REPORT RESUMES

ED 012 370 AA 000 122

MULTIVARIATE PROCEDURES FOR STRATIFYING SCHOOL DISTRICTS. FINAL REPORT.

BY- MILLER, DONALD M. AND OTHERS
WISCONSIN STATE DEPT. OF PUB. INSTR., MADISON

REPORT NUMBER BR-5-8043 PUB DATE MAR 67
WISCONSIN UNIV., MADISON
REPORT NUMBER CRP-S-310
EDRS PRICE MF-$0.75 HC-$7.04 176P.

DESCRIPTORS- *STATISTICAL ANALYSIS, *SCHOOL DISTRICTS, *ORGANIZATION, *SCHOOL DEMOGRAPHY, STATISTICAL STUDIES, POPULATION DISTRIBUTION, EDUCATIONAL RESEARCH, *EDUCATIONAL STRATEGIES, DATA PROCESSING, MADISON

RESEARCH WAS CONDUCTED ON THE PROBLEM OF CONSTRUCTING SCHOOL DISTRICT CLASSIFICATION SCHEMES TO DESCRIBE AND DEMONSTRATE AN APPROACH TO CLASSIFICATION AND ACCREDITATION WHICH TAKES INTO ACCOUNT THE INDIVIDUAL CHARACTERISTICS OF EACH DISTRICT. PART 1 OF THE REPORT DISCUSSED THE SCIENTIFIC ISSUES AND CONCEPTS CONCERNING SCHOOL DISTRICTS AS A POPULATION OF ORGANIZATIONAL ENTITIES, ALONG WITH THE TECHNICAL AND STATISTICAL PROBLEMS INVOLVED IN STUDYING A SAMPLE OF THESE ENTITIES. PART 2 PRESENTED THE STATISTICAL METHODS AND TECHNIQUES APPROPRIATE FOR THE STUDY OF A POPULATION OF SCHOOL DISTRICTS AND THE DATA PROCESSING TECHNIQUES NECESSARY FOR MANIPULATION OF AVAILABLE INFORMATION. PART 3 PROVIDED SEVERAL EXEMPLARY APPLICATIONS WHICH DEMONSTRATE THE BROAD UTILITY OF THE APPROACH AND THE FLEXIBILITY OF THE GENERAL PROCEDURES WHEN APPLIED TO SPECIFIC RESEARCH PROBLEMS. (GD)

S-3

MULTIVARIATE PROCEDURES FOR STRATIFYING SCHOOL DISTRICTS

WISCONSIN STATE DEPARTMENT OF PUBLIC INSTRUCTION
THE UNIVERSITY OF WISCONSIN

U. S. OFFICE OF EDUCATION
PROJECT NUMBER 5-8043-2-12-1

MULTIVARIATE PROCEDURES FOR STRATIFYING SCHOOL DISTRICTS.

The Final Report of
U. W. Office of Education Project
"A Study of Certain Characteristic
Patterns of Elementary School Districts in Wisconsin"

Prepared by
Donald M. Miller, Robert F. Conry
David E. Wiley, Richard G. Wolfe

Investigators
Archie A. Buchmiller and Donald M. Miller

Special Assistance on the Project was Provided by
Robert E. Clasen and Robert M. Pruzek

March, 1967

The Research Reported Herein Was Partially
Supported by a Grant from the
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
Project Number 5-8043-2-12-1
to the
WISCONSIN STATE DEPARTMENT OF PUBLIC INSTRUCTION
in Cooperation
with the
UNIVERSITY OF WISCONSIN
School of Education
Instructional Research Laboratory

```
      SUBROUTINE OUTPUT
      DIMENSION T(64), T13(2,64), T14(2,64), T15(3,64), S(11,64)
      DIMENSION KD(6), X(31), MAP(700)
      COMMON H(10)
      COMMON /A/ A(700,10), ID(700)
      COMMON /KEY/ KK(700), LL(700)
      COMMON /NN/ N, NOBS
      IF ( N .GT. 6 ) RETURN
      NS= 2**N
1     DO 12 I=1,NS
      T(I)= 0.0
      DO 11 J=1,2
11    T13(J,I)= T14(J,I)= T15(J,I)= 0.0
      T15(3,I)= 0.0
      DO 12 J=1,11
12    S(J,I)= 0.0
      DO 13 I=1,NOBS
      II= LL(I)
13    MAP(I)= KK(II) + 1
2     DO 25 I=1,NOBS
      READ(1) XXX, X
      K= KK(I) +1
      T(K)= T(K) +1.0
21    IF ( X(13) .GT. 1.5 ) GO TO 211
      T13(1,K)= T13(1,K) +1.0
      GO TO 22
211   T13(2,K)= T13(2,K) +1.0
22    IF ( X(14) .GT. 1.5 ) GO TO 221
      T14(1,K)= T14(1,K) +1.0
      GO TO 23
221   T14(2,K)= T14(2,K) +1.0
23    IF ( X(15) .GT. 1.5 ) GO TO 231
      T15(1,K)= T15(1,K) +1.0
      GO TO 24
231   IF ( X(15) .GT. 2.5 ) GO TO 232
      T15(2,K)= T15(2,K) +1.0
      GO TO 24
232   T15(3,K)= T15(3,K) +1.0
24    DO 241 J=16,26
241   S(J-15,K)= S(J-15,K) + X(J)
25    CONTINUE
3     I2= 0
      REWIND 1
      NP= N+1
      DO 31 I=NP,6
31    KD(I)= 1H
      DO 5 III=1,NS
      I= NS+1-III
      MX= I-1
      DO 33 J=1,N
      JJ= NP-J
      IF ( (MX/2)*2 .EQ. MX ) GO TO 32
      KD(JJ)= 1H+
      GO TO 33
32    KD(JJ)= 1H-
33    MX= MX/2
      PRINT 34, H, KD
34    FORMAT ( 1H1 10A8, 10X, 8HSTRATUM  6A2 / )
      IF ( T(I) .GT. 0.0 ) GO TO 4
      PRINT 35
35    FORMAT ( // 10X, 5HEMPTY )
      GO TO 5
4     I1= I2+1
      DO 41 J=I1,NOBS
      IF ( MAP(J) .NE. I ) GO TO 42
      CONTINUE
```

Classification is one of the fundamental concerns of science. Facts and objects must be arranged in an orderly fashion before their unifying principals can be discovered and used as the basis for prediction. Many phenomena occur in such variety and profusion that unless some system is created among them they would be unlikely to provide any useful information.

SOKAL, 1966

PREFACE

What is the subject matter of this report?

This report describes an empirical approach to the study of organizational and demographic characteristics of school districts. The particular approach set forth is the result of research work which was initiated as "A Study of Certain Characteristic Patterns of Elementary School Districts in Wisconsin". The initial conceptualization and plan of the study were relatively restricted, but as the work progressed and certain methodological problems were solved it became clear that the empirical approach and technical developments had considerable generality in terms of their applications and implications.

For the purpose of stating clearly the broad applications *and implications of the research methodology, the conceptualization of the project was reformulated as "Multivariate Procedures for Stratifying School Districts". This conceptualization is reflected in the organization of the report, which consists of three major parts:*

Part One discusses the scientific issues and concepts concerning school districts as a population of organizational entities, and the technical and statistical problems involved in studying a population or sample of these entities.

Part Two presents the statistical methods and techniques appropriate for the study of a population of school districts, and the data processing procedures necessary for manipulation of available information.

Part Three provides several exemplary applications which demonstrate the broad utility of the approach and the flexibility of the general procedures when applied to specific research problems.

Of what importance is the research reported herein?

A first glance at the body of this document will understandably dismay the non-technical reader. The text is laced with technical terms such as "descriptive complexity", "replicability", and "generalizability". So-called jargon, on first sight, often evinces doubt about the meaningfulness and relevance of the research from the viewpoint of practicing educational administrators. Certainly the non-technical reader cannot be expected to digest and assimilate the various "gymnastics" of statistical manipulations such as those discussed in Part Two..

The apparent complexities of the mathematical and statistical manipulations should not, however, cloud the fundamental issue to which this report is addressed, for these issues have great importance for administrators in various types of educational organizations.

Practicing educational administrators are repeatedly confronted with problems involving a large number of educational collectivities. For example, the director of a Regional Educational Laboratory is concerned with numerous and varied school districts contained within his territory. On one hand, it is impossible for him to attempt to plan an activity in which each individual district will be considered as a discrete entity with its own individual characteristics. On the other hand, it is foolhardy for him to consider dealing with the population of districts as though they were homogenized, or as if each was highly similar to the others. Practicality requires that he find some way of differentiating, for a variety of purposes, among districts so that those having similar characteristics are grouped together while those which are dissimilar are separated. One approach to solving such a problem has been to classify districts according to student enrollment size. Such a classification achieves a simplification of the situation and the complexity is reduced. However, as most practicing administrators are aware, such an approach tends to oversimplify the situation. It fails to recognize other important characteristics of the districts, such as the socio-economic milieu within which they operate, the internal complexities of the agencies, and the demographic character of the school attendance areas.

The need for differentiating among school districts is one of the basic problems faced by state officials responsible for the distribution of state aids to local school districts. The key question posed by this problem is: How can the individuality of each district be recognized in a classification and/or accreditation scheme which will allow effective and equitable distribution of funds, consistent with the objectives and criteria of a state support program? In Wisconsin, for example, a trichotomous classification of districts has served for a long time as the basis for distributing state aids. Historically, use of these three classes has achieved the purpose of supporting, stimulating, and motivating local districts in the improvement of their educational programs. Now, however, over 60% of Wisconsin districts fall in the same category and are treated as if their needs were equal and of the same kind. This report does not provide any specific, improved solution to this general problem. It does, however, sharply focus attention on the problem of constructing district classification schemes, by describing and demonstrating an approach to classification and/or accreditation which takes into account the individual characteristics of each district.

The crux of the classification and/or accreditation problem is that it is not feasible to treat each district as an individual, discrete entity by developing idiosyncratic rules for say, distributing state aid. Nor is it efficient, effective or fair to treat all districts as though there were no differences among them. District needs vary over the years due to changes in the school-community being served and in the organizational character of districts. These comments are not meant to imply that the problem of classifying districts has been neglected, or that it is an unstudied problem. Concern for differentiating among districts is continually evidenced. For example, a practicing administrator may say "You don't launch a new program of in-service teacher training in 'down-state' districts in the same way that you initiate such a program in our largest city". But how can the individuality of districts be empirically recognized and pragmatically taken into account?

The purpose of this prefactory discussion--and indeed of the total report--is to emphasize the importance of recognizing that there are many characteristics which differentiate school districts, and that they should be and can b^ used for practical purposes. Furthermore, it is important to establish operationally and empirically the exact nature of the differential relationships which exist among a set organizational entitites such as school districts. In this vein, the basic function of this report is to describe a particular set of tools useful in the pragmatic consideration of school district individuality. The exact procedures described in this document are not recommended as ultimately ideal solutions, but they do represent exemplary solutions to the basic problem. And, they demonstrate that empirical and systematic approaches can be made to the difficult and complex--but important--task of accounting for individual differences within a population of school districts.

The utilities afforded by an appropriate, empirical classification of school districts are numerous and varied. Several possible uses are discussed in Parts I and II of this report and are illustrated in Part III.

One of the most noteworthy uses of a stratification (classificiation) scheme is in economizing the study of school districts, in terms of time, effort and money. For example, the need for securing information about an entire state educational system does not mean that all districts in the state must be surveyed. A stratified random sampling scheme can be used for gathering information from a fraction of the district prpulation, and only minute losses in the precision of summary data will occur. The use of such efficient sampling techniques is accompanied by significant reduction in the costs of information collection. One example of this type of utility is given in Part III Section D of this report.

From what viewpoint and background was this research undertaken?

The style and tone of this report might understandably suggest to the reader that the research was an exercise in the manipulation of a grand conceptualization invoking a methodological, statistical paradigm. The initial formulation of the research plans arose as a response to a particular problem concerning the selection of a sample of elementary school teachers in the state of Wisconsin (see Part Three). Initial discussions for the purpose of achieving this goal took place in the fall of 1963 between staff members of the Wisconsin State Department of Public Instruction and a research team of the University of Wisconsin Instructional Research Laboratory. During the following twelve months, much of the preliminary planning and preparatory work was accomplished and was culminated by the submission to the U. S. Office of Education of the proposal for the project as it was initiated in June, 1965. This preliminary work involved preparation of the specific details of the research plans, design of the data matrix, methods of applying the analytic results to sampling problems, decisions about variables which would form the input for the computational analyses, and outlines and tests of some of the necessary computer programs.

Thus the approach to investigating school districts described in this report evolved inductively from continuing consideration of a particular problem until the broad, general implications of the approach became apparent to the research team. Several comments concerning the history and developments of the research work are appropriate for the purposes of indicating the favorable conditions under which the project was executed, and of describing some of the logistic difficulties which the researchers encountered.

The planning activities were co-incident with the development of a data processing system in the Wisconsin State Department of Public Instruction (WSDPI). Without this system, the organization, tabulation, and storage of the input data would have been a task of such magnitude that it is quite likely that the work reported here would never have been undertaken. Datamation has exerted a significant impact on the management of education. However, the technical aspects of such systems are frequently not understood by personnel inexperienced in the logistics of automatic computing machinery. Some of these technical difficulties confronted the researchers for the duration of the project. For example, the WSDPI data processing system used IBM machinery, while the University of Wisconsin Computing Center used a CDC 1604 system. These two computer systems are not precisely compatible for writing and reading data stored on magnetic tapes--a problem which can be corrected by improvements in the manufacture of the electronic mechanisms. One specific difficulty of this kind arose when data written with IBM equipment had to be read by a CDC mechanism. This resulted in errors of various kinds during transfer of tape-stored data.

Such difficulties can be solved. But the solutions take time and continuous monitoring of computations, so that mistakes can be identified and corrected. It should be noted, however, that such difficulties are relatively minor when the alternative is considered:. the manual manipulation of thousands of punched cards gathered from several different sources.

Another pertinent historical comment concerns the analytic methods and their computational applications. The procedures described in the report involve the application of complex multivariate statistical techniques for the purpose of summarizing a large quantity of data entries and variables. The analytic design required the computation of exact distributions of factor scores; two computational techniques were necessary to construct the scores. The psychometric basis for one of these techniques depended on work which C. W. Harris completed in 1962. Harris' work extended a basic mathematical formulation of L. Guttman which was published in 1953. The application of these statistical

procedures to the study of school districts, as described in this report, is believed by the authors to be the first such use in educational research.

The researchers wish to make special acknowledgement for cooperation and assistance provided by the following institutional departments:

> The University of Wisconsin Graduate Research Committee; and the University of Wisconsin Computing Center, which operates with the support of The National Science Foundation and the Wisconsin Alumni Research Foundation.
>
> The Data Processing Division, Wisconsin State Department of Public Instruction.

A special note of appreciation is extended to the following individuals, who gave most helpful assistance: Donald E. Russell (Director), Barbara Berg, and Alfred J. Freitag, all of the Data Processing Division, WSDPI; Dale O. Irwin, Director of Research, WSDPI; Philip Lambert (Director), Emmy Alford, Dorothy Hougum, and Mona Meister, all of the University of Wisconsin, Instructional Research Laboratory.

The authors are grateful to Walter Johnson, of the University of Wisconsin Extension Duplicating Service, for his very able assistance with the details of producing this document.

A.A.B.	-	W.S.D.P.I.
D.M.M.	-	U.W. - I.R.L.
R.F.C.	-	U.W. - I.R.L.
R.G.W.	-	U.W. - I.R.L.
D.E.W.	-	now at U. Chicago

March, 1967

TABLE OF CONTENTS

PART ONE

PART TWO

PART THREE

LIST OF TABLES

LIST OF FIGURES

PART I: PURPOSE AND GENERAL APPROACH

Section A - Rationale: The Need for a Multivariate Stratification Scheme

Section B - Objectives: The Desired Outputs of the Project

Section C - Overview: The Algorithm and Its Empirical Perspective

PART I

SECTION A

RATIONALE: THE NEED FOR A MULTIVARIATE STRATIFICATION SCHEME

The Local Education Agency (LEA or school district) is a focus of study or a sampling unit in numerous and varied educational research investigations.[1] There are two general reasons for researchers being concerned with the LEA.

One reason is that district sampling may be a necessary step in an hierarchial sampling plan, the final target of which is schools or persons. Because of administrative considerations, such a plan requires that a district (or several districts) must first be selected in order to clear channels for gaining research access and cooperation. This is especially true if district personnel are to be active in the research process (e.g. see Ryans, 1960, or Carter and Sutloff, 1960).

A second reason involves the direct study of district characteristics. Occasionally differences among districts are stated to be an important factor, or major source of variation. A preliminary goal of some such research is to describe district characteristics in a limited empirical fashion. Some studies have pursued this in depth, endeavoring to define or discriminate among "types" of districts using measures based on some broadly conceived criterion variables. One of the more important and popular kinds of criterion variables is the adaptability of various types of school systems--often defined in terms of organizational transformation or of the adoption of educational innovations. An early and well-known study which endeavored to discriminate among types of districts on an adaptability criterion variable was executed by Mort and Cornell (1941);

[1] Reported investigations are distributed throughout much of the educational research literature; two sources in which they are frequently cited are particular issues of the Review of Educational Research, entitled "Educational Organization, Administration and Finance" (October, 1961 and October, 1964).

a more recent study with a similar purpose has been reported by Carlson (1965).

But a review of relevant literature indicates that approaches to sampling or studying school districts in educational research can be generally characterized as unsystematic or otherwise inadequate (see Cornell, 1960). The problem confronted by this project was to develop an approach to the characterization and sampling of LEAs which would overcome or circumvent certain typical methodological pitfalls. A systematic methodology for studying LEAs is becoming increasingly necessary because LEAs are becoming increasingly important; the local district has become more powerful as the organization through which educational policies and techniques are developed and channeled. Also, the development of a systematic methodology is now feasible because of progress in the applications of two research tools: one, high-speed electronic computing machinery and associated data banks and, two, new multivariate data analysis techniques.

In the remainder of this section, some common research practices are described and exemplified and inferences are drawn which provide a background for discussing the particular multivariate approach developed in Part II and demonstrated in Part III. Section B of this Part is given to a description of project objectives, and Section C provides an overview of the development and use of the methodological algorithm.

This section has been titled, "Rationale: The Need for a Multivariate Stratification Scheme". Establishing such a need has two facets: first, the preceeding paragraphs have indicated that a systematic methodology for studying LEAs is important in educational research; second, the remainder of the section will indicate that such a system is not found in the reported literature. But the concern in the review of literature will not be for "criticisms", in the negative sense of the word. Rather an attempt is made to educe and evolve from the literature criteria for development of a methodology.

Inter-district variation is not always an acknowledged source of variance in educational research, but it is an implicit one in any study which treats districts as sampling units or observational entities. For instance, many studies collect and process data from several districts to ensure "representativeness", but their reports do not include information about the dimensions used to sample districts--if, indeed, dimensions were used--or about the variation of the criterion characteristics among the districts. Typical of sampling descriptions given in such reports is the statement that "Figures below are based on the endorsements given to variations of attitude by fifty northeastern New Jersey secondary school principals" (Berthold, 1951), or "The answers to these questionnaires provide basic information concerning 433 parochial elementary schools in 29 states... Teachers reported on 41 different schools while they attended the regular summer session. There is no reason to believe that this is a skewed sample of parochial school teachers, or that these women are biased in a way that other similar teachers would not be" (Fichter, 1958). When reviewing the literature on the social backgrounds of teachers, Charters (1963, p. 771) concluded that "it is impossible....to compare the results of one study with another."

Whether systematic individual differences among districts affect research results explicitly or implicity, the results of studies which involve the school district as a sampling unit are typically difficult to generalize beyond their particular contexts. Also, it is usually difficult to compare results across such studies. Such limitations are related to three general classes of research manipulations: 1) describing school districts, 2) comparing results of separate studies, and 3) selecting dimensions with which to discriminate among districts. These are three distinct but related methodological issues, and they are respectively related to theoretical concerns for 1) sufficient descriptive complexity, 2) replicability, and 3) generalizability. These three points are given here not only as criticisms of typical research, but also as criteria for the stratification system developed in this study and specifically described in Parts II and III. It is therefore necessary to explicate the importance of each point. The explication is given in the remainder of this section and a summary is presented in Table 1.

SUFFICIENT DESCRIPTIVE COMPLEXITY

Dimensions typically used for describing and discriminating among districts are inadequate: research studies are limited by definitions of descriptive dimensions, and by oversimplified district characterizations. If the study of school districts is to be scientific in any sense, then an empirically rigorous taxonomic system is required for describing and differentiating districts. An adequate classificatory system is no less necessary in the study of institutions than in any other field of scientific inquiry:

> Classification is one of the fundamental concerns of science. Facts and objects must be arranged in an orderly fashion before their unifying principles can be discovered and used as the basis for prediction. Many phenomena occur in such variety and profusion that unless some system is created among them they would be unlikely to provide any useful information (Sokal, 1966).

The act of making discriminations among districts--even if the distinctions are only dichotomous--necessitates the definition of a dimension of discrimination. In the process of identifying, labelling, and working with such a dimension, the researcher is frequently faced with a dilemma. On the one hand, a researcher might choose a replicable and operationally sound dimension which oversimplifies the differentiations among districts, with a corresponding loss of meaningfulness. On the other hand, if he wishes his procedures to reflect more accurately the complexities of district characteristics, he finds it difficult to define dimensions which are operational and manipulable. For instance, an investigator may feel that "district size" is an important distinction to make among respondents in his study. Though enrollment totals are typically employed

TABLE 1

Summary of Methodological Issues

Issue	Hazard	Criterion
1. Description	Insufficient characterization of LEAs and failure to utilize relevant information, resulting in oversimplification and disregard for the complex structure of these organizational entities	Sufficient Descriptive Complexity
2. Comparison	*Ad hoc* categorization of LEAs and ill-defined and unreported selection procedures resulting in the inability to test and accumulate knowledge.	Replicability
3. Selection	*A posteriori* classification of LEAs, resulting in the confounding of district differences with other dimensions peculiar to the particular sample.	Generalizability

as stratification variables in such studies (for instance, see Ryans, 1960; Schunert, 1951; or Barnes, 1961), the investigator may feel that other indices of size--such as the number of teachers employed in the district, or the number and capacity of school buildings--are also important indicators. There would be no problem in such a case if the various indicators of "district size" were perfectly correlated, but this is infrequently true.

A criterion, then, for procedures which characterize, categorize, or stratify school districts is that they incorporate sufficient descriptive complexity and thereby acknowledge the multivariate nature of school district characteristics.

REPLICABILITY

Procedures for categorizing districts are often project-unique. When researchers need to establish procedures for discriminating among districts, they often rely upon _ad hoc_ rational schemes for district characterization. Such schemes are often based on the nature of the particular content being researched, and on the investigator's familiarity with the districts in which he works; consequently, the assignment of districts to strata, types, or clusters is far too often based on the use of impressionistic observations.

An example of a selection scheme which would be very difficult to replicate may be found in Pierce's (1947) study of factors related to adaptability. His sample of districts was recruited from the "Metropolitan School Study Council", which is

> ... an organization of relatively wealthy communities located, with the exception of a small number ..., within the metropolitan area of New York City, including eight communities from within the city ... It was not possible in the final analysis to include every community in each phase of the study due to incomplete data on some of the measures. Conclusions and statistical analyses are based for the most part on data from 48 communities.

> It seems appropriate to point out the richness of this setting for a study of this kind. Probably no other group could afford a better background for the study of factors which may be related to making good schools . . .

Clearly, the selection of the "sample" for this study was determined more by convenience than by an a priori proscribed rationale. It is also clear that any investigator who wished to replicate the study would have to seek out those same 48 communities and their districts; he has no basis for defining an independent, but somehow comparable, sample of districts or communities.

Brunner (1957) later wrote that Pierce had

> . . . shown most elaborately through statistical techniques that what a school is is determined far more by what the community is than by what goes on within the four walls of the school building (p. 79).

Such a conclusion seems hasty, in view of the ambiguity of the school selection procedure used in the study, and in view of the fact that Pierce did not directly compare the impact of community characteristics with that of "internal" school characteristics.

A study by Duncan and Kreitlow (1954) is another example of one which would be difficult to replicate because the manner of defining the sample is unclear. In this study,

> . . . data were obtained by personal interview in 38 rural neighborhoods located in . . . [various parts] . . .of Wisconsin . . . The neighborhoods were selected so as to constitute 19 matched pairs, one in each pair being homogeneous in ethnic and religious characteristics and the other heterogeneous in these respects. The two neighborhoods in each pair were matched on . . . [socio-demographic] characteristics. The 19 pairs represent a range of agricultural land types, types of school system and specific major ethnic-religious groups in the state (p. 350).

If another investigator wished to replicate this study in a different location, he would need much more information about the identification and classification of "neighborhoods" than that provided by these authors, such as: What are the characteristics of the population of neighborhoods from which the 38 were selected? Exactly what was the selection procedure used in identifying the 38? What are the operational definitions of and the criterion cut-offs for "homogeneity" and "heterogeneity"? What was the attrition rate; that is, how many neighborhoods had to be excluded from the study because they didn't 'match up' properly?

In a study such as Duncan and Kreitlow's, a more attractive alternative to the sampling problem is to: 1) identify a population of 'neighborhoods', 2) develop a theory-based but operationally-defined scheme for classifying neighborhoods as 'homogeneous' or 'heterogeneous', 3) partition the entire population into these categories, 4) draw random samples of neighborhoods from each of these two clusters, and 5) use demographic variables as covariates, if statistical control over them is desired.

Although it is usually methodologically sound for the categorization or stratification of districts to be based upon variables which have apparent theoretical relevance to the content of the research, it is not desirable for these procedures to be largely determined by the idiosyncratic aspects of individual studies, for the cumulation of research findings is attenuated or prohibited. The criterion of replicability, then, for a general system which characterizes, categorizes, and stratifies school districts is that the system allows the cross-study comparison and integration of results which are not now possible. That is, the applications of the system should be replicable. Of course, the system must permit the selection of dimensions (within the system) considered relevant for particular research interests.

GENERALIZABILITY

Prior selection of districts to be studied often determines dimensions for categorizing districts. For purposes of this discussion, generalizability is equated with the avoidance of bias due to sample selection procedures. One type of biasing condition consists of identifying a few (sometimes only two) districts which will grant access to do research, executing the research and finding differences between or among districts, and then comparing districts on the basis of descriptive characteristics in an effort to find some hindsightful way to explain the results.

A recent study of problem solving among elementary school teachers (Turner, 1964) describes a procedure for selecting school districts which may impose stringent limits on the generalizability of the outcomes. Districts which might have cooperated in the research were selected on an "invitational" basis. At first, 25 Indiana systems were invited to participate in the research; twelve of these either did not accept the invitation or were excluded because they did not employ any inexperienced teachers, who were to be the primary respondents for the research. Other districts had to be discarded from some subsequent analyses because their few beginning teachers did not stay for the two-year duration of the study. The basis for selecting the districts to whom the original 25 invitations were sent is not made clear in the report. After the study was well under way, the investigators developed a method for "typing" the 13 participating districts, based on equalized property valuation, and the ratio of working-class to middle-class children in a district. The purpose of this phase of the project was to determine the relationship between the "type" of district and selected characteristics of teachers in the districts.

Inferences about such relationships might have been stronger had the investigators first identified a population of districts and "typed" each of them, and then selected districts from types (preferably at random) until they had secured enough LEAs which employed first-year teachers. It would remain to persuade the sampled districts to cooperate. As it happened, the criteria of having the right kind of teachers and being willing to cooperate were applied prior to the sampling of districts, and the "typology" procedure was thus applied to a small and probably biased sub-set of the population of districts. G:ab-group samples are not likely to be representative with respect to measured characteristics, because "grabability" is doubtless correlated with other important qualities which differentiate districts.

Limitations on generalizability also occur when cases from a population of districts are eliminated because data are inconvenient. For example, study by Terrien and Mills (1955) contains a sequence of eleminations of districts which certainly biased the outcomes of the research,but in some unspecified way:

> Of the 1747 elementary districts in California, data were secured on 732: Because of the fact that in organizations of less than ten persons one person does several jobs of both an administrative and non-administrative character, it seemed justifiable to remove from consideration those 468 elementary districts in which the total organization numbered less than ten .. Of the 245 high school districts, data were secured on 100... four districts were less than ten persons in size, and these were removed....

Such a practice would be less serious if relevant differences between the "eliminated" group and the "retained" group could be described. However, such information is not made available; in fact, the possibility of bias by differential elimination cannot be detected because selection and elimination procedures are not fully discussed.

Such limitations are serious because dependent variables are frequently correlated with relative availability of data. Almost always, some differences can be found among districts which are correlated with and appear to explain the results of the research. Such policies are variously known as "taking advantage of chance", "fishing in polluted data", "exploitation of the grab group", or "*a posteriori* hypothesizing". By any name, the practice violates a host of principles of sampling procedures and scientific and statistical logic.

A criterion, then, for a system which characterizes, categorizes, and stratifies school districts is that the application of the system be unbiased and generalizable. The advantage of such a system is that its availability might encourage analytic procedures more sound than the search for descriptive schema which will "discriminate" among a few districts selected for study because of their proximity, or willingness or ability to cooperate.

PART I

SECTION B

OBJECTIVES: THE DESIRED OUTPUTS OF THE PROJECT

The goal of the research described in this report was to provide for the empirical characterization and categorization of a defined population of school districts.

The two dominant objectives were therefore to develop one general system for characterization and stratification of LEAs and to illustrate the utility and applicability of the system. The system was conceived and formulated as a general methodological algorithm[1] for measuring, stratifying, and sampling districts from a population. Because the investigators considered districts to have multivariate complexity, multivariate analyses have been the bases of the methodology. The basic outputs of the algorithm, therefore, are multivariate descriptions and measures of districts and multivariate sampling paradigms.

Three kinds of criteria shaped the development of the algorithm: A) Fundamental Scientific Criteria, B) Research Utilization Criteria, and C) Practicality Criteria. Detailed discussion of these three kinds of criteria are given in this section, and a summary is given in Table 2. The derived algorithm is operationally described in Part II, and illustrations and applications are given in Part III.

FUNDAMENTAL SCIENTIFIC CRITERIA

The scientific criteria have already been discussed in Section I.A. They are a) sufficient descriptive complexity, b) replicability, and c) generalizability.

[1] The definition for algorithm in this report will be that given in the Random House Dictionary of the English Language: "any particular procedure for solving a certain type of problem", where, in this case, the problem is the empirical stratification of a population of school districts.

TABLE 2

Summary of Criteria

Kind	Criterion
A. Scientific Criteria	1. The dimensions should reflect the multidimensional complexity of districts.
	2. Applications of the algorithm should be replicable.
	3. Outcomes of those applications should be generalizable and unbiased.
B. Utilization Criteria	4. The algorithm should allow precise and concise descriptions of districts.
	5. The algorithm should allow comparison and categorization of districts.
	6. Analytic comparison with outside variables should be possible.
	7. The algorithm should facilitate sophisticated sampling for a wide range of studies concerned with school district variability.
C. Practicality Criteria	8. System data inputs should be extant and immediately available.
	9. System outputs should be meaningful and directly useable.
	10. Special system outputs should be obtainable for particular research interests.

In Section I.A they were stated to be guiding principles for the design of the algorithm. Here, further comments will be made concerning certain practical aspects of the criteria.

Sufficient Descriptive Complexity. School districts are political, geographical, administrative, sociological entities. Within practical limitations, the array of input information should be selected and prepared to reflect and indicate as many aspects as possible of school district characteristics. And in summarizing and compositing the input information, care should be taken to avoid oversimplification.

Replicability. To make comparitive and longitudinal studies possible, it is desirable to be able to update the algorithm periodically. For input availability and structure and the composition of the population change across time. Also, it should be possible to transplant the algorithm to other populations where similar inputs are available. Then, for example, studies made in Wisconsin could be compared to studies made elsewhere.

Generalizability. The algorithm should be comprehensive. That is, all the units in the target population (for this study, all the elementary school districts in Wisconsin) should be included. And all dimensions arising from the algorithm should apply to all the units in an unbiased way.

RESEARCH UTILITY CRITERIA

The use of multivariate methods in the exhaustive application of a comprehensive paradigm to a population of districts should yield dimensions which have utility for these general research processes: description, comparison, logical analysis, and sampling. In the following paragraphs these processes are briefly discussed: Their requirements are stated as criteria in the development of an algorithm; and, as illustrations, previews are given of the ways in which the present algorithm meets the criteria.

Description. An algorithm should uncover meaningful dimensions on which the units vary, dimensions which are intrinsically valuable for developing and verifying theoretical notions on the nature of school districts. In the present study, the wide range of input information was reduced to a small number of independent and meaningful dimensions.

Comparison. It should be possible to partition the population of districts into clusters such that the clusters are homogeneous internally and heterogeneous with respect to one another. One clustering specified by the present algorithm is described in Section III.B. Furthermore, the present algorithm allows specifying new clusterings based on any selection of the characterizing dimensions. Qualitative and quantitative comparison can then be made between clusters or "types" of districts.

Logical Analysis. It should be possible to analyze the dimensions derived within the algorithm as possible sources of variation in understanding and explaining the distributions of other variables across the districts. Deming (1953) makes a useful distinction between analytic and enumerative uses of research data: "In the analytic problem, the action is to be directed at the underlying causes that have made the frequencies of the various classes of the population what they are, in order to govern the frequencies of these classes in time to come." Cornell (1954) demonstrates that most educators are interested in analysis. The present algorithm can be utilized to examine relationships between the complex, derived dimensions for characterizing and differentiating districts (independent variables) and other dimensions which have intrinsic value for particular research interest (dependent variables). This utility, of course, will depend on the meaning or interpretability of the derived dimensions and on their theoretical relevance to the dependent variables.

Sampling Techniques. The algorithm should facilitate sophisticated sampling for a wide range of research applications in which school district variability is of concern. This concern may be for the need to define an especially efficient or especially generalizable sample. Or this concern may be for defining a sample which is highly variable. The present algorithm allows the selection of stratified samples which meet these needs, as illustrated in Section III.D.

PRACTICALITY CRITERIA

While methodological criteria are concerned with the various kinds of research goals for which an algorithm should be useful, criteria of practicality or feasibility are related to the probability that the system will be implemented. If the establishment of the system is extraordinarily cumbersome, involved, and tedious, there is little chance that it will be implemented, no matter how imposing is its justification on theoretical or methodological bases. Because of the availability of automatic computing machinery, implementation need not be prevented by the necessity of massive amounts of purely mechanical operations. But implementation requires information to be available and well designed procedures for treating it to be designed. Three practicality criteria, then, for an algorithm which characterizes, categorizes, and stratifies school districts are: a) availability of inputs, b) meaningfulness of outputs, and c) variety of outputs. These criteria are explained in the paragraphs which follow: they are at first stated negatively, as possible obstacles to the development of an algorithm; then, as illustrations, previews are given of the practical solutions possible within the present algorithm.

Availability of Inputs. There would be little advantage in developing an algorithm which requires the planning and executing of a special large-scale data collection venture solely for the purpose of stratifying. The data for the present scheme are collected and filed in a data library which is at present sparingly used for any research work. The set of inputs is comprised of this data and certain programmed transformations of it.

Meaningfulness of Outputs. Multivariate analyses sometimes produce composites which are highly efficient representations of original data but which are very difficult to interpret. If a multivariate system of data reduction yields essentially uninterpretable results, it is very unlikely that it will be employed as a methodological tool for field research. In the present study, two different multivariate techniques were tried, and the one which produced more interpretable dimensions was selected. Also, the computer program which derives the stratification clusters was designed to produce summary characteristics of the districts in each cluster; thus stratifications are not presented as a bare structure but rather as annotated typology.

Variety of Outputs. No single stratification, or set of dimensions, is sufficient for use in all research problems. The present algorithm allows researchers to select from a variety of dimensions those considered relevant to their particular interests. An illustration of such selection is given in Section III.C.

PART I

SECTION C

OVERVIEW: THE ALGORITHM AND ITS EMPIRICAL PERSPECTIVE

An algorithm for characterizing and categorizing the population of Wisconsin elementary school districts has been designed and performed. It was inspired by the needs discussed in Section 1.A and its development was guided by the criteria and objectives stated in Section 1.B. The algorithm is a series of operations, and these operations are exhaustively specified in Part II. In Part III the applications of the outputs of the algorithm are presented. In this section an overview of the algorithm is given and is followed by a perspective on the design, operation, interpretation, and evaluation of the algorithm.

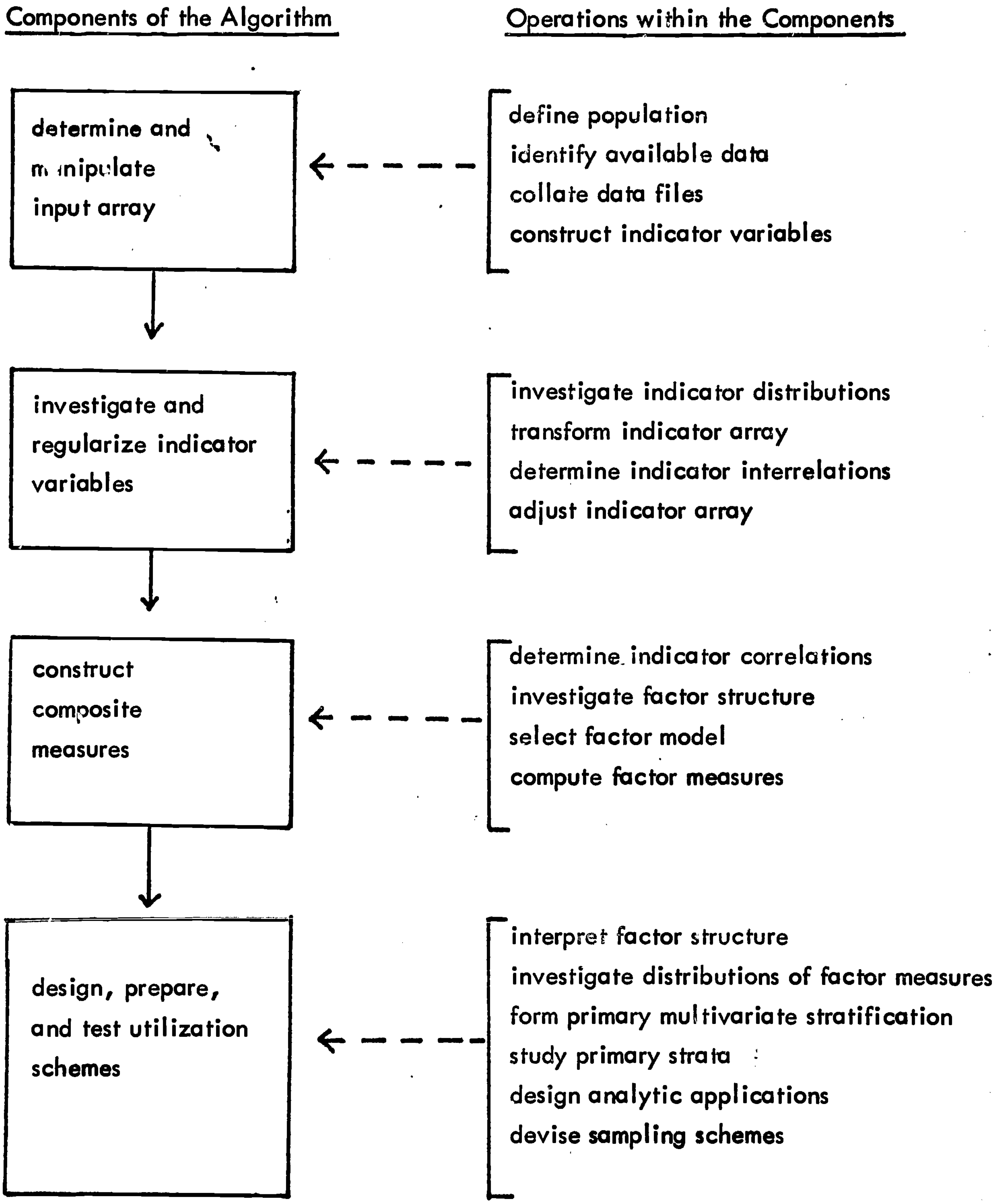
OUTLINE OF THE ALGORITHM

The general algorithm is summarized in Table 3. It has four major components and each component has a list of associated operations. The summary outlines the series of operations which have been performed with respect to the population of Wisconsin elementary school districts. The summary is general: a complete specification of the algorithm is presented in Part II. In the paragraphs which follow, some comments are made about the major components of the algorithm, and references are made to the relevant sections of Part II.

Determine and Manipulate Input Array. The algorithm does not involve a special data generation program, for use is made of banks of data created and maintained for other purposes, such as accounting. Therefore, when the population is defined, the next step is to search out and identify what data is available and to organize and collate it for the purpose of determining what can be extracted from it. The process of extraction requires ingenuity in manipulation and is guided by substantive theorizing. Indicator variables result from the extraction. For the present study, the results of the search for data and the specification of the indicator variables constructed are presented in Section II.A.

TABLE 3

SUMMARY OF THE ALGORITHM



Components of the Algorithm

Operations within the Components

determine and manipulate input array

- define population
- identify available data
- collate data files
- construct indicator variables

investigate and regularize indicator variables

- investigate indicator distributions
- transform indicator array
- determine indicator interrelations
- adjust indicator array

construct composite measures

- determine indicator correlations
- investigate factor structure
- select factor model
- compute factor measures

design, prepare, and test utilization schemes

- interpret factor structure
- investigate distributions of factor measures
- form primary multivariate stratification
- study primary strata
- design analytic applications
- devise sampling schemes

Investigate and Regularize Indicator Variables. The distributions and the interrelations of the indicators must be displayed and checked. The later operations require complete data and at least an approximation to normally distributed, linearly interrelated indicators. Thus transformation and adjustment of the indicator variables may be necessary. For the present study, some values of some indicators were missing--that is, unavailable or uncomputable. The investigation and resolution of this problem is discussed in Section II.B.

Construct Composite Measures. The variables derived from the source data are conceived of as indicators of the important dimensions along which the organizational units vary. So when the indicator array has been produced and regularized, the algorithm calls for a investigation of the multivariate structure of the indicators. From the indicator correlation matrix, a number of multivariate models are calculated. The most meaningful and parsimonious is selected, and the measures corresponding to the independent, uncorrelated factors of this model are computed. That is, from the many indicator variables, a few important composite measures are derived. The techniques employed are discussed in Section II.C.

Design, Prepare, and Test Utilization Schemes. The output of the previous component operations of the algorithm provide the means for aiding four basic kinds of research processes. One, the data's structure within the selected multivariate model and the distributions of the associated measures allow description of the characteristics of the population. Two, by transforming and crossclassifying the measures, a multivariate stratification of the units is obtained; thus the units and types of units may be compared. Three, the measures or transformed versions of them may be combined with outside variables in the process of logical analysis. Four, stratifications of the units derived from various crossclassifications of the measures determine sampling plans. The programs and techniques for accessing and arranging the output measures in the ways required for the realization of these schemes are discussed in Section II.D, and illustrations of their performance are given in Part III.

EMPIRICAL PERSPECTIVE

The various criteria stated in Section I.B were ideals toward which the investigators aimed the project. The algorithm produced is a single and somewhat restricted realization of the ideals. The investigators' approach has intentionally been methodological and pragmatic. An alternative stylistic approach would have been substantive. It might have involved, for example, a more thorough logical analysis of input variables and an exhaustive theory of the organizational nature of school districts. Knowledge of individual differences among school districts is limited, and because the substantive approach requires prior systematic knowledge, the investigators felt their approach should be methodological.

The specification and resolution of methodological issues were performed, however, within a substantive perspective. That is, the decision process in the development of the algorithm was assisted and guided by the investigators ultimate concern for providing substantively useful results. In the paragraphs which follow, notes are made about the pragmatic resolutions that were made at five crucial decision points in the project: A) selection of input, B) selection of population, C) selection of analysis, D) definition of stratification. These notes provide a perspective for how the decisions were made and for how the decisions should be evaluated.

Selection of Input. The main criteria for input was that it be already gathered, that it be easily retrieved and manipulated, and that it reflect a variety of district characteristics. The concern for available and manipulable data was implied by the pragmatic orientation of the project. But the data obtained were not used directly: elaborate transformations were performed. The data obtained were enumerative and, for the purposes of the project, were not suitable as direct input for analyses. Since even the most sophisticated analyses cannot transform trivial observations into useful generalizations, some structure had to be imposed on the raw information.

The decisions leading to the definition and construction of variables were made from a substantive and theoretic perspective, within the limits of project resources. The meaning of the algorithm's outputs and applications must be evaluated in terms of the viability of these decisions, as they were limited by the pragmatic attitude toward the collection of data. One of the purposes of the illustrations in Part III is to demonstrate the substantive validity of the outputs.

Selection of Population. The population of school districts in Wisconsin is really the sum of three populations: elementary districts, secondary districts and elementary/secondary districts. Variables have differential meanings when applied to the three populations. Consider a variable, the student/staff ratio in a district. For an elementary district the ratio concerns, usually, teachers in self-contained classrooms; for secondary districts, the ratio concerns departmentalized teachers; and for elementary/secondary districts, the ratio concerns a mixture of the two kinds of teachers. That is, what is conceived as a single variable derived for the total population is really three separate variables when the composing populations are considered.

A substantive research style would have led to a careful redefinition of such a variable, or to an initial partitioning of the population. The pragmatic style of this project led the investigators to find a partial solution to the problem: namely, the districts with secondary schools only (there were about 50) were excluded from the research. The defined population consisted then of 632 districts which operated elementary schools, and the variables which were selected for the study dealt primarily with the qualities of the elementary schools. Even so, the results of the algorithm were slightly skewed by the continuing presence of an elementary and elementary/secondary distinction. This is discussed further in Section III.A.

Selection of Analysis. As was mentioned above, the input spectrum was limited; the variables that could be constructed formed a small, biased sample of the set of variables that theoretically and substantively might have been desired. This limitation on the array of variables implied the necessity of emphasizing the data-analytic processes of the analysis rather than the statistical or hypothesis-testing possibilities. The multivariate techniques used were chosen to provide reduction of the data in hand and to impose structure on it. The choice between alternative data analyses, however, was dictated by the meaningfulness of their outputs.

Selection of Calculations. The calculations of the descriptions are the product of a certain set of variables, a certain population, and a certain mode of analysis. But the outputs of the algorithm characterize the districts only at a certain point in time. Because the characteristics of individual districts change over time, the analyses should be repeated periodically. For example, an important educational trend in Wisconsin is the increasing consolidation of local education agencies, and this tendency certainly affects the organizational characteristics of districts. From a substantive perspective, then, the selection of calculations for the present study is temporally limited.

Definition of Stratification. Multivariate procedures were considered to be appropriate because the substantive perspective recognized the multivariate complexity of local education agencies. The interaction of the use of multivariate procedures with the objective of stratification has resulted in a multivariate stratification scheme. A "stratification" of a population typically amounts to partitioning the population into mutually exclusive groups by defining (often arbitrary) "cutting points" on a continuous stratifying variable. Essentially, this results in a small set of ranks, and each member of a sample or a population is assigned one of these ranks. An example would be the socio-economic stratification of families in a midwest community according to family income. Each family would be assigned one of these ranks.

Such a univariate stratification is inadequate for use in the general algorithm. First, many important aspects of the entities would not be measured by that single stratifyin variable. Second, the generality of applicability of a particular univariate stratification is limited. A desirable alternative to univariate stratification is to use several carefully composed variables in stratifying. The stratification developed in this study is defined in Section II.D and demonstrated in Section III.B. It is a multivariate stratification, and did indeed use several (five) carefully composed variables to partition the population of districts into mutually exclusive classes. A "stratum" or "cluster", in the multivariate sense, is located with respect to many variables, instead of just one variable. Therefore, it is not possible to say that a stratum in a multivariate system is genotypically "higher" or "lower" than some other stratum. A multivariate stratum is a position in a typology rather than in a hierarchy.

PART II: DATA PROCESSING AND STATISTICAL TECHNIQUES

Section A - Variables: Formation, Definition and Codification

Section B - Missingness: Investigation and Resolution of Missing Data

Section C - Factorization: Reducing and Orthogonalizing the Variables

Section D - Data Processing: Techniques of Computing Scores and Coding Districts

PART II

SECTION A

VARIABLES: FORMATION, DEFINITION AND CODIFICATION

STRUCTURE AND SOURCES OF DATA

Hierarchy of Data. The immediate analytic requirement was to obtain a set of measures for each school district. Data were available from three distinct levels: (1) school district, (2) school, and (3) employee. The array of data formed a hierarchy, since the schools were within districts, and the employees were within schools.

In computing measures on a district, the information used could be district data, distributional characteristics of schools, or employee data. In this section exact data available at each level will be discussed, the classes of variables are indicated, and constructed variables are listed.

Available Data and Their Sources. There were three files of information. The first was the Wisconsin State Department of Public Instruction (WSDPI) "District/School Tape", which had a record for each school district and a record for each school. The second was the WSDPI "Employee Tape", which had a record for each school employee. The third file was the "Valuation Deck", which contained a card for each school district; these were cards punched from records obtained from WSDPI Division of State Aids and Statistical Services. Listed below are data contained in the three files, arranged by the hierarchy of district/school/employee.

For each school district five types of information were available. They are listed in Table 4. The first column in this table denotes the five types of information. The first three types were coded on the "School/District Tape", and the other two types had been punched on the "Valuation Deck". The second column in Table 4 gives the coding records used by the WSDPI, and the third column gives the recoding scheme used to transform these data into input for computations.

TABLE 4

Information Available by School District

Type of Information	WSDPI Coding Record	Recoding
1. Kind of administrative structure of the district	1. City of Milwaukee	1
	2. City Unified	1
	3. City Common	1
	4. City F. D.	1
	5. County Unified	2
	6. County Common	2
	7. County U. H. S.	2
2. Scope of grades taught in the district	1. K-12	1
	2. 1-12	1
	3. 9-12	(Not Used)
	4. K-8	2
	5. 1-8	2
3. Class of state financial aid to the district	1. Integrated	1
	2. Basic with Integrated	2
	3. Basic	3
4. Assessed valuation	Dollars	(Not Used)
5. Equalized valuation	Dollars	(Unchanged)

Note. - Source: Wisconsin SDPI "School/District Tape", except that Types 4 and 5 were taken from the "Valuation Deck".

For each school two types of information were available as denoted in column one of Table 5. They both were taken from the "School/District Tape". The second and third columns give the WSDPI coding and recoding schemes.

For each employee there were eleven types of information available. They are listed in column one of Table 6. These data were on the "Employee Tape" in coded form as given in column two of the table and recoded according to the scheme given in column three.

TABLE 5

Information Available by School

Type of Information	WSDPI Coding	Recoding
1. Type of School	1. Four-year high school	(Not Used)
	2. Six-year high school	(Not Used)
	3. Senior high school	(Not Used)
	4. Junior high school	(Not Used)
	5. Elementary school	(Not Used)
2. Enrollment	Counts	(Unchanged)

Note. - Source: Wisconsin SDPI "School/District Tape"

In addition to the [illegible]les given in Tables 1, 2 and 3, the three files contained information sufficient to locate employees within schools, and schools within districts.

The data which were available from these three sources formed seven classes of variables, and allowed the construction of 31 variables for computational purposes. The classes, abbreviated titles, and identification code numbers of these 31 variables are presented in Table 7. Complete descriptions of the 31 variables are given in Appendix M.

The remaining paragraphs of this section describe the procedures for generating the variables.

CONSTRUCTION OF VARIABLES

Teacher Characteristics. Because the stratification of the school districts was intended to be used in studies of district sub-units (e.g. schools, teachers and students), it was necessary to include input data derived from characteristics of these sub-units. Twelve of the constructed variables deal directly with characteristics of the elementary teachers in each district. As these variables had to be attributes of districts, they had to

TABLE 6

Information Available by Employee

Type of Information	WSDPI Coding Record	Recoding
1. Credential	1. 1 year license	8
	2. 2 year license	2
	3. 3 year license	3
	4. 2 yecr term certificate	7
	5. 3 year term certificate	6
	6. 4 year term certificate	5
	7. 5 year term certificate	4
	8. Life certificate	9
	9. 1 year permit	1
	10. 1 year special license	0
2. Degree	1. less than 2 years	1
	2. 2 years (diploma)	2
	3. 3 years	3
	4. Bachelor's	4
	5. Master's	5
	6. 6 years	6
	7. Doctor's	7
	8. other	0
3. College Conferring Degree	Name	(Not Used)
4. Division of Time	1. Percent Elementary 2. Percent Secondary	(Used for Tabulation Only)
5. Months Employed	Months	(Not Used)
6. Salary	Dollars	Dollars
7. Local Teaching Experience	Months of teaching service in the local district	(Unchanged)
8. Total Teaching Experience	Months of total teaching service in or out of district	(Unchanged)
9. Sex	No record	
10. Position	School staff position codes differentiated various types of administrators and teachers, and non-professional employees. WSDPI codes used were: 27 Secondary Teachers 32 Junior High School Teachers 42 Elementary Teachers 95 Non-professional employees	(Used for Tabulation Only)
11. Grades or Subjects Taught	For elementary teachers, the range of grades taught is indicated. Otherwise, the codes indicate the courses taught.	Grade Span

Note. - Source: Wisconsin SDPI Preliminary Report Forms and "Employee Tape".

TABLE 7

Abbreviated Titles of Recoded Variables

Class of Variable	Variable
A. Elementary Teacher Characteristics:	1. Mean credential
	2. Mean degree
	3. Mean salary
	4. Mean local experience
	5. Mean total experience
	6. Mean grade spread
	7. Log-variance credential
	8. Log-variance degree
	9. Log-variance salary
	10. Log-variance local experience
	11. Log-variance total experience
	12. Log-variance grade spread
B. Classifications of Administrative Structure:	13. Kind
	14. Scope
	15. Class
C. Enrollment:	16. Secondary
	17. Elementary
D. Employee Counts:	18. Full-time elementary teachers
	19. Full-time junior high teachers
	20. Full-time secondary teachers
	21 Other teachers
	22. Other professional employees
E. School Counts by Size:	23. One-room
	24. Two-rooms
	25. Three or more rooms
F. Valuation:	26. Equalized valuation
G. Ratios:	27. Valuation/student
	28. Students/school
	29. Students/staff
	30. Staff/school
	31. Valuation/school

Note. - The number designations for the thirty-one variables as given in column two are the codes used throughout the text and for all tables and appendices.

describe distributions of district characteristics, rather than characteristics of individual teachers. Therefore, means and variances of teacher characteristics within districts were used as inputs. In the computation of variances, logarithms were used in an attempt to eliminate non-normal distributions.

The "Employee Tape" was searched for all records of full-time elementary teachers withih each district. For each such teacher, six codes were examined: "Credential", "Degree", "Salary", "Local Experience", "Total Experience", and "Grades Taught" (See Table 3). Six temporary variables were constructed. The first and second of these were recoded versions of Credential and Degree; the records were arranged in order of increasing values, based on preference ratings according to WSDPI criteria: the highest ratings corresponded to the highest numeric codes. The third, fourth, and fifth of the temporary variables were copied directly from Salary, Local Experience, and Total Experience. Salary is in dollars; experience is in months. The sixth temporary variable was constructed from the "Grades Taught" code. The Grade Spread was determined by counting the number of grade levels a teacher was responsible for in his classroom instruction; for example, if a teacher taught Grades 1-3, then the recoded value was 3.

The arithmetic means of the six temporary variables were computed across the teachers within each district. These district means became Variables 1 to 6. The logarithms (base e) of the variances of the six temporary variables were computed and formed Variables 7 to 12. These twelve variables are listed in Table 7.

Wisconsin SDPI Classifications. The WSDPI assigns to each district a code for each of three classifications, called "Kind", "Scope", and "Class" (See Table 4). The codes were read from the district records on the "District/School Tape" and recoded to produce Variables 13 to 15. Variable 13, Kind, is "1" for city-based districts and "2" for county-based districts. Variable 14, Scope, is "1" for districts with one or more high schools, and "2" for districts with no high schools. Variable 15, Class is coded according to WSDPI criteria for distributing state aids. Districts receiving the class of integrated state aids were assigned a code of "1", those receiving basic with integrated aids were assigned a code of "2", those receiving basic aid were coded "3".

Note that codes for these three variables are inverted with respect to the preference ratings assigned by the WSDPI. Low values of the variables correspond to high WSDPI ratings.

Student Enrollment. The student enrollment of each school was coded in that school's record on the "District/School Tape", By summing the enrollments of the schools in a district, Variables 16 and 17 were computed. For Variable 16, the sum was taken over secondary schools, so that variable is total Secondary Enrollment. For Variable 17, the summing was taken over elementary schools, so that variable is designated as Elementary Enroll ment.

Teacher Counts. The records on the "Employee Tape" were examined for all employees in a district. Non-professional employees were ignored. The "Division of Time" and "Position" codes were considered (See Table 6). From the "Division of Time" codes it was determined whether an employee was full-time. From the "Position" codes it was determined which of the following positions an employee held: (a) elementary teacher, (b) junior high school teacher (c) high school teacher, or (d) non-teacher, usually administrator. A teacher conceivably could have any or all of the four positions.

Variables 18 to 20 are concerned with full-time teachers within each district. Variable 18 is the count of elementary teachers in a district. Variable 19 is the count of full-time junior high school teachers in a district; and Variable 20 is the count within a district of full-time high school teachers. None of the teachers counted in these three variables had non-teaching poisitions, but if they had more than one teaching position they were counted in the "higher" classification. For example, a teacher who taught both junior and senior high school was counted with senior high school teachers, Variable 20.

Variables 21 and 22 are counts of the other professional employees; that is, they are counts of employees who are not full-time professionals or who had non-teaching duties. Variable 21 is the count of those employees who did some teaching. Variable 22 is the count of professional employees who did not teach.

School Counts. The number of schools was derived by counting the school records within a district on the "District/School Tape". No information was directly available on the physical size of the school plants. So information indicating school size was estimated by counting the number of teachers (on the "Teacher Tape") who were assigned to a school. Variable 23 is the count of schools in a district with just one teacher, and is, by inference, the Number of One-Room Schools; Variable 24 is the number of schools in a district with exactly two teachers and hence is the Number of Two-Room Schools; and Variable 25 is the count of the remaining schools within a district and is the Number of Three-or-More-Room Schools.

Valuation. The equalized valuation of each district was obtained from the "Valuation Deck" and designated as Variable 26. Equalized valuation was used rather than assessed valuation, because the assessment formulas vary from district to district.

Ratios. Five variables were formed by taking ratios of selected combinations of enrollment variables, staff variables, and the district valuation variable. The sum of Variables 16 and 17 is the total number of students in a district. The sum of Variables 18 to 22 is the count of the total professional staff in a district. The sum of Variables 23 to 25 is the number of schools in a district.

Variable 27 is the ratio of valuation to the total number of students; that is, the dollar Valuation per Student. Variable 28 is the number of Students per School. Variable 29 is the number of Students per Staff member. Variable 30 is the number of Staff per School. Variable 31 is the dollar Valuation per School.

INITIAL DATA MATRIX

A list of 31 variables has now been described, and their codes have been specified. Each of these variables describes a characteristic of elementary school districts and the whole set of variables describes seven classes of district aspects. Full descriptions of the 31 variables are provided in Appendix M. The means and standard deviations of these variables are given in Appendix B .2, and their intercorrelations are given in Appendix B .3.

As indicated in Section I.C, the population of concern in this project was the group of all Wisconsin school districts which had reported elementary enrollments greater than zero. There were 632 such districts; 44.5% of them consisted of elementary schools only, and 55.5% of them were made up of elementary schools and high schools. The initial data matrix could not be directly used as input for the computations. The input data for the computing algorithm needed to consist of a complete matrix with dimensions 31 x 632; the intersection of a row and a column in this matrix represented the 'score' of a particular district on a certain variable.

One problem had to be solved before the matrix could be satisfactorily used as input: there were missing values in the matrix. It was not possible to obtain directly a measure on every variable for every district. This meant that some of the correlations between pairs of variables, as given in Appendix B.3, were based on data from fewer than 632 districts. Appendix B.1 indicates which correlations were based on subsets of the district population; an entry in this appendix gives the number of districts which had no missing data for either of the corresponding pair of variables.

The causes and solution of this missing data problem are fully discussed next in Section II.B.

PART II

SECTION B

MISSINGNESS: INVESTIGATION AND RESOLUTION OF MISSING DATA

The computations required by the analytic algorithm, presented in Section II.C, were designed for the purpose of achieving the research objectives discussed in Part I. In order to perform these computations, it was necessary that the input data matrix be complete; that it, it could have no missing entries. But there were missing entries in the initial data matrix, and some procedure had to be employed to substitute for missing values.

Discussions in this section will first consider the general problem of missing data, then the procedure of estimating values to replace missing data, and finally the nature of the input data matrix which was used for computational purposes.

THE PROBLEM OF MISSING DATA

Origins of Missing Data. The first twelve variables (Teacher Characteristics) were the only ones for which entries were missing. It has been indicated that Variables 1 to 12 were measures of the six characteristics of the full-time elementary teachers in a district; Variables 1 to 6 were the district means for these characteristics, and Variables 7 to 12 were the log-variances of those characteristics. There had to be at least one full-time elementary teacher in a district before a mean could be defined, and at least two such teachers in a district before a log-variance could be defined.

There were three types of missing data. The first type occurred when a district had only one full-time elementary teacher. There were ninety of those districts; for those ninety districts Variables 7 to 12, being log-variances, could not be defined. This first type of missing data accounted for almost all missing values.

In the second type of missing data, a few full-time elementary teachers had improper code designations for some of the characteristics. Their codes, then, could not be included in computing the means or log-variances of those district characteristics. For two of the districts, this resulted in fewer than two codes on which to base particular means and log-variances. This second type of missing data accounts for only a few missing entries for specific variables.

In a certain sense, there was a third type of missing data. When a district had two or more full-time elementary teachers, all of whom had the same value for some characteristic, the variance of the characteristic was zero. And the logarithm of zero is undefined.

The three types of missing data, their missingness characteristics, and their replacement techniques are summarized in Table 8. A detailed discussion of the replacement techniques is given in the remainder of this section.

Rationale for Replacement of Missing Data. As noted in the preceding paragraphs, and in Table , certain kinds of data were missing for three basic reasons. In order to prepare a proper input data matrix, it was necessary to place values in cells where entries were missing in the initial data matrix. It was desired that a rational procedure be used for placing values in these empty cells. The substitution of a value for a missing entry is rational if the meaning of the substituted value has the same meaning as computed values in corresponding cells for other districts; that is, if the substituted value is an indicator for the same underlying variable.

Of the three types of missing data, Type C, Intra-District Equivalence, presented the most trivial and elementary problem for the rational substitution of values in empty cells. Consequently, its treatment will be presenred first. Thereafter are discussed the replacement procedures for missing data Types A and B, One-Teacher Districts and Improper Code Designations. These two types of missingness posed more serious methodological problems, and their arguments for replacement are developed in detail.

TABLE 8

Types, Characteristics and Replacement Techniques of Missing Data

Type	Characteristic	Replacement Technique
A. One-Teacher District	Only one full-time elementary teacher in a district, with valid characteristic codes; therefore, log-variances were not computable for Variables 7 to 12.	Regression estimates: For districts with no missing data, the relationships among Variables 13 to 31 and 7 to 12 were determined.
B. Improper Code Designation	More than one full-time elementary teacher in a district, some of which had one or more invalid characteristic codes, causing the district's value on one of the Variables 1 to 12 to be missing.	Regression Estimates: Same as for A.
C. Intra-District Equivalence	More than one full-time elementary teacher in a district, all of whom had the same value for a particular characteristic.	The lowest value of the log-variance for that characteristic, for any district, was substituted.

In the third type of missing data, Intra-District Equivalence, all teachers in a district had the same value for a characteristic. The measure to be taken was variability of the characteristic; low values of the measure were to correspond to low variability, and high values to high variability. To be useful in later analyses, it was important to maintain the low-to-high scaling of the measure. Because logarithmic functions were being used as measures of variability, it was necessary to use log-variance values as substitutes for missing data. Hence, the low-to-high scale characteristic was maintained for the substitutions. Consequently, when missing data of Type C was encountered, that is, when a variance constructed for one of Variables 7 to 12 was zero, the lowest log-variance for that variable which had been computed for any district with a valid entry was substituted. Treated this way, the replacement scheme for missing data Type C did not seem to raise any methodological or substantive issues. It was programmed to be carried out automatically without further investigation.

There were more serious difficulties associated with missing data Type A, where a district had only one full-time elementary teacher. The same general rationale existed for substituting values; that is, the substituted values and the computed values should be indicators for the same substantive variable. In particular, the substituted values should be estimates of the values that would have been obtained for log-variances of teacher characteristics if there had been more than one teacher.

As there was no way to directly estimate hypothesized values for missing data, a procedure needed to be developed which would yield rational predicted values for empty cells. Of the techniques available in the field of statistical applications to problems of prediction, regression analysis is considered to be the most justifiable procedure. These considerations implied that the relationships should be found between the variables for which there were no missing data and the variables which did exhibit missingness; and that those relationships should be used to estimate values for the missing entries. Before these relationships were computed, prior investigation was necessary to assure that the relationships between Variables 7 to 12 and Variables 13 to 31 were relatively constant over all districts. Discussion in the following paragraphs reports these analytic investigations.

ESTIMATING VALUES

Properties of Missing Data. The missing data to be investigated, Types A and B, occurred in such patterns that, for purposes of data analysis, missingness could not be considered to be random; that is, missingness was not independent of the processes being studied, for the districts with missing data had few full-time elementary teachers. Having few full-time elementary teachers is a characteristic of a district which might be expected to correlate with other characteristics, such as size or organization.

In order to measure the relationships between missingness in Variables to 1 to 12 (means and variances of teacher characteristics) and Variables 13 to 31 (district characteristics), a correlational analysis was performed. All districts were included in this analysis. There were 31 attributes: the first 12 were dummy variables constructed from the first 12 original variables, and the last 19 were the last 19 original variables. The dummy variables were computed by coding '1' for entries which were not missing and '0' for entries which were missing. The dummy variables may be called "non-missingness" variables. Substituting these dummy variables for the first twelve original variables resulted in a re-definition of the initial 31 x 632 data matrix. In the re-defined matrix, Variables 1 to 12 were dummy variables, and Variables 13 to 31 were the original variables. Using this re-defined matrix, means, standard deviations, and correlations of the variables were calculated and are presented in Appendix A.

The means of the attributes in Appendix A.1 are,considering the method of constructing the dummy variables, just the proportions of districts for which the corresponding original variables were not missing. Thus Variables 1 and 2 were missing for just a few districts--actually, 4 and 3 districts, respectively. Variables 3 to 6 were nowhere missing. And each of Variables 7 to 12 were missing for about 14% of the districts. The standard deviations of these dichotomous dummy variables are not useful, and the means and standard deviations of the last 19 variables are discussed later.

Intercorrelations of the attributes in the re-defined matrix are presented in Appendix A.2. The intercorrelations of Variables 7 to 12 are all approximately equal to 1.00, and it is clear that almost all the missing data was of the first type; that is, most districts had all of Variables 7 to 12 missing, or none of Variables 7 to 12 missing.

From examining the correlations between the dummy variables and original variables, it is clear that missingness is indeed related to other district characteristics. Districts without missing data tended to have high schools (r = -.46 with Scope), integrated state aid (r = -.54 with Class), many students per school (r = .46 with Students per School), and many staff members per school (r = .47 with Staff per School). These rather high correlations evoke some doubts about the substantive accuracy of assuming that relationships between Variables 13 to 31 and Variables 1 to 12 were the same for districts with missing data as they were for districts without missing data. This condition was observed, but investigations of its seriousness were beyond the limits of the project. Though this problem could not be corrected, it was still possible to maximize the use of available information. This maximization will be elaborated in following paragraphs.

Maximum Information Correlations. The next step leading to the actual replacement of the missing data was to perform another correlation analysis. The purpose of this analysis was to provide the best possible estimates of the true intercorrelations of the 31 input variables. In this analysis, the information in the original data matrix was used. But since there were missing entries in the matrix, standard correlation procedures could not be followed. For each coefficient (mean, standard deviation, or correlation) computed, only non-missing data could be used; the analysis used maximum information by using all non-missing data for the computations of the coefficients. This resulted in different numbers of districts being involved in computations of the different coefficients. The results of the maximum information analysis are presented in Appendix B.

In Appendix B.1 is given a matrix of counts of districts. Each entry in the matrix corresponds to a pair of original variables, and is the number of districts for which neither variable was missing. Each diagonal entry corresponds to a single variable, and is the number of districts for which that variable was not missing. As expected, the intercounts for Variables 13 to 31 are uniformly 632, which is the number of districts in the population. This indicates that there were no missing data for Variables 13 to 31. The intercounts of Variables 1 to 6 and the crosscounts of Variables 1 to 6 and 13 to 31 are about constant at 632: this indicates that there are some missing data of the Type B, Improper Code Designations. The intercounts of Variables 7 to 12 and the crosscounts of Variables 7 to 12 with the other variables are about constant at 542; missing data Type A, One-Teacher Districts in 90 districts. There were few missing data of the Type B in this segment of the matrix.

The maximum information means and standard deviations are presented in Appendix B.2. Each mean and standard deviation is based on all districts for which the variable was not missing; that is, it is based on the number of districts indicated in the corresponding diagonal entry of the count matrix of Appendix B.1. The maximum information correlations are presented in Appendix B.3. Each correlation coefficient is based on all districts for which neither correlate was missing; that is, it is based on the number of districts indicated in the corresponding entry of the count matrix of Appendix B.1.

In a sense, the maximum information coefficients provide a criterion of adequacy for the replacement of missing data technique. After substituting values for the missing data, changes in the means, standard deviations and correlations should be negligible. Substantial changes in these values would indicate systematic bias in the replacement procedure; the definitions—that is, the meanings--of the variables would have been changed. Later it will be indicated that the coefficients were not significantly changed by the replacement operation. Since missingness was shown to be related to the district characteristics, some changes in relationships among variables would be expected when missing data substitutions were made and all 632 districts included in the analysis.

Ha⁄i ıg established maximum information coefficients as criteria for the results of replacing estimated values for missing data, the operations for computing the substitution estimates were undertaken.

The Substitutions Operation. The general strategy for finding values to substitute for missing data involved two distinct phases: first, for districts with no missing data, the relationships connecting Variables 13 to 31 with Variables 1 to 12 were determined; second, for districts with missing data, the values of Variables 13 to 31, which were never missing, were manipulated according to the derived relationships. This produced estimates of Variables 1 to 12 for districts with missing data, which were then substituted for the missing values. Multiple regression procedures were used to determine the relationships between variables with missing data and variables with no missing data. Twelve linear regression equations were required; each of Variables 1 to 12 was predicted by the set of Variables 13 to 31.

The first phase, then, involved finding the twelve regression equations. This necessitated performing a correlation analysis. The attributes for the analysis were the 31 original variables. No values were missing, since the entitites were those districts for which no data was missing--there were 539 such districts. The results of the correlation analysis are presented in Appendix C. In Appendix C. 1 are the means and standard deviations, and in Appendix C.2 are the correlations. It should be noted that there are discrepancies between these coefficients and those of Appendix B, where all information is used. This is caused by the fact that about 90 districts have been omitted, and those 90 districts are not randomly scattered, but rather are all districts with few elementary teachers. For example, the districts in the reduced sample have, on the average, about 134 more elementary students; that is, the average elementary enrollment of the entire population of 632 districts is 798, whereas the average elementary enrollment is only 932 when the 90 single-teacher districts are omitted.

The regression equations were determined from the correlations given in Appendix C. The standard form of the equations is presented in Appendix D. Each row contains the beta coefficients for predicting one of the Variables 1 to 12 from the nineteen Variables 13 to 31. The beta coefficients are followed by the squared multiple correlation coefficient for the equation. The inference is that the regression equations are able to account for from 13% (Variable 8) to 53% (Variable 2) of the variation of Variables 1 to 12. Although this is hardly perfect prediction, a significant amount of the variation is being predicted. If the prediction were perfect, then Variables 1 to 12 would be strictly redundant.

From the regression equations of Appendix D, and the never-missing values of Variables 13 to 31, the regression estimates of Variables 1 to 12 were computed for those 90 districts that had missing data. These regression estimates were then substituted for missing data.

The replacement scheme adopted for missing data Type A, One-Teacher Districts, also seemed satisfactory for resolving the problem of missing data Type B, Improper Code Designations, of which there were only two cases.

THE INPUT DATA MATRIX

As a result of the substitution procedures outlined above, the data matrix was complete: all empty cells had been filled. The remaining step was to determine whether the maximum information coefficients of Appendix B had been substantially altered. For this purpose, a correlation analysis was performed, with all 632 districts as entities, and the original variables as attributes, with regression estimates substituted for missing data. The results of this analysis are presented in Appendix E. Appendix E.1 contains the means and standard deviations of the variables, and Appendix E.2 contains their correlations.

The difference between the coefficients of Appendix B and those of Appendix E are small. In general, differences occur in the second or third decimal place, and they occur primarily in the means, standard deviations, and correlations of Variables 7 to 12. Any change of importance would have occurred among these variables, because estimated data was provided for a special subset of the population of districts. As important changes did not occur among these variables, further investigation was not warranted. Because of the small differences caused in correlations by using regression estimates to replace missing data, the data matrix with these estimates substituted was considered to be a satisfactory Input Data Matrix, and served as the basis of all subsequent analyses.

PART II

SECTION C

FACTORIZATION: REDUCING AND ORTHOGONALIZING THE VARIABLES

The 31 variables of the input data matrix were transformed into a smaller number of uncorrelated (orthogonal) variables for stratification. Multivariate techniques were employed in reduction and orthogonalization. In this section is first discussed why these operations were necessary. Then the two computing algorithms used are presented and discussed. Finally, the reduced variables derived from the two algorithms are discussed, interpreted, and compared.

METHODOLOGICAL ISSUES

The Need for Summarization and Reduction. One of the purposes of this study was to obtain a set of variables which could be used for stratifying districts. As described in the previous sections of this Part, the original information available for this purpose was manipulated to produce the input data matrix of 31 variables. For the purpose of district stratification, this data matrix might possibly have been used in any one of several ways: 1) ·ll 31 input variables might be used as a set of stratifying dimensions; 2) a sub-set of the 31 variables might be selected and used for stratification; or 3) an efficient summarization a.d reduction of all 31 variables might be derived. From the methodological viewpoint stated in Part I, the third of these alternatives, summarization, was considered the most rational choice.

Use of all of the 31 variables as a set of stratifiers was impractical, since these were too many variables for construction of a meaningful or useful stratification scheme. For example, if each variable was dichotomized at its median, 2^{31} district categories would have been defined. Since there were only 632 districts, most of these categories would have been null, and only a few districts would have been grouped together in even the largest category. Thus no efficient categorization, or stratification of districts, would have occurred.

Selection of a sub-set of the 31 input variables would require justification for the inclusion or exclusion of each variable. However, each variable is assumed to be an important indicator of a district characteristic. All of these importan characteristics should be somehow included in a stratification scheme; to exclude some of these characteristics from the stratification would be to ignore the acknowledged multivariate complexity of the districts. Furthermore, as discussed in Part I, the selection of a small sub-set of the 31 variables would be equivalent to the formulation of an ad hoc theory for stratification.

Summarization and reduction of the 31 variables is desirable in order to simplify the characterization of districts, and yet maintain sufficient descriptive complexity. This procedure implies that no single variable contains sufficient informat on to be accepted as a stratifying dimension. Rather, each variable is an "indicator" of an underlying dimension. The co-relationships of several indicators, or variables, when taken together as a composite would be expected to define an underlying dimension. For example, the enrollment variables (Variables 16 and 17) and the number-of-employee variables (Variables 18 through 22) might be hypothesized as indicators of the same underlying factor, say, district size. Thus it was considered that the factors hypothesized as underlying the 31 variables would be more fundamental, and hopefully more meaningful dimensions than the original variables for the purpose of stratification.

Orthogonalization of the Variables. The reduction of the number of variables was accomplished using multivariate techniques. These techniques allowed the mathematical derivation of composite variables which were based on the original 31 variables. It was possible to derive composite variables which were orthogonal (uncorrelated) to one another. This orthogonalization was particularly useful in the stratification, and also presented statistical advantages in comparing the composite variables with other (outside) variables.

Multivariate Analysis. The need for reduction in the number of variables and the advantages of orthogonality implied the use of multivariate analysis; in particular they implied the use of certain techniques generally classified as factor analysis. Harris (1955) discusses two possible types of factor analysis models: communality and non-communality models. With general communality-type models, the factors lie outside the variable space, and the factor scores are estimatable but not computable. But with certain non-communality models, the factors lie within the variable space, and factor scores for entities are exactly computable. For this study, the scores for the factors derived from the original 31 variables were to be used for stratification, so it was necessary to use a model wherein the factor scores were computable. Therefore only non-communality analytic models were considered.

The two forms of non-communality analysis considered were Hotelling's (1933, 1935) component analysis and Harris' (1962) version of Guttman's (1953) image analysis. The concepts of image analysis are based on certain linear transformations of the original variables; the concepts of component analysis are based directly on the original variables. For the present study, it was decided to try both forms of analysis. For each form, the analysis consisted of two operational phases.

First, the matrix to be factored was defined and its unrotated factors were computed. Second, the normal varimax rotation procedure (Kaiser, 1958) was applied to the unrotated factors, and a new set of factors--the rotated factors--was obtained. For each form of analysis, the first phase was designed to bring forward certain theoretic qualities of the variables, and the second phase was designed to provide interpretable factors.

ANALYSES

The analyses for both the component and image models are presented here, together with discussion of their theoretic properties and their substantive interpretations for this study. The results from the application of the two models are then compared.

Component Analysis. The inputs to the component analysis consisted of the 31 by 632 input data matrix and its correlation matrix R (given in Appendix E.2). The first objective of the component analysis was to find an orthogonal basis for the variable space. This meant finding a factor matrix F_r which corresponded to the correlations between the original variables and a set of factor scores. These factors have the following properties:

[1] distributions of factor scores are uncorrelated;

[2] they are linear combinations of the original variables;

[3] they span the space of the original variables.

Such an F_r would satisfy the relationship $F_r F_r' = R$. In this study, R was non-singular, and there had to be 31 factors in order for [3] to be satisfied. Hotelling (1933, 1935) gave a technique for finding an F_r which also has the property:

[4] the first factor accounts for a maximum amount of variance in the variables; and each succeeding factor accounts for a maximum of the remaining variance, all given that each factor is uncorrelated with all the preceeding factors.

That F_r is called the principal component factor matrix. Property [4] allowed as much variance to be accounted for in as few variables as possible. Although the property was lost in the rotation of the basis, it was possible to see how much compression of the 31 variables could be achieved.

The computation of F_r proceeded as follows. First a latent root and vector resolution of the correlation matrix was obtained. The latent roots are presented in Appendix F.1; they are considered to be the diagonal entries of a diagonal matrix called M^2. The latent vectors are presented in Appendix F.2; each column is the unit-length latent vector for the corresponding latent root in M^2. The matrix of latent vectors is called Q. It follows from the definition of latent roots and vectors that $R = QM^2Q'$. Hotelling (1933, 1935) showed that the factor matrix F_r, that corresponds to the factor scores which satisfy properties [1] through [4] above, is given by $F_r = QM$. It can be seen that $F_r F_r' = (QM)(QM)' = QM^2 Q' = R$. The matrix F_r is given in Appendix F.3. Each row corresponds to an original variable, and each column corresponds to a factor; the factors are arranged in decreasing order of variance accounted for. Appearing before each row is the row sum of squares, which is always 1.0 since all the variance in each variable is accounted for by the set of factors. Above each column appears the column sum of squares which is the amount of variance accounted for by the factor and which is equal to the corresponding latent root. The amounts sum to 31, and are additive since the factors are uncorrelated. For example, the first factor accounts for about 30% of the variance, and the first five factors account together for about 70% of the variance.

The factors given in F_r had the advantage of compression, but they were not very interpretable; that is, they could not be readily identified with single substantive dimensions. So a rotation was performed. An orthonormal matrix T_r has the property that $T_r T_r' = T_r' T_r = I$, the identity matrix; then a rotation of F_r can be written

F_rT_r since $(F_rT_r)(F_rT_r)' = F_rT_rT_r'F_r = F_rF_r' = R$. Such rotations can provide more interpretable factors, but some of the compression is always lost. But interpretability was essential for the stratification, and the normal varimax rotation procedure (Kaiser, 1958) was applied to secure a T_r. With the normal varimax procedure, the variance of the squared, row-normalized entries of F_rT_r is maximized, and Kaiser claims that a) interpretable, simple-structure factors result, and b) those factors are relatively invariant under changes in the variable selection. Quality[b]is nice to contemplate, but impossible to test. Quality[a]can be verified by examining the rotated factor matrix.

The varimax transformation matrix T_r is presented in Appendix G.1. The rotated factor matrix F_rT_r is given in Appendix G.2. Its rows correspond to the original variables, and its columns correspond to the rotated factors. The rows are bordered by row sums of squares, which are uniformly 1.0 since all the variance in each variable is accounted for by the factors. The columns are bordered by column sums of squares, which equal the variances accounted for by the factors; the factors have been arranged in decreasing order of variance accounted for. Since the factors are uncorrelated, the variances are additive. The first factor accounts for about 25% of the variance, and the first five factors account for about 50% of the variance. It can be seen that the compression has been relaxed, since for the unrotated factors the corresponding figures were 30% and 70%.

The factor matrix (Appendix G.2) gives the correlations between the factors and the original variables. But its entries are not the weights needed to compute the factor scores from the original variables. (See Glass, 1966.) The weights are the entries of the matrix $QM^{-1}T$ (Kaiser, 1962), which is presented in Appendix G.3. Actually, the columns of the matrix have been normalized for printout purposes, but proportionality within columns is correct.

Image Analysis. The input to the image analysis was the 31 by 632 input data matrix and its correlation matrix R (given in Appendix E.2). But conceptually, the image analysis dealt with a transfomed set of variables--the images of the original variables. In image analysis, each variable is conceptually partitioned into two parts: the image variable, which is the original variable as predicted by linear regression from the other original variables; and the anti-image variable, which is the regression residual. Guttman (1953) has shown that in the limit, as a universe of content becomes permeated with variables, image analysis approaches communality-type factor analysis. That is, the image and anti-image variables are approximations, within the original variable space, of the common and unique parts of the variables in a communality type analysis. To the extent that the approximation is accurate--and to the extent that the accuracy is associated with the intercorrelations of the anti-image variables--using factors derived from the image variables may lead to a more reasonable model for the data. Each variable has a certain portion of unexplained (unique or anti-image) variance; it is not necessary to consider each variable as being entirely contained in a common factor space.

Another important property of image analysis is that the results are scale-free. With an analysis such as component analysis, the matrix factored is the correlation matrix, which is the covariance matrix for variables with variances of exactly 1.0. If any rescaled version of the correlation matrix, corresponding to variables with different variances, is factored, different results are obtained. But Guttman (1960) noticed that image analysis produces the same results no matter what scaling is applied to the image covariance matrix. As Kaiser (1963) states: "We are freed from the traditional agnostic confession of ignorance implied by standardizing the [variables]. Here, standardization is merely a convenience to which we are in no way tied." The rotation procedure employed later (normal varimax) preserves this scale-free property.

Because of certain matrix identities, it was not necessary to construct the 31 by 632 matrix of image variables in order to perform the image analysis. Rather, the computation could proceed directly from the correlation matrix R of the original variables. (See Kaiser, 1963.) First the inverse R^{-1} of R was obtained. The reciprocals of the diagonal entries of R^{-1} were considered to be the diagonal entries of a diagonal matrix S^2; these were the variances of the anti-image variables and are presented in Appendix H.1. Each of these entries is the proportion of the variance in the corresponding original variable which was not predictable from the other original variables. Some variables in the present study (for example, 16 to 20) were almost entirely predictable; others varied from about 11% to about 67% unpredictable. The precise transformation from original to image variables is given in Appendix H.2. The matrix $W = 1 - S^2R^{-1}$ which appears there has rows corresponding to original variables and columns corresponding to image variables. Each column is the regression equation for predicting the corresponding original variable from the other thirty variables; that is, for calculating the image variable from the original variables. For printout purposes, the matrix has been column-normalized, but proportionality within columns is correct. The image variables do not equal the original variables; they are non-singular linear transformations of the original variables, and the variance of an image variable is that part of the variance of the corresponding original variable which is predictable from the other original variables.

As a test of the fit to pure factor analysis, the correlation matrix of the anti-image variables was produced. It appears as Appendix H.3, and it was computed as $SR^{-1}S$. The unique parts of pure factor analysis are assumed to be perfectly uncorrelated; so to the extent that image analysis is an approximation of pure factor analysis, the anti-image variables should be uncorrelated. That is, the matrix in

Appendix H.2 should be essentially diagonal. It is diagonal to a certain extent; that is, most of the entries are rather small--less than, say, 0.2. But there are some rather large entries, on the order of 0.9. This is somewhat disconcerting, but then again, the pairs of variables indicated by very large entries are very highly correlated, and they load on the same factor.

The covariance matrix G of the image variables was the matrix factored, and the factorization scheme employed was that of Harris (1962). First the matrix $S^{-1}RS^{-1}$ was formed, and its latent roots and vectors were determined. The latent roots are presented in Appendix I.1 and are considered to be the diagonal entries of the diagonal matrix B_r^2; the corresponding latent vectors are presented in the columns of Appendix I.2 and are considered to form the columns of the matrix X. By the definition of latent roots and vectors, $S^{-1}RS^{-1} = XB_r^2X'$. Harris (1962) showed that $S^{-1}GS^{-1} = XB_g^2X'$, where $B_g^2 = (B_r^2 - I)^2 B_r^{-2}$. So the Harris factor matrix for G is $F_g = SXB_g$ and $F_g F_g' = (SXB_g)(SXB_g)' = SXB_g^2X'S = SS^{-1}GS^{-1}S = G$. F_g appears as Appendix I.3, where it is bordered by row and column sums of squares. The row sums of squares are the image variable variances. The column sums of squares are the amounts of image variable variance accounted for by the factors, and these amounts are additive, since the factors are uncorrelated. Each set of sums of squares sums to 22.4 which is about 72% of the 31 units of variance that were in the original variables.

Note that the Harris factors of G were obtained rather than, say, the principal factors based directly on the latent roots and vectors of G. Certain features of the Harris factorization provide insight into the data. Note that the first 19 roots of G, which correspond to Harris roots of R greater than 1.0, decrease monotonically; and the last 12 roots of G, which correspond to the Harris roots of R which are less than 1.0, increase monotonically. Harris (1962) showed that:

[a] if image analysis is regarded as a first approximation to canonical factor analysis, then the first set of factors correspond to the real canonical correlations;

[b] the number of factors in the first set is the strongest lower bound, devised by Guttman (1954), on the number of common factors in the set of variables;

[c] small roots in the second set imply high off-diagonal entries in the anti-image correlation matrix; that is, poor approximation to communality-type factor analysis.

The ıst point is implicit in Harris' (1962) formulation, and is made explicit in Kaiser (1963). From a formula given by Kaiser (1963, page 164) it was calculated that the second set of factors accounts for about 68% of the sum of squares of the off-diagonal entries of the anti-image correlation matrix.

Harris (1962) suggests that incomplete image analysis, employing only the first set of factors, is desirable. That is, he suggests that the first set of factors should form the basis for the entire image-analytic model of a given set of data, on the grounds that they are the part of the original variables which maximize the approximation to communality-type factor analysis. For three reasons, however, his advice was not followed. First, the omission of the second set would have left still about 32% of the sum of squares in the off-diagonal regions of the anti-image correlation matrix. Significant entries would have remained there. Second, most of the Harris-Guttman reasoning assumes an application to the kinds of variables obtainable from psychological testing, for which approximation to a universe of content is conceptually reasonable. It was not clear that the demographic and enumeration variables of this study could be said to approximate a universe of content; they were selected according to availability rather than according to conscious permeation of a theoretical universe. They were skewed rather than normal. Third, for purposes of comparison, it was considered useful to maintain compatibility with the component analysis, in which all factors were retained. Consequently, all of the image factors were used, and the rotation destroyed the properties of the Harris factorization; any complete factorization would have led to the same rotated factor matrix.

In order to obtain a substantively interpretable basis for the factor space, the normal varimax rotation procedure was applied.[1] The transformation matrix T_g obtained is presented in Appendix J.1, and the rotated image factor matrix $F_g T_g$ is presented in Appendix J.2. The rows and columns of $F_g T_g$ are bordered by sums of squares. The row sums of squares are again the variances of the image variables. The column sums of squares are the variances accounted for by the factors and have been arranged in decreasing order. Since the factors are uncorrelated, these variances are additive. The first factor accounts for about 36% of the image variance, and the first five factors together account for about 87% of the image variance. This is not the most compressed basis possible for the factor space, but it will be shown that the factors obtained in this study are quite interpretable. Because the factor scores were to be computed, the factor weight matrix was calculated. It appears as Appendix J.3 and was computed as $SXB_r^{-1}T_g$; the columns correspond to image factors and give the linear functions for calculating the factor scores from the original variables. For printout purposes, the columns have been normalized, but the proportionality within each column is correct.

[1] The application parallels that made to the components analysis, q.v.

Comparison of Component and Image Models. A rotated component analysis and a rotated image analysis were performed. For the purpose of stratification, one of the analyses needed to be selected. The theoretic qualities of the analyses were to be considered in making the choice, but it was also necessary to consider the substantive interpretability of the factors. Ultimately, the strata were to be used for answering substantive questions. The factor matrices were used for interpretation and comparison. Appendix G.2 is the factor matrix for the component, Appendix J.2 for the image models. The first five factors of each are abstracted in Table 9, where the name and loading of each high-loading variable on each factor is given. The image factors were finally selected and named, and in Table 9 the names assigned to the image factors are given. Also, the correlations between the factor scores of the two analyses were computed, as $T_r' M^{-1} Q' RS^{-1} XB_r^{-1} T_g$, and appears as Appendix K.1; the rows correspond to the rotated component factors and the columns to the rotated image factors.

Theoretical considerations led to the choice of image factors for stratification. First, the image factors were scale-free, and no assumptions had to be made concerning the relative metric properties of the original variables. Second, the image factor model is an approximation to the more reasonable communality models, for which it is not necessary to assume that each variable is entirely contained in a common-factor space.

A theoretical advantage of the component analysis would have been parsimony; that is, the unrotated component analysis extracts the most variance in the fewest factors. For the present data, this is indeed true if we consider the unrotated component factors and the unrotated image factors; the first five component factors account for more variance than the first five image factors. But for the present data, this advantage

TABLE 9

A Comparison of Rotated Factor Structures

Factor One: Numerical Size	Factor Two: Organizational Complexity	Factor Three: Teacher Experience	Factor Four: School Unit Size	Factor Five: Economic Power
IMAGE ANALYSIS				
Secondary enrollment (99)	Staff per school (91)	Mean total experience (74)	One-room schools (58)	Valuation per student (71)
Elementary enrollment (99)	Students per school (90)	Mean local experience (73)	Variance grade-spread (52)	Scope (50)
Other professional employees (99)	Valuation per school (75)	Mean degree (59)	Two-room schools (48)	Class (32)
Equalized valuation (99)	Mean salary (72)	Variance local experience (57)	Scope (-35)	Mean grade-spread (25)
Full-time elementary teachers (98)	Class (-70)	Mean credential (54)	Valuation per school (-32)	
Full-time junior high teachers (97)	Mean grade-spread (-66)	Variance total experience (39)		
Full-time secondary teachers (97)	Mean degree (65)			
Three-or-more-room schools (95)	Variance salary (56)			
Part-time teachers (40)	Variance local experience (46)			
	Scope (-46)			
	Part-time teachers (41)			
	Mean credential (38)			
	Variance total experience (38)			
COMPONENT ANALYSIS				
Factor One	Factor Two	Factor Three	Factor Four	Factor Five
Secondary enrollment (99)	Staff per school (89)	Mean local experience (91)	Valuation per student (96)	Variance total experience (87)
Elementary enrollment (99)	Students per school (88)	Variance local experience (57)	Scope (28)	Variance local experience (35)
Other professional employees (99)	Valuation per school (86)	Mean total experience (38)	Class (26)	
Equalized valuation (99)	Mean salary (55)	Mean credential (23)	Valuation per school (21)	
Full-time elementary teachers (98)	Class (-49)			
Full-time junior high teachers (97)	Mean grade-spread (-45)			
Full-time secondary teachers (97)	Mean degree (44)			
Three-or-more-room schools (95)	Variance salary (33)			
Part-time teachers (38)				

is lost when the factors are rotated. The first five rotated component factors account for about 51% of the variance of the original variables. The first five rotated image factors account for about 87% of the variance of the image variables, and for at least 60% of the variance of the original variables. So, in terms of rotated or meaningful structure, the image factors are more compressed and parsimonious.

The greater parsimony of the image factors leads us to the most important reason for choosing them as stratifying dimensions: they are more interpretable. It can be seen from the crosscorrelations of the two sets of factors (Appendix K.1) that the first and second rotated component factors are essentially equal to the first and second rotated image factors. The first and second image factors have been named "Numerical Size" and "Organizational Complexity". In the image analysis, most of the information of the input data matrix variables has been compressed into the first five factors. Factors One to Five are truly clusters of variables. And, as will be seen in Part III, the factors are reasonable and meaningful. On the other hand, rotated component factors Three to Sixteen are all of essentially the same magnitude in accounting for variance. That is, beyond the first two factors in the component analysis, no real clustering of the variables has occurred. That is, most of the factors, specific factors, each with just one high-loading variable. The first two factors would be useable for stratification, but the others are not substantially different from the input variables.

PART II

SECTION D

DATA PROCESSING: TECHNIQUES OF COMPUTING SCORES AND CODING DISTRICTS

The first three sections of Part II discussed conceptual and empirical problems of the input variables, of missing data, and of factorization. This Section D presents the procedures and problems of using the computer to manipulate input data to provide substitutions for missing data, and to derive the composite variables. Furthermore, it is demonstrated how the computer was programmed to assign codes for the composite variables to loc-l education agencies, so that the district population could be stratified.

DATA REDUCTION PROGRAMS

Manipulating Input Data. The original sources of data were the "School/District Tape", the "Employee Tape", and the "Valuation Deck". These sources were described in Section II.A. The first computer program called,program SSVAR, had as input these three files. The essential function of the program was to merge the information from the three files, and to produce as output a new tape called the "Variable Tape".

For each district, the program assembled the following records: each district's record, from the "School/District Tape"; all of the school records for each district, also from the "School/District Tape"; all of the employee records for each district, from the "Employee Tape"; the equalized property valuation for each district, from the "Valuation Deck". The program then checked whether each district had any elementary enrollment. If the district did have elementary students enrolled, the construction of the variables proceeded. Districts with no elementary enrollment were omitted from further consideration. Construction of the variables involved crossreferencing the collected records, extracting codes and recoding, counting and summing, and forming means and variances. As the program proceeded from one district to the next, all of the data were accumulated in a large table in the computer storage. Special codes were stored in this table when missing data were detected for a district. Finally, when the last district record was encountered, the program went into its second phase.

Substituting for Missing Data. The second phase of program SSVAR was concerned with testing and replacing the missing data. The operations and results have been discussed in Section II.B. In this phase, program SSVAR went through four steps. First, it performed three correlational analyses and produced as output the matrices appearing in Appendices A to C. These analyses provided tests for the characteristics of the missing data. But the program was set up to continue with the replacement operation. In the second step, the regression coefficients were computed; they are presented in Appendix D. Third, the regression estimates for the missing data were computed and substituted into the internal table. In the final stage, a correlational analysis was performed which was based on the matrix with missing data replaced by regression estimates; the results of this analysis appear in Appendix E.

The program SSVAR punched the final correlation matrix on cards, together with the final means and standard deviations of the variables in the input data matrix. The input data matrix was written on magnetic tape in a binary format. Fourty-four numbers were written in a record for each district. The first of these numbers was the WSDPI district identification code; the next 31 numbers were the district's values for the 31 variables of the input data matrix; the final 12 numbers in the record were repeats of the first 12 variables, but with special codes for missing data instead of regression estimates. Thus it could be determined, from a tape printout, which values were regression estimates and which were actual data. The result was the "Variable Tape". A listing of its contents was made for purposes of checking and reference.

Factor Analyses. After the output from program SSVAR was checked and decisions were made about the results, the next program, called program SSFAC, was run. This program had as input the "Variable Tape" and the punched means, standard deviations and correlations produced by program SSVAR. Program SSFAC performed the components analysis, the image analysis, and the cross-correlational analysis. Then the program computed the rotated components and rotated image factor scores, and added them to the "Variable Tape". The modified "Variable Tape" then contained 106 numbers in a record for each district: these included the original 44 numbers, plus 31 rotated components factor scores and 31 rotated image factor scores. Finally, the program produced a listing of all the factor scores, to correspond to the listing of the original variables, for purposes of checking and reference.

Limitations of the Programs. The above description of the programs SSVAR and SSFAC is oversimplified. Although the programs did indeed perform the operations indicated, they were complicated programs, and involved a variety of difficult problems.

First, there were problems of tape compatibility. The tapes obtained from the WSDPI had been prepared on IBM computers, and the processing described here took place on the University of Wisconsin CDC computers. Although the IBM and CDC tape reading/writing schemes are theoretically compatible, there are subtle differences in intensity and al ignment of the magnetic recordings. Tapes written on IBM equipment are difficult to read on CDC equipment. The initial runs of program SSVAR were unsuccessful because of tape reading errors. Severa duplicate tapes from the WSDPI were tried. Finally, the University's auxilary IBM 1460 computer was programmed to copy the WSDPI tapes. The 1460 had been adjusted by IBM engineers to have recording characteristics which are a compromise of the IBM and CDC peculiarities. The copies produced by the 1460 could then be read by the CDC computers.

Second, there were coding problems within the WSDPI tapes. For some yet unknown reason, one particular school did not have a school record on the "School/ District Tape". This caused the merging operation of program SSVAR to halt. A special patch was placed in the program to skip the employees of that school.

Third, the entire operation was performed originally for all districts in the state. But upon examination of the results, it was apparent that the inclusion of districts with no elementary schools was warping the factor structure and stratification. So the programs were rerun with the additional instructions to omit districts which included only high schools.

The point of the above notes is that the programs were specially written for the data received from the WSDPI. They are special programs for operating on special data. Program SSVAR accepts data only in the particular format of WSDPI tapes. And program SSFAC operates only on the 31 variables defined for this study, and only with the particular inputs from the first program. The programs are not useful or useable for any other study. The only case where these programs could be used without modification would be a replication study, using WSDPI information for another year, say, 1967. Even then, the programs would need slight adjustments to meet the particular peculiarities and problems of the 1967 data. The contributions of this project are the production of the results for the particular data, and the development of a general methodology.

The Data Bank. The hard result of the computer processing is the "Variable Tape". This tape contains the original 31 variables; the codes for real data and for data which are regression estimates; the rotated components factor scores; and the rotated image factor scores. The tape is written in such a form (floating-point binary) that it can be used only on CDC computers, but CDC computers could be used to transfer the data onto cards or onto IBM-compatible tape. For analyzing the outside variables described in Section III.C, for example, the first five rotated image factor scores were punched on cards for input to a standard statistical program. The stratification program described below, on the other hand, operates directly on the "Variable Tape". The factor scores give composite measures of the qualities of the districts. The original variables give direct indices of the qualities, such as enrollment, of the districts. In summary, the "Variable Tape" is a bank of data on the elementary school districts of Wisconsin.

STRATIFICATION

The data bank provided the desired empirical basis for classifying districts. The object of classifying districts was to form relatively homogeneous sub groups of districts. The composite variables were used, then, to build a multivariate stratification of the population of Wisconsin elementary school districts. There are several possible ways in which the data bank could be used to construct a multivariate stratification. These possibilities are most distinctly differentiated by the choice of points for segmenting distributions of factor scores. For purposes of this project and the related illustrations, each of a few selected factor score distributions was simply dichotomized.

The stratification, then, does not utilize the composite variables in their full, continuous generality. Rather, strata are defined by dichotomizing selected factor scores. This splitting of distributions is necessary for purposes of classifications; for example, it might be used for dividing the districts into groups of high or low Organizational Complexity. It should be noted that classification is a special case of measurement: it is measurement with variables which have only two values, zero and one. The dichotomized scores contain less information, or fewer distinctions between districts, than the parent distributions of the scores.

A special computer program, identified as SSRAT, was prepared to perform stratifications. There were three sources of input to SSRAT: they were the "Variable Tape", a deck of cards specially prepared from the "School/District Tape", and certain control information. The tape contained the image and components factor scores, as well as the original variables. The cards contained the actual name of each school district, as well as its Kind, Class, and Scope codes. The control information included a code to indicate whether the rotated image or the rotated components factors were to be used, codes to indicate which factors were to be used, and codes for each factor to indicate whether that factor was to be dichotomized at its mean or at its median. The program first read the control information and the deck of name cards. Then it read the "Variable Tape", and stored the selected factor scores; and finally the scores were dichotomized.

The SSRAT program then produced its first important output, which was a complete list of the districts ordered by their WSDPI code numbers. This output contained four types of information for each of the 632 districts: the WSDPI code; the district values for Kind, Scope, and Class; the district scores on the selected stratifying factors; and a code that indicated which classification group the district had been assigned to.

A code for a classification group is called a stratum pattern, and is in the form of a series of pluses and minuses. A plus indicates that a district had a score on the associated factor which was greater than the dichotomization point (mean or median of the distribution); if a district had a score below the dichotomization point, it was assigned a minus for that position in the pattern. For instance, if there are only two factors used in a stratification, then all the possible stratum patterns are : + + , + - , - + , and - -. Each district would have one and only one of these patterns, because it would be in one and only one classification group.

In the case of using only two stratifying variables, the output[1] might look like this:

ID	Name	Kind	Scope	Class	Pattern	Factor Scores	
0056	Wellington	3	4	2	+ -	2.34	- 1.67
0063	Dunedin	2	1	1	- -	-0.47	- 1.22
0070	Timaru	3	1	4	+ -	1.46	- 0.05
0168	Christchurch	4	2	3	+ +	3.11	2.24

The first column is the district identification code, and the listing is ordered by that code. The other columns are, from left to right: the name of the district; the Kind, Scope, and Class of the district; the stratum pattern; and the factor scores. In the actual stratification, five factors were used, instead of two.

The second important output of SSRAT contained a section for each stratum. In these sections were lists of the districts in the strata, and summaries of the characteristics of the strata. The program obtained the list of the districts in each stratum by sorting the information in the first output according to the stratum pattern. The summary of characteristics was then obtained after rereading the "Variable Tape" and retrieving and accumulating the original variables separately for each stratum. For example, the

[1] This is a fabricated example.

total elementary enrollment in the stratum is determined as the sum of Variable 17 over all districts with that stratum pattern. The format of this summary is given in Table 10; the data given in this table are the same summaries taken over all the districts in the state. Selected aspects of Table 10 are presented for all 32 stratum summaries in Appendix L, which permits direct cross-strata comparisons of summary characteristics.

The program SSRAT also produced two less important outputs. The first of these was a table of counts, which shows how many districts were assigned plus and how many were assigned minus on each factor. The second output was a list of districts ordered by stratification pattern codes. This list contained the same information as the first important output, described above: district identification code, name, Kind, Scope, Class, and Stratum pattern. The formats of the two lists were identical; the only difference between them was the method for ordering districts.

TABLE 10

District Summary Characteristics as Printed by Program SSVAR

DISTRICTS
(Listings by Stratum)

Total	Kind City	County	Scope With HS	Without	Class Integrated	Basic +	Basic
632	128	504	351	281	388	16	228

	Enrollment			Staff						Schools				Val.
	Sec.	Elem.	Total	Elem.	JHS.	HS.	AD/T	AD	Total	1-R	2-R	3+-R	Total	(M)
Σ	308137	504334	812471	18219	3902	11212	1362	2190	36885	424	291	2031	2746	21316.0
$\bar{x}$	487.6	798.0	1285.6	28.8	6.2	17.7	2.2	3.5	58.4	.7	.5	3.2	4.3	33.7

Valuation/Student	Student/School	Student/Staff	Staff/School	Valuation/School
26235.986	295.874	22.027	13.432	7762555.608

PART III: SOME APPLICATIONS OF THE ALGORITHM OUTPUT

Section A - Illustration: Characterization and Description

Section B - Illustration: Categorization and Comparison

Section C - Illustration: Logical and Statistical Analysis

Section D - Illustration: Sampling

PART III

SECTION A
ILLUSTRATION: CHARACTERIZATION AND DESCRIPTION

Discussion in this section will center on reporting the exte:.t to which the algorithm provides efficient description of the dimensions along which school districts vary. The known, raw indicators were the original 31 variables defined and discussed in Section II.A. The analytic procedures discussed in Section II.C derived the five factors[1] from the original 31 variables; these factors were weighted sums or composites, and accounted for at least 60% of the variance of the input data. That is, the five factors compressed the input information and therefore were summary dimensions of school district variation. These dimensions also had the special quality of being uncorrelated.

As summary dimensions, the factors are of intrinsic interest in understanding and describing the variation among school districts. So it is necessary to reach an understanding of the factors as quantitative measures on the districts. The discussion in this section is broken into two parts. First, the contents of the factors will be explored by examining the relationships between the factors and the original variables. Second, the distributions of the factors will be investigated for the purpose of understanding the empirical representation of their contents.

FACTOR CONTENTS

The compositions of the factors are defined by the columns of the rotated image factor matrix, Appendix J.2, which is abstracted in Table 9. Descriptions of the factor contents are given in the following paragraphs. Special attention was given to the variables

[1] Consistent with Section II.C, only the first five rotated image factors are treated here.

which had high loadings on a factor, and consideration was made of the kinds of districts which had simultaneously high values for those variables. Labels were assigned to the factors: the labels are for reference both to the operationally defined factors and to the undefined constructs which they represent. In choosing the labels, considerations had to be given to the fact that the factors were uncorrelated. The factors and their labels are the results of a particular set of data and a particular analysis. For example, if ten more variables had been available, then the present factors might have been augmented or their structure might have been altered.

Factor One, Numerical Size. This factor was the largest one in the sense that it accounted for more variance than any other factor. The name was chosen because districts with high scores on this factor tend strongly to have large numbers of students and teachers at all grade levels; they also have relatively high numbers of large schools and are supported financially by a large total equalized property valuation.

It would have been possible to employ statistical techniques to partial the influence of size out of the input variables before computing the factor analysis; then no general size factor would have appeared. One of the objectives of this investigation, however, was to produce results which would be useful for a variety of purposes, and because size is often an important variable in studies of LEAs, no attempt was made to eliminate its effects. The orthogonal factorization ensured that this size factor is uncorrelated with the other image factors.

Factor Two, Organizational Complexity. A district with a high score on this factor tends to have relatively high numbers of students and teachers per school. It also tends to have a greater degree of economic autonomy than districts with lower scores: teachers in such a district receive relatively high salaries, schools in the district have large equalized property valuation, and the district receives a high level of state aid. The teachers in such a district are quite variable with respect to salary and experience, but on the average they have more advanced academic degrees and teaching credentials than their colleagues from districts with a lower Organizational Complexity score. An elementary teacher from the district with the higher score on this factor is likely to specialize in teaching at a single grade level, but in the district there will be several teachers who have non-teaching duties, such as counseling, or who teach only part time. Finally, districts with high scores on this factor are more likely to include a high school than their less complex counterparts. The relationship between the presence of a high school and score on Organizational Complexity is discussed further in the notes on factor score distributions.

Factor Three, Teacher Experience. The six variables with significant loadings on this factor are all measures of the characteristics of the teachers within districts. Districts with high scores on this factor have teachers who have high degrees and credentials, and whose averages for both local and total experience are high. There is also a relatively large variance among the teachers in the district with respect to the length of time they have taught, both locally and in total.

Factor Four, School Unit Size. A district with a high score on the School Unit Size factor has a relatively high number of one-room and two-room schools; it has a small total equalized property valuation per school; it probably has a high school; and it is likely that the teachers in its schools are teaching two or more grade levels.

Factor Five, Economic Power. Districts with high scores on this factor have a high ratio of property valuation per student, and are likely to be elementary-only districts. Such districts are considered economically powerful in supporting their existing school program.

FACTOR SCORE DISTRIBUTIONS

Each computed factor score hud a mean of exactly 0 and a variance of exactly 1; that is, factor score distributions were standardized. For the purpose of illustrating the distributions, each is expressed here as a frequency polygon. The abcissa is identical for all the distributions and has been coded by mapping the range of standard scores into 26 intervals. In Table 11 is presented the coding scheme and the real limits of the intervals; the table also includes the exact interval frequencies for each factor, and the bottom line gives the distribution medians which were used in the stratification. The frequency polygons for the five factor scores are given in Figures 1, 2, 4, 5, and 6. Figure 3 is the distribution of district scores on Factor Two, Organizational Complexity, distinguishing between those districts which include secondary schools as well as elementary schools and those which include just elementary schools. Three peculiarities of the distributions are discussed in the following paragraphs.

Leptokurtosis of Factor One. The distribution of district scores on Factor One, Numerical Size, is the most non-normal of the five distributions. More than 78% of the entire district population resides in the interval coded number 9. If the distribution were normal, 9.9% of the population would be contained in this interval. But the leptokurtosis of the distribution does not imply that there are no reliable differences in size among three-quarters of Wisconsin's elementary school districts. The pile-up at the median is an artifact of the inclusion in the population of two or three large school systems. The largest district in the state is a metropolitan system which had a Numerical Size score of 23.03. The probability of so large a score occuring in a normal distribution is one in several billion. This one district accounts for 84% of the variance of the factor.

TABLE 11

INTERVAL CODES AND EXACT FREQUENCIES OF FACTOR SCORE DISTRIBUTIONS

Interval Number	Lower Limit	Upper Limit	FREQUENCIES				
			I	II	III	IV	V
1	-2.25	-2.01	0	2	4	2	0
2	-2.00	-1.76	0	4	9	1	2
3	-1.75	-1.51	0	27	20	5	2
4	-1.50	-1.26	0	51	30	12	9
5	-1.25	-1.01	0	36	37	36	37
6	-1.00	-0.76	0	33	47	64	65
7	-0.75	-0.51	0	39	47	73	93
8	-0.50	-0.26	37	59	65	85	79
9	-0.25	-0.01	494	61	66	107	84
10	0.00	0.24	69	84	65	82	74
11	0.25	0.49	9	69	71	38	58
12	0.50	0.74	4	46	59	26	47
13	0.75	0.99	1	27	35	28	20
14	1.00	1.24	3	26	20	19	15
15	1.25	1.49	8	22	19	12	10
16	1.50	1.74	1	11	8	12	8
17	1.75	1.99	1	13	5	5	6
18	2.00	2.24	1	4	7	3	4
19	2.25	2.49	0	5	8	4	3
20	2.50	2.74	1	6	1	4	1
21	2.75	2.99	0	3	2	2	3
22	3.00	3.24	0	2	3	2	0
23	3.25	3.49	0	1	1	1	3
24	3.50	3.74	0	1	0	0	2
25	3.75	3.99	0	0	0	2	1
26	4.00	23.03	3	0	3	7	6
Medians of Factor Score Distributions			-0.103	0.008	-0.024	-0.156	-0.156

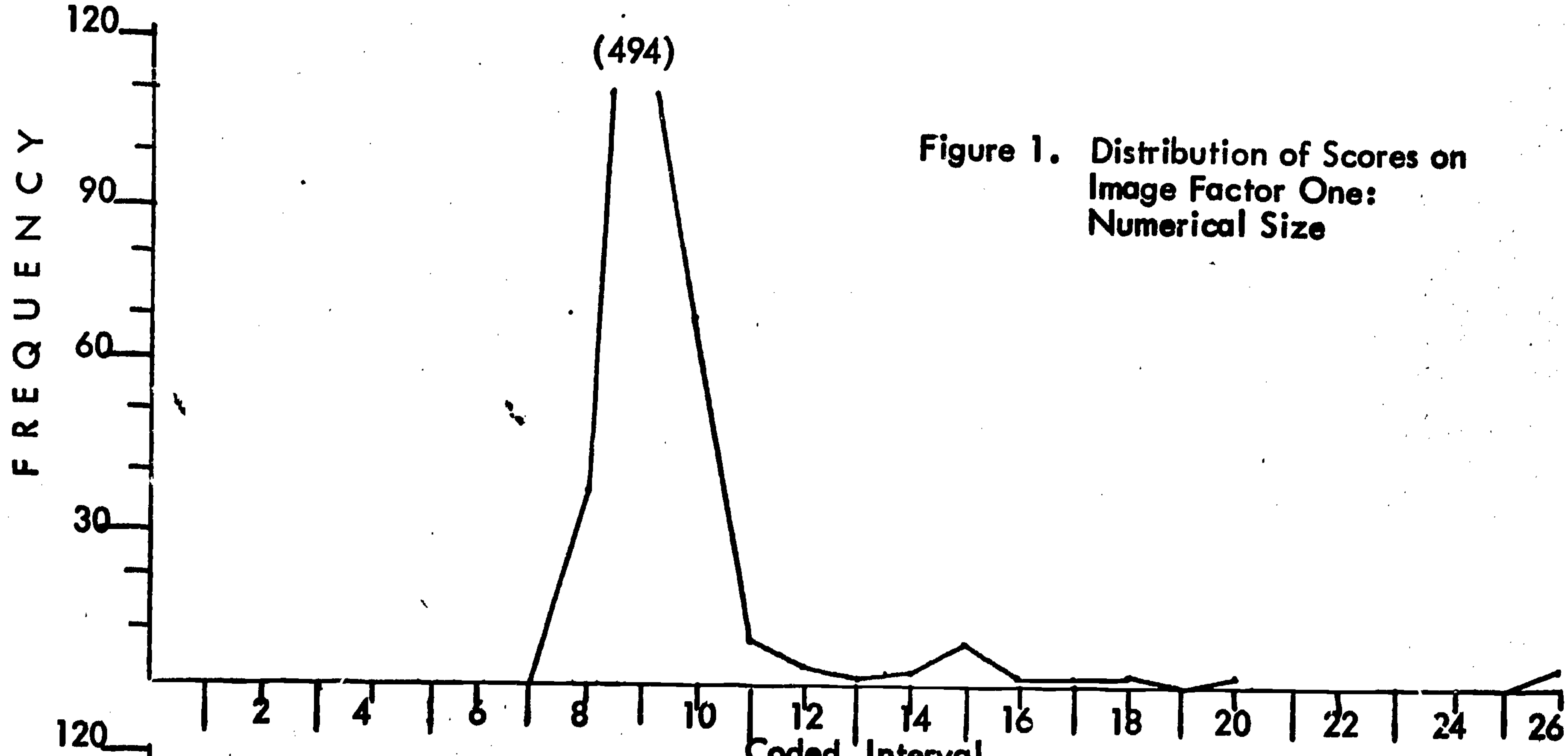


Figure 1. Distribution of Scores on Image Factor One: Numerical Size

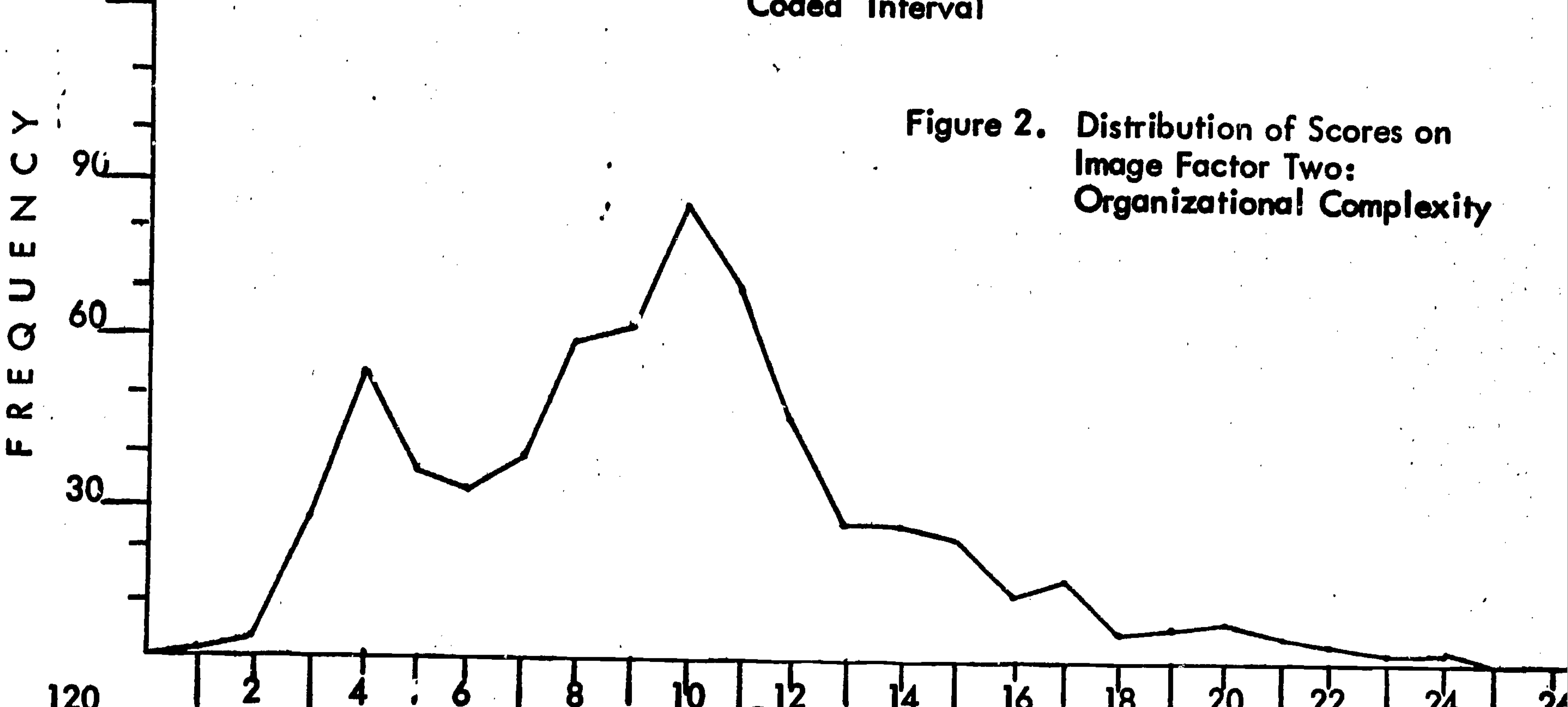


Figure 2. Distribution of Scores on Image Factor Two: Organizational Complexity

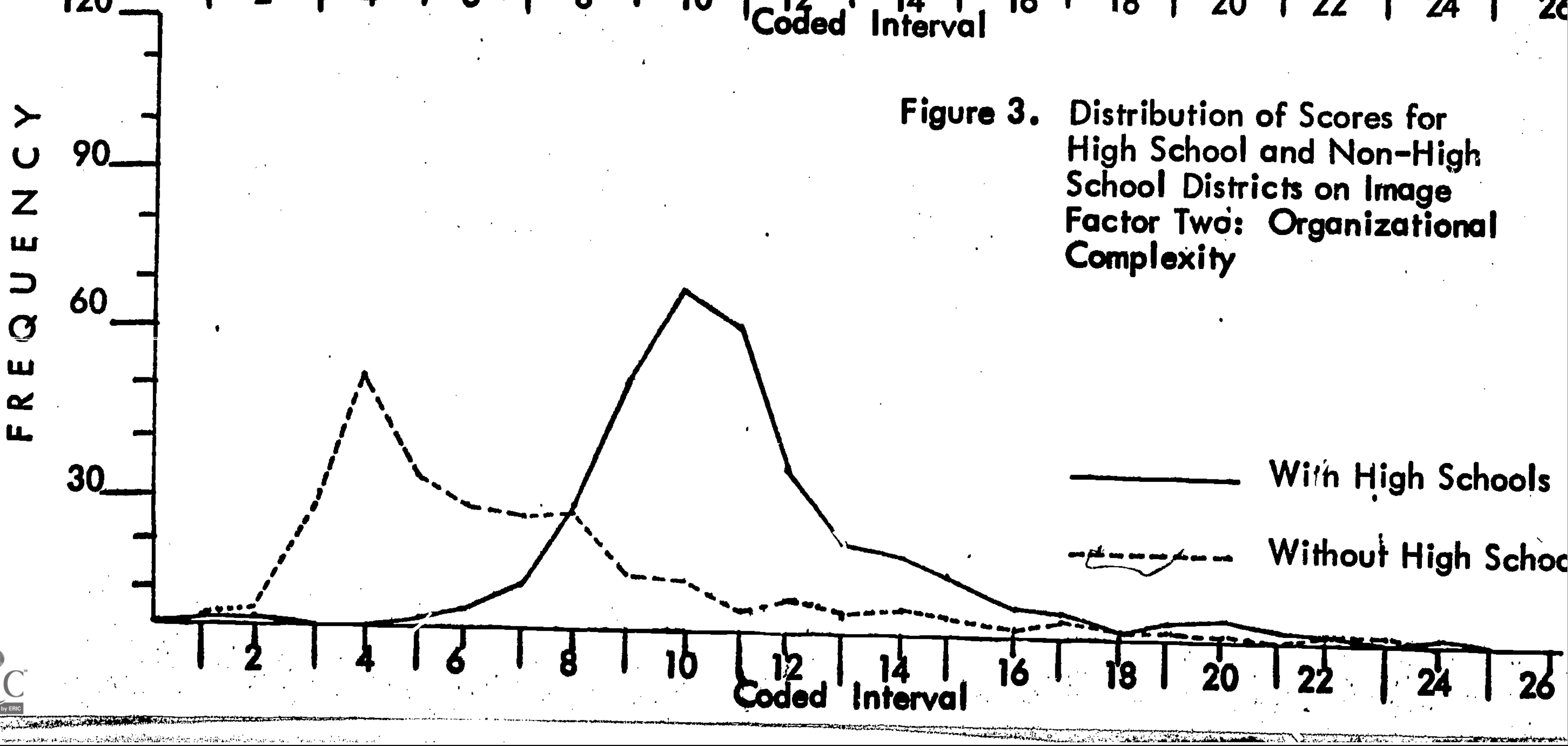


Figure 3. Distribution of Scores for High School and Non-High School Districts on Image Factor Two: Organizational Complexity

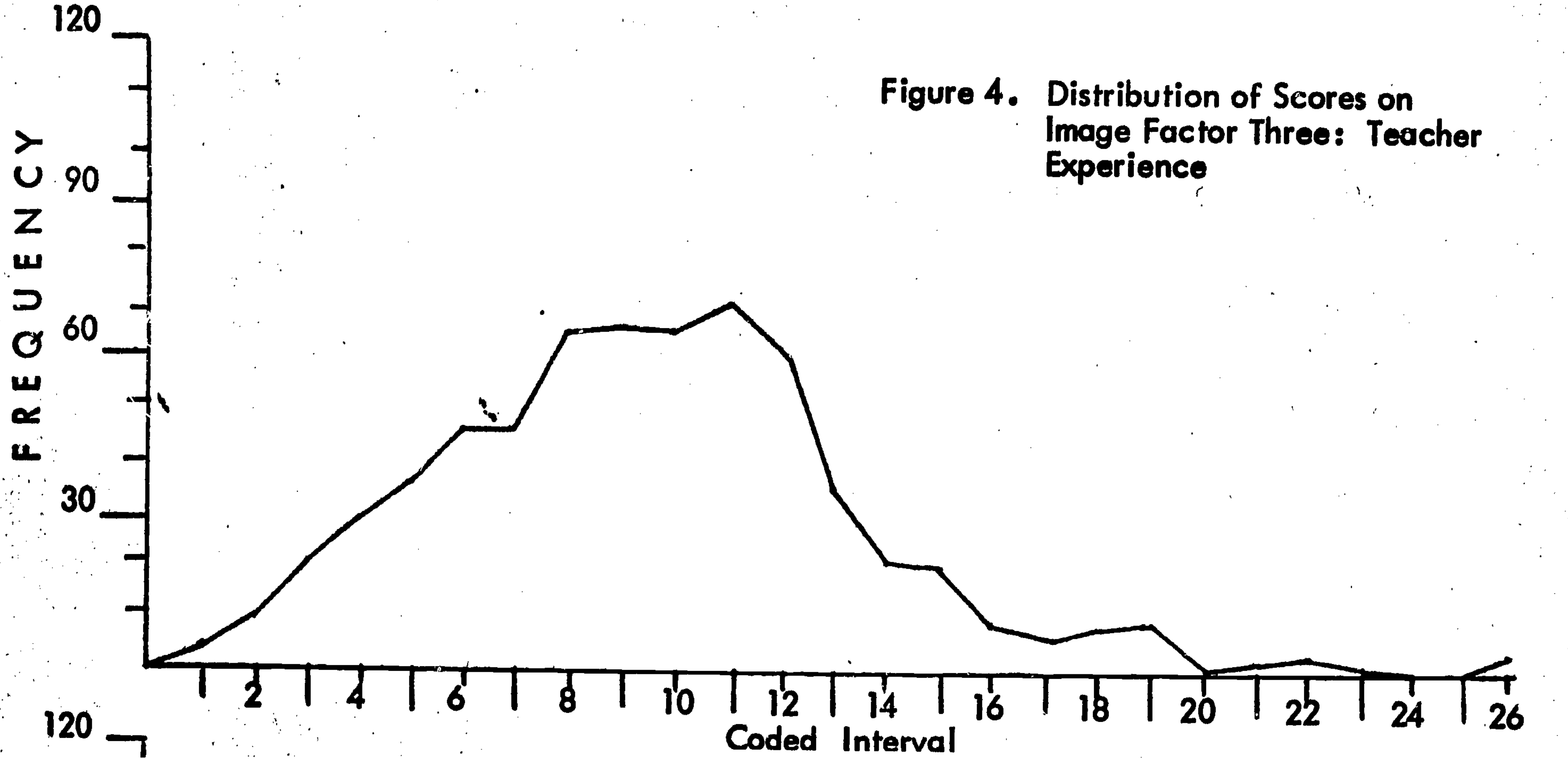


Figure 4. Distribution of Scores on Image Factor Three: Teacher Experience

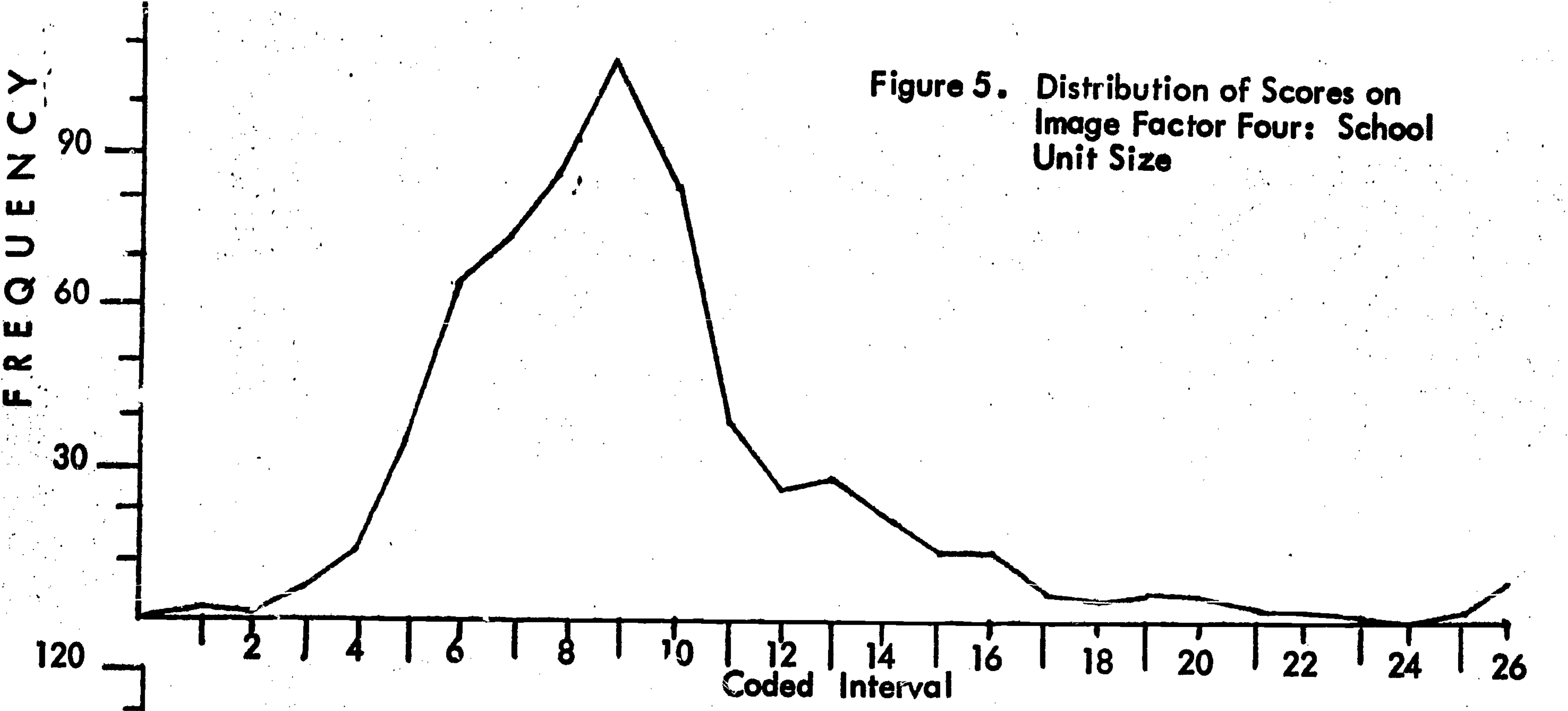


Figure 5. Distribution of Scores on Image Factor Four: School Unit Size

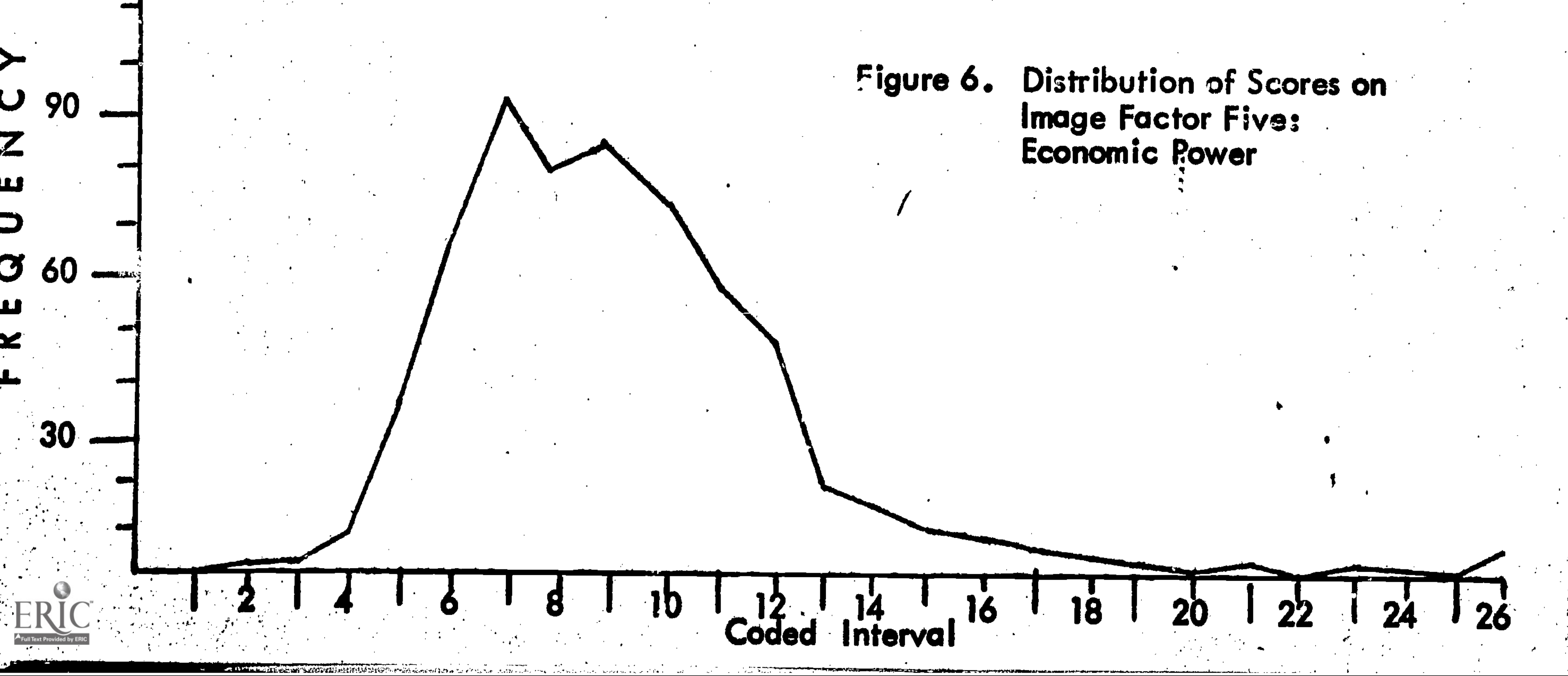


Figure 6. Distribution of Scores on Image Factor Five: Economic Power

It would have been possible to have excluded the large metropolitan districts from the input population or to have made normalizing transformations on their observations. Then these districts would not have been such radical outliers. But given the results as computed with the unaltered data, a researcher still has the choice of excluding the large districts. He may feel that the metropolitan school systems comprise a separate population and should be treated separately from all other districts; or he may believe that they are extreme cases, but are, nevertheless, members of the specified population. For example, in the stratification illustrated in Section III.B, the outliers were not omitted, and consequently the Numerical Size scores in strata associated with "High" Numerical Size may vary from -.119 to 23.03. But the discrimination between "High" and "Low" is adequate for the research purposes of that stratification.

Bimodality of Factor Two. It can be seen in Figu 2 that the distribution of Factor Two, Organizational Complexity, has two frequency peaks--it is bimodal. In Figure 3, separate frequency polygons of Factor Two are displayed for the subpopulation of districts with high schools and the subpopulation of districts without high schools. These distributions have similar shape, but their peaks are located at different points. By comparing Figure 2 with Figure 3, it is apparent that the bimodality of Factor Two is a result of the mixture of the two differently centered subpopulations. This suggests, as is reasonable, that Organizational Complexity is manifested differently for districts with high schools than for districts without high schools. Stratifying the total population of districts at the median of Organizational Complexity produces more a distinction between those districts with high schools and those without than it does a distinction, within the subpopulations, between more or less complex organizations.

Skewness. All the factors are somewhat skewed to the right. This is doubtless a consequence of skewness in the input data, for many of the input variables were enumerative. Demographic enumeration tallies tend to be skewed. All the distributions should be examined carefully before cutting points for a stratification are chosen.

PART III

SECTION B

ILLUSTRATION: CATEGORIZATION AND COMPARISON

The algorithm produced measures on the school districts which, as explained in the previous section, characterize the districts and which are considered important dimensions of school district variation. The measures were the input for the stratification scheme presented in Section II.D. The stratification scheme produced a categorization of the district population which is useful in certain practical research processes such as stratified sampling--these applications are discussed in Section III.D. But also, the categorization provides a framework for substantive comparison of school districts and school district types. The discussion below is broken into two parts. First, further explanation and interpretation is made of the categorization. Second, the summary characteristics of the categories or strata are presented, and notes are made concerning the kinds of substantive inference possible.

CATEGORIZING THE DISTRICTS

The dimensions or measures used for categorization were the first five rotated image factors which are discussed in Sections II.C and III.A. For each dimension the scores were points on a continuum, and no two districts had exactly the same score. In particular, no two districts had the same profile of scores. So while the multivariate analysis resulted in a clustering of the original 31 variables, it did not provide a direct clustering of the 632 districts. To provide such a clustering of districts, a transformation was performed on the factor measures: each factor was coded plus for districts with scores above the median score and minus for districts with scores below the median score. Each district was then identified by a pattern or profile of five pluses or minuses. After this reduction in the amount of information in the factor score specifications, there were districts with the same factor pattern.

In fact, there are only $2^5 = 32$ possible patterns, and so the population was partitioned into 32 categories. A stratum or category of districts is comprised of all the districts in the population which have a unique pattern of five pluses and minuses. Note that the categories correspond to the cells in a full-factorial experimental design with five dichotomous factors.

An objective in categorizing the districts was to put together those districts which were alike and to separate those which were different. However, the discussion in Section III.A on the distributions of the factors demonstrates that there remains considerable variability among, say, the 316 above-median districts for any one factor. By the time five such dichotomizations are superimposed on one another, the within-stratum variability is reduced, but it is still to be acknowledged. The question of whether the within-stratum variability is greater than the between-stratum variability is an empirical one and could be investigated by analysis of variançe procedures. But interpretation of such an analysis would be difficult since the factors are no longer orthogonal after being dichotomized. This is demonstrated in an example in Section III.C.

Within the algorithm it is possible to produce other stratifications. For example, a subset of the factors could be selected, or trichotomization could be used instead of dichotomization. Note that the stratification produces strata of approximately equal size because the factors are uncorrelated.

COMPARING THE CATEGORIES

The clustering and separating of districts according to the 2^5 design provide a framework within which substantive comparisons can be made. Comparison at the level of individual districts is too detailed for general substantive purposes, but efficient comparison among strata can be based on summary characteristics of the strata and the population. In Appendix L such summary characteristics are presented, and in Table 12 two pieces extracted from the Appendix are displayed. The extract on the right of Table 12 corresponds

TABLE 12

Stratum Characteristics: Extracts from Appendix L.

			STATEWIDE CHARACTERISTICS
+	Factor One	- Numerical Size	
+	Factor Two	- Organizational Complexity	
-	Factor Three	- Teacher Experience	
+	Factor Four	- School Unit Size	
+	Factor Five	- Economic Power	
17	Number of Districts		632
35.3	Percent with county-based administration		79.7
94.1	Percent which have high schools		55.5
94.1	Percent which receive integrated aid		61.3
2,331.1	Total Enrollment Per District		1,285.6
1,370.8	Elementary enrollment per district		798.0
960.3	Secondary enrollment per district		487.6
107.9	Total Staff Per District		58.4
51.0	Elementary teachers per district		28.8
46.3	Secondary teachers per district		23.9
10.6	Other professionals per district		5.7
12.3	Number of Schools Per District		4.3
5.9	Schools per district with only one or two rooms		1.1
6.4	Schools per district with three or more rooms		3.0
59,670.6	Equalized Valuation Per District *		33,727.8
189.6	Students per school in the stratum		295.9
21.6	Students per staff in the stratum		22.0
8.8	Staff per school in the stratum		13.4
25.6	Valuation per student in the stratum *		26.2
4,853.6	Valuation per school in the stratum *		7,762.6

* (Dollars x 1000)

from fifth column of first page

from fold-out tab on fourth page

to the fold-out tab on the fourth page of Appendix L. This contains the five factor identifications and the names of a series of twenty characteristics. To the right of these names appear the values of these characteristics for the entire state population. The thirty-two columns in the body of Appendix L contain the values of the characteristics as summarized for each of the thirty-two strata. One of these columns, for stratum [++ - ++], appears in Table 12 and is explained below. But the general form for a column is this: at the top appears the pattern of pluses and minuses for the stratum; next is the number of districts in the stratum and the proportions of districts with certain administrative features; then are given stratum averages related to enrollment, staff, school buildings, and valuation.

In stratum [++ - ++] there are 17 districts and all have factor scores above the median for Numerical Size, Organizational Complexity, School Unit Size, and Economic Power; and all have factor scores below the median for Teacher Experience. Thirty-five percent of the districts have county-based administration; ninety-four percent of them have high schools; ninety-four percent receive integrated aid. And, for example, there is an average of 51 elementary teachers per district in the stratum; across the stratum, there are 21.6 students per staff member; across the stratum, there is $25,600 valuation per student.

Of particular interest in stratum [++ - ++] is the large number of one and two-room schools. In fact, there are 101 such schools in the stratum. Yet the districts in the stratum are predominately city-based, have high schools, and have an about average amount of valuation per student. This suggests, as was verified by examining the districts in the stratum, that the districts are located in rural regions of the state but are centered in the business service centers of these regions. The one and two-room schools apparently are presently maintained in the rural outlying areas of the district.

There are 715 one and two-room schools in Wisconsin, and this stratum--which is to say, this kind of district--accounts for almost 1/7th of them. Appendix L makes possible the substantive comparison of this stratum with other strata having large numbers of one and two-room schools. Stratum[+ - ++ -]contains 59 one and two-room schools and its districts are mostly county-based, have low valuation per student, and centered in small villages. Stratum[-- ++ -] includes 48 one and two-room schools, and its districts are similar to those of stratum [++ - ++ -] , but they are smaller--that is, less of the rural area surrounding the villages is included. Thus by examining and comparing the strata and by making use of the multivariate descriptions, it can be seen that there are several kinds of districts which continue to have one and two-room country schools.

PART III

SECTION C
ILLUSTRATION: LOGICAL AND STATISTICAL ANALYSIS

The algorithm produced measures on the districts which may be analyzed as sources of variation explaining the distributions of other variables across the districts. In particular, the factor scores[1] and functions of them may be used as independent variables in multiple regression on a dependent variable. When the scores are dichotomized, then the multiple regression is equivalent to analysis of variance.

In this section, two illustrations of statistical analysis using the factor scores are given. In the first, the dependent variables concerned ESEA Title I fund allocation to the districts. The objective was to determine the differential prediction of the allocation variables from the district-characterizing measures. In the second, the dependent variable was viewpoint-productiveness of teachers in interviews leading to the construction of an item pool. The approach was to select teachers for interviewing according to an experimental design based on the dichotomized factor scores,and then to analyze their relative viewpoint-productiveness according to analysis of variance. Both illustrations are brief reports. In the first, the kinds of substantive inferences possible are emphasized; in the second, the major focus is on the kinds of experimental manipulations possible.

COMPENSATORY EDUCATION

Problem. School districts in Wisconsin, like districts all across the country, are eligible to receive federal funds under Title I of Public Law 89-10, the Elementary and Secondary Education Act. The districts are to use their funds to support locally

[1] That is, the first five rotated image factor scores.

initiated programs of compensatory education for educationally disadvantaged children. The amount of money for which a district is eligible is the product of two factors: a dollar allocation rate and the number of disadvantaged children in the district. The allocation rate is based on county census figures and reflects the proportion of county families with a basic annual income less than $2,000.

The objective of the analysis was to determine the extent to which the dollar rate, the number of disadvantaged children, and the total Title I allocation are predictable from the five district-characterizing factors developed in this project.

Procedures. Values on the three dependent, Title I, variables were available one year later than the data inputs for the algorithm, and several consolidations of smaller districts had taken place. Also, there were a few districts which were not eligible for Title I funds because they had no disadvantaged children. Due to the disappearance and ineligibility of certain districts, the population base for this analysis was a subset of 527 of the 632 districts which were input to the algorithm.

Two analyses were performed: in the first, the actual factor scores were used; in the second, the dichotomized scores determined for the stratification were used. The five district-characterizing factors, actual or dichotomized, were listed as independent variables and the three Title I indices were listed as dependent variables for input to a multiple regression analysis computer program. The program first computed the two-factor and three-factor interactions of the independent variables--that is, their two-way and three-way products. There were then a total of 25 possible independent variables: the five factors, the ten two-way interactions, and the ten three-way interactions. The analysis for the dichotomized scores is similar to a non-orthogonal analysis of variance.

Results. The intercorrelations of the five factor scores and the three Title I variables are given in Table 13. The independent and dependent variables are separated by vertical and horizontal dashed lines. The intercorrelations from the analysis based on the actual factor scores are given in the upper right triangle, and the intercorrelations from the analysis based on the dichotomized scores are given in the lower left triangle. Note that the actual factor scores are only slightly correlated; had the population size not been reduced, they would have been perfectly uncorrelated.

In Table 14 are displayed the statistics describing the degree to which the dependent, Title I, variables are predictable from various sets of the independent, district-characterizing, variables. The upper half and lower half of the table present, respectively, the results of the analysis using the actual and dichotomized factor scores. The left, middle, and right thirds of the table present, respectively, the results for the number of disadvantaged children, the allocation rate, and the total dollar allocation. Each of the 2 x 3 = 6 sections of the table is concerned, then, with either actual or dichotomized factor scores and with one of the three dependent variables.

In a section there are three columns of coefficients which describe the predictability of the dependent variable from various subsets of the actual or dichotomized independent factors. These subsets are: each of the five factors separately, the five factors together, the five factors along with the two-way interactions, and the five factors along with the two-way and three-way interactions. The entries in the column headed R are multiple correlation coefficients of prediction. The entries in the column headed R^2 are coefficients of determination--that is, proportions of the variance in the dependent variable accounted for by the subsets of independent variables. The entries in the third column are labelled ΔR^2 and are the absolute differences between the R^2 for the particular subsets of independent variables and the R^2 for the subset consisting of the five factors together. These ΔR^2 coefficients facilitate comparing the relative predictiveness of the various subsets of independent variables.

TABLE 13

Intercorrelations[1] of
Stratifying Dimensions
and Title I Allocation Variables

	1	2	3	4	5	6	7	8
1. Numerical Size	---	-010	-003	-009	006	964	137	971
2. Organizational Complexity	062	---	-057	-054	019	007	352	-013
3. Teacher Experience	-185	073	---	-015	-018	068	009	054
4. School Unit Size	332	030	017	---	009	070	-024	038
5. Economic Power	-068	-095	-053	-131	---	-055	-048	-038
6. Eligible Children	163	070	096	022	010	---	128	996
7. Dollar Rate	021	200	-013	-070	-071	128	---	129
8. Allocation	137	043	080	-002	024	996	129	---

[1] The intercorrelations of the continuous variables are given in the upper right segment of the matrix; intercorrelations of dichotomized variables are given in the lower left segment. Decimal points have been omitted.

TABLE 14

Relationships Between Stratifying Dimensions and Title I Variables

Independent Variables	Children R	Children R^2	Children ΔR^2	Rate R	Rate R^2	Rate ΔR^2	Allocation R	Allocation R^2	Allocation ΔR^2	
1. Numerical Size	.964	.930	.015	.137	.019	.128	.971	.943	.007	CONTINUOUS DISTRIBUTIONS
2. Organizational Complexity	.007	.000	.945	.352	.124	.023	-.013	.000	.950	
3. Teacher Experience	.068	.005	.940	-.009	.000	.147	.054	.003	.947	
4. School Unit Size	.070	.005	.940	-.024	.001	.146	.038	.001	.949	
5. Economic Power	-.055	.003	.942	-.048	.002	.145	.038	.001	.949	
All Five Factors	.972	.945		.384	.147		.975	.950		
All Five Factors and All Two-Factor Interactions	.986	.972	.027	.402	.162	.015	.993	.986	.036	
All Five Factors, and All Two-Factor and All Three-Factor Interactions	.987	.974	.029	.419	.176	.029	.994	.987	.037	
1. Numerical Size	.163	.027	.021	.021	.001	.051	.137	.019	.017	DICHOTOMIZED DISTRIBUTIONS
2. Organizational Complexity	.071	.005	.043	.200	.040	.012	.043	.002	.034	
3. Teacher Experience	.096	.009	.039	.013	.000	.052	.080	.006	.030	
4. School Unit Size	.022	.001	.047	.070	.005	.047	.002	.000	.036	
5. Economic Power	.010	.000	.048	.071	.005	.047	.023	.001	.035	
All Five Factors	.220	.048		.227	.052		.189	.036		
All Five Factors and All Two-Factor Interactions	.318	.101	.053	.313	.098	.046	.296	.087	.051	
All Five Factors and All Two-Factor and All Three-Factor Interactions	.391	.153	.105	.361	.130	.058	.380	.144	.108	

Table 15 gives the standardized regression coefficients from the analyses which involved multiple regressions. It is concerned only with the first two of the dependent variables--the total allocation is not treated--and only with the analysis using the actual factor scores. For each dependent variable, three columns are given. In the first of these are the regression coefficients for the five stratifying dimensions, when only those dimensions were used. In the second column are given the coefficients for the five dimensions and the two-way interactions, for the corresponding linear model; and the third column gives the coefficients for the regression model which used all 20 interactions along with the five basic dimensions. Significant coefficients ($p \leq .05$) are indicated. Also given are the multiple correlation coefficient (R) and coefficient of determination (R^2) associated with each of the regressions.

Discussion. In the following five paragraphs, comments are made concerning features of the results. These comments are not intended exhaustively to explain Title I allocation, but rather to suggest the kinds of inferences that may be made with logical and statistical applications of the algorithm.

1. It appears that the two dependent variables used in the allocation of Title I funds are related differentially to two stratifying dimensions. The number of disadvantaged children is best predicted by Factor One, Numerical Size; the dollar rate is best predicted by Factor Two, Organizational Complexity.

2. The number of disadvantaged children is more predictable than the dollar rate. This is probably because the variation in the number of children is larger than the variation in the allocation rate. Allocation rate varies by county rather than by district, and all the districts in a county were assigned the same rate.

3. The results for total allocation are very similar to those for number of disadvantaged children, just as the correlation between the total allocation and the number of disadvantaged children is high. The high correlation is probably due to the difference in variability between the two factors which are multiplied together to determine the total allocation.

TABLE 15

Standardized Regression Coefficients for the Relationship Between Stratification Dimensions and Title I Allocation Variables

	Number of Qualified Children			Rate of Allocation		
Variable	Factors Only	With 2-way	With 3-way	Factors Only	With 2-way	With 3-way
1. Numerical Size	.97*	.80*	.85*	.14*	.04	-.51
2. Organizational Complexity	.03	.06*	.07*	.36*	.31*	.32*
3. Teacher Experience	.07*	.10*	.11*	.03	.02	.08
4. School Unit Size	.08*	.08*	.09*	.00	.00	.00
5. Economic Power	-.06*	-.06*	-.06*	-.06	-.07	-.10
1 x 2		-.12*	-.34*		.06	1.06
1 x 3		.14*	.21		.09	.45
1 x 4		.05	.12		-.08	-.27
1 x 5		-.05*	.05		-.02	-.11
2 x 3		.03*	.03		-.07	-.10
2 x 4		.01	.02		-.02	.04
2 x 5		-.03	-.01		.11	.14
3 x 4		.00	.02		.01	.00
3 x 5		-.01	-.03		.01	.07
4 x 5		-.01	.01		.07	.12
1 x 2 x 3			.01			-.64
1 x 2 x 4			-.27*			.44
1 x 2 x 5			.01			-.05
1 x 3 x 4			.08*			-.09
1 x 3 x 5			-.07*			.07
1 x 4 x 5			.10			-.05
2 x 3 x 4			.00			.06
2 x 3 x 5			-.01			.04
2 x 4 x 5			-.02			-.20
3 x 4 x 5			-.01			-.07
R	.97	.98	.99	.38	.40	.42
R^2	.94	.97	.97	.15	.16	.18

* The associated t - value is significant beyond $p \leq .05$

4. When actual factor scores are used, the inclusion of interactions adds little to the prediction. When dichotomized scores are used, the interactions have considerable utility.

5. There is a considerable loss of precision when dichotomized scores are used. Also, the dimensions are not orthogonal after dichotomization. The entries above the diagonal in the upper-left quadrant of Table 13 are not exactly zero because several districts which supplied inputs in the construction of factors were not included in this analysis. These entries, are, however, very close to zero. Certain intercorrelations of the dichotomized scores depart significantly from zero.

TEACHER VIEWPOINTS

Problem. Attitude and viewpoint item pools are often generated by collecting and collating information from the responses of several informants. In order to understand the nature and composition of such pools it is necessary to ascertain the differential contributions of the sources providing information. This is exemplified in a study (Miller, et. al., 1967) of which a major objective was to secure viewpoints of teachers concerning ways of facilitating classroom learning. The viewpoints were obtained by means of tape-recorded, open-ended interviews, and content analysis of the recordings led to the formation of an item pool.

One dependent measure was viewpoint-productiveness, as measured by the number of statements which resulted from an interview. The analytic objective was to determine how the number of discrete statements obtained from a teacher's interview protocol varied as a function of district factors, teacher factors, and interview factors. Altogether, seven factors were to be studied, and each of these factors had two levels. The first four factors were the first four district-characterizing dimensions; factors five and six were two teacher variables, grade level taught and length of teaching experience; factor seven was the order in which the two sections of the interview schedule were administered.

Procedures. In order to determine the effects of the seven factors in the differential contributions of the teachers to the item pool, the selection of teachers to be interviewed and the administration of the schedule was directed according to an experimental design. The seven factors provided a basis for a 2^7 design; there were seven factors each at two levels, defining 128 cells or treatment combinations. If there were to be replicates for the purpose of estimating error, at least two teachers would have had to be interviewed within each of these cells. This would have necessitated conducting at least 256 interviews, which was impractical given the limitations of time and cost.

So in order to be able to estimate the effects of all seven factors without having to conduct 256 depth interviews, it was decided to employ a 2^{7-3} fractional factorial design[1], which allowed reduction in the number of treatment combinations needed. Two teachers were interviewed in each of the 16 treatment combinations defined by a 2^{7-3} design, and the total number of interviews called for was therefore reduced to 32. As is shown below, however, interpretations of the results of the consequent analysis are more tentative than they would have been in the full, 2^7, case.

For the purpose of developing the fractional factorial design matrix, the levels of each factor were coded as + for "high" and - for "low". The seven factors for the study then were:

FACTOR		CODED +	CODED -
District Factors	1 Numerical Size	Above Median	Below Median
	2 Organizational Complexity	"	"
	3 Teacher Experience	"	"
	4 School Unit Size	"	"
Teacher Factors	5 Grade Level Taught	Intermediate	Primary
	6 Total Teaching Experience	10 years or more	less than 10 years
Interview Factor	7 Order of Interview Schedule	Order A-B	Order B-A

[1] Fractional factorial designs are described in detail in Box and Hunter, 1961. Their utility in educational research is discussed in McLean, 1966.

In the study of the relationships between the seven factors and a dependent variable, there are 128 sources of variation: the grand mean, 7 main effects, 21 two-factor interactions, 35 three-factor interactions, 35 four-factor interactions, 21 five-factor interactions, 7 six-factor interactions, and 1 seven-factor interaction. If there were to be only 16 treatment combinations, each would have to estimate the combined effect of eight of these 128 sources of variation. By using a fractional factorial design, the pattern of the confounding of the sources of variation could be specified, within certain limits, in advance.

This illustration is of the methodology for performing the selection and analyses. Therefore, the results presented in the following paragraphs are the parameters for the particular design constructed. The technical details of the construction are not given.

Results. The present design was structured so that the 1 x 4 interaction would not be confounded with any main effect. This was desirable because, as is noted in Section III.B, the dichotomized factor scores for factors One and Four were substantially correlated. Furthermore, no main effect was confounded with any two-factor interaction which was considered possibly potent. The following are the confounding relationships for the main effects:

$$
\begin{aligned}
I &= 12345 = 1236 = 2347 = 456 = 157 = 1467 = 23567\\
1 &= 2345 = 236 = 12347 = 1456 = 57 = 467 = 123567\\
2 &= 1345 = 136 = 347 = 2456 = 1257 = 12467 = 3567\\
3 &= 1245 = 126 = 247 = 3456 = 1357 = 13467 = 2567\\
4 &= 1235 = 12346 = 237 = 56 = 1457 = 167 = 234567\\
5 &= 1234 = 12356 = 23457 = 46 = 17 = 14567 = 2367\\
6 &= 123456 = 123 = 23467 = 45 = 1567 = 147 = 2357\\
7 &= 123457 = 12367 = 234 = 4567 = 15 = 146 = 2356
\end{aligned}
$$

The first line in the list contains the principal generators of the design. And, for example, the second line indicates that the first main effect was confounded with the 2 x 3 x 4 x 5 interaction, the 2 x 3 x 6 interaction, etc.

The actual design matrix,according to which the teachers were selected to be interviewed,is given as Table 16. Each row of the table shows a treatment combination. For row 6, for example, it was necessary to select--at random--two districts which were above the median on the dimensions Numerical Size and Teacher Experience and below the median on the dimensions Organizational Complexity and School Unit Size. In each of these districts, it was necessary to identify all the teachers who had less than 10 years teaching experience and who taught in the intermediate grades. One of these teachers was selected at random from each of the two districts, and both persons were interviewed in the same schedule order, segment A followed by segment B.

After the interviews had been conducted and the content analyses completed, the statistical analysis could be computed. The structure of the ANOVA summary table is as follows:

SOURCE	SUMS OF SQUARES	DF	MEAN SQUARES	F
A. Numerical Size	SS_A	1	$SS_A/1$	MS_A/MS_I
B. Organizational Complexity	SS_B	1	$SS_B/1$	MS_B/MS_I
C. Teacher Experience	SS_C	1	$SS_C/1$	MS_C/MS_I
D. School Unit Size	SS_D	1	$SS_D/1$	MS_D/MS_I
E. Grade Level	SS_E	1	$SS_E/1$	MS_E/MS_I
F. Teaching Experience	SS_F	1	$SS_F/1$	MS_F/MS_I
G. Schedule Order	SS_G	1	$SS_G/1$	MS_G/MS_I
H. All Other Controlled Sources	SS_H	8	$SS_H/8$	MS_H/MS_I
I. Error	SS_I	16	$SS_I/16$	

The analysis could be modified to test specifically the effect of the 1 x 4 interaction,along with the effects confounded with it.

TABLE 16

Design Matrix for 2^{7-3} Fractional Factorial
Used in the Study of Teacher Viewpoint-Productiveness

Run	1 Numerical Size	2 Organizational Complexity	3 Teacher Experience	4 School Unit Size	5 (1234) Grade Level	6 (123) Personal Experience	7 (234) Schedule Order
1.	+	+	+	+	+	+	+
2.	+	+	+	-	-	+	-
3.	+	+	-	+	-	-	-
4.	+	+	-	-	+	-	+
5.	+	-	+	+	-	-	-
6.	+	-	+	-	+	-	+
7.	+	-	-	+	+	+	+
8.	+	-	-	-	-	+	-
9.	-	+	+	+	-	-	+
10.	-	+	+	-	+	-	-
11.	-	+	-	+	+	+	-
12.	-	+	-	-	-	+	+
13.	-	-	+	+	+	+	-
14.	-	-	+	-	-	+	+
15.	-	-	-	+	-	-	+
16.	-	-	-	-	+	-	-

Discussion. The scheme described and parameterized above provides an efficient approach to logically analyzing the effect of the factors on the differential contributions of the teachers to the viewpoint item pool. The actual numerical results are not presented here, for the purpose of this illustration has been to demonstrate the utility of the outputs of the algorithm in designing statistical analyses. The scheme also insures variability of content, since according to the design, teachers are selected from different kinds of districts and with different teaching situations and experiences. This latter aspect of the design is discussed further in the next section.

PART III

SECTION D
ILLUSTRATION: SAMPLING

The multivariate stratification defined within the algorithm provides the framework for many stratified sampling schemes. A distinction may be made among three basic types of sampling objectives: sampling for analytic studies, sampling for enumeration studies, and sampling for maximum variance.

Analytic studies are concerned with determining relationships among subgroups of a population and often involve analysis of variance. An example of an analytic sampling study was outlined in Section III.C. Enumeration sampling is performed to increase the precision with which the population can be described, and an example of such a study is given later in this section. Maximum variance samples are selected when concern is for maximizing the probability of permeating a content domain. For example, if the responses of selected subjects were transformed into items for a factor battery, and a certain type of person in the population was not represented in the selected subjects, then items special to that type of person would be missed. Maximum variance sampling is exemplified by the interview study of elementary teachers' viewpoints outlined in Section III.C. This is explained at the conclusion of this section.

ISSUES IN SAMPLING DESIGN

The basic concepts and techniques of sampling theory are essential in applying the algorithm for obtaining stratified random samples of school districts. Most references on the theory of stratified sampling deal with the properties of the estimates from a stratified sample and with the ways for choosing stratum sample sizes so as to obtain adequate precision. Procedures for constructing strata are usually not discussed in these references. Conversely, this report deals with a particular approach for constructing strata. The technicalities of sampling theory and formulae are not presented in this report. But, of course, careful attention to sampling theory is crucial in ensuring validity of a sampling application of the algorithm. In order to provide perspective, brief mention will be made here of four generally important issues of sampling theory raised in Cochran's text, Sampling Techniques (1953). Other rele ant issues may be found in Deming (1953), Showell (1957), and Moonan (1953).

Stratification. Any research study involving random sampling requires the identification and coding of all the members of the population. Such specification immediately makes feasible simple random sampling, and stratification may seem unnecessary. But stratified random sampling may yield increased precision either in the sense that a certain precision is maintained in every stratum of the population or in the sense that estimates of characteristics of the total population are more precise. Such increase in precision results when the strata are relatively homogeneous with respect to a criterion variable. In such a case, it is advantageous to treat each stratum as a population in its own right, and this is accomplished through stratified random sampling. But the homogeneity of the strata with respect to a particular criterion variable is rarely known in advance, and whether stratification is appropriate must usually be determined by substantive and theoretical hypothesis. Estimates of the gain in precision due to stratification may, however, be made after criterion data are collected.

A secondary reason for choosing stratified random sampling may be administrative and logistic convenience. For example, in educational studies, access to individual schools ordinarily must be obtained from district headquarters, so stratification of schools into districts is naturally imposed by this administrative circumstance. Logistic problems of sampling and of collecting data may differ in various subdivisions of the population.

For example, it may be practical to obtain measures on all schools in small districts, while it may be necessary to further sub-sample within large districts. Perhaps conversely, it may be possible to obtain measures on all schools in a geographically compact urban area while it may be costly to reach all schools in a rural area where schools are farther apart.

Sample Size. The main problem ii. determining sample size is to obtain maximum precision with minimum cost--in terms of available resources. The function relating precision with sample size can usually be estimated; generally it involves unknown population parameters. When estimated it may be substituted into the function relating sample size and cost, and then,given the desired level of precision, the minimum sufficient sample size may be computed. If cost and precision functions markedly differ for subdivisions of the population, then the sample size needs to be separately determined for each subdivision. If several characteristics are to be studied in one sample, there may be conflicting sample size requirements. These conflicts need to be resolved by considering the costs incurred by oversampling and the relative priorities among the characteristics.

Geography. If geographically adjacent units are more alike than units which are far apart, then it may seem reasonable to use geographical boundaries as strata definitions. Although sampling from such strata may be fairly efficient, the stratifying dimensions are probably not directly related to the specific objectives of a sampling study. A sufficiently complex stratification based on theoretically relevant dimensions will, however, result implicitly in geographical differentiation if, indeed, there are genuine differences between geographical regions. The stratified random samples drawn within the present algorithm have, for example, been scattered throughout the state of Wisconsin.

Sampling Unit. The units of analysis--that is, the units which ultimately are measured--in some studies are selected after a series of sampling stages involved different units of classification. In the study outlined in Section III.C, for example, teachers were sampled from a stratified random sample of districts. The district level was chosen for the present algorithm because schools, teachers, and students form a nested hierarchy under districts. The utility of sampling, for example, teachers from a stratified random sample of districts depends on whether the teachers nested in a stratum f districts are relatively homogeneous. When this stage sampling is used, it is possible to compute how much of the variance of a measure obtained on the teachers is due to individual differences among teachers and how much is due to differences in teachers across districts and strata of districts.

ENUMERATION SAMPLE: TEACHING VACANCIES

Problem. Late in the summer of 1966, the Wisconsin State Department of Public Instruction found that it was necessary to determine, or to estimate accurately, the number of teaching positions, at all grade levels, which had not been filled for the approaching academic year. It appeared to be impossible to canvas all the districts in the state in the time available, so it was decided to distribute a questionnaire to a carefully selected sample of districts, to focus the resources of the WSDPI on obtaining complete and reliable information from that sample, and to estimate the state's teaching vacancies from the sample data.

Procedures. The first two factors,[1] Numerical Size and Organizational Complexity, were selected as stratifying variables and were dichotomized at their medians, so there were four strata of districts: ++, +-, -+, and --.

[1] Actually these were the factors of an early, incomplete version of the algorithm.

Results. All but two of the questionnaires were returned in time for analysis, and the projections were computed and disseminated within ten days from the initiation of the survey.

Discussion. There was no way to validate directly the precision of the findings based on this sample, but it was possible to validate it in an indirect fashion. Although the teaching vacancies in districts were not available as a direct check on accuracy, the district enrollment figures were uniformly available. As a kind of check on the sampling accuracy, then, enrollment figures from the same sample of 50 districts was used to project the statewide K-12 public school enrollment in Wisconsin. The resulting projection was accurate within 1.5% of the actual statewide public school enrollment.

SAMPLING CONTENT

Problem. When attitude and viewpoint item pools are constructed from collated information gathered from the responses of a number of informants, it is of major concern that procedures for collection and collation be devised so as to ensure permeation of the desired content domain. One manipulable aspect of the collection process is the selection of respondents. This selection is critical if different informants make contributions to different parts of the content domain.

The content domain of interest in the study (Miller, et al., 1967) outlined in Section III.C was elementary teachers' viewpoints on classroom learning. An underlying assumption of that study was that individual teachers differ in the areas of the content domain to which they might contribute. This was considered plausible because teachers vary with respect to their life experiences, personality characteristics, and teaching backgrounds. Furthermore, it was assumed that teachers from different kinds of school districts would tend to contribute to different areas of the content domain. This was considered plausible because particular kinds of districts attract particular kinds of teachers, and the kinds of educational conditions in a district influence the professional experience of a teacher in a district.

Procedures. The selection of teachers for interviewing was based on the fractional factorial design of Section III.C. Within the design, teachers were stratified according to teaching experience and background, and the selection of districts from which teachers were sampled was based on the district-characterizing dimensions, which include teacher experience characterizing information.

Discussion. A typical approach to selection is to examine those who are most accessible--that is, those who teach in districts which cooperate with research projects. To demonstrate the inadequacy of this routine, Table 17 has been prepared to show the factor scores of ten school districts from which researchers frequently solicit cooperation. Clearly this set of schools is not a representative sample, especially with respect to the first two factors, Numerical Size and Organizational Complexity. If teachers from these often researched districts had been interviewed in the study of viewpoints on the facilitation of learning, and if differences in viewpoints were correlated with Numerical Size or Organizational Complexity, important regions of the content domain might have been missed.

Demonstration of the power of the staged stratified random teacher selection process in permeating the content domain was not accomplished by Miller et al., (1967) due to the great difficulty in quantifying the qualities of the content areas. The selection process was based on substantive assumptions concerning teacher experience and district characteristics. The selection process was intended to provide maximum variance in the sample leading to permeating the content domain; in that sense the sample was intended to be representative, for effort was made to ensure that as many different viewpoints as possible would be presented.

TABLE 17

Factor Scores and Summary Statistics
for a Sample of Often-Researched Districts

District	Pattern	Factor One	Factor Two	Factor Three	Factor Four	Factor Five
A	+++-+	0.92	1.32	0.32	0.63	0.47
B	+++-+	1.42	2.93	0.84	-0.20	2.97
C	++-++	0.09	1.87	-0.68	0.15	0.80
D	+++--	1.39	1.14	0.16	-1.36	-1.34
E	++--+	5.55	1.36	-0.76	-0.37	0.35
F	+-+-+	23.03	-2.04	0.57	-0.64	-0.11
G	++---	1.28	1.72	-0.26	-0.89	-0.56
H	++---	4.40	0.52	-0.47	-2.30	-1.78
I	++---	0.50	2.15	-1.38	-0.88	-0.84
J	-+---	-0.10	1.26	-0.87	-0.22	-0.85
	Highest Value	23.03	2.93	0.84	0.15	2.97
	Lowest Value	-0.10	-2.04	-1.38	-2.30	-1.78
	Mean Value	3.80	1.22	-0.25	-0.73	-0.09

APPENDICES

LIST OF APPENDICES

APPENDIX A

This appendix contains the results of a correlational analysis which was designed to indicate special characteristics of districts with missing data. The initiai 31 by 632 data matrix was temporarily redefined by changing the recordings of the first 12 variables to special dummy variables. The dummy variables have been called "non-missingness" variables, since they were computed to be "1" for entries which were not missing (in the original data) and "0" for those that were.

Based on the redefined data matrix, means, standard deviations, and correlations were computed. The means and standard deviations are presented in Appendix A. 1. The correlations are presented in Appendix A. 2.

Of special interest are the means of the first 12 dummy variables, since they are the proportions of non-missing data. Also the upper left 12 by 12 portion of the correlation matrix indicates the predominance of Type A missing data. And the upper right 12 by 19 portion indicates the special characteristics of the districts with no missing data.

This appendix is explained and interpreted in Section II.B.

Appendix A. 1. MEANS AND STANDARD DEVIATIONS

	MEANS
1	0.993671
2	0.995253
3	1.000000
4	1.000000
5	1.000000
6	1.000000
7	0.852848
8	0.856013
9	0.857595
10	0.857595
11	0.857595
12	0.857595
13	1.797468
14	1.444620
15	1.746835
16	487.558544
17	797.996835
18	28.827532
19	6.174051
20	17.740506
21	2.155063
22	3.465190
23	0.670886
24	0.460443
25	3.213608
26	33727812.816406
27	32552.498190
28	199.202534
29	20.849323
30	9.517363
31	5380376.963135

	STANDARD DEVIATIONS
1	0.079304
2	0.068734
3	0.000000
4	0.000000
5	0.000000
6	0.000000
7	0.354257
8	0.351077
9	0.349465
10	0.349465
11	0.349465
12	0.349465
13	0.401886
14	0.496924
15	0.954249
16	2060.797673
17	3224.220010
18	101.939330
19	36.860129
20	55.195788
21	3.214092
22	20.754031
23	1.660002
24	0.932035
25	7.207828
26	157630550.390625
27	28658.983108
28	158.497818
29	4.198261
30	7.386481
31	5501990.301147

Appendix A. 2. CORRELATIONS: R_n

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	28	-0	-0	-0	-0	19	19	14	14	14	14	-4	-9	-10	2	2	2	1	3	5	1	-0	2	4	2	1	9	-6	9	7
2	28	100	0	0	0	0	17	17	10	10	10	10	-3	-8	-9	2	2	2	1	2	5	1	0	1	3	1	-12	8	8	8	5
3	-0	0	0	0	0	0	0	-0	0	0	0	0	-0	0	-0	0	0	-0	0	-0	-0	-0	0	-0	0	0	0	-0	0	-0	0
4	-0	0	0	0	0	0	0	-0	0	0	0	0	-0	0	-0	0	0	-0	0	-0	-0	-0	0	-0	0	0	0	-0	0	-0	0
5	-0	0	0	0	0	0	0	-0	0	0	0	0	-0	0	-0	0	0	-0	0	-0	-0	-0	0	-0	0	0	0	-0	0	-0	0
6	-0	0	0	0	0	0	0	-0	0	0	0	0	-0	0	-0	0	0	-0	0	-0	-0	-0	0	-0	0	0	0	-0	0	-0	0
7	19	17	0	0	0	0	100	99	98	98	98	98	-21	-46	-55	10	10	11	7	13	27	7	-6	16	19	8	-30	47	10	47	33
8	19	17	-0	-0	-0	-0	99	100	99	99	99	99	-21	-46	-54	10	10	11	7	13	27	7	-6	16	18	8	-31	46	10	47	33
9	14	10	0	0	0	0	98	99	100	100	100	100	-21	-46	-54	10	10	11	7	13	26	7	-6	17	18	8	-31	46	10	47	32
10	14	10	0	0	0	0	98	99	100	100	100	100	-21	-46	-54	10	10	11	7	13	26	7	-6	17	18	8	-31	46	10	47	32
11	14	10	0	0	0	0	98	99	100	100	100	100	-21	-46	-54	10	10	11	7	13	26	7	-6	17	18	8	-31	46	10	47	32
12	14	10	0	0	0	0	98	99	100	100	100	100	-21	-46	-54	10	10	11	7	13	26	7	-6	17	18	8	-31	46	10	47	32
13	-4	-3	-0	-0	-0	-0	-21	-21	-21	-21	-21	-21	100	44	39	-31	-26	-28	-29	-33	-44	-22	-18	-16	-38	-25	15	-38	-11	-35	-27
14	-9	-8	0	0	0	0	-46	-46	-46	-46	-46	-46	44	100	77	-20	-17	-19	-14	-28	-34	-13	-19	-26	-31	-14	47	-44	-1	-44	-7
15	-10	-9	-0	-0	-0	-0	-55	-54	-54	-54	-54	-54	39	77	100	-18	-17	-20	-13	-24	-43	-13	-12	-18	-29	-14	43	-64	-2	-66	-38
16	2	2	0	0	0	0	10	10	10	10	10	10	-31	-20	-18	100	99	99	98	99	48	98	0	5	98	99	-6	38	8	34	32
17	2	2	0	0	0	0	10	10	10	10	10	10	-26	-17	-17	99	100	99	96	98	45	99	-1	4	97	99	-5	37	8	33	32
18	2	2	-0	-0	-0	-0	11	11	11	11	11	11	-28	-19	-20	99	99	100	96	98	48	97	-1	4	97	99	-6	41	8	37	36
19	1	1	0	0	0	0	7	7	7	7	7	7	-29	-14	-13	98	96	96	100	94	46	95	-2	1	95	96	-2	35	7	31	32
20	3	2	-0	-0	-0	-0	13	13	13	13	13	13	-33	-28	-24	99	98	98	94	100	49	97	2	8	97	98	-9	42	8	38	32
21	5	5	-0	-0	-0	-0	27	27	26	26	26	26	-44	-34	-43	48	45	48	46	49	100	38	-0	14	57	43	-15	48	1	48	38
22	1	1	-0	-0	-0	-0	7	7	7	7	7	7	-22	-13	-13	98	99	97	95	97	38	100	-1	1	94	99	-3	31	6	28	28
23	-0	0	0	0	0	0	-6	-6	-6	-6	-6	-6	-18	-19	-12	0	-1	-1	-2	2	-0	-1	100	26	1	-2	-8	-21	5	-22	-22
24	2	1	-0	-0	-0	-0	16	16	17	17	17	17	-16	-26	-18	5	4	4	1	8	14	1	26	100	9	1	-14	-11	6	-13	-17
25	4	3	0	0	0	0	19	18	18	18	18	18	-38	-31	-29	98	97	97	95	97	57	94	1	9	100	95	-12	43	9	39	33
26	2	1	0	0	0	0	3	8	8	8	8	8	-25	-14	-14	99	99	99	96	98	43	99	-2	1	95	100	-2	36	7	32	35
27	1	-12	0	0	0	0	-30	-31	-31	-31	-31	-31	15	47	43	-6	-5	-6	-2	-9	-15	-3	-8	-14	-12	-2	100	-24	-25	-23	15
28	9	8	-0	-0	-0	-0	47	46	46	46	46	46	-38	-44	-54	38	37	41	35	42	48	31	-21	-11	43	36	-24	100	14	98	79
29	-6	8	0	0	0	0	10	10	10	10	10	10	-11	-1	-2	8	8	8	7	8	1	6	5	6	9	7	-25	14	100	2	7
30	9	8	-0	-0	-0	-0	47	47	47	47	47	47	-35	-44	-66	34	33	37	31	38	48	28	-22	-13	39	32	-23	98	2	100	79
31	7	5	0	0	0	0	33	33	32	32	32	32	-27	-7	-38	32	32	36	32	32	38	28	-22	-17	33	35	15	79	7	79	100

APPENDIX B

This appendix contains the results of a correlational analysis of a special kind. The computations were based on the initial data matrix, and the results served as criteria for the adequacy of replacements for missing data. Because some of the data items in the initial data matrix were missing, standard correlational techniques could not be applied. Instead, for each coefficient (mean, standard deviation, or correlation), all non-missing data available for computing the coefficient were used. Thus different numbers of districts are involved in the calculations of the different coefficients. And the results use the maximum amount possible of the information in the initial data matrix.

Appendix B. 1 is a matrix of counts of districts. For each pair of variables, the entry equals the number of districts for which neither of the variables was missing. Each diagonal entry equals the number of districts for which the corresponding variable was not missing. Appendix B. 2 presents the means and standard deviations of the variables in the initial data matrix; each mean and standard deviation is based on a variable's values in all districts for which that variable was not missing. Appendix B. 3 is the correlation matrix of the variables; each correlation coefficient is based on all districts for which neither of the correlates was missing.

This appendix is discussed and interpreted in Section II.C.

Appendix B. 1. COUNTS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	628	628	628	628	628	628	539	541	541	541	541	541	628	628	628	628	628	628	628	628	628	628	628	628	628	628	628	628	628	628	628
2	626	629	629	629	629	629	539	541	541	541	541	541	629	629	629	629	629	629	629	629	629	629	629	629	629	629	629	629	629	629	629
3	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
4	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
5	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
6	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
7	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539	539
8	541	541	541	541	541	541	539	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541	541
9	541	541	542	542	542	542	539	541	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542
10	541	541	542	542	542	542	539	541	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542
11	541	541	542	542	542	542	539	541	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542
12	541	541	542	542	542	542	539	541	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542	542
13	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
14	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
15	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
16	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
17	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
18	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
19	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
20	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
21	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
22	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
23	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
24	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
25	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
26	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
27	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
28	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
29	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
30	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
31	628	629	632	632	632	632	539	541	542	542	542	542	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632

Appendix B. 2. MEANS AND STANDARD DEVIATIONS

	MEANS		STANDARD DEVIATIONS
1	5.549394	1	1.746399
2	3.138771	2	0.706045
3	5012.482243	3	839.014015
4	47.416935	4	30.421660
5	113.412368	5	53.507813
6	1.632026	6	2.287443
7	1.941661	7	0.767724
8	-0.792777	8	0.838533
9	12.839001	9	1.427974
10	7.530149	10	1.405401
11	8.914971	11	1.11775
12	-1.673442	12	1.8028
13	1.797468	13	0.4018
14	1.444620	14	0.4969
15	1.746835	15	0.954249
16	487.558544	16	2060.797673
17	797.996835	17	3224.220010
18	28.827532	18	101.939330
19	6.174051	19	36.860129
20	17.740506	20	55.195798
21	2.155063	21	3.2140 2
22	3.465190	22	20.754031
23	0.670886	23	1.660002
24	0.460443	24	0.932035
25	3.213608	25	7.207828
26	33727812.816406	26	157630550.402344
27	32552.49819 3	27	28658.983093
28	199.202534	28	158.497818
29	20.849323	29	4.198261
30	9.517363	30	7.386481
31	5380376.962402	31	5501990.301270

Appendix B. 3. CORRELATIONS: R_m

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	61	51	50	56	-41	22	-1	27	48	33	12	-27	-39	-42	15	14	15	13	18	27	11	1	10	21	13	-23	37	10	37	27
2	61	100	69	44	41	-53	26	-8	47	50	36	12	-41	-50	-62	23	22	24	20	27	40	17	1	9	31	20	-20	60	9	61	47
3	51	69	100	39	30	-57	15	-18	50	44	29	-4	-42	-41	-58	29	28	31	26	33	46	23	-5	8	37	27	-12	68	11	68	61
4	50	44	39	100	66	-34	11	7	16	75	38	11	-24	-39	-34	14	13	14	12	17	23	11	3	12	19	12	-18	23	1	24	15
5	56	41	30	66	100	-19	18	10	9	57	56	15	-16	-31	-26	4	3	3	2	6	13	2	7	12	8	2	-12	10	5	11	3
6	-41	-53	-57	-34	-19	100	-27	-7	-49	-41	-35	12	26	54	66	-13	-13	-15	-10	-17	-35	-10	12	-9	-23	-11	37	-62	-10	-62	-43
7	22	26	15	11	18	-27	100	33	32	26	44	18	-16	-28	-30	4	4	5	3	6	14	3	10	-1	8	3	-9	21	-11	22	11
8	-1	-8	-18	7	10	-7	33	100	25	18	27	19	10	-14	-13	-7	-6	-7	-7	-6	-6	-5	10	3	-5	-7	3	-13	-11	-12	-14
9	27	47	50	16	9	-49	32	25	100	42	31	21	-26	-32	-41	16	15	18	14	19	25	12	7	-1	22	14	-4	40	-0	41	32
10	48	50	44	75	57	-41	26	18	42	100	61	22	-33	-48	-50	18	16	18	16	21	30	14	11	7	24	15	-24	36	2	36	21
11	33	36	29	38	56	-35	44	27	31	61	100	19	-21	-38	-43	11	10	11	9	13	21	8	9	4	15	9	-13	28	-5	28	17
12	12	12	-4	11	15	12	18	19	21	22	19	100	-17	-37	-31	2	2	1	-2	5	4	1	43	28	5	-0	-17	-12	5	-13	-16
13	-27	-41	-42	-24	-16	26	-16	10	-26	-33	-21	-17	100	44	39	-31	-26	-28	-29	-33	-44	-22	-18	-16	-38	-25	15	-38	-11	-35	-27
14	-39	-50	-41	-39	-31	54	-28	-14	-32	-48	-38	-37	44	100	77	-20	-17	-19	-14	-28	-34	-13	-19	-26	-31	-14	47	-44	-1	-44	-7
15	-42	-62	-58	-34	-26	66	-30	-13	-41	-50	-43	-31	39	77	100	-18	-17	-20	-13	-24	-43	-13	-12	-18	-29	-14	43	-64	-2	-66	-38
16	15	23	29	14	4	-13	4	-7	16	18	11	2	-31	-20	-18	100	99	99	98	99	48	98	0	5	98	99	-6	38	8	34	32
17	14	22	28	13	3	-13	4	-6	15	16	10	2	-26	-17	-17	99	100	99	96	98	45	99	-1	4	97	99	-5	37	8	33	32
18	15	24	31	14	3	-15	5	-7	18	18	11	1	-28	-19	-20	99	99	100	96	98	48	97	-1	4	97	99	-6	41	8	37	36
19	13	20	26	12	2	-10	3	-7	14	16	9	-2	-29	-14	-13	98	96	96	100	94	46	95	-2	1	95	96	-2	35	7	31	32
20	18	27	33	17	6	-17	6	-6	19	21	13	5	-33	-28	-24	99	98	98	94	100	49	97	2	8	97	98	-9	42	8	38	32
21	27	40	46	23	13	-35	14	-6	25	30	21	4	-44	-34	-43	48	45	48	46	49	100	38	-0	14	57	43	-15	48	1	48	38
22	11	17	23	11	2	-10	3	-5	12	14	8	1	-22	-13	-13	98	99	97	95	97	38	100	-1	1	94	99	-3	31	6	28	28
23	1	1	-5	3	7	12	10	10	7	11	9	43	-18	-19	-12	0	-1	-1	-2	2	-0	-1	100	26	1	-2	-8	-21	5	-22	-22
24	10	9	8	12	12	-9	-1	3	-1	7	4	28	-16	-26	-18	5	4	4	1	8	14	1	26	100	9	1	-14	-11	6	-13	-17
25	21	31	37	19	8	-23	8	-5	22	24	15	5	-38	-31	-29	98	97	97	95	97	57	94	1	9	100	95	-12	43	9	39	33
26	13	20	27	12	2	-11	3	-7	14	15	9	-0	-25	-14	-14	99	99	99	96	98	43	99	-2	1	95	100	-2	36	7	32	35
27	-23	-20	-12	-18	-12	37	-9	3	-4	-24	-13	-17	15	47	43	-6	-5	-6	-2	-9	-15	-3	-8	-14	-12	-2	100	-24	-25	-23	15
28	37	60	68	23	10	-62	21	-13	40	36	28	-12	-38	-44	-64	38	37	41	35	42	48	31	-21	-11	43	36	-24	100	14	98	79
29	10	9	11	1	5	-10	-11	-11	-0	2	-5	5	-11	-1	-2	8	8	8	7	8	1	6	5	6	9	7	-25	14	100	2	7
30	37	61	68	24	11	-62	22	-12	41	36	28	-13	-35	-44	-66	34	33	37	31	38	48	28	-22	-13	39	32	-23	98	2	100	79
31	27	47	61	15	3	-43	11	-14	32	21	17	-16	-27	-7	-38	32	32	36	32	32	38	28	-22	-17	33	35	15	79	7	79	100

APPENDIX C

This appendix contains the results of a correlational analysis which was performed as a necessary prerequisite for the regression analysis, which in turn was necessary for replacing the missing data. This analysis was based on the initial data matrix. However, before input to the analysis, the initial data matrix was temporarily modified by omitting from it the entries for all the districts for which there were any missing data. This reduced the dimensions of the initial data matrix from 31 by 632 to 31 by 539. The reduced data matrix had no missing entries, so it could be analyzed according to ordinary correlational techniques.

In Appendix C. 1 appear the means and standard deviations of the 31 variables, based on the 539 districts with no missing data. In Appendix C. 2 appear the correlations of the varicbles.

Because districts have been omitted, this correlation matrix differs from the maximum information correlation matrix, Appendix B. 3. But since all the coefficients in this matrix are based on the same districts, unlike those of the maximum information correlation matrix, regression analysis could proceed from them.

This appendix is discussed and interpreted in Section II.B.

Appendix C. 1. MEANS AND STANDARD DEVIATIONS

	MEANS
1	5.816363
2	3.263983
3	5173.263965
4	51.238409
5	116.700587
6	0.748498
7	1.941661
8	-0.784994
9	12.862589
10	7.544315
11	8.918223
12	-1.659862
13	1.762523
14	1.348794
15	1.530612
16	571.682746
17	932.012987
18	33.623377
19	7.239332
20	20.801484
21	2.513915
22	4.063080
23	0.632653
24	0.521336
25	3.768089
26	39371783.858398
27	28936.453205
28	229.901673
29	21.026479
30	10.968411
31	6133231.986572

	STANDARD DEVIATIONS
1	1.386845
2	0.582853
3	772.474792
4	28.063402
5	43.862761
6	0.831027
7	0.767724
8	0.830277
9	1.370490
10	1.383897
11	1.111709
12	1.798665
13	0.425537
14	0.476589
15	0.866020
16	2220.711595
17	3473.788151
18	109.673648
19	39.816884
20	59.233129
21	3.350476
22	22.419175
23	1.790132
24	0.988341
25	7.669907
26	180919462.738281
27	22163.118453
28	151.789172
29	3.188124
30	7.046192
31	5616370.680664

Appendix C. 2. CORRELATIONS: R_p

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	66	49	52	62	-25	22	-3	27	48	32	11	-27	-34	-35	16	14	15	14	18	24	12	5	5	19	14	-14	30	2	30	21
2	66	100	72	37	42	-48	26	-10	47	50	36	11	-43	-45	-60	25	23	26	23	28	38	19	5	3	31	22	-16	59	0	60	44
3	49	72	100	30	27	-46	15	-19	50	43	30	-5	-39	-26	-47	29	27	31	27	32	41	23	-3	-0	34	27	-3	62	6	62	57
4	52	37	30	100	73	-19	11	7	16	75	37	10	-21	-33	-24	13	11	12	12	15	19	11	6	9	17	11	-16	12	-7	13	6
5	62	42	27	73	100	-13	18	10	11	58	56	15	-17	-35	-29	3	2	2	1	5	12	1	10	12	6	1	-14	5	-6	6	-2
6	-25	-48	-46	-19	-13	100	-27	-5	-48	-40	-35	14	19	36	54	-13	-13	-14	-12	-15	-30	-9	20	20	-18	-11	34	-61	-0	-62	-38
7	22	26	15	11	18	-27	100	33	32	26	44	18	-16	-28	-30	4	4	5	3	6	14	3	10	-1	8	3	-9	21	-11	22	11
8	-3	-10	-19	7	10	-5	33	100	23	18	27	18	11	-13	-11	-7	-6	-8	-8	-7	-7	-5	10	4	-6	-7	4	-14	-11	-13	-15
9	27	47	50	16	11	-48	32	23	100	44	34	19	-27	-31	-41	16	16	18	14	19	25	13	6	-0	22	15	-4	40	1	41	32
10	48	50	43	75	58	-40	26	18	44	100	61	22	-33	-48	-49	18	17	18	16	21	30	14	11	8	24	15	-24	35	0	35	21
11	32	36	30	37	56	-35	44	27	34	61	100	19	-21	-39	-43	11	10	11	9	13	21	8	9	4	15	9	-13	27	-7	28	17
12	11	11	-5	10	15	14	18	18	19	22	19	100	-16	-36	-30	2	1	1	-2	5	3	1	43	29	5	-0	-17	-13	4	-14	-17
13	-27	-43	-39	-21	-17	19	-16	11	-27	-33	-21	-16	100	39	34	-30	-25	-27	-28	-31	-40	-21	-19	-13	-36	-24	13	-33	-13	-29	-22
14	-34	-45	-26	-33	-35	36	-28	-13	-31	-48	-39	-36	39	100	70	-18	-14	-16	-12	-25	-26	-11	-25	-22	-26	-11	52	-29	6	-29	9
15	-35	-60	-47	-24	-29	54	-30	-11	-41	-49	-43	-30	34	70	100	-15	-14	-16	-11	-20	-35	-11	-18	-11	-23	-12	43	-52	6	-54	-25
16	16	25	29	13	3	-13	4	-7	16	18	11	2	-30	-18	-15	100	99	99	98	99	47	98	1	4	98	99	-4	38	10	33	30
17	14	23	27	11	2	-13	4	-6	16	17	10	1	-25	-14	-14	99	100	99	96	98	45	99	-0	2	97	99	-3	37	10	33	31
18	15	26	31	12	2	-14	5	-8	18	18	11	1	-27	-16	-16	99	99	100	97	98	47	97	-0	3	98	99	-3	41	10	37	34
19	14	23	27	12	1	-12	3	-8	14	16	9	-2	-28	-12	-11	98	96	97	100	94	45	95	-2	-0	96	96	-0	36	9	32	31
20	18	28	32	15	5	-15	6	-7	19	21	13	5	-31	-25	-20	99	98	98	94	100	48	97	2	6	97	98	-7	41	10	36	30
21	24	38	41	19	12	-30	14	-7	25	30	21	3	-40	-26	-35	47	45	47	45	48	100	38	2	10	55	42	-10	42	-2	42	32
22	12	19	23	11	1	-9	3	-5	13	14	8	1	-21	-11	-11	98	99	97	95	97	38	100	-1	0	95	99	-1	32	8	28	27
23	5	5	-3	6	10	20	10	10	6	11	9	43	-19	-25	-18	1	-0	-0	-2	2	2	-1	100	28	2	-2	-15	-21	7	-22	-21
24	5	3	-0	9	12	20	-1	4	-0	8	4	29	-13	-22	-11	4	2	3	-0	6	10	0	28	100	7	-0	-13	-21	9	-24	-24
25	19	31	34	17	6	-18	8	-6	22	24	15	5	-36	-26	-23	98	97	98	96	97	55	95	2	7	100	96	-9	40	11	35	29
26	14	22	27	11	1	-11	3	-7	15	15	9	-0	-24	-11	-12	99	99	99	96	98	42	99	-2	-0	96	100	1	36	9	32	34
27	-14	-16	-3	-16	-14	34	-9	4	-4	-24	-13	-17	13	52	43	-4	-3	-3	-0	-7	-10	-1	-15	-13	-9	1	100	-15	-18	-13	33
28	30	59	62	12	5	-61	21	-14	40	35	27	-13	-33	-29	-52	38	37	41	36	41	42	32	-21	-21	40	36	-15	100	14	98	76
29	2	0	6	-7	-6	-0	-11	-11	1	0	-7	4	-13	6	6	10	10	10	9	10	-2	8	7	9	11	9	-18	14	100	-3	5
30	30	60	62	13	6	-62	22	-13	41	35	28	-14	-29	-29	-54	33	33	37	32	36	42	28	-22	-24	35	32	-13	98	-3	100	76
31	21	44	57	6	-2	-38	11	-15	32	21	17	-17	-22	9	-25	30	31	34	31	30	32	27	-21	-24	29	34	33	76	5	76	100

APPENDIX D

This appendix presents the results of the regression analysis which provided formulas for replacing missing data. The regression analysis was computed from the correlation matrix which appears as Appendix C. 2 and which is based on districts with no missing data. Regression analysis requires a complete data matrix.

The analysis yielded the normal beta coefficients presented in Appendix D. 1. Each row in the 12 by 19 array gives the normal beta coefficients for the 19 variables 13 through 31 in predicting one of the twelve variables 1 through 12. The coefficients are least-squares best sets for the reduced sample of districts which had no missing data. After each row appears the squared multiple correlation coefficient of the prediction--that is, the proportion of the variance of the variable (1 through 12) which is predictable from the 19 variables (13 through 31).

The formulas based on the beta coefficients were used in computing replacement values for the districts with missing data.

This appendix and its application are explained in Section II.B.

Appendix D. 1. NORMAL BETA COEFFICIENTS AND SQUARED MULTIPLE CORRELATIONS

	BETA COEFFICIENTS																			SQUARED MULTIPLE CORRELATIONS
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1	-9	-23	-5	-150	153	-130	60	73	1	-19	-0	3	40	-23	-4	-85	13	81	27	0.205766
2	-13	-10	-25	-132	51	-48	44	59	-1	10	4	4	65	-42	8	-37	7	67	15	0.534084
3	-10	8	-18	-169	-49	10	39	107	3	38	1	6	76	-41	5	-29	10	50	31	0.515940
4	-7	-37	9	-163	61	-72	62	48	9	34	-4	5	31	4	-14	-35	-4	11	32	0.159770
5	-6	-37	-9	-98	99	-90	40	24	7	24	-3	4	9	-14	2	-60	6	39	16	0.159021
6	-4	5	18	84	-148	160	-8	34	0	16	12	10	-114	-20	22	38	-6	-79	-11	0.558276
7	-4	-16	-8	-118	90	-61	8	-12	-1	5	11	-3	63	19	8	50	-16	-26	-5	0.143468
8	13	-18	-24	-62	120	-83	14	-37	-1	5	4	-2	24	16	24	32	-8	-37	-16	0.129864
9	-1	-11	-12	-69	37	-64	-25	-26	-11	-6	11	0	152	10	19	-40	10	76	2	0.300577
10	-7	-32	-12	-74	115	-102	23	-14	2	-22	3	3	58	18	-4	-44	7	50	18	0.331498
11	0	-23	-22	-9	56	-75	-4	-34	-0	-6	5	0	54	18	9	-5	-2	16	6	0.228530
12	-5	-24	-25	-250	106	-3	68	140	1	38	22	7	-17	-78	1	-11	5	-42	26	0.336061

APPENDIX E

This appendix presents the final correlational analysis on the input data. The analysis is based on the complete data matrix with regression estimates substituted for missing data.

In Appendix E. 1 appear the means and standard deviations of the variables, and in Appendix E. 2 appears the matrix of variable intercorrelations.

The correlation matrix, R, given in Appendix E. 2, served as input to the factorization procedures.

This appendix is further discussed in Section II. B.

Appendix E. 1. MEANS AND STANDARD DEVIATIONS

	MEANS
1	5.544972
2	3.135900
3	5012.482243
4	47.416935
5	113.412368
6	1.632026
7	1.879863
8	-0.791359
9	12.654994
10	7.313257
11	8.782369
12	-1.760688
13	1.797468
14	1.444620
15	1.746835
16	487.558544
17	797.996835
18	28.827532
19	6.174051
20	17.740506
21	2.155063
22	3.465190
23	0.670886
24	0.460443
25	3.213608
26	33727812.816406
27	32552.498190
28	199.202534
29	20.849323
30	9.517363
31	5380376.963135

	STANDARD DEVIATIONS
1	1.741822
2	0.705652
3	839.014015
4	30.421660
5	53.507813
6	2.287443
7	0.735184
8	0.795903
9	1.414085
10	1.409000
11	1.089274
12	1.685647
13	0.401886
14	0.496924
15	0.954249
16	2060.797673
17	3224.220010
18	101.939330
19	36.860129
20	55.195788
21	3.214092
22	20.754031
23	1.660002
24	0.932035
25	7.207828
26	157630550.390625
27	28658.983108
28	158.497818
29	4.198261
30	7.386481
31	5501990.301147

Appendix E. 2. CORRELATIONS: R

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	60	51	50	56	-41	22	-3	30	48	33	14	-28	-39	-42	15	14	15	13	18	27	11	1	10	21	13	-23	37	10	37	27
2	60	100	69	44	42	-53	27	-6	49	52	39	15	-41	-51	-62	23	22	24	20	27	40	17	1	9	31	20	-20	61	9	61	47
3	51	69	100	39	30	-57	21	-14	56	52	38	3	-42	-41	-58	29	28	31	26	33	46	23	-5	8	37	27	-12	68	11	68	61
4	50	44	39	100	66	-34	15	6	23	71	40	13	-24	-39	-34	14	13	14	12	17	23	11	3	12	19	12	-18	23	1	24	15
5	56	42	30	66	100	-19	15	7	12	46	45	14	-16	-31	-26	4	3	3	2	6	13	2	7	12	8	2	-12	10	5	11	3
6	-41	-53	-57	-34	-19	100	-27	-2	-46	-48	-39	-8	26	54	66	-13	-13	-15	-10	-17	-35	-10	12	-9	-23	-11	37	-62	-10	-62	-43
7	22	27	21	15	15	-27	100	35	36	30	47	20	-19	-33	-36	6	6	7	4	9	18	4	8	3	11	5	-6	27	-16	28	17
8	-3	-6	-14	6	7	-2	35	100	25	15	26	18	10	-12	-10	-6	-6	-7	-7	-6	-6	-5	9	3	-5	-6	16	-12	-16	-11	-12
9	30	49	56	23	12	-46	36	25	100	48	38	24	-31	-41	-50	18	17	20	15	22	31	14	5	4	26	16	-4	48	6	49	39
10	48	52	52	71	46	-48	30	15	48	100	65	26	-37	-57	-59	20	19	21	17	24	36	15	8	13	29	17	-30	47	9	47	30
11	33	39	38	40	45	-39	47	26	38	65	100	21	-26	-46	-51	13	12	13	10	16	27	9	7	8	20	11	-13	37	-3	38	25
12	14	15	3	13	14	-8	20	18	24	26	21	100	-19	-38	-32	3	3	3	-1	7	7	2	42	29	7	1	-15	-5	8	-6	-11
13	-28	-41	-42	-24	-16	26	-19	10	-31	-37	-26	-19	100	44	39	-31	-26	-28	-29	-33	-44	-22	-18	-16	-38	-25	15	-38	-11	-35	-27
14	-39	-51	-41	-39	-31	54	-33	-12	-41	-57	-46	-38	44	100	77	-20	-17	-19	-14	-28	-34	-13	-19	-26	-31	-14	47	-44	-1	-44	-7
15	-42	-62	-58	-34	-26	66	-36	-10	-50	-59	-51	-32	39	77	100	-18	-17	-20	-13	-24	-43	-13	-12	-18	-29	-14	43	-64	-2	-66	-38
16	15	23	29	14	4	-13	6	-6	18	20	13	3	-31	-20	-18	100	99	99	98	99	48	98	0	5	98	99	-6	38	8	34	32
17	14	22	28	13	3	-13	6	-6	17	19	12	3	-26	-17	-17	99	100	99	96	98	45	99	-1	4	97	99	-5	37	8	33	32
18	15	24	31	14	3	-15	7	-7	20	21	13	3	-28	-19	-20	99	99	100	96	98	48	97	-1	4	97	99	-6	41	8	37	36
19	13	20	26	12	2	-10	4	-7	15	17	10	-1	-29	-14	-13	98	96	96	100	94	46	95	-2	1	95	96	-2	35	7	31	32
20	18	27	33	17	6	-17	9	-6	22	24	16	7	-33	-28	-24	99	98	98	94	100	49	97	2	8	97	98	-9	42	8	38	32
21	27	40	46	23	13	-35	18	-6	31	36	27	7	-44	-34	-43	48	45	48	46	49	100	38	-0	14	57	43	-15	48	1	48	38
22	11	17	23	11	2	-10	4	-5	14	15	9	2	-22	-13	-13	98	99	97	95	97	38	100	-1	1	94	99	-3	31	6	28	28
23	1	1	-5	3	7	12	8	9	5	8	7	42	-18	-19	-12	0	-1	-1	-2	2	-0	-1	100	26	1	-2	-8	-21	5	-22	-22
24	10	9	8	12	12	-9	3	3	4	13	8	29	-16	-26	-18	5	4	4	1	8	14	1	26	100	9	1	-14	-11	6	-13	-17
25	21	31	37	19	8	-23	11	-5	26	29	20	7	-38	-31	-29	98	97	97	95	97	57	94	1	9	100	95	-12	43	9	39	33
26	13	20	27	12	2	-11	5	-6	16	17	11	1	-25	-14	-14	99	99	99	96	98	43	99	-2	1	95	100	-2	36	7	32	35
27	-23	-20	-12	-18	-12	37	-6	16	-4	-30	-13	-15	15	47	43	-6	-5	-6	-2	-9	-15	-3	-8	-14	-12	-2	100	-24	-25	-23	15
28	37	61	68	23	10	-62	27	-12	48	47	37	-5	-38	-44	-64	38	37	41	35	42	48	31	-21	-11	43	36	-24	100	14	98	79
29	10	9	11	1	5	-10	-16	-16	6	9	-3	8	-11	-1	-2	8	8	8	7	8	1	6	5	6	9	7	-25	14	100	2	7
30	37	61	68	24	11	-62	28	-11	49	47	38	-6	-35	-44	-66	34	33	37	31	38	48	28	-22	-13	39	32	-23	98	2	100	79
31	27	47	61	15	3	-43	17	-12	39	30	25	-11	-27	-7	-38	32	32	36	32	32	38	28	-22	-17	33	35	15	79	7	79	100

APPENDIX F

This appendix presents the outputs from a factorization of the correlation matrix, R, given in Appendix E. 2. The factorization scheme is that of principal components.

In Appendix F. 1 are given the complete set of latent roots of the correlation matrix; they are considered to be the diagonal entries of the diagonal matrix M^2. The corresponding latent vectors of R are given as the columns of Appendix F. 2; this matrix is denoted Q. The factor matrix for the principal components is given in Appendix F. 3. It was computed according to the formula: $F_r = QM$. In the appendix the factor matrix is bordered by row and column sums of squares. The row sums of squares are uniformly 1.0 because all the variance of each variable is accounted for in the factors. The column sums of squares are the factor variances, and are additive since the factors are uncorrelated. Note that the column sums of squares are equal to the corresponding latent roots, given in Appendix F. 1, and are arranged in order of decreasing magnitude.

This appendix is further discussed in Section II. C.

Appendix F. 1. LATENT ROOTS OF R: M^2

1	10.599676
2	5.491264
3	2.596522
4	1.734891
5	1.613978
6	1.137623
7	0.982707
8	0.796910
9	0.731679
10	0.671310
11	0.620614
12	0.542390
13	0.513268
14	0.475593
15	0.435506
16	0.369534
17	0.352517
18	0.295970
19	0.277043
20	0.239874
21	0.160089
22	0.129978
23	0.105080
24	0.062545
25	0.023133
26	0.016350
27	0.009915
28	0.008315
29	0.004258
30	0.000893
31	0.000572

Appendix F. 2. LATENT VECTORS OF R: Q

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-14	19	-4	15	-29	8	-2	-1	-37	-37	18	5	-16	-8	6	5	61	9	-29	-2	1	7	-5	0	-1	1	0	-0	0	-0	-0
2	-19	21	6	8	-8	18	3	-1	-28	-14	13	-1	2	-8	-15	-15	-30	56	45	31	-5	-0	6	-3	1	0	0	-1	-0	-0	-0
3	-21	17	17	7	-4	25	3	12	-8	0	3	-16	11	19	-16	1	-8	-6	14	-79	-22	-3	2	2	2	1	1	-0	-0	0	-0
4	-13	18	-16	9	-45	-5	3	2	11	29	6	-3	4	15	47	4	-22	4	-5	1	-8	53	-11	0	0	0	-0	0	0	0	-0
5	-8	18	-19	10	-52	2	-0	-6	-5	-11	-9	3	-8	-18	-21	-6	-30	-56	8	10	10	-31	5	1	-1	-1	0	-0	-0	-0	0
6	16	-22	-12	-5	-12	23	2	-25	8	-0	-3	11	-4	-3	-7	-81	2	9	-14	-18	7	15	4	0	-1	1	3	3	-1	0	0
7	-9	15	-8	-42	8	-10	2	-22	2	-63	-16	5	10	33	31	0	-24	3	-8	-7	7	-7	1	1	1	-0	-1	0	0	0	-0
8	1	6	-20	-53	-1	-14	-25	25	8	-1	30	-17	-21	-48	15	-9	6	-0	22	-17	-12	1	3	0	-1	-0	-1	1	-0	0	-0
9	-15	17	3	-24	14	19	-26	19	-6	11	41	-7	16	45	-20	-20	7	-31	-9	31	11	11	0	-1	3	0	-0	-1	1	-0	0
10	-18	23	-12	-2	-17	-7	-11	-1	27	33	-5	-1	6	20	16	-12	23	33	-6	-3	16	-63	4	-2	-0	-1	0	-1	0	-0	-0
11	-14	20	-12	-25	-13	-11	-8	-9	31	-2	-44	1	9	1	-53	9	27	11	10	7	-14	33	-2	-1	0	1	-0	0	0	0	0
12	-5	10	-37	-4	25	18	-19	-6	-28	22	-15	67	-18	7	9	9	1	-7	16	-9	-14	-1	-6	4	0	-0	-0	-0	-1	0	-0
13	16	-8	7	-10	-14	-34	-20	32	-46	12	-34	-18	-34	28	-10	-13	-15	13	-18	-1	-6	-2	6	5	-2	-5	1	0	-1	-1	-1
14	17	-21	20	-5	-20	22	-15	5	7	-10	-14	-5	-16	21	15	4	23	-8	53	0	30	1	-43	-1	-6	-12	-5	0	-2	-0	0
15	19	-25	4	0	-23	13	-8	0	13	-12	5	4	6	22	12	2	13	-5	11	20	-74	-18	24	-3	1	4	5	3	-2	0	0
16	-25	-24	-6	-1	-2	-3	-1	-0	-1	-1	1	-0	3	-1	-1	0	-0	0	-1	1	-2	-1	-1	5	18	-29	-2	-7	14	33	-78
17	-25	-25	-5	-2	-2	-4	-3	1	-5	-0	-2	-1	2	1	-0	1	-0	-0	2	-0	3	-1	-6	-18	-7	30	17	-12	21	72	37
18	-25	-23	-4	-2	-1	-2	-2	1	-4	-0	-3	-2	1	1	1	-2	-0	-1	1	1	-1	-1	-4	2	-42	52	-34	-17	32	-36	-21
19	-24	-25	-4	-1	-4	-2	-1	-2	1	-1	2	-1	2	0	-0	2	2	2	3	2	2	-0	4	82	32	-3	-6	-2	7	-8	28
20	-26	-21	-7	-0	0	-4	0	0	-3	0	-0	1	5	-2	-2	-3	-2	-1	-6	-1	-9	-1	-3	-36	4	-60	-40	-5	23	-17	35
21	-20	2	2	2	8	11	32	7	36	-6	13	3	-77	19	-10	4	-8	3	-7	4	-3	1	3	-7	10	5	2	-2	1	-3	1
22	-23	-26	-6	-2	-4	-6	-5	0	-8	1	-2	-1	6	-1	-0	3	1	-1	4	-2	7	-1	-6	-30	38	17	54	33	11	-41	-2
23	0	3	-38	3	22	36	-6	-35	-16	11	-18	-65	-13	-0	9	9	3	-5	-3	10	-8	-3	-3	1	-1	-1	1	1	-2	0	0
24	-3	5	-32	16	15	19	28	70	9	-19	-30	-5	18	-5	14	-20	8	-4	2	10	-0	-0	-3	2	3	1	1	1	-2	-0	0
25	-27	-19	-7	0	1	-3	2	2	3	-2	6	1	-1	3	-4	5	-2	2	-2	-0	1	1	4	17	-69	-28	30	35	-28	0	2
26	-24	-25	-4	-3	-4	-2	-4	0	-5	1	-3	-0	4	-1	2	2	2	-1	2	-1	7	2	11	-18	17	17	-23	-22	-81	-2	-1
27	7	-12	12	-39	-22	53	2	15	-1	9	-6	10	7	-17	-9	27	-20	20	-42	4	1	-11	-23	1	-0	2	-1	1	-2	-0	0
28	-23	13	31	1	13	-4	-5	-6	1	3	-15	0	-2	-16	13	-22	-1	-11	-8	10	-25	-6	-33	5	-8	-13	37	-55	-7	-15	2
29	-4	0	0	40	9	9	-74	10	30	-25	-4	0	-8	-9	-1	3	-20	10	-18	-3	2	5	-3	-0	3	3	-5	7	-0	1	-0
30	-22	15	32	-3	11	-6	4	-7	-4	7	-14	-0	-4	-13	12	-21	2	-13	-5	11	-23	-7	-31	2	9	15	-32	59	-9	15	-2
31	-18	6	38	-10	-0	26	-9	1	-2	9	-30	3	-8	-14	26	0	11	-12	5	6	20	12	66	-4	-3	-5	4	2	12	1	1

Appendix F. 3. UNROTATED PRINCIPAL COMPONENTS FACTOR MATRIX: F_r =QM

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
		1060	549	260	173	161	114	98	80	73	67	62	54	51	48	44	37	35	30	28	24	16	13	11	6	2	2	1	1	0	0	0
1	100	-47	44	-7	20	-37	8	-2	-1	-37	-30	14	3	-11	-6	4	3	36	5	-16	-1	0	3	-2	0	-0	0	0	-0	0	-0	-0
2	100	-63	48	10	11	-10	19	3	-1	-24	-11	11	-1	2	-5	-10	-9	-18	31	24	15	-2	-0	2	-1	0	0	0	-0	-0	-0	-0
3	100	-67	40	27	10	-5	26	2	10	-7	0	3	-12	8	13	-11	1	-5	-3	7	-39	-9	-1	1	0	0	0	0	-0	-0	0	-0
4	100	-42	42	-26	12	-58	-5	3	2	10	24	5	-2	3	10	31	3	-13	2	-3	1	-3	19	-4	0	0	0	-0	0	0	0	-0
5	100	-28	41	-31	13	-66	3	-0	-5	-4	-9	-7	2	-6	-12	-14	-4	-18	-31	4	5	4	-11	2	0	-0	-0	0	-0	-0	-0	0
6	100	53	-51	-19	-7	-16	24	2	-22	7	-0	-3	8	-3	-2	-5	-49	1	5	-7	-9	3	6	1	0	-0	0	0	0	-0	0	0
7	100	-29	35	-14	-56	10	-11	2	-19	2	-52	-13	4	7	23	21	0	-14	1	-4	-3	3	-2	0	0	0	-0	-0	0	0	0	-0
8	100	3	15	-33	-70	-1	-15	-24	22	7	-1	23	-12	-15	-33	10	-6	3	-0	12	-8	-5	0	1	0	-0	-0	-0	0	-0	0	-0
9	100	-50	40	6	-31	18	20	-26	17	-5	9	32	-5	12	31	-13	-12	4	-17	-5	15	4	4	0	-0	0	C	-0	-0	0	-0	0
10	100	-58	53	-20	-2	-21	-8	-11	-1	23	27	-4	-1	5	14	11	-7	13	18	-3	-2	6	-23	1	-0	-0	-0	0	-0	0	-0	-0
11	100	-45	48	-19	-33	-16	-12	-7	-8	26	-1	-35	1	6	1	-35	6	16	6	5	3	-6	12	-1	-0	0	0	-0	0	0	0	0
12	100	-15	24	-60	-6	31	20	-19	-5	-24	18	-12	49	-13	5	6	5	1	-4	8	-4	-6	-0	-2	1	0	-0	-0	-0	-0	0	-0
13	100	51	-19	11	-14	-18	-36	-19	29	-39	10	-27	-13	-24	19	-7	-8	-9	7	-9	-0	-2	-1	2	1	-0	-1	0	0	-0	-0	-0
14	100	55	-50	33	-6	-25	23	-14	5	6	-8	-11	-3	-12	15	10	2	13	-4	28	0	12	0	-14	-0	-1	-2	-0	0	-0	-0	0
15	100	62	-58	6	0	-29	14	-8	0	11	-10	4	3	4	15	8	1	8	-3	6	10	-30	-6	8	-1	0	1	0	0	-0	0	0
16	100	-82	-56	-10	-1	-2	-3	-1	-0	-1	-1	1	-0	2	-1	-1	0	-0	0	-0	0	-1	-0	-0	1	3	-4	-0	-1	1	1	-2
17	100	-81	-58	-8	-2	-3	-5	-3	1	-4	-0	-2	-1	1	0	-0	0	-0	-0	1	-0	1	-0	-2	-4	-1	4	2	-1	1	2	1
18	100	-82	-55	-6	-2	-1	-3	-2	1	-3	-0	-2	-1	1	1	0	-1	-0	-1	1	0	-0	-0	-1	0	-6	7	-3	-2	2	-1	-0
19	100	-78	-58	-6	-2	-5	-2	-1	-2	1	-1	2	-1	2	0	-0	1	1	1	2	1	1	-0	1	20	5	-0	-1	-0	0	-0	1
20	100	-84	-50	-11	-0	1	-4	0	0	-3	0	-0	0	3	-2	-1	-2	-1	-0	-3	-0	-4	-1	-1	-9	1	-8	-4	-0	2	-1	1
21	100	-65	5	4	3	11	12	31	6	31	-5	10	2	-55	13	-7	2	-5	1	-4	2	-1	0	1	-2	2	1	0	-0	0	-0	0
22	100	-76	-61	-10	-3	-5	-6	-5	0	-7	1	-2	-1	5	-0	-0	2	0	-0	2	-1	3	-0	-2	-7	6	2	5	3	1	-1	-0
23	100	0	7	-61	4	28	38	-6	-32	-13	9	-14	-48	-9	-0	6	6	2	-3	-1	5	-3	-1	-1	0	-0	-0	0	0	-0	0	0
24	100	-10	12	-51	21	19	20	28	62	7	-15	-24	-3	13	-3	9	-12	5	-2	1	5	-0	-0	-1	1	0	0	0	0	-0	-0	0
25	100	-87	-45	-12	0	2	-3	2	2	2	-2	5	1	-1	2	-3	3	-1	1	-1	-0	0	0	1	4	-10	-4	3	3	-2	0	0
26	100	-79	-60	-6	-4	-5	-3	-4	0	-5	1	-3	-0	3	-0	1	1	1	-0	1	-1	3	1	3	-4	3	2	-2	-2	-5	-0	-0
27	100	24	-28	20	-51	-28	57	2	13	-0	7	-4	8	5	-12	-6	16	-12	11	-22	2	0	-4	-7	0	-0	0	-0	0	-0	-0	0
28	100	-74	31	50	1	16	-5	-5	-5	1	2	-12	0	-2	-11	8	-13	-1	-6	-4	5	-10	-2	-11	1	-1	-2	4	-5	-0	-0	0
29	100	-11	1	0	52	11	9	-73	9	26	-21	-3	0	-5	-6	-1	2	-12	6	-9	-1	1	2	-1	-0	0	0	-0	1	-0	0	-0
30	100	-71	34	52	-4	14	-7	4	-6	-4	6	-11	-0	-3	-9	8	-13	1	-7	-2	6	-9	-3	-10	1	1	2	-3	5	-1	0	-0
31	100	-58	15	61	-13	-0	28	-9	1	-2	7	-24	2	-5	-10	17	0	6	-6	3	3	8	4	21	-1	-0	-1	0	0	1	0	0

APPENDIX G

The principal component factorizntion of the correlation matrix R was given in Appendix F. The factor matrix presented there provided the basis for deriving the normal varimax orthogonal factorization of R.

The factor matrix F_r given in Appendix F. 3 was subjected to the normal varimax orthogonnl rotation procedure, and the matrix $F_r T_r$ was derived, where T_r is an orthonormal matrix and is presented in Appendix G. 1. The matrix $F_r T_r$ is the rotated component factor matrix, and is presented as Appendix G. 2. The rows and columns are bordered by row and column sums of squares. The row sums of squares are uniformly 1.0, because the factors account for all the varicnce in each variable. The colum. sums of squares are the factor variances and are additive since the factors are uncorrelated. Note that the factors have been arranged in order of decreasing variance. Because the factor scores for the rotated component factors were to be computed, the rotated component factor weight matrix, $QM^{-1}T$, was computed. This matrix is given in Appendix G. 3. The columns have been normalized and give the normal weights for computing the factor scores from the original variables.

This appendix is further discussed in Section II. C.

Appendix G. 1. NORMAL VARIMAX TRANSFORMATION MATRIX: T_r

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-74	-44	-19	10	-15	-6	-15	-0	16	-5	-10	15	-3	-1	-15	-9	-15	13	-13	-10	6	-6	-0	-0	-0	-0	-0	-0	0	-0	-0
2	-65	35	27	-16	22	13	20	8	-22	7	16	-11	1	5	18	17	6	-18	16	10	-9	8	-1	-0	0	-0	0	0	-0	-0	-0
3	-13	61	-20	15	-14	-38	-5	-21	18	-32	-9	8	0	-38	2	-17	2	-7	2	7	3	-5	3	0	-0	0	-0	0	0	0	-0
4	-3	-5	10	-40	-23	-5	15	-55	-5	17	-42	-10	41	3	-21	9	3	-4	7	5	0	-1	-3	-0	0	-0	0	0	-0	0	0
5	-5	17	-51	-26	-13	26	-30	1	-22	16	9	-15	8	23	13	-49	9	-11	-6	-4	-10	-4	-3	-0	1	0	0	0	-0	0	0
6	-7	10	-6	61	-12	17	9	-17	25	19	-11	-34	10	37	19	3	11	21	15	18	7	-4	10	1	0	0	0	0	0	0	0
7	-3	-3	-0	-1	-8	-19	-1	-26	-16	29	2	-21	-75	-6	-24	0	31	0	4	2	-5	-6	-3	0	1	0	0	-1	-0	-0	0
8	1	-1	2	15	-9	-6	-1	25	5	70	-22	32	10	-36	20	-6	7	-24	0	10	-1	-1	2	-1	0	0	-0	-0	-0	0	-0
9	-5	-7	19	0	34	-29	-39	9	9	8	3	-46	31	-16	-6	-8	36	9	-26	-7	8	15	0	2	1	0	0	1	-0	0	0
10	-0	12	38	6	4	24	-38	1	-12	-18	-63	11	-25	12	13	-14	-6	-1	-13	-1	-8	19	2	-0	-0	0	-0	-0	0	-0	-0
11	-0	-30	5	-8	-44	-15	19	29	-16	-29	-16	-37	-4	-18	43	-8	15	-6	15	6	2	-3	-8	1	1	-0	0	-0	-0	-0	0
12	-1	-3	-4	11	0	68	4	-17	-3	-5	5	-18	0	-65	-8		3	12	-2	-14	3	-1	1	-0	0	-0	0	0	0	-0	0
13	4	-7	6	6	10	-18	-15	-21	-16	17	10	-34	-7	-13	18	-8	-79	-5	3	10	4	4	-2	-1	-1	-1	0	0	0	-0	-0
14	0	-22	20	-14	3	6	-8	-48	25	-5	33	28	-7	-1	49	-20	20	-3	-6	19	13	12	-1	0	1	0	-0	-0	-0	0	-0
15	-1	21	48	-6	-55	7	3	15	22	12	29	-8	-0	8	-24	-25	-12	-3	-19	-19	11	6	8	0	-1	0	-0	0	0	0	-0
16	1	-19	1	26	8	8	5	-9	4	-18	0	-12	3	8	-19	-7	4	-86	-13	4	-1	-10	10	1	0	-0	-1	-0	0	-0	-0
17	-0	3	-14	-18	31	1	59	7	29	6	-25	-14	-18	2	7	-32	-9	4	-33	-9	11	21	6	1	-1	0	-0	-0	0	-0	-0
18	0	-14	14	17	16	-5	12	-0	-13	-3	3	10	10	-4	-27	-57	5	7	60	-1	-7	28	-1	1	1	-0	-0	0	-0	-0	0
19	1	-0	-6	-36	8	13	-26	19	61	1	-9	-17	-16	-3	-8	10	-8	-13	47	16	7	-5	7	2	-1	0	-2	-0	0	-0	-0
20	-0	4	-1	4	5	-8	-4	-13	4	7	-6	1	-3	8	26	9	2	-14	26	-86	21	-5	-2	2	0	0	0	0	0	0	0
21	1	-8	-3	0	-8	-7	0	-8	18	1	6	-5	1	-5	11	9	-2	1	-2	-18	-85	21	33	1	-2	-1	-3	-1	1	-0	-0
22	-0	-2	26	-5	20	-1	6	-0	2	-0	-5	-0	3	-2	6	-28	0	14	0	-3	-19	-85	16	0	0	-0	-0	0	0	0	0
23	-0	-3	-6	-8	-2	-3	-3	2	-23	-2	0	4	-1	-2	0	4	2	5	4	5	32	6	90	5	1	0	0	-1	3	-0	0
24	-0	1	-0	0	-0	1	0	-0	-1	1	0	2	-0	0	-0	1	-2	0	-1	1	-2	-1	-5	97	22	10	-7	2	-1	0	3
25	-0	-0	0	0	-0	-0	-0	-0	-1	0	0	0	1	-0	0	-0	2	0	-0	0	1	-1	-1	24	-82	-51	6	-10	4	-4	1
26	-0	-0	-0	0	0	-0	-0	-0	-2	0	-0	-1	1	-0	-0	-0	1	0	-0	0	2	-1	-1	-2	-34	50	-77	-18	4	3	-10
27	-0	0	-0	0	-0	-0	0	-0	-1	0	-0	0	-0	0	-0	0	0	1	-0	0	2	0	1	-5	23	-56	-58	44	-30	-2	-7
28	-0	1	0	0	-0	-0	-0	0	-0	0	0	0	1	0	-0	-0	-0	1	-0	-0	0	-0	-0	-2	28	-31	-13	-87	-21	-9	-7
29	0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	0	-0	0	-0	-0	-0	0	-0	-0	0	3	2	-19	25	18	4	-93	5	2
30	0	-0	0	0	0	-0	0	0	-0	0	-0	-0	0	-0	0	-0	0	-0	0	-0	0	0	-0	-1	2	-9	-7	-8	1	86	48
31	-0	0	0	-0	-0	0	-0	-0	0	0	0	0	-0	0	-0	0	-0	0	-0	0	-0	-0	-0	3	0	-3	11	1	2	49	-86

Appendix G. 2. ROTATED PRINCIPAL COMPONENTS FACTOR MATRIX: $F_r T_r = QMT_r$

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
		814	388	158	125	111	108	105	105	105	105	104	103	103	102	101	97	86	80	68	48	28	28	14	6	2	1	1	1	1	0	0
1	100	8	22	23	-9	7	5	87	-2	-8	4	7	-7	4	0	7	24	5	-9	14	7	-2	3	0	0	-0	0	-0	-0	-0	-0	-0
2	100	13	44	19	-6	9	6	26	-5	-14	4	8	-14	3	3	16	17	9	-12	72	12	-5	3	0	0	0	-0	0	-0	-0	-0	-0
3	100	19	55	18	0	10	-2	19	-11	-8	7	4	-14	4	0	24	11	12	-14	18	64	-5	3	1	0	0	-0	-0	-0	-0	0	0
4	100	9	10	91	-6	10	3	17	2	-8	4	3	-6	-1	1	4	27	4	-8	8	5	-1	-1	0	0	0	0	0	0	-0	0	0
5	100	-0	1	38	-3	19	4	25	2	-7	4	4	-4	2	2	-0	86	2	-1	10	5	-2	2	-0	-0	-0	0	0	0	-0	0	0
6	100	-5	-45	-14	20	-12	-2	-12	-3	17	-6	-9	4	-4	7	-15	-2	-8	78	-9	-8	5	-2	-0	-0	-0	0	-0	-0	-0	0	0
7	100	2	15	4	-2	17	8	6	18	-8	0	93	-6	-10	4	11	3	4	-5	4	1	-2	1	0	-0	-0	-0	-0	0	-0	-0	-0
8	100	-5	-9	4	8	11	7	-2	96	-4	1	16	5	-7	3	11	2	-1	-1	-2	-3	-1	2	-0	0	0	0	-0	0	0	-0	-0
9	100	11	33	9	2	11	12	8	16	-10	1	13	-9	3	3	86	0	6	-11	10	9	-3	4	0	0	0	-0	0	-0	0	-0	-0
10	100	12	29	57	-15	35	12	13	10	-16	5	7	-12	5	4	19	10	9	-8	7	6	-5	51	0	0	0	0	-0	0	0	-0	-0
11	100	7	22	19	-3	87	8	7	14	-12	3	21	-7	-3	3	10	18	6	-9	5	4	-4	5	0	0	-0	0	0	-0	-0	-0	-0
12	100	2	-6	5	-6	7	94	4	8	-12	13	8	-6	4	20	9	3	1	-1	3	-1	-3	2	-0	0	0	-0	0	0	0	-0	0
13	100	-21	-20	-9	4	-6	-7	-7	6	12	-7	-6	91	-5	-10	-8	-3	-13	3	-7	-5	2	-2	-0	-0	-0	-0	-0	-0	0	-0	-0
14	100	-12	-21	-18	28	-16	-19	-11	-7	78	-14	-12	17	4	-10	-13	-9	-7	16	-10	-4	5	-4	1	0	-0	0	-0	0	-0	-0	0
15	100	-8	-49	-12	26	-18	-19	-10	-9	40	-10	-11	11	4	-10	-14	-7	-12	21	-15	-9	51	-5	-0	0	-0	0	0	-0	0	0	0
16	100	99	10	3	-1	2	0	2	-1	-3	2	1	-6	2	0	2	1	6	-1	2	2	0	1	-1	2	-1	-3	3	1	-1	-0	3
17	100	99	10	2	-1	1	1	2	-0	-0	1	1	-1	2	-0	2	1	4	-2	2	2	-2	1	-1	-5	-2	2	-3	1	-1	3	-0
18	100	98	14	2	-1	2	0	2	-2	-1	2	1	-3	1	1	3	0	6	-1	2	2	-0	1	-1	-1	1	10	-3	-1	-1	-0	-0
19	100	97	9	2	1	1	-3	2	-1	1	-2	-0	-6	1	-1	1	-0	6	0	2	1	1	1	1	22	0	-0	-0	-0	-0	-0	-0
20	100	97	13	4	-3	3	3	3	-1	-9	4	2	-7	1	1	3	1	6	-1	3	3	1	1	-2	-8	-1	-2	10	-0	-0	-0	-0
21	100	38	27	3	-6	7	2	6	-2	-7	8	5	-16	-3	0	7	3	84	-7	6	5	-3	2	0	0	0	0	0	-0	-0	-0	-0
22	100	99	5	2	-0	1	1	1	0	1	-1	0	-0	0	-0	1	1	-1	-2	1	1	-2	0	-0	-6	-5	-6	-4	-1	-3	-2	-2
23	100	0	-16	1	-3	3	19	1	3	-6	11	4	-9	3	95	2	2	0	4	1	0	-2	1	-0	0	-0	0	0	0	0	0	0
24	100	3	-10	5	-6	2	12	3	1	-8	97	0	-6	2	11	1	3	6	-3	2	2	-2	1	-0	0	0	-0	0	-0	-0	-0	-0
25	100	95	12	5	-4	4	2	4	-1	-9	4	3	-10	2	-0	7	1	14	-6	4	4	-2	1	0	2	13	1	-0	0	-0	-0	-0
26	100	99	10	2	2	1	0	2	-1	1	-1	0	-2	1	-1	1	0	2	-1	1	1	-1	0	5	-3	-4	-0	-1	-0	6	-0	-0
27	100	-2	-6	-8	96	-3	-6	-7	9	15	-6	-2	3	-13	-3	2	-2	-4	11	-2	0	3	-2	1	0	-0	0	-0	-0	0	-0	-0
28	100	27	88	6	-13	9	-4	7	-4	-14	-6	8	-10	8	-9	11	1	9	-10	8	3	-2	2	-13	0	1	-0	0	7	-0	0	0
29	100	6	5	-0	-11	-2	4	3	-7	2	2	-8	-3	98	3	2	2	-2	-2	1	1	0	1	0	0	0	-0	0	0	0	0	0
30	100	22	89	7	-13	9	-4	8	-4	-14	-7	8	-8	-4	-10	11	1	10	-11	8	4	-2	3	-12	0	-1	0	0	-6	0	-0	-0
31	100	23	96	4	21	5	-3	6	-5	14	-7	3	-6	4	-6	8	-1	6	-6	4	5	-1	1	33	0	0	-0	-0	0	0	0	0

Appendix G. 3. ROTATED PRINCIPAL COMPONENTS FACTOR WEIGHT MATRIX: $QM^{-1}T_r$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-2	-6	-8	4	3	-2	96	2	1	-1	-4	1	-2	0	-2	-19	-1	5	-17	-8	-2	-2	-3	-0	0	1	-0	0	-0	-0	0
2	-2	-8	-3	-1	1	-1	-11	3	6	-1	-2	4	-1	-1	-6	-7	-2	3	95	-12	8	1	3	-1	-1	-0	0	1	0	-0	0
3	-4	-1	-3	-3	-2	4	-4	7	-1	-2	1	3	-1	-0	-10	-2	-4	6	-10	93	7	-1	-3	2	-2	-1	-1	0	0	0	0
4	-3	-0	95	3	-5	-0	-9	-2	5	-2	2	3	3	1	-1	-29	-2	7	-4	-5	-6	-47	-4	-0	-0	0	0	-0	-0	0	0
5	2	6	-26	-2	-17	-0	-20	-1	2	-1	1	-1	-3	0	6	92	-0	-5	-11	-5	1	13	1	1	1	0	0	1	0	-0	-0
6	-1	11	6	-9	4	-2	5	2	-10	4	2	3	2	-6	6	-5	2	95	3	7	-9	-4	3	-1	3	-2	-3	-2	0	0	-0
7	1	-7	2	2	-16	-4	-4	-14	3	1	97	4	9	-3	-6	1	-1	2	-2	2	-2	4	2	1	-1	-0	1	-1	-0	-0	0
8	1	10	-2	-10	-9	-4	2	96	5	1	-14	-7	4	-1	-15	-0	1	3	5	12	7	-4	1	-0	1	0	0	-1	0	-0	0
9	-2	-11	-1	-5	-2	-8	-2	-12	5	1	-6	3	-3	-2	96	5	-1	7	-8	-18	-2	-10	2	1	-3	-0	0	1	-0	0	-0
10	-1	-3	5	3	-6	-2	-3	-3	3	-0	1	2	-1	-0	-3	-3	-2	-0	-0	-2	2	81	-1	-1	-0	0	1	1	-0	-0	-0
11	-1	-9	-4	-1	96	-3	3	-8	6	-1	-15	1	1	-1	-2	-15	-1	4	3	-1	5	-27	-1	-0	-1	0	-0	-0	-0	0	-0
12	-0	7	-1	1	-3	97	-2	-3	11	-9	-4	3	-4	-15	-9	0	1	-2	-2	6	9	-4	-6	3	0	-0	-0	-0	1	-0	0
13	6	7	3	1	1	2	1	-6	-12	3	4	98	4	8	3	-1	12	4	5	4	3	3	2	3	4	-2	2	1	0	-1	0
14	3	6	4	-15	6	9	1	3	94	6	3	-10	-6	4	4	2	-0	-11	6	-3	-35	3	-26	-3	7	-1	8	-0	2	-0	-0
15	1	1	-1	-4	3	3	0	2	-4	1	1	-1	-2	2	1	0	3	-5	5	5	89	4	1	-3	1	-3	-5	-1	0	0	-0
16	35	-4	-1	-0	0	-0	0	1	-0	-0	0	1	-1	-1	0	0	-2	-1	0	-0	-1	1	0	-14	-10	-8	-18	-2	-18	-18	86
17	36	-4	-1	0	-0	0	-0	0	1	-0	0	2	-1	-0	0	0	-2	-2	0	-0	-1	1	-1	-13	-10	-11	-19	-2	-17	81	-5
18	34	-3	-1	-1	0	0	-1	-0	-0	-0	0	2	-0	-0	0	0	-5	-1	0	-0	-0	-0	1	-12	-20	87	-2	7	-23	-40	4
19	33	-3	-1	-0	-0	0	-0	0	0	-0	0	2	-0	-0	-0	1	-4	-1	0	0	-1	0	-1	93	-17	-3	21	2	-3	9	-29
20	34	-3	-0	-0	-0	-0	-0	0	1	-1	0	2	-0	0	[illegible]	0	-4	-1	0	0	-1	0	1	-10	-9	8	85	2	-13	7	-37
21	-13	-9	-1	2	-2	1	-1	1	0	-6	-1	10	4	[illegible]	[illegible]	0	98	2	-1	-4	5	-2	2	-1	-10	-1	-2	2	-0	-2	-2
22	38	-2	-0	-0	-0	0	-1	0	1	-0	0	0	[illegible]	[illegible]	0	0	-5	-1	-0	-0	-1	-0	-1	-16	-13	-34	-31	-0	-27	-35	-17
23	-0	13	1	2	-1	-16	0	-1	4	-7	-3	8	-3	97	-2	0	0	-7	-2	-3	6	-1	-5	0	2	-2	-1	-0	1	0	-0
24	-1	13	-1	2	-0	-9	-1	0	7	98	1	4	[illegible]	-6	1	-1	-8	6	-2	-7	3	0	-3	2	-0	-2	-2	-1	1	0	-1
25	31	-5	-1	1	0	-1	1	-0	1	-1	0	1	-1	0	-0	1	-3	-1	-0	0	-1	0	0	-15	93	-28	-15	-12	14	-0	-2
26	36	-4	-0	-0	0	-0	0	-0	1	-0	1	1	-1	-0	0	0	-3	-1	0	0	-1	0	-0	-12	-5	-10	-15	2	87	-3	-0
27	-1	-1	3	97	-1	2	5	-10	-21	2	2	2	13	2	-6	-2	3	-14	-2	-7	-11	8	-22	-0	1	-0	-1	-1	2	-0	-0
28	-7	57	-1	1	-4	4	-2	4	3	7	-3	3	-1	7	-7	4	-6	11	-9	-12	10	-3	-38	1	1	-2	-4	68	7	-10	-7
29	-3	-3	3	12	1	-4	-2	4	-8	-2	9	4	98	-2	-4	-3	5	3	-1	-1	-5	-4	-2	0	-1	0	-0	-9	0	0	0
30	-8	60	0	-1	-5	4	-3	5	4	6	-4	3	-2	7	-7	4	-6	11	-10	-13	11	-3	-40	1	7	-8	-5	-71	6	9	7
31	-7	44	-1	-2	-4	3	-3	4	-1	5	-3	3	-2	5	-6	4	-5	10	-8	-10	10	-3	75	-1	1	0	4	1	-12	1	-0

APPENDIX H

An image analysis was performed on the correlation matrix R given in Appendix E.\2. In image analysis,each original variable is conceptually partitioned into two parts: the image variable, which is the original variable as predicted in linear regression by the other original variables; and the anti-image variable, which is the regression residual. The analysis then proceeds on the basis of the image variables.

In Appendix H. 1 are presented the variances of the anti-image variables. These are considered to be the diagonal entries in the diagonal matrix S^2 , and each is the proportion of the variance of an original variable which is not predictable from the other original variables. In Appendix H. 2 is given the matrix transformation, $I - R^{-1} S^2$, from original to image variables. Each column corresponds to an image variable and gives the linear equation for computing the image variable from the original variables. The columns in the appendix have been normalized. In Appendix H. 3 is given the correlation matrix, $SR^{-1} S$, of the anti-image variables.

The anti-image variables are approximations to the unique factors in pure factor analysis, and so they are expected to be essentially uncorrelated.

This appendix is further explained in Section II. C.

Appendix H. 1. VARIANCES OF ANTI-IMAGE VARIABLES: S^2

1	0.477839
2	0.351161
3	0.292145
4	0.294976
5	0.377501
6	0.368327
7	0.629175
8	0.620136
9	0.456239
10	0.234209
11	0.401067
12	0.600189
13	0.523727
14	0.190296
15	0.187413
16	0.000828
17	0.001197
18	0.003468
19	0.006127
20	0.003283
21	0.308779
22	0.004168
23	0.636060
24	0.667299
25	0.014584
26	0.005915
27	0.431822
28	0.012577
29	0.462151
30	0.012060
31	0.111527

Appendix H. 2. ORIGINAL TO IMAGE VARIABLE TRANSFORMATION MATRIX: $W = I - R^{-1}S^2$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	0	32	4	1	27	-2	8	-0	-1	3	-9	1	0	-1	2	-0	0	-0	0	0	-0	-0	-0	0	0	0	-4	-0	1	0	1
2	26	0	9	3	13	-0	5	-8	6	-2	-6	1	-1	-1	-5	-0	0	-0	0	0	-0	-0	1	0	1	-1	4	-0	1	1	0
3	12	31	0	3	8	-4	-5	-25	21	3	1	-6	-1	4	-5	-1	-0	0	0	1	1	0	1	2	1	-1	5	0	0	-0	2
4	2	7	2	0	51	-5	1	-0	-8	46	-29	-2	0	-2	5	-0	-0	0	0	0	0	0	-0	0	0	0	1	0	-3	-1	0
5	31	19	3	32	0	4	-6	3	-6	-12	36	0	0	-1	-0	0	0	-0	-0	-0	0	-0	-1	0	-1	-0	1	-0	2	0	-1
6	-7	-1	-5	-9	13	0	-1	-4	-4	6	-8	3	-1	4	3	0	-1	1	-0	0	0	-0	3	-3	-5	-1	7	1	-2	-1	-1
7	7	5	-1	0	-4	-0	0	17	5	-5	22	2	-1	-1	1	-0	0	-0	0	-0	0	0	2	-0	1	0	-1	0	-3	-0	-0
8	-0	-8	-8	-0	2	-1	19	0	14	4	6	-0	1	-0	-4	-0	0	-0	0	0	0	0	-0	0	-1	-0	7	0	-1	-0	-0
9	-1	11	12	-6	-7	-2	9	24	0	11	-5	5	0	-1	2	0	-0	-0	-0	-0	-1	-0	2	-1	3	1	6	-0	2	1	-1
10	6	-4	2	51	-22	3	-14	10	15	0	57	3	-2	-2	-2	0	0	-0	-0	-0	0	-0	0	0	0	0	-8	-1	3	1	0
11	-9	-8	0	-17	33	-3	30	8	-4	29	0	-0	1	-0	-2	0	-0	0	-0	-0	-0	0	0	-0	1	0	4	0	-1	-0	0
12	2	3	-6	-2	0	2	5	-1	9	4	-0	0	-1	-3	-5	-1	0	0	0	0	0	0	5	2	-1	-1	-1	-0	1	-0	2
13	0	-7	-2	1	0	-2	-6	8	1	-4	3	-1	0	2	-1	-1	1	-0	0	0	-1	-0	-5	-0	-3	-0	-1	-1	0	1	-0
14	-3	-4	6	-6	-5	6	-4	-3	-2	-5	-1	-6	2	0	22	0	1	0	-0	-1	2	1	-4	-1	-5	-3	-0	-1	2	-0	7
15	7	-26	-9	11	-2	4	7	-19	6	-5	-9	-8	-1	20	0	1	-1	1	-0	0	-1	-0	-9	-3	-4	-2	10	0	2	-1	-1
16	-14	-21	-69	-48	36	39	-64	-48	39	20	34	-76	-53	29	46	0	40	-32	86	92	57	32	33	77	54	25	33	43	-20	-44	1
17	59	32	-27	-36	30	-53	27	40	-17	53	-44	24	75	49	-56	32	0	85	-24	-15	48	82	-13	-20	10	57	41	59	-26	-52	-48
18	-50	-15	19	14	-10	54	-9	-19	-8	-31	6	1	-16	10	47	-20	67	0	14	8	-26	-39	57	37	62	38	1	-27	17	34	-3
19	11	16	18	14	-12	-4	3	24	-30	-3	-10	22	16	-6	-4	55	-19	14	0	-34	-19	-7	-1	-26	9	13	1	-12	5	14	-10
20	13	7	48	20	-25	16	-1	13	-36	-21	-7	44	14	-60	23	71	-14	10	-41	0	-23	-12	16	-10	17	40	16	-16	11	25	-27
21	-2	-4	3	2	1	1	1	2	-6	0	-2	1	-2	6	-3	2	3	-2	-1	-1	0	-4	1	0	7	-1	-1	-3	1	4	1
22	-22	-5	20	26	-4	-3	3	16	-10	-29	6	15	-16	23	-12	17	53	-32	-6	-8	-44	0	4	-12	-11	31	-12	-36	16	35	-5
23	-1	3	1	-1	-2	2	5	-0	2	0	1	5	-2	-2	-5	0	-0	1	-0	0	0	0	0	-0	-4	-3	-3	-0	1	-1	2
24	0	0	3	0	0	-3	-2	0	-3	0	-1	3	-0	-1	-3	1	-0	1	-1	-0	0	-0	-0	0	-2	-3	-2	-0	-1	-1	1
25	5	26	9	4	-20	-37	22	-17	51	4	15	-6	-19	-30	-25	5	1	9	1	2	14	-2	-43	-13	0	-29	-22	2	-5	-9	12
26	1	-37	-16	8	-2	-9	14	-8	14	5	4	-22	-4	-24	-14	3	9	8	3	7	-3	8	-52	-31	-43	0	-59	-11	-2	-4	79
27	-9	11	4	1	2	4	-3	19	9	-8	9	-1	-1	-0	5	0	0	0	0	0	-0	-0	-3	-1	-2	-3	0	-0	-5	-1	7
28	-23	-18	7	12	-26	14	42	24	-20	-19	10	-3	-12	-10	2	10	16	-10	-4	-5	-18	-15	-9	-1	5	-19	-6	0	63	44	15
29	3	6	1	-7	8	-3	-16	-5	5	6	-4	2	0	2	2	-0	-1	0	0	0	0	0	1	-1	-1	-0	-10	5	0	-5	1
30	17	45	-5	-21	18	-24	-23	-36	36	26	-6	-13	10	-4	-19	-10	-14	12	5	7	19	15	-17	-13	-23	-8	-29	45	-65	0	12
31	9	1	10	3	-8	-3	-4	-6	-5	2	3	9	-1	19	-3	0	-2	-0	-1	-1	1	-0	10	5	6	25	38	3	3	2	0

Appendix H. 3. CORRELATIONS OF ANTI-IMAGE VARIABLES: SR $^{-1}$S

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	-26	-10	-2	-33	7	-9	0	1	-5	10	-3	-0	2	-5	1	-3	5	-1	-1	1	2	1	-0	-1	-0	10	4	-3	-3	-5
2	-26	100	-20	-4	-14	1	-4	8	-9	2	6	-3	6	2	13	1	-1	1	-1	-1	2	0	-3	-0	-4	3	-8	2	-5	-6	-1
3	-10	-20	100	-4	-7	11	4	23	-28	-4	-1	16	5	-10	14	7	3	-4	-5	-10	-5	-5	-3	-9	-4	4	-10	-3	-2	2	-12
4	-2	-4	-4	100	-49	13	-1	0	11	-61	26	4	-2	6	-12	3	3	-2	-3	-3	-2	-4	1	-0	-1	-1	-2	-3	12	6	-3
5	-33	-14	-7	-49	100	-14	5	-3	9	19	-36	-0	-1	4	1	-2	-2	1	2	3	-1	0	2	-0	4	0	-3	5	-10	-4	5
6	7	1	11	13	-14	100	1	4	6	-9	8	-7	8	-13	-8	-6	10	-17	2	-5	-4	1	-8	14	24	4	-14	-8	12	14	6
7	-9	-4	4	-1	5	1	100	-23	-9	11	-29	-6	7	2	-4	3	-1	1	-0	0	-1	-0	-6	2	-4	-2	3	-7	17	4	2
8	0	8	23	0	-3	4	-23	100	-28	-8	-8	2	-10	2	14	2	-2	2	-3	-1	-2	-2	1	-0	3	1	-21	-5	5	7	4
9	1	-9	-28	11	9	6	-9	-28	100	-18	6	-16	-1	2	-6	-3	1	1	6	5	8	2	-5	5	-15	-3	-14	6	-9	-10	4
10	-5	2	-4	-61	19	-9	11	-8	-18	100	-46	-7	7	5	5	-1	-5	4	1	3	-0	5	-1	-0	-1	-1	14	5	-11	-7	-2
11	10	6	-1	26	-36	8	-29	-8	6	-46	100	0	-3	1	7	-2	2	-1	1	1	1	-1	-1	1	-3	-0	-9	-2	4	1	-1
12	-3	-3	16	4	-0	-7	-6	2	-16	-7	0	100	4	12	17	10	-4	-0	-8	-12	-2	-4	-19	-10	3	8	2	2	-7	7	-14
13	-0	6	5	-2	-1	8	7	-10	-1	7	-3	4	100	-6	2	13	-22	8	-11	-7	7	9	16	0	19	3	3	12	-2	-9	2
14	2	2	-10	6	4	-13	2	2	2	5	1	12	-6	100	-44	-4	-9	-3	2	18	-16	-7	7	5	19	9	0	6	-7	2	-32
15	-5	13	14	-12	1	-8	-4	14	-6	5	7	17	2	-44	100	-6	9	-13	2	-6	8	4	18	11	14	5	-16	-1	-6	10	5
16	1	1	7	3	-2	-6	3	2	-3	-1	-2	10	13	-4	-6	100	-24	25	-92	-87	-29	-24	-4	-18	-12	-5	-3	-23	5	24	-0
17	-3	-1	3	3	-2	10	-1	-2	1	-5	2	-4	-22	-9	9	-24	100	-82	31	17	-29	-72	2	5	-3	-15	-5	-38	7	33	17
18	5	1	-4	-2	1	-17	1	2	1	4	-1	-0	8	-3	-13	25	-82	100	-31	-15	27	58	-15	-18	-29	-17	-0	30	-8	-37	2
19	-1	-1	-5	-3	2	2	-0	-3	6	1	1	-8	-11	2	2	-92	31	-31	100	87	26	14	1	17	-6	-8	-0	17	-4	-21	8
20	-1	-1	-10	-3	3	-5	0	-1	5	3	1	-12	-7	18	-6	-87	17	-15	87	100	23	17	-4	5	-8	-17	-3	17	-5	-26	16
21	1	2	-5	-2	-1	-4	-1	-2	8	-0	1	-2	7	-16	8	-29	-29	27	26	23	100	50	-3	-2	-30	4	2	35	-6	-38	-4
22	2	0	-5	-4	0	1	-0	-2	2	5	-1	-4	9	-7	4	-24	-72	58	14	17	50	100	-1	6	6	-15	3	44	-9	-42	3
23	1	-3	-3	1	2	-8	-6	1	-5	-1	-1	-19	16	7	18	-4	2	-15	1	-4	-3	-1	100	1	24	18	9	5	-3	9	-15
24	-0	-0	-9	-0	-0	14	2	-0	5	-0	1	-10	0	5	11	-18	5	-18	17	5	-2	6	1	100	13	19	6	1	4	12	-12
25	-1	-4	-4	-1	4	24	-4	3	-15	-1	-3	3	19	19	14	-12	-3	-29	-6	-8	-30	6	24	13	100	26	10	-4	5	20	-15
26	-0	3	4	-1	0	4	-2	1	-3	-1	-0	8	3	9	5	-5	-15	-17	-8	-17	4	-15	18	19	26	100	16	16	1	6	-62
27	10	-8	-10	-2	-3	-14	3	-21	-14	14	-9	2	3	0	-16	-3	-5	-0	-0	-3	2	3	9	6	10	16	100	3	25	11	-46
28	4	2	-3	-3	5	-8	-7	-5	6	5	-2	2	12	6	-1	-23	-38	30	17	17	35	44	5	1	-4	16	3	100	-58	-92	-18
29	-3	-5	-2	12	-10	12	17	5	-9	-11	4	-7	-2	-7	-6	5	7	-8	-4	-5	-6	-9	-3	4	5	1	25	-58	100	59	-7
30	-3	-6	2	6	-4	14	4	7	-10	-7	1	7	-9	2	10	24	33	-37	-21	-26	-38	-42	9	12	20	6	11	-92	59	100	-13
31	-5	-1	-12	-3	5	6	2	4	4	-2	-1	-14	2	-32	5	-0	17	2	8	16	-4	3	-15	-12	-15	-62	-46	-18	-7	-13	100

APPENDIX I

The image covariance matrix, G, was factored according to the Harris factorization scheme. The matrix G is not given here, but it equals $R-2S^2+S^2F^{-1}S^2$.

The Harris factorization begins with the determination of the Harris roots of R, which equal the latent roots of $S^{-1}RS^{-1}$. They are given as Appendix I. 1, and are considered to be the diagonal entries of the diagonal matrix B_r^2. The corresponding Harris roots of G, which equal the latent roots of $S^{-1}GS^{-1}$, are $B_g^2 = (B_r^2-1)^2B_r^{-2}$ and are also given in the Appendix I. 1. The Harris vectors of R and G, which equal the latent vectors of $S^{-1}RS^{-1}$ and $S^{-1}GS^{-1}$, form the columns of the matrix X, which is presented in Appendix I. 2.' The Harris factors of G are given in Appendix I. 3, which is computed as $F_g = SXB_g$ and is called the unrotated image factor matrix. In the appendix the rows and column are bordered by row and column sums of squares. The row sums of squares are the image variable variances, and the column sums of squares are the image variances of the factors. The order of the factors follows the order of the Harris roots.

The appendix is further discussed in Section II. C.

Appendix I. 1. HARRIS ROOTS OF R AND G: B_r^2 and B_g^2

	B_r^2		B_g^2
1	3261.584970	1	3259.585277
2	156.135910	2	154.142315
3	18.912700	3	16.965574
4	15.477748	4	13.542357
5	8.479642	5	6.597571
6	6.500032	6	4.553877
7	4.949156	7	3.151210
8	3.825505	8	2.086908
9	3.161966	9	1.478225
10	2.602987	10	0.987161
11	2.335022	11	0.763283
12	2.009707	12	0.507292
13	1.752313	13	0.322987
14	1.589986	14	0.218922
15	1.321496	15	0.078214
16	1.267657	16	0.056514
17	1.221154	17	0.040052
18	1.104365	18	0.009863
19	0.953511	19	0.002267
20	0.891548	20	0.013193
21	0.845286	21	0.028317
22	0.802138	22	0.048806
23	0.762615	23	0.073892
24	0.738060	24	0.092964
25	0.650423	25	0.187884
26	0.594352	26	0.276857
27	0.538834	27	0.394692
28	0.432827	28	0.743218
29	0.416901	29	0.815550
30	0.312838	30	1.509379
31	0.263364	31	2.060393

Appendix I. 2. HARRIS VECTORS OF R AND G: X

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-0	-4	-11	2	15	-16	-7	-19	2	-21	-5	27	2	23	-5	16	-31	-19	12	-18	-1	-40	56	8	4	1	9	18	7	-2	-1
2	-1	-8	-12	1	14	-10	-9	-8	-9	-28	-7	35	4	-1	-6	8	-6	10	1	-40	37	-7	-59	0	-11	-3	-14	-1	2	-1	-2
3	-1	-10	-8	-0	13	-7	-22	-8	-16	-30	5	17	19	-22	-9	-39	1	17	-23	8	-6	30	27	-7	35	10	-34	-2	-1	-5	6
4	-0	-4	-19	2	29	-31	-10	-25	28	8	-1	-19	3	-22	22	24	10	-19	-15	15	13	23	-7	-8	3	-2	6	38	32	1	7
5	-0	-2	-16	3	23	-28	-9	-24	22	0	-4	27	-17	33	6	2	11	15	5	31	-37	11	-14	[illegible]	-10	-10	4	-34	-26	-0	-5
6	0	8	9	-4	-12	2	-12	-1	16	24	-17	23	-29	-17	-16	17	6	34	-0	-38	-17	20	21	-35	-23	16	-9	12	10	7	9
7	-0	-3	-7	3	8	-2	3	29	1	17	11	22	-1	26	-27	4	-12	-45	-21	-5	25	47	6	-9	-1	21	22	-14	-10	-1	2
8	0	1	-5	3	8	-6	6	34	-3	30	17	13	13	12	31	22	-15	-2	-29	-3	-23	-31	-13	-25	28	8	-36	-2	5	-1	4
9	-0	-6	-8	2	11	-1	-8	22	-24	7	13	19	40	-31	-2	8	-28	10	4	8	-34	14	-3	14	-38.	-22	25	4	15	5	-7
10	-1	-8	-23	3	30	-23	-2	-4	4	35	5	-33	3	-36	-10	-10	-9	6	7	-30	13	-18	11	1	1	6	5	-34	-35	-3	-7
11	-0	-5	-14	3	18	-15	-0	20	3	33	12	2	-10	29	-29	-50	11	15	19	3	3	-12	-7	12	-3	-7	-9	33	30	3	3
12	-0	0	-13	6	7	3	2	14	-24	11	-24	7	-24	-12	9	20	-17	-9	63	9	-0	27	-4	6	35	-9	-18	5	-5	-9	4
13	1	3	14	7	2	-3	11	1	14	4	2	-2	17	17	45	-19	-39	30	16	14	39	17	14	-19	-28	11	-9	4	-9	-12	-5
14	1	8	38	-8	-4	-7	-26	-15	2	7	27	8	-13	-13	-10	-19	-27	-12	7	5	1	-10	-18	-41	23	-31	25	17	-18	2	-0
15	1	12	24	-9	-15	-3	-22	-23	24	23	15	13	20	-10	-15	4	-7	-31	17	10	-1	-1	-5	33	-18	16	-50	-8	-2	4	3
16	-61	6	-23	-34	-30	-14	-1	-2	-4	4	11	4	-10	-1	9	-5	-3	-2	-4	-4	2	1	2	3	-5	-12	3	-9	9	-46	22
17	-50	6	24	37	29	16	12	-8	-4	4	19	8	-14	-2	10	4	4	5	-5	-7	10	5	10	16	2	-19	-4	-13	5	38	27
18	-30	-3	14	3	37	49	-17	-2	22	9	-31	6	13	-3	-5	-8	1	-4	1	4	-7	-10	-9	-5	9	16	12	-2	12	-37	-27
19	-22	4	9	-72	18	-8	23	6	1	-2	-17	4	6	-1	2	-3	-5	3	3	4	2	3	-1	1	9	10	1	3	-3	48	-20
20	-30	-5	-39	28	-45	3	-27	0	16	6	-14	1	17	2	-2	-5	-10	3	6	9	7	-5	-5	-11	14	7	8	-0	0	43	-24
21	-1	-5	-12	-7	6	18	-18	3	6	-17	47	-18	-47	1	2	12	-31	17	-9	1	-8	9	-7	25	2	25	-2	7	5	1	-31
22	-27	11	21	30	-4	-43	33	7	-12	-18	9	5	-5	-18	-11	-1	13	-9	12	8	-11	-7	-5	-14	-7	35	-1	8	2	-12	-37
23	0	2	-10	3	4	6	-2	8	-18	8	-30	16	-36	-25	-9	-4	-12	-6	-34	47	22	-26	3	3	-27	-4	-14	13	-20	2	3
24	-0	1	-13	3	3	10	-6	-3	-9	-5	-7	1	-16	-4	46	-44	4	-44	2	-31	-30	5	-7	-2	-24	11	-3	14	-14	13	2
25	-14	-4	-25	-10	17	27	-7	3	-9	-17	32	-13	16	8	-16	19	27	-6	25	13	3	-7	6	-40	-25	4	-18	19	-27	4	11
26	-23	4	20	8	4	-21	-22	11	-8	-6	-29	-38	11	29	-15	12	-16	8	-17	-14	-17	15	-7	16	-4	-9	-13	32	-35	-5	5
27	0	3	13	-5	-1	-11	-36	29	1	6	11	27	4	-8	30	8	41	16	7	2	18	-8	9	29	11	18	26	18	-29	-4	3
28	-6	-65	11	-2	-12	4	10	-23	-19	27	7	10	-1	7	8	8	17	-3	-9	-8	6	11	11	2	-2	-24	-14	13	-10	-6	-41
29	-0	-1	-1	-0	-0	4	1	-47	-50	28	0	-1	9	13	-1	3	-8	14	-2	9	-5	-3	-15	1	6	50	22	7	1	3	24
30	-6	-68	14	0	-6	-5	10	16	30	-16	-5	-1	-4	-12	-6	-4	-15	-1	11	9	-12	-9	-11	1	0	24	8	1	-5	4	44
31	-2	-17	23	-7	6	-19	-50	15	-30	-6	-10	-22	-12	12	9	3	7	-10	9	7	10	-8	7	-19	-17	5	-4	-37	38	9	-3

Appendix I. 3. UNROTATED IMAGE FACTOR MATRIX: $F_g = SXB_g$

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
		893	522	269	25	89	62	51	83	44	31	19	15	12	8	4	2	1	0	0	1	1	2	3	2	8	7	13	19	20	5	16
1	52	-16	-37	-32	4	27	-24	-9	-19	2	-14	-3	14	1	7	-1	3	-4	-1	0	-1	-0	-6	10	2	1	0	4	10	4	-2	-1
2	65	-24	-59	-30	1	22	-12	-10	-7	-7	-17	-3	15	1	-0	-1	1	-1	1	0	-3	4	-1	-10	0	-3	-1	-5	-1	1	-1	-2
3	71	-30	-65	-18	-1	19	-9	-21	-6	-10	-16	2	7	6	-6	-1	-5	0	1	-1	1	-1	4	4	-1	8	3	-12	-1	-1	-3	5
4	71	-14	-24	-42	4	40	-37	-9	-20	19	4	-1	-7	1	-6	3	3	1	-1	-0	1	1	3	-1	-1	1	-0	2	18	16	1	6
5	62	-4	-14	-40	6	37	-37	-10	-21	17	0	-2	12	-6	9	1	0	1	1	0	2	-4	1	-2	-1	-3	-3	2	-18	-14	-0	-5
6	63	15	63	22	-9	-18	3	-13	-1	12	15	-9	10	-10	-5	-3	3	1	2	-0	-3	-2	3	4	-6	-6	5	-4	6	5	5	8
7	37	-7	-29	-21	8	16	-4	4	33	1	13	8	12	-0	9	-6	1	-2	-4	-1	-0	3	8	1	-2	-0	9	11	-9	-7	-1	2
8	38	6	8	-15	10	17	-10	8	39	-3	23	11	7	6	4	7	4	-2	-0	-1	-0	-3	-5	-3	-6	10	3	-18	-1	3	-1	5
9	54	-19	-48	-22	4	19	-2	-9	22	-19	4	8	9	15	-10	-0	1	-4	1	0	1	-4	2	-1	3	-11	-8	11	2	9	4	-6
10	77	-21	-47	-46	5	37	-24	-1	-3	2	17	2	-11	1	-8	-1	-1	-1	0	0	-2	1	-2	1	0	0	1	1	-14	-15	-2	-5
11	60	-14	-38	-36	6	30	-21	-0	19	3	21	7	1	-4	9	-5	-8	1	1	1	0	0	-2	-1	2	-1	-2	-4	18	17	3	3
12	40	-3	3	-42	18	14	5	3	16	-23	8	-16	4	-10	-4	2	4	-3	-1	2	1	-0	5	-1	1	12	-4	-9	4	-3	-8	5
13	48	30	31	41	18	4	-4	15	1	13	3	1	-1	7	6	9	-3	-6	2	1	1	5	3	3	-4	-9	4	-4	3	-6	-11	-5
14	81	20	44	68	-13	-5	-7	-20	-10	1	3	10	3	-3	-3	-1	-2	-2	-1	0	0	0	-1	-2	-5	4	-7	7	7	-7	1	-0
15	81	19	66	43	-15	-17	-3	-17	-14	12	10	6	4	5	-2	-2	0	-1	-1	0	1	-0	-0	-1	4	-3	4	-14	-3	-1	2	2
16	100	-100	2	-3	-4	-2	-1	-0	-0	-0	0	0	0	-0	-0	0	-0	-0	-0	-0	-0	0	0	0	0	-0	-0	0	-0	0	-2	1
17	100	-100	3	3	5	3	1	1	-0	-0	0	1	0	-0	-0	0	0	0	0	-0	-0	0	0	0	0	0	-0	-0	-0	0	2	1
18	100	-99	-2	3	1	6	6	-2	-0	2	1	-2	0	0	-0	-0	-0	0	-0	0	0	-0	-0	-0	-0	0	1	0	-0	1	-3	-2
19	99	-97	4	3	-21	4	-1	3	1	0	-0	-1	0	0	-0	0	-0	-0	0	0	0	0	0	-0	0	0	0	0	0	-0	5	-2
20	100	-99	-3	-9	6	-7	0	-3	0	1	0	-1	0	1	0	-0	-0	-0	0	0	0	0	-0	-0	-0	0	0	0	-0	0	3	-2
21	69	-47	-35	-27	-15	8	21	-18	2	4	-9	23	-7	-15	0	0	2	-4	1	-0	0	-1	1	-1	4	0	7	-1	3	2	1	-25
22	100	-99	9	6	7	-1	-6	4	1	-1	-1	0	0	-0	-1	-0	-0	0	-0	0	0	-0	-0	-0	-0	-0	1	-0	0	0	-1	-3
23	36	1	21	-34	8	8	10	-3	9	-18	6	-21	9	-16	-9	-2	-1	-2	-0	-1	4	3	-4	1	1	-9	-1	-7	9	-14	2	4
24	33	-4	11	-42	8	6	18	-9	-3	-9	-4	-5	0	-7	-1	10	-9	1	-4	0	-3	-4	1	-1	-1	-8	5	-1	10	-11	13	2
25	99	-98	-5	-12	-4	5	7	-1	1	-1	-2	3	-1	1	0	-1	1	1	-0	0	0	0	-0	0	-1	-1	0	-1	2	-3	1	2
26	99	-99	4	6	2	1	-3	-3	1	-1	-0	-2	-2	0	1	-0	0	-0	0	-0	-0	-0	0	-0	0	-0	-0	-1	2	-2	-0	1
27	57	6	24	35	-12	-2	-16	-42	27	1	4	7	13	1	-2	5	1	5	1	0	0	2	-1	2	6	3	6	11	10	-17	-3	3
28	99	-40	-90	5	-1	-3	1	2	-4	-3	3	1	1	-0	0	0	0	0	-0	-0	-0	0	0	0	0	-0	-1	-1	1	-1	-1	-7
29	54	-8	-6	-4	-1	-1	6	1	-46	-41	19	0	-0	3	4	-0	0	-1	1	-0	1	-1	-0	-3	0	2	18	9	4	1	3	24
30	99	-36	-92	6	0	-2	-1	2	3	4	-2	-0	-0	-0	-1	-0	-0	-0	-0	0	0	-0	-0	-0	0	0	1	1	0	-1	1	7
31	89	-34	-71	32	-8	6	-14	-29	7	-12	-2	-3	-5	-2	2	1	0	0	-0	0	0	1	-1	1	-2	-3	1	-1	-11	11	4	-2

APPENDIX J

The unrotated image factor matrix was given in Appendix I. 3. It provided the basis for finding the normal varimax orthogonal factorization of the image covariance matrix G.

The normal varimax orthogonal rotation procedure was applied to F_g, given in Appendix I. 3, and the matrix $F_g T_g$ was derived, where T_g is an orthonormal matrix, and is presented in Appendix J. 1. The matrix $F_g T_g$ is the rotated image factor matrix, and is presented in Appendix J. 2. In the appendix, the rows and columns are bordered by row and column sums of squares. The row sums of squares are the variances of the image variables. The column sums of squares are the rotated image factor variances. Note that the factors have been arranged in order of decreasing image variance. The requirement for securing factor scores necessitated the computation of the rotated image factor weight matrix. It was computed as $S^{-1} XB_r^{-1} T_g$. In Appendix J. 3, each column gives the normalized weights for the original variables in computing the factor score.

This appendix is discussed further in Section II. C.

Appendix J. 1. NORMAL VARIMAX TRANSFORMATION MATRIX: T_g

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-99	-14	-4	-1	2	1	-2	-3	3	-1	-1	-0	-0	-1	-0	-0	0	1	0	-1	1	0	-0	0	-0	0	-0	0	-1	0	-0
2	14	-94	-8	19	13	4	-3	-3	3	-7	-7	7	-5	-10	2	1	-0	1	7	2	0	1	-2	4	1	-1	0	-0	0	-0	0
3	5	-1	-43	-58	36	-13	3	-18	39	-12	-0	-7	-2	4	15	-14	0	22	-3	-4	5	-2	3	-2	6	10	5	-8	-9	-1	1
4	0	5	3	19	-15	11	-2	-23	49	5	-1	3	-6	-2	4	16	-2	-1	-3	28	-68	-0	1	-6	16	3	1	1	-15	-4	2
5	-1	5	50	10	7	23	2	12	35	2	-0	16	1	-18	16	-8	15	33	17	-6	27	-2	-2	4	-9	36	11	-14	-18	-4	2
6	0	3	-56	26	-24	-15	5	40	6	9	-8	4	-4	10	-7	15	7	13	7	3	2	-5	9	-6	-28	43	4	0	-9	-6	2
7	3	-10	-15	-12	-67	16	0	-24	19	15	3	-7	0	-16	-28	-15	3	7	-15	-18	29	3	-12	-0	19	-8	5	-12	-2	-6	-0
8	1	14	-35	16	26	54	-55	-7	-8	26	11	9	0	-12	11	-1	9	-8	1	-3	6	1	-10	2	1	-4	7	-1	-0	5	-2
9	2	-19	29	-33	-10	-10	-56	7	3	13	10	-26	5	32	-20	18	16	-8	-6	-5	-9	-8	11	-8	-11	23	11	9	-1	-1	-0
10	1	-11	-0	-1	5	43	35	1	1	14	28	4	43	39	-0	-14	-15	2	-28	6	0	-8	17	-14	-7	13	-14	9	4	-4	3
11	-1	-5	-0	-36	5	28	11	52	-4	6	-3	15	-2	-18	-29	17	-6	17	14	-11	-30	-7	-2	-5	2	-30	8	-17	8	-19	2
12	-0	5	7	10	19	8	2	-14	-0	20	-23	12	-56	39	-33	-10	-31	9	0	-10	7	3	8	10	16	14	-3	2	6	-21	1
13	1	2	-2	-26	-3	15	8	-26	2	-8	-18	44	9	-13	-27	28	6	-30	21	1	9	12	21	9	-21	16	-22	20	-11	18	-6
14	1	-5	6	-15	-6	11	10	7	8	26	16	-30	-42	-27	24	17	-39	-9	-15	4	8	9	9	-3	-38	-3	-14	7	-8	18	4
15	0	4	2	5	8	12	1	5	24	-27	-24	-4	-6	-1	4	40	10	-22	-38	15	31	-14	-14	-18	-1	-5	3	7	27	-36	12
16	0	-3	4	-4	-0	9	2	8	-11	-2	-38	7	-18	8	2	-34	38	-13	-43	1	-14	1	10	-8	-24	-13	-6	-25	-37	9	3
17	0	-0	1	-2	7	-4	-2	-12	-18	-2	13	-11	5	-10	-5	11	-10	-19	5	-2	-4	-29	-24	-0	4	28	-47	-51	-22	-28	-9
18	0	0	0	-2	1	1	2	5	6	-11	5	5	5	13	-13	-13	-25	-21	11	1	3	-16	-54	5	-19	-10	27	29	-47	-4	24
19	0	-0	0	0	0	-2	-0	-0	2	-1	5	1	4	3	-1	2	-10	-2	-6	22	10	8	5	10	-13	-13	26	-1	-18	-27	-84
20	-0	0	1	4	1	-1	1	2	4	-1	-1	2	-16	-18	2	-24	5	-8	11	-28	-6	-43	6	-57	8	4	-13	39	2	1	-29
21	0	1	-1	3	3	-7	-1	1	12	14	10	-12	4	-13	3	-21	24	-7	7	-13	-7	5	17	40	-18	-13	-27	33	-0	-58	14
22	-0	-2	-0	-1	-2	-12	-1	2	9	31	4	8	-5	17	2	6	14	5	28	49	30	-38	18	-7	7	-39	-14	-8	-14	6	12
23	0	-2	2	1	1	-7	-4	-1	6	9	0	3	5	12	7	-6	-2	-12	28	0	7	63	-7	-58	-7	-10	-9	-14	-6	-26	8
24	-1	1	-1	3	9	-11	-1	5	-6	-2	18	7	2	-31	-10	-13	-17	-37	-9	8	6	-4	50	-7	22	15	42	-14	-14	-16	24
25	-0	1	-1	-10	4	22	6	2	-30	-12	-14	-40	-4	-21	-21	-11	11	22	7	51	2	15	-5	-8	19	23	-6	28	-13	-6	4
26	-0	-3	-2	-3	9	10	30	26	15	26	-9	-27	-2	13	10	11	26	-45	14	-16	5	11	-10	17	42	9	5	2	-10	16	-15
27	-1	-4	6	-10	16	-33	12	-6	-10	55	10	13	3	-22	-8	4	17	6	-33	14	-5	6	-35	-8	4	16	7	12	10	-3	-1
28	0	3	-2	18	20	-4	5	8	17	-27	59	-1	-25	2	-44	7	28	4	-13	-3	7	15	-2	-3	-1	-12	-13	-2	-10	19	-4
29	-1	-6	-2	-26	-29	14	2	6	-20	-21	34	32	-40	18	38	-9	18	-12	9	18	-5	1	-9	7	10	15	5	3	13	-14	1
30	0	-2	-2	2	-3	-2	2	-8	-25	-1	4	8	-1	7	22	46	3	30	-17	-30	5	5	15	-6	28	-11	-1	19	-52	-14	7
31	-0	0	-2	4	7	11	33	-44	-19	9	6	-25	-11	4	-5	17	30	3	23	-12	-9	-15	1	-10	-32	-1	42	-13	10	-10	6

Appendix J. 2. ROTATED IMAGE FACTOR MATRIX: SXB_gT_g

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
		813	572	236	137	110	77	57	40	32	29	21	18	16	13	11	10	8	7	6	5	4	3	3	2	1	1	0	0	0	0	0
1	52	7	38	54	8	-7	-5	5	3	-3	-1	3	3	-17	-3	-4	-1	0	-3	2	-1	0	16	-1	-1	0	-0	-0	0	-0	0	-0
2	65	13	65	39	10	-9	-2	2	5	-7	-0	-5	4	-11	-0	-1	-0	-2	-1	4	-3	1	-2	2	14	0	-0	0	-0	0	0	-0
3	71	18	72	29	2	1	-1	7	7	-9	-7	-6	-2	-4	1	-0	2	1	-0	21	4	-0	2	2	1	-0	0	-0	-0	-0	0	0
4	71	8	20	73	6	-10	8	3	3	-5	-9	20	5	1	2	-2	3	21	-2	1	4	0	-2	-1	-1	-0	-0	0	-0	-0	0	0
5	62	-1	10	74	10	-4	3	-3	-1	-1	11	-13	-4	3	-1	3	1	-15	2	-1	-1	-0	-6	1	0	0	-0	0	0	-0	0	0
6	63	-4	-66	-20	2	25	-5	-1	-7	-2	-6	4	-5	-5	27	5	-2	2	1	0	-3	-0	-1	1	-0	0	0	0	0	0	-0	0
7	37	2	31	12	9	-0	30	-8	3	-2	39	2	4	2	-1	-0	0	-1	-1	-0	0	-0	-0	0	0	-0	0	-0	0	-0	0	-0
8	38	-5	-4	4	11	3	59	-9	-2	2	4	1	1	1	-0	-0	-0	0	-0	-0	-0	-0	-0	-0	0	0	-0	0	-0	0	-0	-0
9	54	10	56	13	11	0	18	-0	5	-7	9	5	38	1	-1	0	2	1	-0	-0	-0	0	0	-0	0	-0	0	0	-0	0	0	0
10	77	11	46	57	15	-15	18	3	8	-5	13	-2	3	32	-5	1	-3	-0	-1	-1	-1	0	-0	-0	-1	0	0	0	0	-0	0	-0
11	60	6	38	39	11	-9	36	-2	6	-7	7	36	7	-1	1	1	0	1	-0	-1	-1	-0	1	-0	-0	-0	-0	0	0	-0	C	0
12	40	1	7	9	52	-12	24	3	-3	-4	3	-1	-0	-1	-5	0	-9	2	-1	2	17	-1	-0	-1	-1	0	0	-0	0	-0	0	-0
13	48	-22	-36	-15	-21	8	4	-7	-13	45	-2	-2	-3	-1	-0	-0	-2	-0	1	-0	-0	-0	-0	0	-0	0	0	0	0	-0	-0	0
14	81	-11	-46	-25	-35	50	-16	6	-6	16	-15	-2	-2	-1	2	4	-7	-3	25	-0	-1	1	-2	-0	-0	-0	0	-0	-0	0	-0	-0
15	81	-8	-70	-21	-23	32	-9	4	-3	5	-19	-7	-5	1	14	3	-2	-7	-0	7	-6	-0	-4	15	2	-0	0	-0	0	0	-0	-0
16	100	99	12	4	1	-2	-1	2	3	-7	1	1	0	0	1	-1	-1	-0	-2	0	0	0	-0	-0	-0	-1	-2	0	-0	3	-0	0
17	100	99	12	3	0	-2	-0	3	1	1	1	0	1	-1	0	1	1	0	2	0	1	-3	-1	0	-1	1	1	2	-1	-1	-2	1
18	100	98	16	3	0	-1	-2	2	6	2	1	0	1	0	2	1	-1	1	1	1	2	0	-0	0	-0	-2	7	0	-0	-0	-0	-0
19	99	97	9	4	-4	1	-2	1	6	-11	-0	1	1	1	-0	1	-3	1	2	-1	-7	16	1	-0	1	-0	-0	-0	0	0	-0	-0
20	100	97	17	6	[illegible]	-6	-2	1	3	-7	2	1	1	0	3	-1	5	-1	-5	-2	3	-8	1	1	-1	2	-2	-2	4	0	0	-0
21	69	40	41	11	8	-7	-6	-9	55	-14	3	2	3	1	-1	0	4	0	-0	0	-0	0	0	-0	0	-0	-0	-0	-0	0	0	0
22	100	99	6	3	-1	-1	0	1	-2	4	-0	0	1	-1	-2	1	-2	-1	-1	-1	2	-5	1	-2	1	6	-3	-1	-1	-0	0	-0
23	36	0	-11	4	58	-0	5	1	-3	-1	2	1	1	2	3	0	-6	-2	1	-1	-8	1	0	0	1	-0	-0	-0	-0	-0	0	0
24	33	3	-4	10	48	-7	-3	5	13	-4	1	0	2	-1	-1	-2	25	0	-1	0	-2	-0	-0	-0	-0	-0	0	0	0	-0	-0	-0
25	99	95	20	8	9	-6	0	3	11	-8	3	0	2	1	-1	-3	4	2	-0	3	-3	2	-0	0	-0	-6	1	0	-1	-0	0	-0
26	99	99	11	3	-1	4	-2	2	-1	1	-1	2	-1	1	-1	2	-0	0	-0	0	2	-2	0	1	-0	-0	-1	0	-0	-1	2	0
27	57	-1	-17	-11	-9	71	5	-9	-2	1	2	-1	0	-1	-0	0	-0	0	-3	-0	-0	-0	0	-0	-0	0	-0	-0	-0	-0	-0	-0
28	99	26	90	5	-20	-13	-7	6	6	-1	4	6	-5	6	10	-3	-4	-4	-0	-10	-1	0	-0	3	-4	1	-1	-4	0	0	-0	0
29	54	5	5	2	7	-7	-11	71	-2	-2	-3	-0	-0	-0	-0	0	1	0	0	0	0	0	0	-0	0	0	0	0	0	-0	0	0
30	99	22	91	5	-22	-11	-4	2	-1	-3	8	7	-10	4	10	-3	0	4	-1	-6	-4	-1	-2	2	-5	-1	1	4	-1	0	-0	0
31	89	23	75	2	-32	22	-5	4	1	-4	-5	3	1	3	7	31	-6	-1	2	0	0	0	-2	1	-0	0	0	0	0	-0	0	0

Appendix J. 3. ROTATED IMAGE FACTOR WEIGHTS: $S^{-1} XB_r^{-1} T_g$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-2	-0	21	0	2	-3	1	0	1	-2	3	1	-11	-0	-1	-1	0	-0	-2	-1	-1	18	1	-5	-1	1	1	-1	0	-1	-1
2	-2	9	8	3	0	0	-1	-0	0	-6	-5	0	-8	2	-1	-0	-1	1	-2	-1	-2	-4	2	24	-1	-1	0	1	-0	-1	2
3	-3	12	1	2	2	6	1	-1	-0	-10	-8	-7	-2	5	-3	0	-1	0	16	3	4	1	-1	-5	2	-1	-1	2	-0	0	3
4	-3	-7	36	-4	-1	-0	2	0	-0	-12	23	4	-13	[illegible]	-1	0	22	-0	-2	2	1	-5	1	-3	0	-1	-3	-0	0	-1	1
5	-0	-4	43	-1	1	-4	-2	-1	1	11	-21	-2	1	-0	2	0	-15	1	-2	-1	1	-10	-1	-7	-0	0	1	-0	-0	2	-2
6	2	-7	-0	3	1	1	1	-0	0	2	6	-0	-2	37	2	1	3	0	2	-1	1	3	-6	2	-1	-4	0	-4	-2	1	1
7	-0	-3	1	-1	3	0	1	0	0	27	-4	-1	-1	0	2	-0	4	1	2	1	1	0	4	1	1	-2	-2	-0	1	1	1
8	2	3	-6	-2	-1	24	1	1	-1	-13	-12	-5	-4	5	-0	1	1	1	-2	-1	-1	1	-1	-0	-0	2	-0	1	1	2	2
9	-3	4	-3	-0	1	-3	-1	-1	-0	1	3	21	-2	5	-1	0	-2	-0	-2	-0	1	-2	0	-4	0	-1	2	-0	0	2	-0
10	-0	-1	25	2	2	-0	-1	0	1	10	-23	-2	43	-3	0	-1	-6	0	0	-1	-3	7	-1	6	-0	0	1	-0	0	-1	1
11	-1	-3	2	-1	-1	9	1	0	-1	-7	41	1	-9	1	-0	0	-6	0	2	-1	-1	0	-0	3	0	1	1	1	-0	-1	-2
12	-0	6	-7	7	1	6	0	-0	-0	-5	-3	-5	-2	2	-1	-2	2	1	-0	6	5	-1	3	-1	-0	-3	1	0	-1	-1	-5
13	4	0	2	0	-0	0	-1	1	8	2	-1	0	1	6	-0	0	0	-3	2	-1	5	-0	-1	1	-2	-5	-0	4	1	-2	-1
14	2	1	6	-1	12	-6	0	-0	-1	-0	3	-0	4	-4	-8	1	2	12	-4	1	-6	1	-4	-0	-1	-2	-1	6	2	1	-3
15	4	-7	1	-3	-1	9	1	2	1	-14	-6	-4	4	15	1	-0	-5	-4	5	-5	-2	-3	18	2	2	-1	2	-3	-0	0	-4
16	40	-25	27	-48	-4	41	49	-34	-7	36	4	-41	-44	-14	-47	-63	7	-57	73	63	-47	-70	-45	-3	-73	1	50	-58	83	-12	19
17	58	-35	-50	-5	-49	58	55	-63	-62	13	-2	-1	-49	20	60	52	28	66	8	-40	-29	-34	62	-75	-4	-21	53	-12	-35	-81	70
18	20	-6	23	-24	-10	-29	-30	41	40	16	-17	29	6	19	-6	-33	-14	-30	-1	44	11	14	-30	40	4	88	-20	11	18	25	-26
19	29	-19	1	-5	-7	-12	-16	8	-19	-6	-3	18	13	6	21	20	-5	20	-30	-26	72	25	14	1	30	-7	-13	30	-25	-3	-1
20	16	-23	-2	2	-10	-36	-30	15	-31	4	-18	33	24	40	29	34	-15	18	-46	-12	-13	37	13	5	33	-3	-29	70	-27	-1	-13
21	-5	-2	0	-1	0	-3	-3	11	3	-1	2	1	-2	-3	2	-2	-2	-1	-3	1	-0	1	0	1	4	-3	2	1	-1	2	2
22	36	-0	27	-22	-11	-2	-31	39	44	-12	-23	24	3	-2	-7	-19	-17	-22	-9	-6	-1	26	-38	43	44	-5	-32	-11	2	22	-47
23	0	7	-5	14	4	-4	-1	-1	2	-3	3	-2	1	1	-0	-2	1	0	-0	-7	-3	-2	2	-4	0	-0	-1	2	1	-0	1
24	-0	5	-3	8	3	-3	1	-0	2	-2	3	-1	2	4	1	5	-1	1	-2	-1	5	1	1	1	2	-1	0	-3	1	1	-1
25	14	30	-25	33	17	11	8	2	10	-7	-14	-16	7	1	-17	10	20	7	15	-31	-15	-3	-7	-5	-19	-21	-13	-15	-6	5	-28
26	43	61	-1	70	76	-31	-4	-12	24	-38	69	-46	40	-72	-34	3	-1	-1	-10	6	23	6	29	-5	-18	-31	3	-12	-11	46	24
27	-0	8	2	5	21	-2	1	0	1	8	1	-3	3	-9	-8	1	2	-4	-3	-0	1	0	-2	-2	1	2	-0	-1	-1	-3	0
28	-3	36	-6	8	-8	-20	-23	22	20	-51	9	23	24	5	-16	-11	-59	4	-24	13	20	13	11	-3	3	-12	-32	-7	-2	-1	9
29	-4	-3	-2	0	4	2	15	-0	-0	9	2	-3	-4	-1	-0	1	6	-2	0	-1	-0	-1	-2	1	1	1	2	2	-0	2	-1
30	-5	25	-20	-1	19	15	26	-24	-12	46	5	-44	-17	16	-10	11	61	-2	15	-19	-14	-17	-4	-8	-3	5	32	-0	2	5	-11
31	-8	2	-3	-12	-8	6	-0	1	-5	[illegible]	-11	9	-5	7	29	0	-1	-2	3	-1	-5	-1	-3	2	-0	0	-1	1	1	-4	-3

APPENDIX K

In order to facilitate comparison of the rotated component and the rotated image factor structures, the matrix of crosscorrelations between the rotated image and the rotated component factor scores was computed.

In Appendix K. 1 is given the matrix of crosscorrelations. The columns correspond to the rotated image factors and the rows correspond to the rotated component factors. The crosscorrelation matrix was computed as $T_r{}'M^{-1}Q'RS^{-1}XB_r^{-1}T_g$.

This appendix is discussed further in Section II. C.

Appendix K. 1. CORRELATIONS BETWEEN PRINCIPAL COMPONENTS FACTORS AND IMAGE FACTORS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	100	0	1	1	-1	-0	0	-1	0	0	-0	1	-0	-1	-0	0	-0	-0	0	-0	0	0	-0	0	-1	0	-1	1	-0	-0	-1
2	0	88	-8	-22	6	-8	2	-6	7	-5	4	-18	7	21	16	-3	-0	3	-15	-2	2	-5	-1	-10	2	0	-3	0	7	-2	-3
3	-0	-0	68	-4	-2	3	3	4	1	-17	16	8	25	7	-0	-0	57	-4	-6	8	-1	-14	6	-3	4	-3	-17	-2	4	-8	8
4	0	-2	-1	0	86	6	-2	3	-6	4	-1	2	-4	-9	-2	4	-2	-31	-7	4	2	-6	-6	-1	4	9	-8	-6	-10	-33	0
5	-0	8	19	4	-2	33	0	4	-4	5	76	-1	11	-5	-0	1	-37	4	16	-7	-3	4	-9	13	-3	8	10	9	-4	-7	-15
6	-0	8	-3	45	-4	20	1	-6	-2	-3	-9	-8	-3	1	8	-25	9	10	-10	52	13	-2	11	-2	0	-18	16	5	-19	-17	-46
7	-0	10	38	3	4	-9	4	-1	1	-4	-7	2	-37	-2	-3	-6	-4	-3	-4	-9	-4	79	9	-9	-10	1	14	-2	5	-6	-4
8	0	0	-3	-1	1	84	-0	4	0	-19	-24	-8	-6	5	-2	7	1	4	-25	-8	-3	6	-11	-6	-2	8	-4	5	10	17	19
9	1	-15	0	-24	39	-5	8	2	9	-18	2	3	2	-0	7	-13	0	74	2	3	-8	-1	9	5	8	-3	8	8	27	11	-16
10	-0	3	1	41	1	-9	3	10	10	-6	3	-2	4	6	10	66	-9	5	-6	12	17	4	2	5	15	8	1	-22	45	-3	-3
11	-0	7	2	1	5	20	1	4	4	79	3	0	-12	2	6	-2	13	13	11	6	2	4	33	7	6	-11	-29	-6	12	7	11
12	-0	-11	-2	-13	-1	5	-3	-4	75	5	1	1	-3	8	7	13	3	-12	3	-2	18	-2	-3	-2	-14	-23	14	42	3	-23	3
13	-1	-1	1	3	-1	-3	96	11	4	0	-2	5	1	-2	2	-5	-11	-6	-8	3	3	-2	-2	3	4	-0	-11	10	-6	6	4
14	-0	2	-3	62	7	-8	-4	-7	8	-3	5	-1	-1	6	14	-38	4	4	-0	-46	-11	-10	1	-15	13	10	-7	22	18	-4	19
15	-1	22	-0	9	8	8	-1	-5	-1	4	-7	87	5	9	-17	5	-2	-2	9	-4	3	-7	3	-10	-2	-7	23	1	-1	18	-5
16	-1	-4	56	-1	2	-6	-6	-5	3	18	-27	-11	-19	5	5	2	-47	5	-9	-3	4	-42	-9	-23	-1	2	6	-0	-7	16	-10
17	1	8	-0	5	1	-4	-4	86	-9	5	-1	-8	-6	-3	1	-1	9	3	2	-3	-9	-9	-3	-7	-8	-19	35	5	-4	-2	14
18	1	-20	-2	4	11	-5	1	-1	-5	5	10	-3	2	86	9	-6	-2	1	-1	8	4	12	-23	8	-5	-10	4	-15	-14	11	13
19	-0	17	14	11	-0	-2	-3	-4	-0	-9	-15	2	-25	2	-12	-1	-2	7	-3	-8	-3	-15	3	87	-8	-2	4	6	-5	-8	13
20	-0	15	7	7	6	7	1	0	-4	-19	-18	-12	-6	4	-12	6	-4	6	81	19	8	5	-11	-9	19	-1	-15	19	-7	3	18
21	0	-9	-0	-11	1	9	1	10	-4	-23	-7	-8	7	28	3	-3	-15	-29	14	-16	-2	-8	77	3	6	5	7	0	8	2	-19
22	0	3	13	5	3	2	-0	-1	3	21	-36	-8	78	-6	1	-6	-18	4	4	-7	-6	24	-5	16	-2	2	18	-2	-1	-10	4
23	-0	-3	2	-2	5	4	-3	-0	-6	-10	-4	14	-1	-19	86	-0	1	-13	17	-0	1	3	-6	13	-20	-13	-5	-7	-3	21	0
24	-0	-3	1	-9	-0	1	-0	4	-27	2	-0	-2	3	-3	-2	-14	4	5	-2	-24	90	2	-3	3	-2	-2	4	7	10	-3	-1
25	-1	3	-4	12	1	4	6	12	2	-1	-11	-2	3	10	-12	16	10	15	17	-34	-5	-4	-8	-6	-60	1	-36	-9	-9	-11	-43
26	-0	0	-1	2	-2	-7	-4	13	10	7	-0	8	1	6	11	-3	6	9	-1	26	10	1	11	-1	-26	83	5	19	-12	4	12
27	-1	-3	1	-0	3	-11	-7	-4	-45	5	-1	2	7	9	0	20	-7	-8	-18	16	-13	6	0	-0	-15	-15	-16	72	21	4	-3
28	0	0	0	2	-10	-4	-12	18	7	-22	2	22	6	-2	-1	-19	-41	14	-15	24	8	4	16	-8	-16	-17	-45	-14	2	-30	35
29	0	6	-0	22	25	-10	-2	-13	12	-10	17	-22	13	-17	-21	9	1	1	-8	6	12	3	23	-4	-32	-19	4	-2	-23	56	24
30	0	-1	-2	1	-5	6	15	-29	-26	2	1	-2	-5	2	16	32	6	31	2	-13	-4	-9	22	-18	-13	-7	30	-1	-30	-42	35
31	0	-1	0	-4	1	4	13	-18	-9	4	2	-9	-4	-1	-11	-25	2	-17	19	20	-7	-16	-6	-4	-46	-2	28	-19	59	-9	15

APPENDIX L

This appendix allows comparisons across strata of different values for selected characteristics. The list of characteristics and the associated statewide means are given on the fold-out tab. The strata were formed by first separating each of five rotated image factor score distributions at the median, and then identifying all the districts which were members of each of the 32 unique combinations of above-the-median (+) and below-the-median (-) for the set of five rotated image factor scores. The procedure for forming strata is more fully described in Sectior. II. D.

The appendix is divided into four pages, and eight strata are described on each page. Each column gives the selected values for one of the strata. The pattern of (+) and (-) at the head of the column identifies the stratum. The next four numbers give the size of that stratum and the portion of the districts therein which fall into certain WSDPI classifications. The rest of the entries in the column are stratum averages, taken over all the districts in the stratum, for the selected characteristics.

This appendix is discussed and interpreted in Section III. B.

-	-
-	-
+	+
+	+
+	-
18	22
100.0	86.4
16.7	100.0
16.7	95.5
139.2	613.3
100.6	399.6
38.6	213.6
6.7	30.5
3.9	14.3
2.1	13.4
0.7	2.9
1.8	4.6
1.3	2.0
0.5	2.6
6,227.8	8,872.7
75.9	133.6
20.7	20.1
3.7	6.7
44.8	14.5
3,397.0	1,932.7

Appendix L

Summary Measures for Stratum Characterization

+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+
+	+	+	+	-	-	-	-
+	+	-	-	+	+	-	-
+	-	+	-	+	-	+	-
28	31	10	11	17	35	8	12
17.9	58.1	30.0	27.3	35.3	74.3	50.0	83.3
96.4	100.0	100.0	100.0	94.1	100.0	62.5	100.0
100.0	100.0	100.0	100.0	94.1	100.0	100.0	100.0
.945.1	1,938.1	6,320.9	4,154.0	2,331.1	1,679.4	6,771.1	5,164.0
.222.7	1,160.8	3,405.8	2,383.9	1,370.8	1,082.2	4,142.0	3,165.4
,721.8	777.3	2,915.1	770.1	960.3	597.3	2,629.1	1,998.6
175.9	87.3	291.6	198.5	107.9	78.1	323.4	223.1
80.6	41.5	126.0	94.3	51.0	40.3	172.3	110.4
79.0	38.3	139.7	88.0	46.3	31.1	125.8	94.3
16.3	7.5	25.9	16.2	10.6	6.7	25.4	18.3
12.6	8.0	11.7	9.3	12.3	6.3	11.8	8.8
3.0	2.1	0.4	0.0	5.9	1.1	0.3	0.1
9.6	5.9	11.3	9.3	6.4	5.2	11.5	8.7
.492.9	34,187.1	208,080.0	102,072.7	59,670.6	30,868.6	202,887.5	117,475.0
312.0	243.2	540.2	448.0	189.6	264.8	576.3	590.2
22.4	22.2	21.7	20.9	21.6	21.5	20.9	23.2
13.9	11.0	24.9	21.4	8.8	12.3	27.5	25.5
29.0	17.6	32.9	24.6	25.6	18.4	30.0	22.7
.055.9	4,290.7	17,784.6	11,007.8	4,853.6	4,866.7	17,267.0	13,425.7

+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-
+	+	+	+	-	-	-	-
+	+	-	-	+	+	-	-
+	-	+	-	+	-	+	-
7	20	11	7	45	21	21	21
100.0	85.0	90.9	100.0	95.6	95.2	100.0	100.
14.3	100.0	9.1	14.3	11.1	57.1	0.0	12.
14.3	95.0	9.1	0.0	8.9	52.4	0.0	6.
97.9	884.3	10,963.6	74.8	181.5	417.6	44.3	100.
48.9	564.9	6,770.4	53.6	118.4	279.7	44.3	81.
49.0	319.4	4,193.3	21.0	63.1	138.0	0.0	18.
5.3	41.8	432.5	4.3	8.3	20.1	2.9	5.
2.0	20.9	203.6	2.6	4.6	10.8	2.2	3.
2.7	17.1	181.7	1.3	3.0	7.8	0.0	1.
0.6	3.8	47.2	0.4	0.7	1.6	0.1	0.
2.1	7.0	14.5	1.3	2.6	3.9	1.0	1.
2.0	3.0	0.6	0.7	2.1	2.2	0.7	0.
0.1	4.0	13.9	0.6	0.5	1.7	0.3	0.
3,314.3	12,425.0	352,209.1	2,028.6	4,197.8	6,276.2	2,019.0	1,928.
45.7	127.2	753.8	58.0	70.4	108.3	44.3	80.
18.5	21.2	25.3	17.4	21.9	20.8	15.5	19.
2.5	6.0	29.7	3.3	3.2	5.2	2.9	4.
33.9	14.1	32.1	27.2	23.1	15.0	45.5	19.
1,546.7	1,787.8	24,214.4	1,577.8	1,628.4	1,627.2	2,019.0	1,542.

-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	+	+
+	+	+	+	-	-	-	-	-
+	+	-	-	+	+	-	-	-
+	-	+	-	+	-	+	-	-
9	25	31	31	5	14	30	19
5.6	60.0	93.5	71.0	60.0	85.7	90.0	63.2
7.8	100.0	9.7	100.0	20.0	100.0	16.7	94.7
0.0	100.0	87.1	100.0	100.0	100.0	90.0	94.7
3.8	851.3	577.6	824.0	833.4	697.2	620.3	820.6
2.6	554.3	596.2	528.2	732.6	441.8	488.3	510.3
1.2	297.0	78.2	295.8	100.8	255.4	132.0	310.3
0.0	40.8	28.1	39.7	44.2	33.3	29.4	40.1
9.7	19.3	18.9	18.8	31.6	15.3	19.2	17.7
6.3	17.3	4.0	17.3	5.8	14.3	6.4	18.1
4.0	4.2	5.2	3.6	6.8	3.7	3.8	4.4
4.2	3.8	1.8	2.6	2.4	2.9	1.5	2.3
1.8	1.0	0.0	0.0	0.2	0.6	0.0	0.0
2.4	2.8	1.0	2.6	2.2	2.3	1.5	2.3
22.2	14,528.0	24,145.2	14,845.2	45,740.0	14,757.1	23,640.0	16,473.7
02.2	224.0	325.6	319.3	347.3	238.1	442.9	354.3
21.3	20.9	20.5	20.8	18.9	20.9	21.1	20.5
9.5	10.7	15.9	15.4	18.4	11.4	20.0	17.3
29.5	17.1	41.8	18.0	54.9	21.2	38.1	20.1
73.7	3,823.2	13,609.1	5,752.5	19,058.3	5,039.0	16,118.2	7,113.6

-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
+	+	+	+	-	-	-	-
+	+	-	-	+	+	-	-
+	-	+	-	+	-	+	-
18	22	35	20	12	7	29	9
100.0	86.4	100.0	95.0	100.0	85.7	100.0	100.0
16.7	100.0	2.9	100.0	0.0	100.0	0.0	33.3
16.7	95.5	5.7	45.0	8.3	85.7	0.0	22.2
139.2	613.3	105.2	372.7	82.5	624.3	109.6	189.0
100.6	399.6	103.7	239.9	82.5	424.6	109.6	149.6
38.6	213.6	1.5	132.9	0.0	199.7	0.0	39.4
6.7	30.5	5.4	19.8	3.3	29.0	4.5	9.3
3.9	14.3	4.1	8.5	2.8	13.7	4.3	5.7
2.1	13.4	0.3	9.0	0.0	11.3	0.0	2.8
0.7	2.9	1.1	2.3	0.4	4.0	0.2	0.9
1.8	4.6	1.1	2.9	1.4	4.1	1.0	1.3
1.3	2.0	0.2	0.2	1.1	2.0	0.1	0.1
0.5	2.6	0.9	2.2	0.3	2.1	0.9	1.2
6,227.8	8,872.7	4,711.4	6,285.0	3,875.0	9,685.7	4,520.7	3,811.1
75.9	133.6	99.5	128.5	58.2	150.7	109.6	141.8
20.7	20.1	19.4	18.9	25.4	21.5	24.3	20.3
3.7	6.7	5.1	6.8	2.3	7.0	4.5	7.0
41.8	14.5	44.8	16.9	47.0	15.5	41.3	20.2
3,397.0	1,932.7	4,456.8	2,167.2	2,735.3	2,337.9	4,520.7	2,858.3

			STATEWIDE CHARACTERISTICS
-	-	Factor One - Numerical Size	
-	-	Factor Two - Organizational Complexity	
-	-	Factor Three - Teacher Experience	
-	-	Factor Four - School Unit Size	
+	-	Factor Five - Economic Power	
29	9	Number of Districts	632
100.0	100.0	Percent with county-based administration	79.7
0.0	33.3	Percent which have high schools	55.5
0.0	22.2	Percent which receive integrated aid	61.3
109.6	189.0	Total Enrollment Per District	1285.6
109.6	149.6	Elementary enrollment per district	798.0
0.0	39.4	Secondary enrollment per district	487.6
4.5	9.3	Total Staff Per District	58.4
4.3	5.7	Elementary teachers per district	28.8
0.0	2.8	Secondary teachers per district	23.9
0.2	0.9	Other professionals per district	5.7
1.0	1.3	Number of Schools Per District	4.3
0.1	0.1	Schools per district with only one or two rooms	1.1
0.9	1.2	Schools per district with three or more rooms	3.0
4,520.7	3,811.1	Equalized Valuation Per District*	33,727.8
109.6	141.8	Students per school in the stratum	295.9
24.3	20.3	Students per staff in the stratum	22.0
4.5	7.0	Staff per school in the stratum	13.4
41.3	20.2	Valuation per student in the stratum*	26.2
4,520.7	2,858.3	Valuation per school in the stratum*	7,762.6

* (Dollars x 1000)

APPENDIX M

This appendix contains the full descriptions of the input variables. The variables are named and numbered to correspond with the list of input variables given in Table 7.

The definitions and sources of the data represented by the variables are detailed in Section II. A.

APPENDIX M

Names and Descriptions of Initial Input Variables

Variable 1: Mean Credential

Elementary school teachers hold different kinds of teaching credentials. Ten kinds of credentials are recognized in Wisconsin, and for this study they were given preference ratings according to WSDPI criteria: the highest ratings corresponded to the highest numeric codes. These are the kinds of credentials and their numeric codes:

Code	Kind of Teaching Credential
0	1 year special license
1	1 year permit
2	2 year license
3	3 year license
4	5 year term certificate
5	4 year term certificate
6	3 year term certificate
7	2 year term certificate
8	1 year license
9	life certificate

For a district, Variable 1, Mean Credential, is the arithmetic mean of the ratings of all the full-time elementary teachers in that district.

Variable 2: Mean Degree

Elementary school teachers differ with respect to the academic degrees which they have earned. The WSDPI records the highest academic degree held by a teacher, and for this study the degrees were given preference ratings according to WSDPI criteria: the highest ratings corresponded to the highest numeric codes. These are the degrees, together with their numeric codes:

Code	Kind of Academic Degree
1	less than 2 years (no diploma)
2	2 years (diploma)
3	3 years
4	Bachelor's
5	Master's
6	6 years
7	Doctor's
0	other

For a district, Variable 2, Mean Degree, is the arithmetic mean of the numeric codes of all full-time elementary teachers in that district..

Variable 3: Mean Salary

The WSDPI records the salary, in dollars per school year, of each teacher in Wisconsin. Variable 3 for a district is the arithmetic mean of the salaries of all the full-time elementary teachers in that district.

Variable 4: Mean Local Experience

The WSDPI records the teaching experience, in months, of each teacher in Wisconsin The record is given for both local experience and total experience. Local experience is given by the total number of months a teacher has been teaching in the district where he is currently employed. For a district, Variable 4 is the arithmetic mean of the total months of local teaching experience of all the full-time elementary teachers in that district.

Variable 5: Mean Total Experience

This variable for a district is the arithmetic mean of the total number of months of teaching experience, both local and elsewhere, or all the full-time elementary teachers in that district.

Variable 6: Mean Grade Spread

A teacher might be responsible for a classroom group which includes students from several grades; for instance, some of his students may be in grade 2, some in grade 3, and some in grade 4. Another teacher might have only fifth grade students in a self-contained classroom. Variable 6 reflects the spread or range of grades of the students for which a teacher is responsible. A teacher whose students are all at the same grade level will be assigned a score of "1". If the students are from two grade levels, a score of "2" will be assigned. If a teacher has students in Grades 2 through 4, the Grade Spread Score is "3". For a district, then, this variable is the arithmetic mean of the grade spread scores for all the full-time elementary teachers in that district.

Variable 7: Log-Variance Credential

This variable for a district is the logarithm, base e, of the variance of the credential codes (see Variable 1) of all the full-time elementary teachers in a district.

Variable 8: Log-Variance Degree

This variable for a district is the logarithm, base e, of the variance of the degree code scores (see Variable 2) of all the full-time elementary teachers in that district.

Variable 9: Log-Variance Salary

This variable for a district is the logarithm, base e, of the variance of the salaries in dollars (see Variable 3) of all the full-time elementary teachers in that district.

Variable 10: Log-Variance Local Experience

This variable for a district is the logarithm, base e, of the variance of the years of local teaching experience (see Variable 4) of all the full-time elementary teachers in that district.

Variable 11: Log-Variance Total Experience

This variable for a district is the logarithm, base e, of the variance of the total years of teaching experience (see Variable 5) of all the full-time elementary teachers in that district.

Variable 12: Log-Variance Grade Spread

This variable for a district is the logarithm, base e, of the variance of the spread in grades taught (see Variable 6) by all the full-time elementary teachers in that district.

Variable 13: Kind

This variable indicates the general kind of administrative structure un'er which a school district operates. The WSDPI uses a seven-category coding scheme (See Table 1, Part II). Each category indicates a slightly different tax base. For research purposes the WSDPI scheme was reduced to a two-category scheme, wherein a "1" was assigned to city-based districts, and a "2" to county-based districts. These codes are inverted with respect to preference ratings assigned by the WSDPI (see Table 1). The lower value corresponds to the higher WSDPI rating.

Variable 14: Scope

This variable designates whether or not a district has a high school. The variable is "1" for districts with one or more high schools and "2" for districts with no high schools. Note that these codes are inverted with respect to preference ratings assigned by the WSDPI (see Table 1). The lower value corresponds to the higher WSDPI rating.

Variable 15: Class

This variable indicates the level of state aid a district receives. The WSDPI distributes state aid according to three classifications, each of which has a specific set of criteria. The three classes of state financial aids are: "Integrated", the class of districts which receives the highest rate of aid distribution; "Basic with Integrated", the class which receives the second highest rate of aid distribution; and "Basic", the class receiving the lowest rate of aid. For research purposes, numeric codes were assigned to the three classes as follows: "1" for Integrated, "2" for Basic with Integrated, and "3" for Basic.

Variable 16: Secondary Enrollment

This variable is the total student enrollment of all the secondary schools in the district. It is coded "0" if there is no secondary school in the district.

Variable 17: Elementary Enrollment

This variable is the total student enrollment of all the elementary schools in the district. Since only districts with some elementary students were included in this study, the variable is always greater than zero.

Variable 18: Number of Elementary Teachers

This variable for a district is its total number of all full-time elementary school teachers.

Variable 19: Number of Junior High School Teachers

This variable for a district is its total number of all junior high school teachers. Included are teachers who teach some elementary students, as well as some junior high.

Variable 20: Number of High School Teachers

This variable is the total number of all high school teachers in a district. Also included are teachers who teach some elementary or junior high grades, as well as some high school.

Variable 21: Number of Other Teachers

This variable for a district is the total number of all teachers in the district who have not been counted in Variables 18, 19, or 20. Included are part-time teachers, and administrators who teach.

Variable 22: Number of Other Professionals

This variable for a district is the total number of all professional staff in the district who were not counted in variables 18, 19, 20, or 21. Included are those professional employees with no teaching duties. Note that the sum of Variables 18 to 22 is the total number of professional employees in a district.

Variable 23: Number of One-Room Schools

This variable is the total number of schools in a district with only one teacher, and is, by inference, the number of one-room schools in that district.

Variable 24: Number of Two- oom Schools

This variable is the total number of schools in a district with exactly two teachers, and is, by inference, the number of two-room schools in that district.

Variable 25: Number of Three-or-More-Room Schools

This variable for a district is the total number of its schools with three or more teachers, and is, by inference, the number of three-or-more room schools. Note that the sum of Variables 23 to 24 is the total number of schools in a district.

Variable 26: Equalized Valuation

This variable is the equalized valuation in dollars of a district. It is the standardized property wealth of a district, as determined by the Wisconsin State Department of Taxation. The variable affects the distribution of state financial aid to the school district.

Variable 27: Valuation/Student

This variable is the ratio of the equalized dollar valuation of a district to the number of students in that district. Note that this variable is equal to the ratio of Variable 26 to the sum of Variables 16 and 17.

Variable 28: Students/School

This variable is the ratio of the number of students in a district to the number of schools in that district. Note that this variable is the ratio of sum of Variables 16 and 17 to the sum of Variables 23, 24, and 25.

Variable 29: Students/Staff

This variable is the ratio of the number of students in a district to the number of professional employees in that district. Note that Variable 29 is the ratio of the sum of Variables 16 and 17 to the sum of Variables 18 to 22.

Variable 30: Staff/School

This variable is the ratio of the number of professional employees in a district to the number of schools in that district. Note that Variable 30 is the ratio of the sum of Variables 18 to 22 to the sum of Variables 23 to 25.

Variable 31: Valuation/School

This variable is the ratio of the equalized dollar valuation of a district to the number of schools in that district. Note that this variable is the ratio of Variable 26 to the sum of Variables 23 to 25.

References

Barnes, R. E. A survey of status and trends in departmentalization in city elementary schools. *J. educ. Res.*, 1961, *55* (6), 291-292.

Berthold, C. A. *Administrative concern for individual differences*. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1951.

Box, G. E. P. , & Hunter, J. S. The 2^{k-p} fractional factorial designs. Part 1. *Technometrics*, 1961, *3* (3), 311-351.

Box, G. E. P., & Hunter, J. S. The 2^{k-p} fractional factorial designs. Part 2. *Technometrics*, 1961, *3* (4), 449-458.

Briner, C. Educational organization, administration, and finance. *Rev. educ. Res.*, 1964, *34* (4) 395-494.

Brunner, E. de S. *The growth of a science: a half-century of rural sociological research in the United States*. New York: Harper, 1957.

Carlson, R. O. *Adoption of educational innovations*. Eugene, Oregon: Univer. of Oregon, Center for the Advanced Study of Educational Administration, 1965.

Carter, R. F., & Suttloff, J. *Communities and their schools*. U.S.O.E. Project 308, final report. Stanford: Stanford Univer., 1960.

Charters, W. W. The social background of teaching. In N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963. Pp. 715-813.

Cochran, W. G. *Sampling techniques*. New York: Wiley, 1953.

Cornell, F. G. Sample surveys in education. *Rev. educ. Res.*, 1954, *24* (5), 359-374.

Cornell, F. G. Sampling methods. In C. W. Harris (Ed.), *Encyclopedia of educational research*. (3rd ed.) New York: Macmillan, 1960. Pp. 1181-1184.

Deming, W. E. On the distinction between enumerative and analytic surveys. *J. amer. statist. Ass.*, 1953, *48* (3), 244-255.

Duncan, J., & Kreitlow, B. *Selected cultural characteristics and the acceptance of* educational programs and practices. *Rural Sociol.*, 1954, *19*, 349-358.

Fichter, J. H. *Parochial school: a sociological study*. Notre Dame: Univer. of Notre Dame Press, 1958.

Glass, G. V., & Maguire, T. O. Abuses of factor scores. *Amer. educ. Res. J.*, 1966, *3* (4), 297-304.

Gregg, R T. Educational organization, administration, and finance. *Rev. educ. Res.*, 1961, *31* (4), 347-443.

Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, *18* (4), 277-296.

Guttman, L. The determinancy of factor score matrices with implications for five other basic problems of common-factor theory. *Brit. J. statist. Psychol.*, 1955, *8* (2), 65-81.

Guttman, L. The matrices of linear least-squares image analysis. *Brit. J. statist. Psychol.*, 1960, *13* (3), 109-118.

Harris, C. W. Separation of data as a principle in factor analysis. *Psychometrika*, 1955, *20* (1), 23-28.

Harris, C. W. Some Rao-Guttman relationships. *Psychometrika*, 1962, *27* (3), 247-263.

Hotelling, H. Analysis of a complex of variables into principal components. *J. educ. Psychol.*, 1933, *24* (4), 417-441 & 498-520.

Hotelling, H. Simplified calculation of principal components. *Psychometrika*, 1935, *1* (1), 27-35.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, *23* (3), 187-200.

Kaiser, H. F. Formulas for component scores. *Psychometrika*, 1962, 27 (1), 83-87.

Kaiser, H F. Image analysis. In C. W. Harris (Ed.), *Problems in measuring change.* Madison, Wisconsin: Univer. of Wis. Press, 1963. Pp. 156-166.

McLean, L. D. Phantom classrooms. *Sch. Rev.*, 1966, *74* (2), 139-149.

Miller, D. M., Baker, Joan F., Conry, Julianne L., Conry, R. F., Pratt, A. D., Sheets, S. E., Wiley, D. E., & Wolfe, R. G. *Elementary school teachers' viewpoints of classroom teaching and learning*. U.S.O.E. Project 5-1015-2-12-1, final report. Madison, Wisconsin: Univer. of Wis., Instructional Res. Lab., 1967.

Moonan, W. J. On the problem of sample size for multivariate simple random sampling. *J. exp. Educ.*, 1954, *22* (3), 285-288.

Mort, P. R., & Cornell, F. G. *American schools in transition*. New York: Teachers Coll , Columbia Univer., 1941.

Pierce, T. M. *Controllable community characteristics related to the quality of education*. New York: Teachers Coll., Columbia Univer., 1947.

Ryans, D. G. *Characteristics of teachers: their description, comparison, and appraisal*. Washington: American Council on Education, 1960.

Showell, M. How much stratification? *Int. J. Opin. Attitude Res.*, 1957, *5* (2), 229-240.

Schunert, J. The association of mathematical achievement with certain factors resident in the teacher, in the pupil, and in the school. *J. exp. Educ.*, 1951, *19* (3), 219-238.

Sokal, R. R. Numerical taxonomy. *Sci. Amer.*, 1966, *215* (6), 106-118.

Stein, J. (Ed.) The Random House dictionary of the English language. New York: Random House, 1966.

Terrien, F. W., & Mills, D. L. The effect of changing size upon the internal structure of organizations. Amer. sociol. Rev., 1955, 20 (1), 11-13.

Turner, R. L. Problem solving proficiency among elementary school teachers. U.S.O.E., C.R. Project 1262, final report. Bloomington, Indiana: Indiana Univer., Inst. of educ. Res., 1964.