DOCUMENT RESUME

ED 251 506 TM 840 803

AUTHOR	Bachelor, Patricia; Buchanan, Aaron
TITLE	Using Cluster Analysis to Solve Real Problems in Schooling and Instruction.
INSTITUTION	Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	SWRL-TR-85
PUB DATE	17 May 84
CONTRACT	NEC-00-3-0064
NOTE	44p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Cluster Analysis; Cluster Grouping; Computer Software; Correctional Education; *Evaluation Methods; *Instructional Improvement; *Mathematics Skills; Quantitative Tests; *Remedial Mathematics; Research Design; Secondary Education; Statistical Studies; *Student Evaluation
IDENTIFIERS	Evaluation Problems; Pre Algebra; Training and Employment Prerequisites Survey

ABSTRACT

This report describes cluster analysis methodology and illustrates its potential merits for educational research by describing a study designed to identify the natural subgroups existing among students beginning a secondary level remedial mathematics course. Cluster analysis forms groups of relatively homogeneous subjects represented in large data blocks that contain several different observations for each subject. While recently developed software packages make the computation more manageable, identifying the optimal number of clusters, and making sense out of them, requires considerable knowledge of the subjects and variables being clustered. Cluster analysis was applied to data from the two Prealgebra surveys of the Training and Employment Prerequisites Survey given to almost 1,500 wards of the California Youth Authority enrolled in remedial mathematics classes. The Prealgebra Surveys cover the skills and concepts most common to mathematics instruction up through grade 7. Results indicate that even fairly well prepared students will be unsuccessful in traditional remediation programs which focus first on mastery of all types of complicated computational skills. Instead, general mathematics instruction should consist of one course that redevelops basic concepts and skills for handling fractions and decimals, and another course on more advanced topics in general mathematics applications or an introduction to algebra where the requirements for complicated forms of computation are carefully controlled. (BS)

SWRL EDUCATIONAL RESEARCH AND DEVELOPMENT

Using Cluster Analysis to Solve Real Problems in Schooling and Instruction

TR 85 May 17, 1984

SWRL

SWRL EDUCATIONAL RESEARCH AND DEVELOPMENT

TECHNICAL REPORT 85

May 17, 1984

USING CLUSTER ANALYSIS TO SOLVE REAL PROBLEMS IN SCHOOLING AND INSTRUCTION

Patricia Bachelor and Aaron Buchanan

ABSTRACT

The basic issues of the methodology of cluster analysis as applied to educational research are discussed. The relative merits of the technique are far-reaching and deserving of more attention on the part of educational researchers and practitioners. An application of the technique on school and instructional data are presented.

USING CLUSTER ANALYSIS TO SOLVE REAL PROBLEMS IN SCHOOLING AND INSTRUCTION

Patricia Bachelor and Aaron Buchanan

> Consider that you have the opportunity to redesign coursework in mathematics for students who have completed seven or eight years of school (not counting preschool or kindergarten) but have not seen a lot of success in mathematics. Whatever else they may have done, they have fallen seriously behind the pace of instruction by the end of grades 7 or 8, and they are not usually thought to be promising candidates at this point for regular instruction in a first year-long course in algebra. These are students who typically flow into a ninth-grade course in general mathematics and, if they can get by requirements for graduation with only one year of mathematics, are unlikely to take any more mathematics--ever.
>
> Are these students basically alike? Should they all take about the same course in mathematics at this point in their instructional history? Or are there enough differences within the group to offer two or three different kinds of course/options that will build upon what, if anything, students know already. In other words, do their performances on different kinds of tasks, all taken from fairly straightforward things that they should have had several opportunities to learn at earlier grade levels, show mostly likenesses that can all be built upon in about the same way? Or do they suggest basic differences that would be more adequately addressed by two or three course/options.

Improving the delivery of instruction so that it better fits past accomplishments and present weaknesses of students who will benefit from it is a very practical problem, but one that is also very complicated. The reason this problem is so complicated has everything to do with formal tests that are regularly used as part of school instruction, with the information that is present--but not always obvious--in test results, and with how this information can be organized to tell us something about how to reshape instruction intended for whole populations of students who are at a common point in school. The main problem is how to organize what can be huge amounts of data in ways that tell us more about the structure of the population of students that the data represent. Educational groups have not made much headway in solving this problem, but researchers and practitioners in other areas of science have, and cluster analysis is one of the tools they have used to help them.

Cluster analysis is a family of statistical procedures designed to "create" clusters within a set of data. It is used extensively in many areas of science to organize things according to their likenesses and differences within large blocks of complicated data, and it has the potential to be applied successfully to some very practical problems in schooling and instruction. Wherever there are large sets of data consisting of many observations such as test scores for individual students, classrooms, or schools, cluster analysis has great potential to assist in sorting out groups of students, classrooms or schools that appear within the data to be more alike than different. It is especially useful when we have several observations for each student, classroom, or school. Consider, for example, the observation of students' performances in mathematics. These days, the tests that schools use to observe mathematics performances provide scores for individual students on several different mathematics objectives.[1] However, most of the

[1]Sometimes, these scores are for a variety of mathematics subskills rather than very specific mathematics objectives, but the difference has more to do with nomenclature than with the substance of what is tested. In current practice, "subskills" and "objectives" are both represented by the same kinds of test items grouped within the test in about the same way.

grouping that schools do for purposes of instruction is still based on a single, overall mathematics score. The result is a "high," a "medium," and a "low" group based only on the overall scores of individual students; all of the details represented by the high performances and low performances of individual students on different objectives is "averaged" away. Schools don't pay much attention to different scores made by different students on different objectives, because all of these differences are too hard to keep track of when decisions about the grouping of students are being made. Cluster analysis could help to simplify this problem and, at the same time, give schools a lot more power to respond to the diversity of needs and accomplishments residing within a large group of students, but people who run schools--and other people who advise them--don't know about it.

If cluster analysis is not a practical tool for schools to use, it is partly because educational researchers have not provided much leadership. Cluster analysis is simply not a procedure that's used very much in educational research. It was not really used very widely in any area of science prior to the time that the computer became well-established as a tool for processing huge quantities of research data. Since then, cluster analysis has become a practical tool in taxonomic sciences, where, for example, there is a lot of interest in identifying different classes of flora and fauna that are precisely alike in some respects and precisely different in others. It is also used widely in information sciences, where researchers use a technique called co-citation analysis to identify new fields of scientific endeavor based on observations of the topics that are covered by articles published in scientific journals and the references that are made in the text of these articles to works of other authors which are published elsewhere. These days, researchers in education could be using cluster analysis to study not only large populations of students but also populations of schools and classrooms, based on observations of likenesses and differences that are collected into large data bases on a fairly routine basis--but they don't. As a tool for making sense out of information buried in large data bases, the kind that exist in growing abundance in local school districts, state departments of education, and federal bureaus, cluster analysis has yet to have much impact on education.

Several months ago, SWRL staff began to use cluster analysis to shed more light on the structure of large populations of students representing fairly broad units of instruction. What we have been looking at directly is the instructional accomplishments of different students, different classrooms, and different schools based on the proficiencies they appear to demonstrate on formal assessments for different kinds of school subject matter such as reading, mathematics, composition and science.

The instructional accomplishments which we observed were taken directly from the raw percentages for different skill areas within, say, a mathematics assessment, that are commonly reported as test results. For an individual student, these results are the percentage of assessment items for a particular skill area that the student answers correctly; for classrooms and schools, the results on a particular skill area are averages of the results obtained by individual students within a classroom or within a school.

What we wanted to find out by using cluster analysis was the following:

> To what extent do the accomplishments of students, or classrooms, or schools form meaningful clusters that suggest that, based on their accomplishments, we are not dealing with one large group of students, classrooms, or schools that are all basically alike but are faced with several smaller groups that are obviously quite different?

We were not doing research on cluster analysis, as such. Rather, we were applying procedures for doing cluster analysis--procedures that already exist in packaged software that is almost universally available to researchers. In fact, more than applying procedures for doing cluster analysis, which are as easy as the software package is "user-friendly," we were grappling with the problem of interpreting results in ways that had direct and obvious implications for shaping instruction for students as a population. That's the hard part, because there simply hasn't been enough use of cluster analysis on questions regarding schooling and instruction for us to have many precedents to follow. But that's getting ahead of the story. To fully appreciate the problems inherent in interpretation of the results of cluster analysis , one needs to be fairly well grounded in what cluster analysis--in this case, cluster analysis software--is designed to do in the first place.

<u>What cluster analysis does</u>

The general objective of cluster analysis is to partition a set of subjects into subgroups that are as homogeneous as possible. Sometimes, the differences between these subgroups are large and obvious and we assume they are meaningful; other times, the differences are so small that it would be impossible to make a case that one subgroup should be thought of as being any different from the other. In the latter case, the "mathematics" of cluster analysis gives us subgroups that are different in a strictly technical sense, but, realistically, the ways in which they are different don't appear to be worth much concern.

To the extent that cluster analysis generates subgroups that are meaningful, a careful look at these subgroups and how they relate to one another will allow the user to:

- Identify natural clusters within a mixture of subjects that may represent several different populations
- Construct a useful scheme for putting subjects into different classes
- Find out whether or not classes that are believed to be present within a certain population are actually there
- "Snoop" within a population for unsuspected clusters

Cluster analysis forms groups or clusters of relatively homogeneous subjects that are represented in large blocks of data that contain several different observations (e.g., "scores") for each subject. These groups are formed on the basis of how "close" together individual subjects are in the data base. At the heart of cluster analysis, "closeness" is measured mathematically in different ways depending on the method for cluster analysis that is being used. In most software packages closeness is measured by some form of what is known as Euclidean distance between points in the data base or by a form of what's known as sums of squares. Either way, these methods are designed to form clusters so that distances between individual subjects are as small as possible within clusters and as large as possible between clusters. The thing to keep in mind is that methods for clustering data are designed to create clusters whether or not any meaningful clusters actually exist. For example, the cluster method will, at some point, partition a data base containing, say, 40 subjects into ten clusters even though the data base may not actually contain any cluster that's meaningful, much less 10. In other words, the different methods for cluste. analysis are designed to grind out clusters according to some mathematical specifications that are built into the method. It is the researcher's job to decide when there is enough difference between clusters within a certain configuration for the clusters themselves to be meaningful.

How Cluster Analysis Works

What really happens is this: the data in a cluster analysis are assembled into a matrix, or a rectangular array of numerical entries, corresponding to the observations made on each variable for each subject. The rows represent **N** subjects while the columns represent n variables, such as test scores on different skill areas in mathematics. A complete row of "scores" for each skill area may be considered the subject's profile. Next, the data matrix is transformed into a square **NxN** matrix of measures of distances that exist between each pair of subjects.[2]

---

[2]Usually these are Euclidean distances, sums of squares, or correlation coefficients. And at this point, the researcher using a software package has already selected the method for determining distance available in the package that is most appropriate for the kind of data that will be clustered. However, there does exist a large body of measures that can be used in cluster analysis which may be divided into four major groups: one, distance measures which are generally used with continuous or ordinal data but can also be applied to binary or qualitative data; two, association coefficients generally used with binary or qualitative data; three, correlation coefficients, such as Pearson's r for continuous data or phi for binary data; four, probabilistic similarity coefficients based on information statistics.

Actual clustering begins as soon as the matrix of distance measures is complete. (This all takes the computer but a few seconds to do.) There are several methods for forming clusters, but most software packages only use the most popular ones, and these are known as hierarchical methods.[3] The hierarchical methods can be classified as divisive or agglomerative. Agglomerative, techniques which are the most common, start by treating all N subjects as individual clusters and then proceed step-by-step to form new clusters from subjects that are closest together. At each step, the two entities that are closest are combined to form a new cluster. Sometimes this means combining two individual subjects that are not already in a cluster; sometimes it means combining an individual subject with a cluster that has been formed previously; and sometimes it means combining two clusters that have been formed previously. The result is to form bigger and bigger clusters until, finally, there is only one huge cluster that contains all N subjects. Divisive techniques operate in the reverse. This process starts with all N subjects in one cluster and then divides this cluster into two clusters, then three, and so on until there are N clusters that each contain one subject.

[3]Clustering techniques are broadly classifiable into nonhierarchical methods and hierarchical methods. The nonhierarchical--or single level--procedures are of two basic types. One technique involves iterative partitioning of subjects into multiple clusters. Typically, some form of optimizing criterion is applied to relocate subjects from one cluster to another after the initial assignment. The method begins with a predetermined number of clusters and, through various iterative processes, tries to find a revised classification which will make the distances from subject to subject within each cluster as small as possible in combination with maximizing the distances between clusters.

The use of cluster analysis involves almost no assumptions about the underlying structure of the data base, which makes it a very desirable instrument for many different types of applied research. However, there are two important issues that need to be addressed early in the design of the research. These are (a) how to know when you have an optimal number of clusters and (b) how to interpret these clusters once you've got them. Software packages usually provide some kind of statistic intended to "measure" differences that exist between clusters in a particular configuration. Whatever else goes into the decision is based on practical considerations of how many clusters can reasonably be handled within the interpretation--especially when the purpose of interpretation is to generate practical implications for how to improve schooling and instruction.

While recent advances and interest in cluster analysis have lead to software packages that make the computational drudgery that accompanies cluster analysis more tractable, the identification of an optimal number of clusters and making sense out of clusters in an optimal solution still require a good deal of outside knowledge about the subjects that are being clustered and the variables (e.g., scores on different skill areas in mathematics) on which the clustering is based. There do exist some guidelines for selecting the number of clusters. Unfortunately, these guidelines are mostly tied to work in areas other than schooling and instruction, and they don't provide much help.

Proceeding to the second phase of analysis, let's suppose that, in looking at performances in mathematics, we suspect that three clusters are optimal. First, we need to get some practical sense of how different the clusters are; second, we need to provide some explanation of why these differences occur, and, better still, what implications they have for improving instruction. Until now, there has not been enough work done on cluster analysis to provide us with much of any model to follow, but we do know that there are some areas where interpretations could easily go wrong. In our hypothetical case here, where we suspect that three clusters are optimal it would be easy to try to portray them as "high," "average," and "low" clusters. The trouble with these labels is that they are consistent with our intuitions that most instructional groups have students at three different levels of ability, but they are not consistent with the regularity that we know exists in the effects of instruction over a long period of time. All students are taught about the same things at about the same point of time. Moreover, they tend to learn about the same things in the same order, especially in mathematics. For several reasons, some students learn some things sooner than others, and so we are likely to see basic differences in the things--especially the number of things--that different subgroups of students can do. Differences in ability undoubtedly play some role in how soon different students can do what they've been taught, but it's only one of many things that affect instruction, and it isn't a very useful consideration in deciding what to teach and when to do it.

Often, in what we call a 3-cluster solution in cluster analysis, there may be two clusters that both contain a series of high and low performance on different skill areas. Skill areas that show high performances for one cluster may show low performances for the other, and vice versa. Sometimes, it may be hard to see any meaningful differences between the two clusters, and in this case, it may be necessary to look at a 4-cluster or 5-cluster solution to actually "see" fairly clear differences among two or three major clusters. Generally, if we try to interpret a solution containing too few clusters then we run the risk that one large cluster masks differences between variables which differentiates between the clusters of subjects. If, on the other hand, there are too many clusters, then the interpretation becomes clouded and differences between subjects begin to overwhelm similarities that exist between clusters of subjects.

Some traditional statistical methods may at first seem helpful in "testing" a decision about an optimal number of clusters, but quite frequently these are not valid for use within the cluster analysis methodology. For example, ordinary significance tests, such as F-tests, are not valid for testing differences between clusters. Since clustering methods attempt to maximize the separation between clusters, the assumptions behind the usual significance test are drastically violated. Also, most valid tests for clusters either have intractable sampling distributions or involve null hypotheses for which rejection is vacuous.[4]

[4]For clustering methods based on distance matrices, a popular null hypothesis is that all permutations of the values in the distance matrix are equally likely. Using this null hypothesis one can do a permutation test or a rank test. The trouble with permutation hypotheses is that with any real data, the null hypothesis is totally implausible even if the data do not provide any useful information.

Application of cluster analysis to Real School Data: An Illustration

Over the past several months, we have had several opportunities to apply cluster analysis to real data on school accomplishments resulting from administration of some assessment instruments that have been developed by SWRL staff. In one case, cluster analysis was used to see what kind of natural clusters that elementary schools might fall into based on average school performances on different objectives included in a science assessment. Most recently, we have been using cluster analysis to look at natural subgroups that exist among students who are just beginning a formal remedial course in general mathematics at the secondary level, and, so far, this has provided us with the clearest illustration of the potential that exists for using cluster analysis to solve complicated, but very practical problems, that go with interpretation of achievement data so that there are clear implications for designing school coursework.

The data for our illustration came from results of what we call the Training and Employment Prerequisites Survey (TEPS). This survey was given to wards who are enrolled in remedial mathematics classes provided as part of its education program by the California Youth Authority. In mathematics, the TEPS series consists of two surveys, Prealgebra A and Prealgebra B, which cover the most salient skills and concepts that are most common to grade-by-grade instruction in mathematics up through about grade 7.

Prealgebra A represents skills and concepts that schools would cover in regular instruction by about the middle of grade 4. This includes:

- use of whole numbers up to one-hundred thousand,
- use of simple common fractions for part of a set or region,
- addition and subtraction of whole numbers,
- multiplication and division of 2-digit and 3-digit whole numbers by a number that is ten or less,
- simple measurement skills involving length, time and money,
- recognition of simple geometric shapes,
- solution of basic types of word problems involving addition, subtraction and multiplication.

Prealgebra B continues with:

- use of large whole numbers,
- use of equivalent fractions,
- simple relationships between fractions, decimals, and percents,
- computation with whole numbers up through multiplication and division by 2-digit numbers,
- computation with fractions, decimals, and percent,
- measurement skills that are a little more advanced than ones covered in Prealgebra A, (but they still deal with length, time, and money),
- basic kinds of word problems involving computation that is a little more advanced than the computation required by word problems in Prealgebra A.

Altogether, we had data from almost 1500 students. About two-thirds of them took Prealgebra B, because their instructors felt that Prealgebra A was too elementary to show a broad range of mathematical tasks that they could or could not perform. The rest of the students took Prealgebra A because, just prior to the time of our assessment, th had been working on the very simplest of mathematical skills and concepts.

Cluster analysis was applied only to data from students who had been in instruction two months or less. This distinction is important in data coming from education programs provided at correctional institutions. Courses do not usually have fixed beginning and ending points, as they do in regular high schools, because students are coming and going on a continuous schedule. Our main interest was to see what kinds of clusters might exist among students who are entering a course in remedial instruction, so we did not look at data from students who had been in a mathematics course for more than two months. From the data set for students who had only been in remedial work for a short time, we generated several random samples, each consisting of data from about 25 students. Ten samples were generated from results on Prealgebra A and ten from results on Prealgebra B. Five samples for each survey were used to look for basic patterns in the kinds of clusters that seemed to occur naturally and the other five were used to try to verify the patterns that occurred in the first five samples. All of the data processing for the cluster analysis was carried out using CLUSTER programs that are part of a software package from SAS.[5]

Analysis of Results from Prealgebra B

The examples we are providing here come from results of processing three of our ten random samples using CLUSTER. The first example from sample 3 is shown in Tables 1 and 2. It includes data from 25 students who took Prealgebra B. In the terminology of cluster analysis, the 25 students are cases and the nine skill areas on Prealgebra B, each involving about five to seven mathematics problems, represent the variables. The performances (percent correct) for each student on each skill area constitute the data on which CLUSTER is performed. The results of CLUSTER contain a lot of different kinds of information about

[5]SAS User's Guide: Statistics. 1982 Edition. SAS (Statistical Analysis System) Inc., Box 8000, Cary, North Carolina.

Table 1

Cluster Analysis Pre-Algebra A (Sample 3)
Name of Observation or Cluster

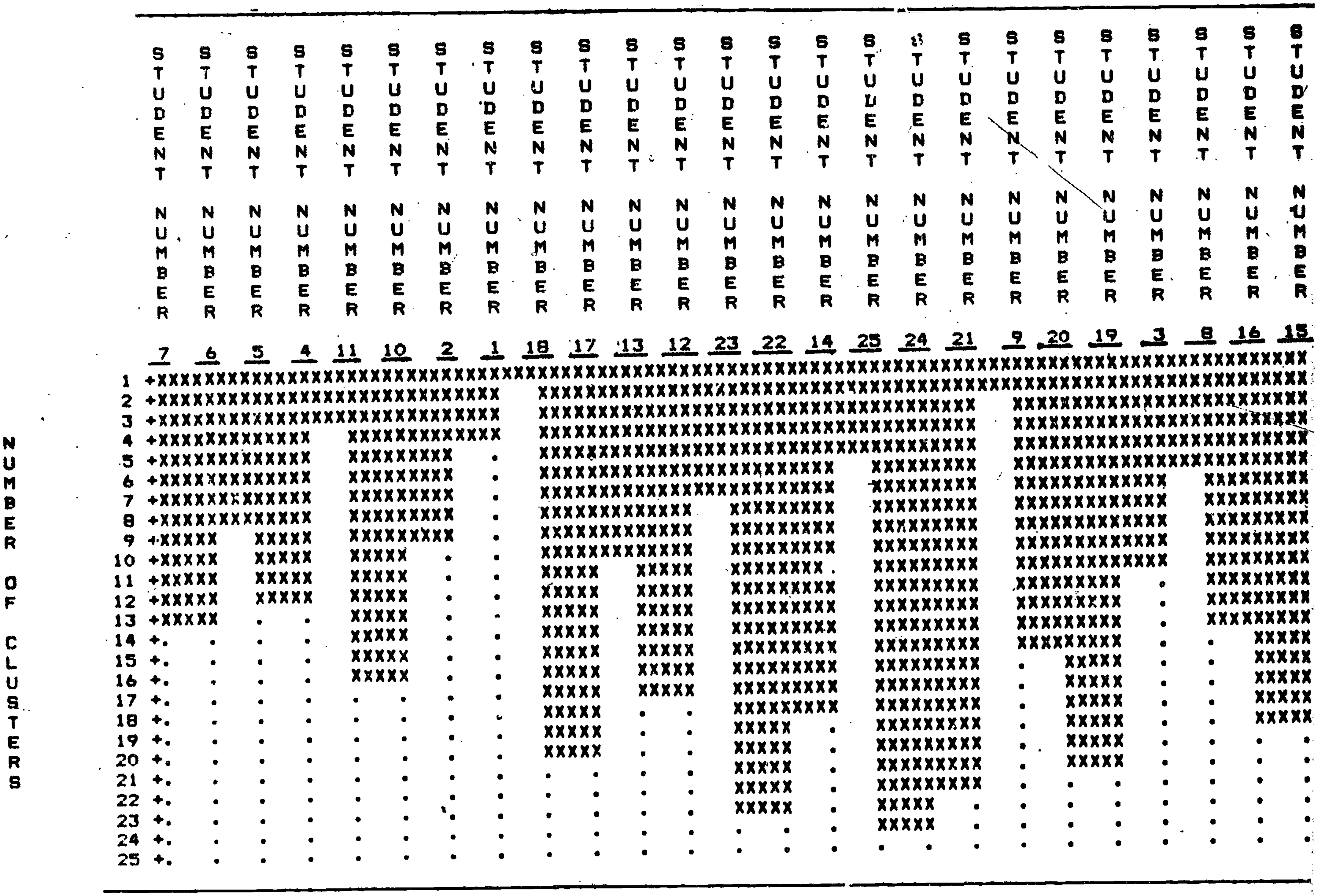

Table 2

Cluster Analysis Pre-Algebra B (Sample 3)
Ward's Hierarchical Cluster Analysis

Simple Statistics

	MEAN	STD DEV
WHOLENO	0.62400	0.29620
FRACTION	0.68667	0.27352
DECIMALS	0.36800	0.29257
COMPWHLE	0.82857	0.24046
COMPFRAC	0.54857	0.31047
COMPDECI	0.53500	0.23805
COMPPERC	0.63200	0.31979
MEASURE	0.60000	0.35119
PROBLEMS	0.75200	0.31770

Cluster Means of Various Solutions by Skill Area

SOLUTION	N	WHOLENO	FRACTION	DECIMALS	COMPWHLE	COMPFRAC	COMPDECI	COMPPERC	MEASURE	PROBLEMS
4 clusters	4	.15000	.41667	.30000	.64286	.285714	.343750	.20000	.10000	.10000
	4	.55000	.37500	.20000	.57143	.214286	.375000	.45000	.30000	.70000
	10	.88000	.90000	.540000	.97143	.857143	.712500	.82000	.86000	.94000
	7	.57143	.71429	.257143	.87755	.448980	.482143	.71429	.68571	.88571
3 clusters	8	.35000	.39583	.250000	.60714	.250000	.359375	.32500	.20000	.40000
	10	.88000	.90000	.540000	.97143	.857143	.712500	.82000	.86000	.94000
	7	.57143	.71429	.257143	.87755	.448980	.482143	.71429	.68571	.88571
2 clusters	8	.35000	.39583	.250000	.60714	.250000	.359375	.32500	.20000	.40000
	17	.75294	.82353	.423529	.93277	.689076	.617647	.77647	.78824	.91765
1 cluster	25	.62400	.68667	.368000	.82857	.548571	.535000	.63200	.60000	.75200

different cluster arrangements within the data set, as is true with most statistical packages that are available these days. The basic components that one would be most likely to use in interpreting CLUSTER results are shown in Tables 1 and 2.

Table 1 shows the optimal ways to separate all 25 students into first one cluster, then two, then three clusters and so on. The key is to look at the rows that start on the left side of the table. For example, the best arrangement for two clusters includes the first eight students in one cluster, reading across the top of the table, and the other 17 students in the second cluster. In the 3-cluster solution, the cluster with 17 students splits in two clusters, one with 10 students and the other cluster with nine, while the cluster with eight students stays intact. Finally, at the bottom of the table, we are left with 25 clusters where each student is defined as an individual cluster. Looking from the bottom up, 25 students in 25 clusters doesn't have much significance if one is thinking about shaping different kinds of courses to fit basic differences in past accomplishments--differences that are buried within a set of performance data in mathematics. In real schooling and instruction, it makes more sense to look at the top of the diagram in Table 1 to see what kinds of differences exist between clusters when you go from one cluster, that includes all students, to a maximum of four or five clusters.

For real applications to schooling and instruction, it helps to think of cluster analysis in reverse. Instead of forming clusters of things that are most alike, it's easier to think about "forcing" clusters to split. You begin with one cluster, that includes all 25 students, and then "force" it to split into two, ostensibly at the weakest point in the linkages between students' performances that, altogether, constitute the entry-level accomplishments that a course of instruction might be designed to build upon. You then force the weakest of the two clusters to split, now forming three clusters, and then force the weakest of these to split to form four clusters, and so on. Looking at things in this way, one is asking a question that is very basic to design of instructional coursework:

> within a population of students, are all students about the same, or are there different subgroups of students who are quite different in terms of past school accomplishments as these accomplishments are now represented by different levels of performance on different skill areas of mathematics.

Table 2 provides information about different clusterings that helps to sort out various strengths and weaknesses within this group of 25 students. First, there are the simple statistics which show mean performance levels for the entire sample of 25 students on each of the nine skill areas. The sample as a whole had a relatively high performance level on computation with whole numbers (COMPWHLE) where, on the average, they answered about 83% of the items correctly. On the other hand, the performance level on decimals was quite low. Here, students answered, on the average, about 37% of the items correctly. Mainly, the problems in this skill area required students to give equivalent decimals and percents for common fractions, such as 3/20, or mixed numbers, such as 5 3/5. The bottom of Table 2 shows the same kind of statistic, mean performance levels on each skill area, for each of the clusters in

"1-cluster," "2-cluster," "3-cluster," and "4-cluster" solutions. The means for the 1-cluster solution are the same as the simple statistics for the entire sample that are shown at the top of the Table. For the other solutions, these cluster means can show us where we are dealing with several subgroups rather than a single group.

The 2-cluster solution in Table 2 shows us that our sample of 25 students is composed of at least two subgroups. One subgroup with 17 students in it has performance levels on each skill area that are consistently 30% to 40% higher than the remaining cluster of eight students. Only on DECIMALS and COMPDEC (computation with decimals) was the spread between the two clusters less than 30 percentage points. In fact, except for DECIMALS, the differences between these two clusters are so great that it would be foolish to try to design a single course in mathematics that could possibly build on the past accomplishments of these two groups of students; they are simply too different. On the other hand, would two courses take care of things, or are other differences still buried in either of these two clusters? The answers are no and yes--no, two courses wouldn't be sufficient because, yes, there are still subgroups that are quite different buried within the larger cluster of 17 students.

Consider the 3-cluster solution, where the cluster splits into two smaller clusters of seven and ten students respectively. Most of the "spreads" in cluster means that resulted from splitting the cluster of 17 students into two smaller ones are of the nature of 20 percentage points or less--but not all. There are huge differences between these two new clusters when it comes to performances on whole numbers (WHOLENO)--88% compared to 57%, decimals (DECIMALS)--54% compared to 26%, and computation with fractions (COMPFRAC)--86% compared 45%. Were it not for such large differences on these three skill areas, it is conceivable that a common course in mathematics could build adequately on the profile of accomplishments represented by these two clusters, even if students from the two clusters were taught in the same course. An instructor would have to take some precautions to be sure that the cluster of seven students received a little extra review on topics related to most of the skill areas, but that should not be an insurmountable difficulty. As it is, the differences between these two clusters on whole numbers, decimals, and especially, computation with fractions, are so large that it is inconceivable that an instructor could, in the same course, build adequately on past accomplishments of the 17 students taken as a single group.

The 4-cluster solution shows that additional differences also exist within the cluster of eight students who demonstrated relatively low levels of performance in our 2-cluster solution. When this cluster breaks into two smaller clusters, each containing four students, there are fairly large spreads in two skill areas: whole numbers (WHOLENO) and problem solving (PROBLEMS). However, there are reasons to be cautious at this point. The spreads in performance levels for the two new clusters occur on whole numbers (15%), computation with percent (20%) and problem solving (10%). The tasks in each of these skill areas requires considerably more reading than tasks in the other skill areas, so the new "performance spreads" we see in the 4-cluster solution could all be tied to students who have little proficiency for reading English.

The 3-cluster solution for sample 3 contains some patterns among the twenty-seven cluster means that suggest very clearly the kinds of proficiencies that the 25 students in this sample do have for instructors to build upon. First, there is the cluster of ten students who are already fairly well prepared. Except for numeration with decimals and percent (DECIMALS) and computation with decimals (COMPDEC), students in this well-prepared cluster, on the average, were successful more than 80% of the time on problems in each of the other skill areas. By contrast, there is an "under-prepared" cluster of eight students who were successful less than about 40% of the time on problems in all skill areas except computation with whole numbers (COMPWHLE). In-between, there is a cluster of seven students who were relatively successful on problems dealing with:

fractions (FRACTIONS)
computation with whole numbers (COMPWHLE)
computation with percent (COMPPERC)
measurement (MEASURE)
problem solving (PROBLEMS)

and relatively unsuccessful on problems dealing with:

numeration with decimals and percent (DECIMALS)
computation with fractions (COMPFRAC)
computation with decimals (COMPDEC)

The difference between clusters in the 3-cluster solution is clearest if one first looks at numeration for decimals and percent (DECIMALS) and computation with whole numbers (COMPWHLE). On these two skill areas, the relationships that exist between clusters are fairly obvious. Keep in mind that numeration for decimals and percent comes relatively late in the traditional mathematics textbook series intended for study through grade 8, while computation with whole numbers has been around since about grade 3. In other words, the students who took Prealgebra B had undoubtedly had less opportunity to learn numeration with decimals and percents and more opportunity to learn computation with whole numbers than any other topics covered in the survey.

Note that in the 3-cluster solution, fewer than about 50% of the students in all three clusters were successful, on the average, on problems dealing with numeration for decimals and percents (DECIMALS), while more than 50% of the students in all three clusters were successful, on the average, on problems dealing with computation involving whole numbers (COMPWHLE). Using 50% as a kind of watershed for looking at the various cluster means, we get a pattern of "pluses" and "minuses" as shown in Table 3, where "+" represents cluster means that are above 50% and "-" represents cluster means that are 50% or lower. Obviously, instructors who plan to teach a general mathematics course that either deals with or applies the arithmetic of fractions and decimals would have a fairly extensive background of residual skills to build upon if the course were taught to students in CLUSTER Y and almost no residual skills to work with if they were trying to offer the course to students in CLUSTER X.

What is even more compelling about the pattern of pluses and minuses in Table 3 is that it is repeated, almost identically, for each of the other nine samples of about 25 students selected at random from the

Table 3

Prealgebra B

Cluster Means Above 50% (+) and at or Below 50% (-) in a 3-Cluster Solution (Sample 3)

	WHOLENO	FRACTION	DECIMALS	COMPWHLE	
Cluster X	-	-	-	+	
Cluster Y	+	+	+	+	
Cluster Z	+	+	-	+	
	COMPFRAC	**COMPDEC**	**COMPPERC**	**MEASURE**	**PROBLEMS**
Cluster X	-	-	-	-	-
Cluster Y	+	+	+	+	+
Cluster Z	-	-	+	+	+

population of students taking Prealgebra B. These patterns are shown for all ten samples in Table 4. From this table, it is clear that the three clusters of students we identified in sample 3 are different from each other in almost exactly the same way in every one of the ten samples we analyzed.

The pluses and minuses for our three clusters reprepresent different patterns of proficiencies that students who are channeled into remedial coursework bring with them for instructors to work with. They are present in our population of remedial students no matter how many times we chose a cross section of this population for our analysis. Moreover, these three cluster types were not originally part of a large subpopulation of students that existed at only one institution in the California Youth Authority. In other words, the various students who made up the

Table 4

Prealgebra B

Cluster Means Above 50% (+) and at or Below 50% (-) in a B-Cluster Solution Across all Samples

SAMPLE	WHOLENO	FRACTION	DECIMALS	COMPWHLE	COMPFRAC	COMPDEC	MEASURE	PROBLEMS
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
CLUSTER X	+ - - - -	- - - - -	- - - - -	- + + + +	- - - - -	- - - - -	- - - - -	- - - - -
CLUSTER Y	+ + + + +	+ + + + +	- + + + +	+ + + + +	+ + + + +	+ + + + +	+ + + + +	+ + + + +
CLUSTER Z	- + + + +	+ + + + +	- - - - -	+ + + + +	- - - - -	- - - - -	+ + + + +	+ + + + +
ALTERNATE SAMPLE	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
CLUSTER X	- - - - -	- - - - -	- - - - -	+ + - + +	- - - - -	- - - - -	- - - - -	- - - - -
CLUSTER Y	+ + + + +	+ + + + +	- + + - +	+ + + + +	+ + + + +	+ + + + +	+ + + + +	+ + + + +
CLUSTER Z	+ + + + +	+ + + + +	- - - - -	+ + + + +	- + - - -	- - - - -	+ + + + +	+ + + + +

"well-prepared" cluster, Cluster B, in each of our samples--there were close to 100 of them altogether--came from many different classrooms in many different institutions. They were not "good" students who all came from the same school. This is important, because we need some assurance that, when we combined results for all students from different institutions to form a population and drew out random samples of about 25 students who had been in instruction for two months or less, we were not merely dividing up an intact group of "good" students across our various samples.

Analysis of Prealgebra A

Some results of cluster analysis on Prealgebra A are shown in Table 5. These are cluster means for 2-cluster, 3-cluster, and 4-cluster solutions for the 32 students who made up sample 3. As you can see, the structure of this sample is quite different from structures we saw within the samples of students who took Prealgebra B. For one thing, little is gained by splitting this sample into more than two clusters, which consist of 15 students and 7 students respectively. In the 3-cluster solution, the cluster of seven students, splits off a single student and leaves a new cluster of only six students. The same thing happens when we look at four clusters; a single student is split from the cluster of six students. While the performances of this individual student would most certainly be of concern to his instructor, they don't tell us much about designing general purpose coursework. We would still need to design coursework to meet the needs of two different subgroups. All that would be gained by looking at a 3-cluster or a 4-cluster solution would

Table 5

Prealgebra A

Cluster Means of Various Solutions by Skill Area

Sample 3

		NUMERATION	FRACTIONS	ADD BASIC	ADD OPR	MULT BASIC	MULT OPR	MEASURE	GEOMETRY	PROBLEMS
1	15	.89524	.92000	.96667	.90000	.98889	.93333	.82667	.96000	.93333
2	5	.82857	.48000	.93333	.83333	.83333	.77143	.68000	.72000	.84000
3	1	.71429	1.00000	.83333	.83333	.83333	.42857	.40000	.60000	1.00000
4	1	.71429	.80000	1.00000	.66667	.16667	.28571	1.00000	.80000	.80000
1	15	.89524	.92000	.96667	.90000	.98889	.93333	.82667	.96000	.93333
2	6	.80952	.56667	.91667	.83333	.83333	.71429	.63333	.70000	.86667
3	1	.71429	.80000	1.00000	.66667	.16667	.28571	1.00000	.80000	.80000
1	15	.89524	.92000	.96667	.90000	.98889	.93333	.82667	.96000	.93333
2	7	.79592	.60000	.92857	.80952	.73810	.65306	.68571	.71429	.85714

be an opportunity to see more clearly the nature of performances in the smaller cluster by weeding out some results that may represent aberrations in the mainstream of what individual students in this cluster bring with them to instruction.

Another thing that is clearly different about Prealgebra A is the fact that the performances of students in different clusters are a lot more alike than they were in Prealgebra B. In the 4-cluster solution, performances in the cluster with 15 students are fairly close to performances in the cluster with five students except for two or possibly three skill areas. There is a difference of over 40 percentage points on FRACTIONS and differences of a little over 20 percentage points on GEOMETRY and about 15 points on MEASUREMENT. Still, the performances of students in the smaller cluster are relatively high on geometry and measurement, especially when compared to performances we observed among different clusters of students who took Prealgebra B.

Clearly, most students who took Prealgebra A were fairly proficient on the residual skills that Prealgebra A represents. These are all basic concepts and basic forms of computation that elementary school textbooks routinely cover quite thoroughly by the middle of grade 4. Notice that the well-prepared cluster in sample 3 is by far the largest one. It is also the largest cluster in each of the other samples of students, except one, that we looked at from Prealgebra A. The size of clusters in the 4-cluster solutions for all ten samples of data from Prealgebra A are shown in Table 6. The well-prepared cluster is two to three times as large as the other major cluster in all samples except Sample 2 where the two large clusters are the same size.

Table 6

Prealgebra A

Cluster Sizes of the "Well-prepared" Cluster
Under the Four Cluster Solution

Sample	1	2	3	4	5	1A	2A	3A	4A	5A	Mean
N	17	10	15	19	18	11	13	18	15	19	15.5
Total Sample Size	26	24	22	29	30	19	24	33	25	29	26.1
% of Sample	65%	42%	68%	66%	60%	58%	54%	55%	60%	66%	59%

Looking back at Table 5, we see that there is no truly "under-prepared" cluster like the one that appeared consistently in all samples of data from Prealgebra B. Only in the 4-cluster solution do we see a cluster where students were successful, on the average, less than 50% of the time, and, even then, it only occurred for the skill area that involved recognition of common fractions. Using 50% as a kind of watershed, as we did earlier with results from Prealgebra B, we get a pattern of pluses and minuses for the two main clusters in sample 3 as shown in Table 7. This pattern is not as consistent as the patterns we saw for Prealgebra B, which were all almost identical, but, as Table 7 shows, they are very similar. In almost all samples, there was one cluster where students were relatively unsuccessful on recognition of simple fractions (FRACTIONS). In four out of the ten samples, there was a cluster where students were not very successful with multiplication and

Table 7

Prealgebra A

Cluster Means Above 50% (+) and at or Below 50% (-) for the Two Main Clusters in Sample 3

	NUMERATION	FRACTION	ADDBASIC	ADDOPER	MULTBASIC
Cluster X	+	+	+	+	+
Cluster Y	+	-	+	+	+

	MULTOPER	MEASUREMENT	GEOMETRY	PROBLEMS
Cluster X	+	+	+	+
Cluster Y	+	+	+	+

division beyond basic facts (MULTOPER), but all other major clusters showed moderate to high levels of success.

Implications for Course Design

The combination of results of cluster analysis of data from Prealgebra A and Prealgebra B have enormous implications for designing coursework for students who are typically headed for remediation at the secondary level. Each of our samples contained a substructure composed of two or three clusters of students who bring quite different sets of residual skills for instructors to build upon. More significant is the fact that each of our samples of students who took the same survey, Prealgebra A or Prealgebra B, contained clusters that represented the same relative strengths and weaknesses. Recall that all of the students in our population were at about the same point in instruction--they were near the beginning of a course in general mathematics after having been out of school for at least several months--but they were from seven different correctional institutions that did not share a common program of instruction. In fact, each institution had been designated to receive wards who were in a certain age range or who represented a different need for supervision and security. The education programs varied widely from one institution to another. The regularities that we saw among clusters of students in one sample after another represent a history of past instruction among students whose backgrounds are highly irregular. The fact that these inputs to instruction are so regular tells us a lot about how common are the effects of past instruction in the elementary grades. Students who have either had more experience or more successful experiences bring more residual skills than other students to current coursework, but, more important, their residual skills are all about the same.

In looking at means for different clusters on different skill areas, we saw patterns of pluses and minuses, based on whether means were above or below 50%, that were very regular across all of our samples from Prealgebra B and almost all samples from Prealgebra A. What is even more dramatic is the fact that the combination of these means for the population as a whole yields the same pattern of pluses and minuses as shown in Tables 8 and 9.

What the data in these two tables show most clearly is that differences between clusters represent real differences in the population based on past instruction. For example, clusters that have means below

Table 8

Prealgebra B

Weighted Average of Cluster Means Above (+) and at or Below (-) 50% for Three Major Clusters

	WHOLENO %	FRACTIONS %	DECIMALS %	COMPWHLE %	COMPFRAC %
Cluster X	87(+)	91(+)	57(+)	97(+)	86(+)
Cluster Y	59(+)	68(+)	21(-)	88(+)	35(-)
Cluster Z	47(-)	41(-)	19(-)	70(+)	22(-)

	COMPDEC %	COMPPERC %	MEASURE %	PROBLEMS %	AVERAGE NUMBER IN CLUSTER
Cluster X	76(+)	87(+)	86(+)	97(+)	8.5
Cluster Y	42(-)	67(+)	70(+)	83(+)	8.9
Cluster Z	35(-)	38(-)	32(-)	37(-)	7.7

Table 9

Prealgebra A

Weighted Average of Cluster Means Above (+) and at of Below (-) 50% for Two Major Clusters

	NUMERATION %	FRACTIONS %	ADDBASIC %	ADDOPER %	MULTBASIC %
Cluster X	89(+)	87(+)	97(+)	92(+)	96(+)
Cluster Y	70(+)	49(-)	94(+)	85(+)	75(+)

	MULTOPER %	MEASURE %	GEOMETRY %	PROBLEM %	AVERAGE NUMBER IN CLUSTER
Cluster X	89(+)	83(+)	91(+)	95(+)	17.5
Cluster Y	55(+)	74(+)	71(+)	78(+)	6.7

50% may vary a great deal between 0% and 50%, but they vary in about the same way for pluses and minuses that, we would be inclined to believe, represent about the same cluster of students on the same skill area. In other words, we don't have a situation where minuses on, say, computation with decimals represent levels of performance that, based on cluster means, are lower for our B clusters than they are for our C clusters in the ten samples from Prealgebra B.

In order for new coursework to take maximum advantage of the patterns of residual skills represented in Tables 8 and 9 it will be necessary to rethink some issues that are basic to remedial instruction. As it is now, all of the students represented by these data are caught in a fairly painful cycle of remediation, and there is little likelihood that they will ever break loose. Certainly, another course in general mathematics, whatever it's called, will make little difference in what these students will be able to do once the course is finished. A course intended to force-march students to mastery on all possible types of mathematics problems that might reasonably be classified under each of the skill areas in Prealgebra B will not achieve much real success, even among fairly well-prepared students in the A cluster. What will be achieved instead are a series of increments in what students will be able to do, and most of these will be related to computation with whole numbers and, perhaps, computation with decimals and percent. What will not be achieved is any kind of closure on the basic skills that lie at the core of each skill area.

It would be more productive to base the design of new coursework on two considerations:

1. What kind of course can build most effectively on the residual skills that students bring with them from elementary school, especially given the fact that time is limited to one or two semesters (or one to four quarters)?

2. How can coursework be redesigned so as to break the cycle of remediation for the fairly well-prepared cluster of students that we know exists?

What these two considerations do for the redesign of mathematics coursework prior to a first course in algebra is, more than anything else, to shift first priorities for instruction away from mastery of complicated skills for doing accurate and precise computation with large whole numbers, fractions, and decimals. Such skills are important, but data from national surveys of how mathematical skills develop within the population of school age children and adolescents make one fact abundantly clear: 20 to 30% of the population do not master the full set of computation skills on whole numbers, common fractions, and decimals, no matter how many times they are recycled through remedial coursework. To make matters worse, most remedial coursework is sequenced so that redevelopment of computation skills comes first, as a prerequisite to problem solving and other, more formal kinds of applied mathematics, such as accounting and personal or business economics. The effect is to guarantee two things: first, most students will not complete, much less get past, redevelopment of computation skills in the remediation cycle, and second, they still won't have reliable use of any of the operations on whole numbers, fractions, and decimals, except for addition and subtraction whole numbers and money values. Their success in the remediation cycle will be limited to tacking bits and pieces onto parts of computation skills they already have, which means they will still have no real power to handle numbers with any degree of confidence or reliability.

What is more important, given the clusterings of students and their residual skills that we have seen here, is to redesign coursework around options--options that require reliable use of arithmetic operations, including such long-neglected skills as approximation and estimation, but do not depend upon highly accurate skills for doing precise computation. For example, our analysis of Prealgebra B suggests three different kinds of course options for three different clusters of students. Students in our well-prepared cluster, Cluster X, need a course that briefly reviews the relationship between common fractions, decimals, and percents including the equivalence of different ways to express the same quantity. Beyond this review, the course should focus on more advanced topics that require use and interpretation of numbers without requiring much by-hand calculation. The beginning of a specially designed first course in algebra should not be excluded as an option. The fact that "advanced" coursework in general mathematics doesn't exist now should not be taken as evidence that it can't. More likely, it shows the lack of any real challenge to instructors and to course developers because redevelopment of computation skills provided not only the focus of general mathematics but, in practice, it also defined the boundaries.

Students in CLUSTER Y need a course that covers many of the same topics that are covered now in the middle half of a general mathematics textbook. They may need a brief review of computation with whole numbers, but redevelopment of this skill is not necessary. They do need to redevelop computation skills with fractions and decimals. Instruction in this area should focus first on things like rounding and estimation, before extensive work on basic forms of computation, so that students in this cluster get a maximum amount of power to handle applications that require computation, especially applications that involve percent. Contrary to many of our assumptions about the inability of remedial students to solve word problems, the indications from cluster analysis are that students in CLUSTER Y are, in fact, fairly proficient in handling basic word problems. What this means is that a course in problem solving that involves general mathematics already has at least a modest number of residual skills to build upon--it doesn't have to start from scratch.

Students in the under-prepared cluster, CLUSTER Z, show only limited evidence of the residual skills covered in Prealgebra B. They can obviously do some forms of computation with whole numbers, probably addition and subtraction, but little else is in place. Students who have scores in the different skill areas that look about like the ones in the profile in Table 8 should take Prealgebra A for instructors to get a more complete profile of the residual skills that they actual have to build upon.

The course options needed for students who took Prealgebra A are a little different. About two-thirds of the students are in a well-prepared cluster, CLUSTER X, and they should be given Prealgebra B to get a better idea of the extent of their residual skills. Otherwise, there would be little alternative but to begin intensive practice on computation with whole numbers for the purpose of being able to compute faster and with greater reliability. Instructors should look carefully at performances of students in Cluster Y from Prealgebra A. They should pay special attention to individual problems where individual students were unsuccessful. There is a good chance that many of these students have limited proficiency with English, which would mean that they could not be very successful on any skill areas that included much besides computation problems. There is also a good chance that some of the students in CLUSTER Y did not complete all of the sections of Prealgebra A, perhaps because they ran out of time. However, what the results of the analysis show most clearly is that students in Cluster Y need a thorough review of tasks that are typically part of "recognizing" common fractions and mixed-numbers and a redevelopment of basic tasks involved in multiplication and division by a 1-digit number. When instruction in these two areas is completed, they should be prepared to begin about the same course as students in CLUSTER Y Prealgebra B.

In summary, using cluster analysis we have shown a need for two basic course clusters in general mathematics: one that includes at least one semester or two quarters of redevelopment work on computation with fractions and decimals, and a second dealing with more advanced topics in the applications of general mathematics or an introduction to algebra where the requirements for complicated forms of computation are carefully controlled, at least at first. A lower-level course that begins as far back as multiplication and division facts doesn't seem to have much potential, although a small number of short well-organized modules dealing with single topics may be very useful for quickly preparing students to do some productive course work with fractions and decimals. General mathematics should be one course that redevelops basic concepts and skills for handling fractions and decimals and another course that uses fairly adequate proficiencies for handling fractions and decimals to learn how to do something else. Based on what we've seen in our analyses, general mathematics should not become mired in mastery of pre-requisite skills that mostly involve whole numbers--but, that is what it does now in too many courses for too many students.