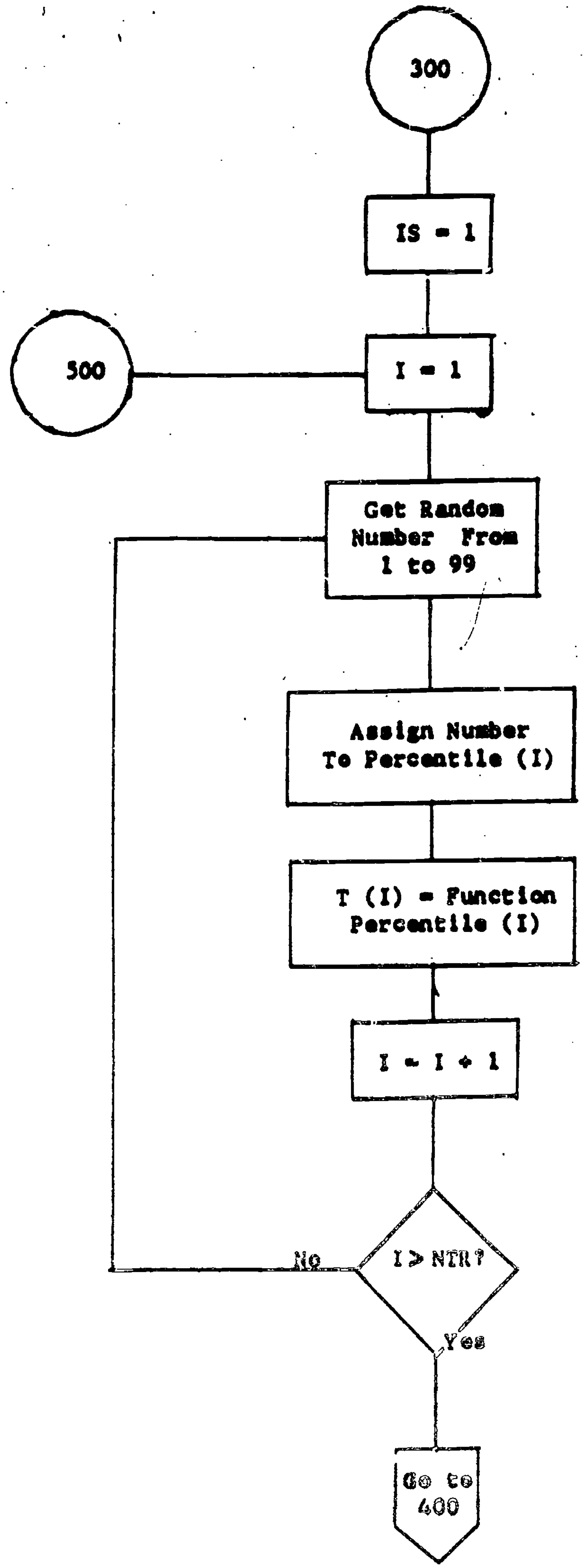
VERY POORLY PREPARED — AVERAGE IN PREPARATION — EXTREMELY WELL PREPARED

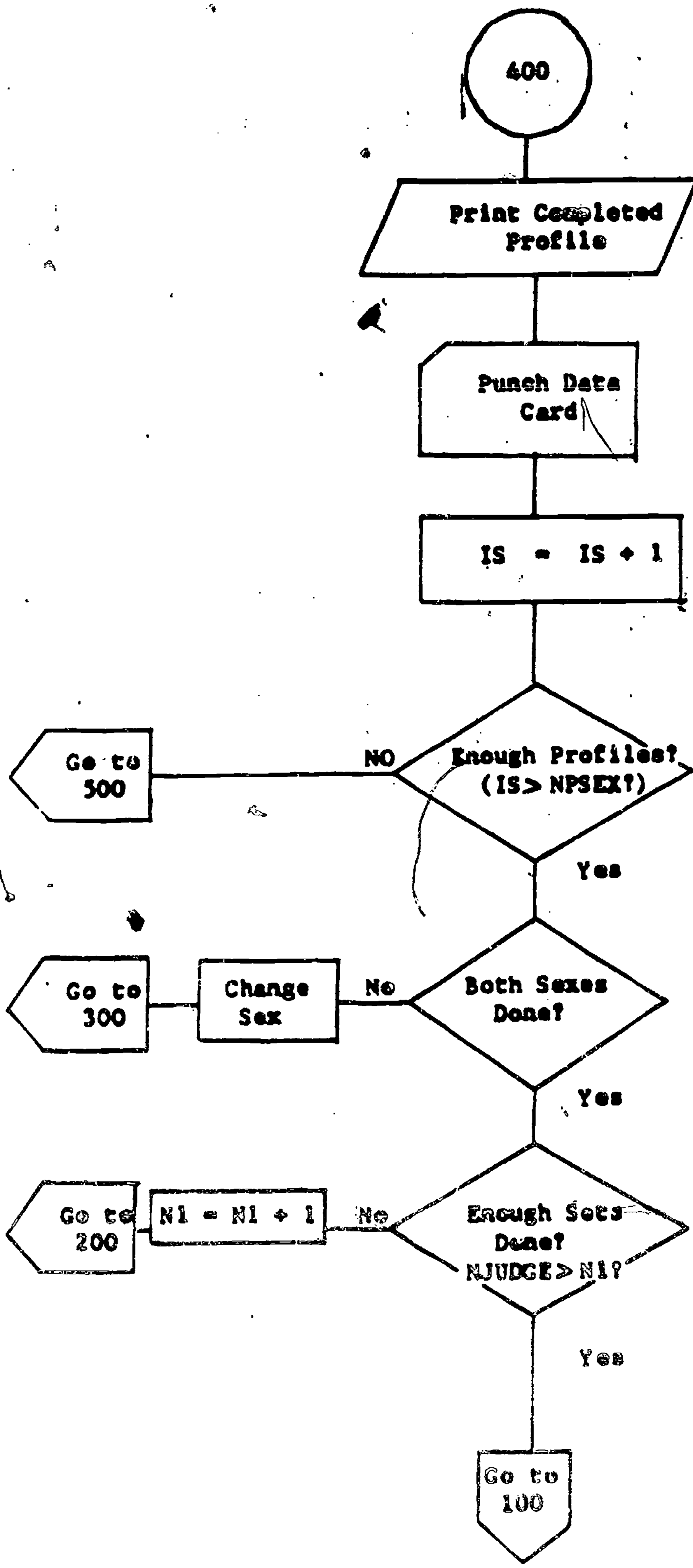
120126

APPENDIX C


Start PROFILE
Set Format Requirements
End of Input File?
Yes
No
End PROFILE
100
Read N1 - First Judge Number
NJUDGE - Number of Judges
NTR - Number of Traits
NPSEX - Number of Profiles/Sex
N - Level Number
Initialize Random Number From Computer Clock
Read NTR Trait Titles
NJUDGE = NJUDGE + N1
Go to 200


200
I = 1
Get Random Number Between 1 and NTR
Number Already Assigned?
Yes
No
Assign Number to Trait (I)
I = I + 1
I ≥ NTR?
No
Yes
Assign Random First Sex
Go to 300


300
IS = 1
500
I = 1
Get Random Number From 1 to 99
Assign Number To Percentile (I)
T (I) = Function Percentile (I)
I = I + 1
I > NTR?
No
Yes
Go to 400


400
Print Completed Profile
Punch Data Card
IS = IS + 1
Enough Profiles? (IS > NPSEX?)
NO
Go to 500
Yes
Both Sexes Done?
No
Change Sex
Go to 300
Yes
Enough Sets Done? NJUDGE > N1?
No
N1 = N1 + 1
Go to 200
Yes
Go to 100

APPENDIX D

DESCRIPTIONS OF THE TRAITS

QUANTITATIVE

The use of numbers in school and in life. This refers to arithmetic computation and reasoning, algebra, geometry, statistics, trigonometry, and, for the better students, perhaps beginning calculus and computer programming. It includes the ability to use math in applied problems, and to read graphs and tables. A high score would imply strong methematical achievement and advancement. A low score would mean low or liited success in numeric operations.

ARTS

Activities, skills, and concepts usually considered esthetic: drawing, and painting, modeling clay, and knowledge of famous artistic work of others; musical performance, composition, and appreciation, performance in drama, interpretative reading, and poetry; photography and film-making; and multi-media presentations. A high score would imply ability in more fields, or perhaps distinction in one or more. A low score would mean a noticeable weakness in esthetic ability.

PHYSICAL

Personal health and hygiene, strength, posture, stamina and agility, as well as performance in sports and gymnastics. Sports may include team sports such as basketball, and individual sports such as tennis or bowling. Included also are such activities as camping and related skills. A high score would imply superior vitality and athletic ability. A low score may mean some physical or sensory weakness.

SOCIAL STUDIES

Knowledge of history, culture, civics, social problems, anthropology, psychology, politics, government, and citizenship, so far as these can be expected to be mastered through the twelfth grade. A high score would mean advanced knowledge of many of the fields associated with the political, social, and cultural behavior of the human race. A low score would imply little knowledge of these facts and principles.

VERBAL

The effective use of language skills; reading, writing, speaking, conversation, prose, and poetry. The ability to write speeches and debates. Vocabulary and verbal reasoning. A high score would imply ability in many of these fields, probably ability to use formal grammar and rhetorical structure, and some ability in foreign language as well. A low score would mean lack of facility to express oneself clearly and with accepted usage, and poor reading ability.

PERSONALITY

This refers to many social behaviors of the student, his habits, and characteristics of his relationships to other people, with some regard to the morality and propriety of those behaviors. Included are social popularity, personal appearance, apparent attitudes toward those around him, ability to meet people and interact constructively, to develop strong and lasting friendships. A high score would imply definite strong personal leadership and character. A low score would imply a more withdrawn, ineffective, or deviate personality.

NATURAL SCIENCES

Ability and achievement in the physical and biological sciences, within the expectations of twelfth-grade students. Included are knowledge, skills, and concepts in general science, health, environment, biology, physics, chemistry, astronomy, botany, physiology, zoology, and similar subjects which might be learned by students of this age. A high score would mean high general learning in these fields, with perhaps distinction in one or more. A low score would imply poor performance in these areas, a lack of understanding of the principles or facts of science.

APPENDIX E

INSTRUCTIONS TO THE EVALUATORS

As I mentioned, we are doing a type of public opinion survey; we are doing a survey of educational values. We are trying to find out what people mean when they use such terms as "well-educated" or "well-prepared". (We intend the terms here to be somewahat synonymous.)

What might we mean when we say that a person is "well-educated" or "well-prepared?" There are many ways of answering this question, of making such a judgment, depending on what qualities and abilities one might regard as important. We are asking you to help us decide, and to let your own values and beliefs be your guide.

So we are asking you to judge fifty (50) twelfth-grade students, 25 boys and 25 girls, to rate them for how "well-prepared" or "well-educated" they are. These boys and girls will be presented to you by means of a student "profile", such as this one, and all that you will know about them is contained in the trait scores, percentiles and T-scores, on the profile. (These students are not real, the profiles were madeup for the purpose of this study.)

(Sample profile presented.)

To help you understand what these profiles mean, here are some rough descriptions of each trait.

(Descriptions of the traits are given to the evaluator in the order printed on the profile.)

You should study these descriptions until you have a fairly solid understanding of what kinds of abilities are included in each one and of how you feel about the importance of each for an educated person. No importance should be attached to the order of the traits. They are arranged in a chance order: for each set of profiles the computer randomly rearranges the trait order.

Since the descriptions of the traits may be somewhat arbitrary, please read them carefully and I will try to answer any questions you may have about them.

(Answer questions.)

Now it is up to you to decide which of the traits are more important for you. It is unlikely that you would consider that they are exactly equal in importance; some will probably be more important than others. Please try to consider the exact importance you would give to each one.

If you are ready then we will continue.

A. (If tokens are first - otherwise, got to B.)
(Before we go back to the sample profile) please spread out the trait cards. Here are 50 pennies. I would like you to assign pennies to the traits, by placing them on the card, according to how important you believe each one is for a twelfth-grade (boy or girl - first one listed on the profiles). If you should decide that only one is really important, then you would put all the pennies on that card. If you think one or more of the traits have no importance, then you would put no pennies on these cards. Most likely, however, you will feel that most of the traits deserve some pennies. Remember, this time you are weighting them for a twelfth-grade (boy or girl, as above).

Do you have any questions? (Answer the questions.)
Okey, please put the pennies on the cards. When you are satisfied with the way you have arranged them, I will record each number. These numbers should add up to 50. (Wait and record.)

Now do the same thing again for a twelfth-grade (boy or girl, opposite sex from above). When you are finished I will again record it.

B. Now you should go back to the profile of the twelfth-grade (girl or boy) which I gave you, to understand in what ways he/she is outstanding, average, or poor. To help you understand the numbers and the bar graphs used, here is a brief explanation of the profile.

There are two scores for each trait. In the column headed %ILE, you will find the percentile score achieved by this student, compared with other twelfth-grade students. Do you know what a percentile is? To refresh you, here is a brief explanation.

In the sample profile the (boy/girl) has a percentile score of ______ for the trait ______. (Pick out a central one.) This means that this particular student scores better than ______% of the other twelfth-graders, and ______% score better than he does. Let us look at a higher one. (Pick out a high one.) What does the ______ after ______ mean? It means that this student is very good indeed in ______, compared with other students, better than ______% of them and only ______% are better than (he/she) is. And if we look at a low one, (Point out low one) like ______, we see that this student is only at the ______ percentile, (he/she) is better than only ______% of the other twelfth-grade students, and ______% are better than (he/she) is. And so on.

Do you have any questions about this part? (Answer questions.)

Another way of describing the students is by the bar graphs on the right. These represent T-scores, which are a transformation of the percentile distribution to the normal curve. To understand these scores we draw a picture of the normal curve above them. You can see from the curve how these scores tend to bunch together toward the middle, as is indicated by the height of the curve, in that area, and how very few scores fall toward the ends. For the percentiles, however, the same number of people get each score. The T-score distribution shows that the one percentile point difference in score at the middle of the curve is not anywhere near as meaningful as a one-point difference would be at the ends.

As I said these scores are just translations of the percentile scores to the normal curve, to bunch the middle scores and spread out the ends. The scores range from about 20 to 80, with the extremes being very rare. The higher the percentile score, the longer will be the bar. You should be able to tell by looking at the bar, compared with the curve above, just about how strong or how weak the student is on any particular trait.

Remember that both the percentile score and the bar graph are just two ways of showing how the student compares with other students on the particular trait in question.

Do you have any further questions? (Answer questions.)

Now that you are familiar with the profile and the descriptions of the traits, I would like to ask you to look again at the profile and to make some decision about how "well-prepared" or "well-educated" you think this student is.

If you value the traits in which (he/she) is high and don't think the low ones are very important, you may decide that (he/she) deserves a high score, perhaps an 8, 9, or even a 10. Then you should make a mark at the point where you want it to count along the line marked YOUR OVERALL JUDGMENT which appears below the profile.

On the other hand, if you regard the low traits as most important, and the others less so, then you may want to give (him/her) a low score, and make a mark in the 1, 2, or 3. Or you might value the middle traits, in which case you might give him a middle score. Decide now how you feel, and make a mark.

The easiest mark for me to interpret would be a vertical line at the point you desire. Remember that there are more than 10 points on the scale.

Any questions? (Answer questions).

Now I will give you the 50 student profiles which will look the same as the one we have been discussing, except that the scores for the traits will be different. Also, the first 25 will be (boys/girls) like this one, and the second 25 will be (girls/boys). On each one please make the same kind of judgment you just made on the sample, and make your mark somewhere along the line. Remember that there are no "right" or "wrong" answers. The only right answer is the one which agrees with your own best judgment about the student presented in the profile.

Any last questions? (Answer questions.)

Are you ready to begin? Then here are the profiles. Try to make judgments at various points along the line. I will be back when you are finished so I can collect the evaluations and give you your check.

(If the tokens come after, go to A.). I will have another short thing for you to do then. (After, for the tokens, the instructions will be the same as above, except that "before we go back to the sample profile" will be deleted. Those three paragraphs will be used.

(When the profiles are finished, they will be collected and the evaluator will be given his stipend. The researcher may discuss the project with the subject, if there is time and interest.)