USING DATA MINING TO DIFFERENTIATE INSTRUCTION IN COLLEGE ALGEBRA

by

RACHEL BECHTEL MANSPEAKER

B.A., Bridgewater College, 2005
M.S., Kansas State University 2008

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Abstract

The main objective of the study is to identify the general characteristics of groups within a typical Studio College Algebra class and then adapt aspects of the course to best suit their needs. In a College Algebra class of 1,200 students, like those at most state funded universities, the greatest obstacle to providing personalized, effective education is the anonymity of the students. Data mining provides a method for describing students by making sense of the large amounts of information they generate. Instructors may then take advantage of this expedient analysis to adjust instruction to meet their students' needs. Using exam problem grades, attendance points, and homework scores from the first four weeks of a Studio College Algebra class, the researchers were able to identify five distinct clusters of students. Interviews of prototypical students from each group revealed their motivations, level of conceptual understanding, and attitudes about mathematics. The student groups where then given the following descriptive names: Overachievers, Underachievers, Employees, Rote Memorizers, and Sisyphean Strivers. In order to improve placement of incoming students, new student services and student advisors across campus have been given profiles of the student clusters and placement suggestions. Preliminary evidence shows that advisors have been able to effectively identify members of these groups during their consultations and suggest the most appropriate math course for those students. In addition to placement suggestions, several targeted interventions are currently being developed to benefit underperforming groups of students. Each student group reacts differently to various elements of the course and assistance strategies. By identifying students who are likely to struggle within the first month of classes, and the recovery strategy that would be most effective, instructors can intercede in time to improve performance.

USING DATA MINING TO DIFFERENTIATE INSTRUCTION IN COLLEGE ALGEBRA

by

RACHEL BECHTEL MANSPEAKER

B.A., Bridgewater College, 2005
M.S., Kansas State University, 2008

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Major Professor
Andrew G. Bennett

Copyright

Abstract

The main objective of the study is to identify the general characteristics of groups within a typical Studio College Algebra class and then adapt aspects of the course to best suit their needs. In a College Algebra class of 1,200 students, like those at most state funded universities, the greatest obstacle to providing personalized, effective education is the anonymity of the students. Data mining provides a method for describing students by making sense of the large amounts of information they generate. Instructors may then take advantage of this expedient analysis to adjust instruction to meet their students' needs. Using exam problem grades, attendance points, and homework scores from the first four weeks of a Studio College Algebra class, the researchers were able to identify five distinct clusters of students. Interviews of prototypical students from each group revealed their motivations, level of conceptual understanding, and attitudes about mathematics. The student groups where then given the following descriptive names: Overachievers, Underachievers, Employees, Rote Memorizers, and Sisyphean Strivers. In order to improve placement of incoming students, new student services and student advisors across campus have been given profiles of the student clusters and placement suggestions. Preliminary evidence shows that advisors have been able to effectively identify members of these groups during their consultations and suggest the most appropriate math course for those students. In addition to placement suggestions, several targeted interventions are currently being developed to benefit underperforming groups of students. Each student group reacts differently to various elements of the course and assistance strategies. By identifying students who are likely to struggle within the first month of classes, and the recovery strategy that would be most effective, instructors can intercede in time to improve performance.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

First, I would like to thank Dr. David Allen, Dr. Todd Cochrane, Dr. Virginia Naibo, and Dr. David Gustafson for serving on my committee. Your advice has been invaluable. Also, thank you to all the members of the Center for Quantitative Research for your support, constructive criticism, and companionship. I am indebted to Physics Education Research group for providing me with my first opportunity to present this research and for teaching me how to properly conduct an interview. Ultimately, many thanks go to my major professor, Dr. Andrew Bennett, without whom I could not have finished this project. Thank you for pushing me to take chances and pursuing an unusual area of study.

I would also like to thank my mother, Dr. Janet Manspeaker, for catching all of my contractions, verb tense errors, and excessive use of pronouns. To my father, Dr. Brian Manspeaker, thank you for the inspirational postcards. Finally, I need to thank my fiancé, Peter, for his encouragement and love (and for taking over my dishwashing duties).

Dedication

I would like to dedicate this dissertation to my grandfather, Dr. Robert Daryl Bechtel. It was from him that I first learned that mathematics was dynamic, powerful, and exciting. Thank you for always listening and inspiring me to follow my own path.

Chapter 1 - Introduction

The main objective of the study is to identify the general characteristics of groups of students within a typical Studio College Algebra class and then adapt aspects of the course to best suit their needs. In an ideal world, every college student would be guided through his or her studies with individual expert instruction from a highly qualified, empathetic educator. Prohibitive costs and a shortage of eligible instructors prevent this tailored attention from being attainable at most public universities. In a College Algebra class of 1,200 students, like those at Kansas State University, the greatest obstacle to providing personalized, profound education is the anonymity of the students. Because classes are so large and rigidly structured, individuals' struggles in the class often cannot be addressed efficiently or effectively.

Although College Algebra instructors may not be able to get to know each of their 1,200 students individually, instructors may be able to learn about their students' academic performance through Data Mining Analysis. Data Mining provides methods for describing students by making sense of the large amounts of information students generate. For example, clustering techniques can be used to look for groups of students who perform similarly on assignments and examinations. Monitoring the progress of how groups of students behave is more feasible than trying to keep track of individual development. Instructors may then take advantage of this expedient analysis to adjust instruction to meet their students' needs.

Using examination problem scores, attendance points, and homework scores from the first four weeks of a Studio College Algebra class, the researcher was able to identify five distinct clusters of similarly behaving students. Data Mining techniques, however, could not reveal the qualitative characteristics of these groups of students. To uncover the motivations, beliefs, and levels of conceptual understanding contributing to these group behaviors, the researcher interviewed prototypical students from each group. Using the results from the quantitative data mining analysis and the qualitative interview analysis, the researcher created descriptive profiles of each of the student clusters. The student groups were then given the following illustrative names: Overachievers, Underachievers, Employees, Rote Memorizers, and Sisyphean Strivers.

The identification of student clusters provides instructors with a reasonable method for assessing their students' needs by simplifying the 1,200 individuals to five "types" of students. Each student group not only behaves distinctly, but the groups have different beliefs about mathematics, reactions to Studio College Algebra, and study habits. Advisors can use this information to help students enroll in the most suitable mathematics course. In order to improve placement of incoming students, new student services and student advisors across campus have been given profiles of the student clusters and placement suggestions. Preliminary evidence shows that advisors have been able to effectively identify members of these groups during consultations and suggest the most appropriate mathematics course for those students.

By identifying students who are likely to struggle within the first month of classes, and the corresponding recovery strategy that would be most effective, instructors can intercede in time to improve performance. In addition to placement suggestions, several targeted interventions are currently being developed to benefit underperforming groups of students. For example, the researcher has developed Problem Solving Workshops targeted to a particular group of students who perform poorly on examinations. These students spend more than adequate time and effort studying the material in College Algebra and were able to demonstrate during interviews that they understood the concepts. However, these students were unable to communicate their knowledge on written examinations. The Problem Solving Workshop is designed to help students become familiar with examination structure and connect examination problems with the appropriate concepts.

With technological advances, we can continually discover new ways to learn about students, provide instant feedback, adapt instruction and placement, and generally offer a high quality educational experience to all scholars.

Motivation

The main objective of the researcher was to more fully understand the students enrolled in mathematics courses and adapt those courses to better suit the students' educational needs. This practice of adjusting course content to reach all students is known in education research circles as Differentiated Instruction. Typically, a college course is structured so that every student is taught the same way. All students attend a lecture, lab, or question and answer

session, complete the same homework assignments, and take the same examinations. Evidence increasingly shows that many students fail to learn in this environment, leading a large number of students to drop or fail classes. At Kansas State University, where this study was conducted, for example, from 2003 to 2009 on average only 80% of first year students return to the university for a second year (*OPA* 3). Equally troubling is the graduation rate. Of the students who enrolled into Kansas State between 2000 and 2004, only 58.8% graduated within 6 years with a bachelor's degree. This number was substantially lower for ethnic minorities. The overall graduation rate of ethnic minorities at Kansas State University over this period of time was 38.4%, with African Americans being lowest graduating group, at 31.2% (*OPA 2*).

The factors influencing failure of college students have been widely studied, and it has been recognized that academics are only partly accountable for these high dropout rates. Educators, however, cannot discount the contributions of their teaching methods and course structure to the success and failure of their students. Because mathematics is a well established field, with introductory courses teaching concepts that have been practiced for centuries, mathematics instructors especially have been slow to break from tradition and attempt to address their students' problems.

College Algebra is the second largest course on the Kansas State University campus, with roughly 25% of incoming freshman enrolled in the fall semester. Several studies show a strong link between success in mathematics courses and high probability of graduation (Parker, 28), (Pederson). Improving the support structures available for students in College Algebra and increasing their chance of success in the class would significantly address retention problems. Increasing retention directly benefits students and increases revenues for the University. For example, at Kansas State University, raising the freshman retention rate from 80% to 81% would add close to $250,000 to the annual budget (assuming students are paying full in-state tuition) (*OPA 1*). In a depressed economy where state and federal budgets for higher education are being cut dramatically, this gives administration strong financial motivation to promote Differentiated Instruction and other innovative educational methods in their first year courses.

Differentiated Instruction is a collection of teaching practices aimed at helping all students reach the goals of understanding content and effective application. Proponents of Differentiated Instruction recognize that students are individuals with different educational backgrounds, ways of processing information, interacting with their peers, and holding diverse

beliefs about the college education experience. These differences greatly affect how students learn and the resources and support they need to realize their full potential. In most educational settings where Differentiated Instruction has been put into practice, the teacher is able to get to know each of his or her students through class time interaction and then adapt instruction accordingly. For example, in an elementary school where each teacher spends 6-7 hours a day with his or her small group of students, it is possible to quickly assess each pupil's abilities, personality traits, and special needs.

The structure of most large first year courses in a university setting does not allow for this type of differentiation. This is mostly due to the large class sizes and limited contact time of most large lecture courses. Before parts of the course can be adapted to fit the needs of students, these needs must be identified. When the average College Algebra lecturer, however, spends one or two hours each week with a class of 350 students, this level of personal attention is impossible. Traditionally, Universities have attempted to provide a smaller-class experience by employing teaching assistants to guide recitation sessions for one or two hours a week. The teaching assistants usually find that they are unable to get to know all but the most vocal students. Also, the rigidity of this setup prevents lecturers and teaching assistants from instituting any meaningful changes during the semester. The course schedule is entirely predetermined and carefully coordinated, so adapting to students' academic needs proves problematic.

College students are thus left to identify their own shortcomings and obstacles to learning, and make their own changes in order to find the best methods to overcome them. Universities often contain many organizations and groups dedicated to providing support for students working to adapt to college life. These assets range from student health organizations, to tutoring groups, to social clubs. Unfortunately, even though resources available to help struggling students exist, they are often not utilized by those students who really need them. Struggling students may be highly capable and determined to succeed, but these students are often unaware that they are headed for academic trouble, and unsure of how to change their path. Therefore, it is not enough for Universities to provide support structures for their students. To truly differentiate instruction, instructors must be able to identify students' individual traits and needs and then direct the students to the most appropriate resources.

Fortunately, new technologies can provide ways of both learning about students and adding adaptable elements to first year college courses. Computer programs can make a course more adaptable, such as providing students with more practice problems or alerting instructors to changes in behavior. "iClicker" technology can allow lecturers to ask multiple-choice questions during class and receive instant feedback from their students, allowing them do adapt their lectures to the level of student understanding. By incorporating online components, such as homework and electronic gradebooks, instructors also can record an enormous amount of information about their students. Every mouse click a student makes, such as activating a homework problem "hint" or replayed a lecture video, can be stored and later analyzed. Techniques for sifting through this data have been applied to the fields of business and engineering for several years now and are fairly well understood. Only recently have these techniques been employed to analyze student-produced data with the intention of understanding student behavior and how to improve instruction.

Research Questions

Improving instruction for College Algebra students through better understanding their behavior and motivation was the general goal for this research project. However, identifying the learning preferences, personality traits, level of conceptual understanding, and progress of 1,200 individuals through personal interaction during a single semester is impossible. To further expound on the goal of recognizing student needs and to clarify the direction of the project, the researcher specified the following questions:

1) How can Data Mining techniques facilitate determining student needs? Which Data Mining strategies would be most efficient and effective at revealing important knowledge about students in a timely manner?

2) Data Mining analysis looks for patterns in quantitative data. However, implementing Differentiated Instruction relies on understanding students' backgrounds, motivations, preferences, and shortcomings. Is it possible to uncover these qualitative student traits with Data Mining techniques?

3) Given the constraints involved with teaching mathematics in large lecture courses, how can instructors use the insight gained from Data Mining Analysis to effectively modify instruction?

Hypotheses

1) Patterns and similarities in student behavior can be efficiently and accurately identified using standard Data Mining techniques known as clustering algorithms.
2) College Algebra students' attitudes and beliefs about mathematics can be revealed by examining their behavior in the course.
3) This information can be used to develop effective math placement strategies, identify students in need of intervention, and improve freshman retention.

The first hypothesis represents a shift in describing student behavior from categorization to clustering. Traditionally, educational researchers have categorized students using a pre-existing framework developed after years of qualitative research and classroom studies. For example, the widely known theoretical model for student learning known as "Multiple Intelligences" was developed by Dr. Howard Gardner to describe a many-facetted approach to attaining knowledge. According to this model, a one-dimensional measure of intelligence is artificially limited. People can build many areas of intelligence, including interpersonal, intrapersonal, musical, kinesthetic, spatial, and the traditionally tested linguistic and logical intelligences (How People Learn, 101). Many instruments have been designed to test for aptitudes in these different intelligences. In general, much of educational research is built around developing useful and accurate educational models and then correctly placing students into the appropriate categories.

Data Mining provides an alternate method for describing students by making sense of the large amounts of information students generate. The set of data mining techniques known as "clustering" finds similarities in subjects based on patterns within the data. Clustering techniques are commonly and very successfully used in the business setting. For example, one of the keys to Netflix's recent success in the DVD rental system is its highly accurate ratings system. Netflix first asks their users to rate several movies they have seen. Then, rather than trying to label each individual as someone who likes "romantic comedies" or "1980's action

movies," Netflix compares user preferences to the preferences of other subscribers. Netflix identifies groups of people who have similar preferences, and then recommend movies based on what that subgroup liked. According to Netflix, the algorithm the company employs is accurate to within half a star on a five-star scale at least 75% of the time. Also, 50% of Netflix users who rent suggested movies rate them with the highest rating (Reinsburg, Feb 2010). In September of 2009, Netflix awarded a prize worth $1 million to a team of data miners who were able to increase the accuracy of their ratings prediction algorithm by 10%. The winning team applied a combination of several clustering techniques to a data set composed of over 100 million movie ratings to reach the goal (Lohr, 21 Sept 2009).

In this study, the researcher aimed to demonstrate that students also may be "clustered" by their behavior in College Algebra. Rather than trying to place students in an appropriate learning category, clustering techniques can identify groups of students who behave similarly. Finding patterns using computer algorithms and little expert knowledge are known as "black box" data mining methods. By using these black box procedures to group students based on the relationships in their behavior and responses to classroom assignments and examinations, one can avoid the problem of determining if preconceived categories are relevant or valid. Also, because students in Studio College Algebra generate large and varied amounts of data, many methods to test the accuracy and efficiency of predictions based on student clustering schemes exist.

In order for information about student clusters to be relevant for differentiating instruction, the data must be gathered and processed in a timely manner. Some initial work is needed to describe the characteristics of each student cluster and so useful analysis is not immediately available to the instructors. Once the student clusters have reliable descriptions, however, the process for identifying members should be easily reproducible and adaptable from semester to semester. Also, instructors are less likely to use clustering analyses if data collection for each semester is lengthy or prohibitive. Therefore, the collected data should be restricted to information normally recorded during the course of the semester, such as attendance points, homework, and examination scores.

Unfortunately, student clusters identified with data mining techniques do not come with the convenient descriptions that are available for "Multiple Intelligences" and other learner categories. Data mining algorithms can identify groups of students who behave similarly, but the

algorithms have more difficulty describing what those behaviors are and motivations for those actions. By confirming the second hypothesis, the researcher aims to show that student clusters based on behavior can be traced back to common levels of preparation, ideas about mathematics and education, cooperative skills, or other sources. Uncovering these motivating factors requires more qualitative research methods, such as interviewing prototypical members of each cluster about their beliefs and reactions to Studio College Algebra.

The third hypothesis is perhaps the most important one: demonstrating that applying this data analysis will have a positive impact on College Algebra students. The researcher explicitly aims to use this analysis to improve placement into appropriate courses and target underperforming groups of students for academic intervention, thus improving retention. Of course, the specific proposals for improved placement and intervention depend on the characteristics of the identified student groups. No inherent theoretical framework on which to build these placement and intervention strategies exists since the clusters are not based on preconceived categories. Therefore, attempts to differentiate instruction based on student clustering results must be carefully considered and monitored closely.

Limitations

Several factors of this study limit the generalization of its results. The study was restricted to students enrolled in the Studio College Algebra Course of Kansas State University. Most other universities of this size do not offer a mathematics course similar to Studio College Algebra. Also, the student demographics at Kansas State are not representative of college students as a whole, and thus may not yield the same student clusters as other universities. Studio College Algebra students were chosen as subjects of the study because they are mainly first year students from a wide range of backgrounds and levels of preparedness. However, these students do not represent a random sample of college students and tend to consist of students with limited mathematical background (with some important exceptions- See Chapter 4: Group UA).

Kansas State University's mathematics department also offers a course called Traditional College Algebra, which is a much larger course geared towards students intending to go on to the engineering calculus sequence. Studio College Algebra is a relatively new course, with the first trial classes running during Fall 2007. Studio College Algebra was designed for students in

business or social sciences who wish to take Statistics or General Calculus as part of their major. Studio College Algebra emphasizes practical applications and modeling with functions over algebraic manipulations. During the four semesters in which the study was conducted, it was determined that the general academic community at the University was unaware of the differences between the two courses, and so the students enrolled in these courses were fairly heterogeneous, with students from all backgrounds and majors equally likely to be enrolled in each course.

In Studio College Algebra, a high number and wide variety of assignment scores are recorded electronically during the first four weeks of the semester. This allowed the researcher to measure many types of behaviors, including performance on timed exams, performance on cooperative assignments, procedural fluency and success with applications, and attendance. Other more advanced math courses at the university do not produce this amount of information, and would thus be harder to analyze with Data Mining methods. Also, the behaviors exhibited by Studio College Algebra students were recorded as quantitative scores. Analyzing courses outside of mathematics with data mining techniques might require ways to convert more subjectively recorded behaviors into quantitative data.

The study was conducted over a period of four semesters (two years). While quantitative data was recorded and analyzed each semester, interviews of group representatives were only conducted during Spring 2009. This was due to budget and time constraints. Investigation of group characteristics revealed that students in the Fall semesters and the Spring semesters had different backgrounds and academic traits. The class sizes in the fall and spring also were very different. Therefore, several steps were made by the researcher to allow for the groups to be comparable. This process is described in detail in Chapter 3.

Summary

Differentiating Instruction in a large introductory mathematics course presents several challenges. Particularly frustrating is the anonymity of the students due to large class sizes and limited contact time. Data mining can provide a method for describing students by finding patterns and similarities in the large amounts of information they generate. This study aimed to show that analysis from Data Mining techniques such as Clustering could give educators insight

into their students' needs and suggest ways to successfully adapt the course to address these problems.

Because Educational Data Mining is a relatively new field, it does not have well established and tested research protocols. While the researcher was guided by her research questions and hypotheses, each stage of the study was heavily influenced by the results and questions raised by the preceding steps. The remainder of this thesis presents a narrative of the Data Mining techniques tested and selected by the researcher, investigation and interpretation of the results of using these algorithms, and actions taken in response to the analysis.

Chapter 2 - Literature Review

In this chapter, we review literature related to Differentiated Instruction, the Educational Data Mining community, and Data Mining techniques used in this study. Please note that all educational theory and mathematical statements in this section have been previously published by other authors.

Differentiated Instruction and Educational Data Mining

What is Differentiated Instruction?

Before the 19th and 20th centuries, education was reserved for the elite. Most knowledge was passed on through personal mentoring or tutoring, and classrooms with more than four or five pupils were rare. As access to education opened up in the industrial era, teachers taught increasing numbers of students, and education became organized into an institution. Now, most students in the United States are taught in classes separated by age with one or two teachers instructing an average class size of 24 (Digest of Education Statistics, 2003 stats). Traditionally, these classes are very teacher oriented, with the structure of the lessons and environment of the classrooms designed to allow the instructor to disseminate information to their students as efficiently as possible (Wormeli, 8). All students were taught the same way, and if a student struggled then he or she was either unintelligent or lazy. More recently, research in cognitive science and educational studies has shown that students differ in readiness, interest, and learning styles as well as innate ability (Tomlinson, 179). Tapping into these differences and using them to motivate learning has been linked to higher student achievement. Many educators have concluded that the traditional "one size fits all" method of teaching is inadequate for their students' needs and have been turning to other teaching strategies.

Differentiated Instruction is a collection of teaching practices designed to help all students meet their learning goals. According to Rick Wormeli, a leading advocate for Differentiated Instruction,

"The two simple charges of differentiation are:

(1) Do whatever it takes to maximize students' learning instead of relying on a one-size-fits-all, whole-class method of instruction and

(2) prepare students to handle anything in their current and future lives that is not differentiated, i.e., to become their own learning advocates." (Wormeli, 9)

This method of instruction is not the same as tiering or tracking, where students are segregated into groups with the same level of proficiency. Often with tiering, students in the lower tiers are given less work or fewer assignments than students in the more advanced tiers. For example, if a quick student finishes an assignment before the rest of the class, a teacher might give him or her another problem set to complete. Studies show that punishing students who learn faster than their peers by giving them more work leads to decreased motivation and poorer future performance. Likewise, struggling students have no incentive to improve if they know it would only lead to more work (Wormeli, 66), (Tomlinson, 180).

In Differentiated Instruction, the preferred method of adapting to students' speeds is to "change the nature of the work, not its quantity." (Wormeli, 10) While every student is expected to reach the same intellectual goals, he or she is are provided with many different tools for achieving those goals. Often, students are given the option to choose assignments that appeal to their personal interests. Other times, students are given problems with varying complexity or different applications. In any case, the goal of Differentiated Instruction is to recognize that students learn in different ways, and to facilitate their learning through the use of different materials, challenges, and support structures.

Although Differentiated Instruction can best be described as a compilation of many techniques and practices, there are a few crucial features to every implementation of Differentiated Instruction. First, to differentiate instruction effectively, one must get to know the students as individuals and identify their particular needs and challenges. Carol Ann Tomlinson provides a table outlining possible students' traits and their implications for learning, which is provided below:

Table 2.1 Categories of Student Variance with Contributors and Implications for Learning

Category of Student Variance	Contributors to the Category	Some Implications for Learning
Biology	Gender Neurological "wiring" for learning Abilities Disabilities Development	High ability and disability exist in a whole range of endeavors. Students will learn in different modes. Students will learn on different timetables. Some parameters for learning are somewhat defined, but are malleable with appropriate context and support.
Degree of Privilege	Economic status Race Culture Support System Language Experience	Students from low economic backgrounds, and representing races, cultures, and languages not in positions of power, face greater school challenges. Quality of students' adult support system influences learning. Breadth/depth of student experience influence learning.
Positioning for learning	Adult models Trust Self-concept Motivation Temperament Interpersonal Skills	Parents who actively commend education positively affect their children's learning. Trust, positive self-concept, positive temperament, and motivation to learn positively impact student learning. Positive interpersonal skills and "emotional intelligence" positively impact student learning.
Preferences	Interests Learning Preferences Preferences for individuals	Student interests will vary across topics and subjects. Students will vary in preference for how to take in and demonstrate knowledge. Students will relate to teachers differently.

(Tomlinson, 17)

Secondly, teachers must continuously assess the success of their teaching strategies as well as the progress of their students. Proponents of Differentiated Instruction stress that there must be ongoing assessment of the students preparedness *before* designing a lesson plan, evaluation of the lesson's effectiveness and student growth *during* the implementation of the lesson (with time for mid-lesson adjustments), and a cumulative measure of student understanding at the *completion* of the lesson. Not all of these assessments should be formal, traditional written exams or count as graded assignments. Most importantly, these assessments should provide both the teachers and the students with an accurate, timely, and useful measure of their progress toward understanding the concepts and execution of applications.

Finally, teachers implementing Differentiated Instruction techniques are encouraged to collaborate with their colleagues and provide their students opportunities to cooperate with each other. By working together, students can benefit from the diversity of thought and approach found among their peers. As a teacher, using Differentiated Instruction in the classroom can be a rewarding experience. However, Differentiated Instruction techniques can also be much more challenging and time consuming than traditional teaching methods. Many resources and support networks are available for educators using differentiated teaching strategies, including ways to partner with other teachers to provide more support opportunities for students.

Most of the available Differentiated Instruction literature focuses on presenting teachers with motivation, resources, and strategies for implementing Differentiated Instruction in their classrooms. However, research on the merits of using differentiated instruction techniques is limited. Dr. Tomlinson and other educators recognize that the "package" of Differentiated Instruction "is lacking empirical validation. There is an acknowledged and decided gap in the literature in this area and future research is warranted." (Hall, et.al.) In particular, research about the implementation of Differentiated Instruction strategies in secondary and post-secondary settings is very scarce. That said, there are many testimonials and small classroom studies indicating that students in Differentiated Instruction settings demonstrate achievement gains, and have a more positive attitude about education (Tomlinson, 184). Also, there is much theoretical support for Differentiated Instruction in its ties to accounting for differences in biological cognitive science, multiple intelligences, and learning styles (Subban, 15).

The Emerging Field of Educational Data Mining

The Educational Data Mining (EDM) community website, www.educationaldatamining.org, defines educational data mining as follows: "Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in (Baker, 3)." General Data Mining and data mining techniques are fairly well established fields with theoretical support going back to the 1950's and practical applications being explored as early as the 1970's. However, these techniques were not applied to the field of education until the early to mid-1990's, when computers became more common in the classroom and universities began offering online courses.

While much of educational research is built on theory based in classroom observation and case studies, EDM researchers aspire to provide solid justification for these theories rooted in large amounts of quantitative analysis. Proponents of EDM emphasize the need to have shared data and verifiable results, providing the basis for a community of researchers. The EDM community began organizing in the early 2000's and the first international Educational Data Mining conference was held in Quebec in June of 2008 (JEDM website).

In October 2009, the first issue of the Journal of Educational Data Mining (JEDM) was published, and the writers anticipated the publication of the first Handbook of Educational Data Mining (Baker, 4).The first article of the first issue of JEDM, "The State of Educational Data Mining in 2009: A Review and Future Visions" outlines and summarizes the methods, goals, and trends of EDM researchers. The authors use the following classification for Educational Data Mining:

- "Prediction
  - Classification
  - Regression
  - Density estimation
- Clustering
- Relationship
  - Association rule mining
  - Correlation mining
  - Sequential pattern mining
  - Causal data mining
- Distillation of data for human judgment
- Discovery with models"

(Baker, 6)

The first three categories are standard Data Mining methods that are well understood and have been applied to many types of data sets. A more detailed description of Clustering and Correlation mining will be given in the following section. The final two categories make use of Data Mining to support or refine educational models and theory. By analyzing the EDM papers published between 1995 and 2005, it can be noted that the two most common approaches used were Relationship mining (with 26 out of 60 papers) and Prediction (with 17 out of 60) (Baker, 7).

Educational Data Mining differs from traditional data mining mainly in its applications. EDM researchers use the vast amounts of information generated by student-computer interactions to study how students learn. One typical application of EDM is to study and improve software learning systems. Also, in recent years, researchers have been analyzing the interactions between students, their peers, their instructors, and the system while enrolled in distance courses and other computer supported collaborative learning environments. Computer adaptive testing is also an active area of study, especially looking at how students can "game the system." Each of these topics accentuates students' use of and interactions with computers and learning software, and using information about those interactions to improve the learning experience. In recent years, there has been more active research on studying the effectiveness of online courses, using Baysian networks and other approaches to look at changes in behavior over time.

Other areas of research in EDM involve using recorded information to learn more about general student behavior, and not just their connections with technology. In particular, Educational Data Miners have been verifying and improving student models. These student models "represent information about a student's characteristics or state, such as the student's current knowledge, motivation, meta-cognition, and attitudes." (Baker, 6) Education models can be used to predict student behavior, identify areas of improvement, and improve curriculum to better serve student needs. Educational Data Mining has also been used to study the factors of student failure and retention issues.

Data Mining Techniques

Data miners use a variety of techniques to look for patterns and salient information in very large and often complex data sets. Often, traditional statistics and data analysis approaches cannot be used on these data sets because of either the sheer number of points or the high dimensional nature of the information. New algorithms are needed to overcome the challenges of the extremely large scale and high number of independent attributes commonly found in modern data sets. This section describes the unconventional Data Mining techniques used in combination with traditional analysis used to find behavior patterns among Studio College Algebra students.

Motivation and Pre-processing

Data Mining techniques are usually used in the pursuit of one of two distinct goals: predictive modeling or deriving descriptive relationships within the data. Predictive modeling is tied to supervised classification techniques, where unlabeled explanatory or independent variables are used to predict the presence of an attribute known as a target or dependent variable that is not part of the collected data set. Usually this target is a class label that is assigned using a predetermined model. In contrast, cluster analysis and anomaly detection use only inherent relationships within the data set such as proximity or correlation (Tan, 9).

One can measure the proximity of two objects in several different ways, with appropriate choices made depending on the nature of the data set. It is often appropriate and convenient to consider data objects with n attributes as points in *n*-dimensional Euclidean space. The coordinates of the data objects/points are given by the attribute values. In this case, the proximity measure between two points $\mathbf{x}$ and $\mathbf{y}$ is generally the Euclidean distance between the two points (or L_2 norm), given by the following formula:

Equation 2.1 Euclidean Distance between two points

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

Note that other norms may be used as measures of proximity, and the L_2 norm is not appropriate for many data sets such as sets containing only binary attributes. The proximity data between all

points in a set is often organized into a symmetric distance matrix, where both the rows and columns are the data points and the entries the distances between each point (Tan, 68).

Just as it is often useful to calculate the similarity or dissimilarity between two objects, one can also measure the relationship between separate attributes. The correlation between two attributes is a measure of the linearity of the relationship between them. The correlation coefficient between two attributes is a value between -1 and 1, with a value of 1 indicating a perfect positive linear relationship between the two attributes, and a value of -1 indicating at perfect negative linear relationship. A value close to zero indicates there is no linear relationship between the two attributes, although a non-linear relationship may exist. The correlation coefficient of attributes **v** and **w** in an *m* point data set is calculated using the following equation:

Equation 2.2 Correlation Coefficient

$$corr(\mathbf{v},\mathbf{w}) = \frac{\mathrm{cov}(\mathbf{v},\mathbf{w})}{s_v * s_w},$$

where,

Equation 2.3 Covariance

$\mathrm{cov}(\mathbf{v},\mathbf{w}) = \frac{1}{m-1}\sum_{k=1}^{m}(v_k - \overline{v})(w_k - \overline{w})$ is the covariance of **v** and **w**,

and

Equation 2.4 Standard Deviation

$s_v = \sqrt{\frac{1}{m-1}\sum_{k=1}^{m}(v_k - \overline{v})^2}$ is the standard deviation of **v** (and similarly **w**).

As with distances between objects, correlation coefficients between attributes are often organized in a correlation matrix (Tan, 77). Highly correlated attributes can be interpreted as providing redundant information about the data points. In clustering methods, one could then reduce the dimensionality of the data without losing important relationships set by disregarding these superfluous attributes.

Often in order to accurately compare the similarity or dissimilarity of objects having more than one attribute, the data must be transformed prior to running data mining algorithms. One such technique for ensuring that one set of attributes with large values does not dominate the results of a calculation is called standardization. During this process, a set of attribute values undergoes the transformation given in the following equation:

Equation 2.5 Standardization of a Variable

$$x' = (x - \bar{x}) / s_x ,$$

where $\bar{x}$ is the mean of the attribute values and s_x is their standard deviation.

The mean (average) value for this transformed data set is now equal to 0, while the standard deviation is equal to 1. In this way, attributes with vastly different scales can be compared with large scale data being valued as more important than data without smaller scale data (Tan, 65). For example, online homework scores in Studio College Algebra are measured on a 10 point scale, while examination questions are worth 5 points each. Without standardization, the difference between students who earned scores of 8 and 10, respectively, on their written homework would be calculated to be twice as much as students who earned scores of 4 and 5 on their exam problem. Thus the homework problem scores would dominate the measure of similarity of these students, when in reality exams scores have equal importance.

Clustering Techniques

Cluster analysis refers to grouping objects based solely on information from the data that describes the objects and their relationships. The main goal is to sort objects into groups so that each object in that group is similar to other members of the group, but dissimilar from objects belonging to other groups. Much of the time, the idea of a cluster is not well defined. The number, size, and make-up of clusters are subject to interpretation of the data and the relative importance of different relationships. In addition, depending on the situation, an object might naturally belong to one cluster exclusively, more than one cluster simultaneously, or have weighted assignments to more than one cluster. This latter clustering assignment setup is known as fuzzy clustering, where each object belongs to a given cluster with a probability between 0 and 1. The three techniques referred to in this paper employ exclusive clustering (Tan, 490-492).

K-means and K-medoids

One of the oldest and most well understood clustering algorithms is the prototype-based, partitional clustering technique known as K-means. This algorithm uses the model of a cluster as "a set of objects in which each object is closer (more similar) to the prototype that defines the cluster than to the prototype of any other cluster." (Tan, 494) K-means defines a cluster prototype as the centroid of a group of points in continuous *n*-dimensional space.

The algorithm for the basic K-means clustering technique is simple and has few steps. First, the user chooses K initial centroids, with K being a predetermined number of clusters. These starting centroids can be chosen arbitrary or intentionally well-separated to avoid non-optimal clusterings. Next, each point is assigned to its closest centroid, forming K initial clusters. Then, the centroids of these initial clusters are updated. The cycle continues by disbanding the former clusters and assigning each point to its closest (updated) centroid. The process is repeated until the centroids do not change, or equivalently, no data points are reassigned to a different cluster (Tan, 497-499).

For data in Euclidean space, we use Euclidean distance (or the L_2 norm) given by Equation 1 above as our proximity function. Determining optimal clustering assignment is expressed by minimizing an objective function. The most commonly used objective function for K-means is the sum of the squared error (SSE). If c_i represents the centroid of the cluster C_i, then the SSE of the clustering can be expressed by the following equation:

Equation 2.6 Total Sum Squared Error Formula

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2$$

Note the SSE of a cluster is found using the formula above, but without summing over every cluster. Once the proximity and objective functions are specified, the centroid of a cluster is defined as the point in *n*-dimensional space that minimizes the objective function of that cluster. The following proof shows that using Euclidean distance and SSE as our proximity and objective functions, respectively, forces the centroids of the clusters to be their geometric mean (Tan, 514).

Proposition 1.1

The centroid that minimizes the SSE of a cluster is its arithmetic mean.

Proof:

The Sum Squared Error of a size m cluster, *C*, can be written as a function of its centroid, $\boldsymbol{c}$. A local minimum of this function will occur if the functions' first partial derivatives exist and equal zero. Therefore, we can find the minimizing centroid by differentiating the SSE, setting it equal to zero, and solving for $\boldsymbol{c}$. Let c_j be the jth coordinate of $\boldsymbol{c}$. Then,

$$\frac{\partial SSE}{\partial c_j} = \frac{\partial}{\partial c_j}\sum_{\mathbf{x}\in C}\left(\sqrt{\sum_{k=1}^{n}(c_k - x_k)^2}\right)^2 = 0 \Rightarrow$$

$$\frac{\partial}{\partial c_j}\sum_{\mathbf{x}\in C}\sum_{k=1}^{n}(c_k - x_k)^2 = 0 \Rightarrow$$

$$\sum_{\mathbf{x}\in C}\sum_{k=1}^{n}\frac{\partial}{\partial c_j}(c_k - x_k)^2 = 0 \Rightarrow$$

$$\sum_{\mathbf{x}\in C} 2(c_j - x_j) = 0 \Rightarrow$$

$$\sum_{\mathbf{x}\in C} 2c_j - \sum_{\mathbf{x}\in C} 2x_j = 0 \Rightarrow$$

$$mc_j = \sum_{\mathbf{x}\in C} x_j \Rightarrow$$

$$c_j = \frac{1}{m}\sum_{\mathbf{x}\in C} x_j$$

Because the choice of the coordinate j was arbitrary, if the jth coordinate of the centroid ***c*** is the mean of the jth coordinates of all the points in the cluster *C*, then the centroid ***c*** must be the geometric mean of the points in the cluster C:

Equation 2.7 Cluster Centroid

$$\mathbf{c} = \frac{1}{m}\sum_{\mathbf{x}\in C}\mathbf{x} \qquad \square$$

Using K-means has many theoretical and practical advantages, as well as some strong disadvantages. By not using a global approach, i.e. calculating all possible clustering scenarios and choosing the optimal one, K-means greatly simplifies and speeds up the calculations involved. The K-means algorithm minimizes the SSE through a gradient descent technique, which starts with an initial solution and then computes the change that would best optimize the objective function. By using Euclidean distance and SSE as the optimizing function, this process will eventually converge to a locally optimal solution (Tan, 498). Unfortunately, this solution is not guaranteed to be globally optimal. However, because of the efficiency of the algorithm, one can ameliorate this problem by running the program several times with different initial centroids and then choosing the clustering with the minimum total SSE (Tan, 502-503).

Other problems prove more difficult to solve. For example, because K-means is based on Euclidean distance, the ideal clusters are Euclidean balls of similar density. If the data set is not naturally made of these ideal clusters, it will be difficult to identify them with this algorithm (Tan, 510). Also, because all distances are squared in the SSE optimization function, outliers have an unduly heavy influence on the formation of clusters (Tan, 506). Finally, the K-means algorithm is not helpful in determining the number of natural clusters a data set contains. Because of these fundamental flaws, a modified version of this algorithm, K-medoids was employed in this study.

The K-medoids algorithm, also known as Partitioning About Medoids (PAM), uses points *within* the data set as representative objects of clusters rather than a geometric center. In this scenario, only the dissimilarities between the points are considered rather than their geometric position. Thus the optimal clusters are chosen to minimize the average dissimilarity between the points in the cluster and their representative point, or medoid. As in K-means, the dissimilarity between two data points is the Euclidean distance between them. However, using the average dissimilarity rather than the Sum Squared Error as the objective function makes PAM significantly different. The algorithm used in PAM is slightly more complex than that of K-means, but this method amends many of the practical flaws of the K-means clustering method.

The PAM algorithm contains two phases: BUILD and SWAP. During the BUILD phase, the medoids are initially assigned to maximize the chance of rapidly obtaining an optimal clustering. Using D_j and E_j to denote object j's dissimilarity with its closest and next closest medoid, respectively, the BUILD algorithm can be summarized thusly:

Step 1: Select the first medoid to be the object for which the sum of the dissimilarities to all other objects is smallest.

Step 2: Consider two non-selected objects i and j. Calculate object j's dissimilarity D_j with the most similar previously selected medoid, and dissimilarity $d(i,j)$ with the object i. If the difference between D_j and $d(i,j)$ is positive, then j will contribute to the decision to select object i as the next medoid. So, we need to calculate $C_{ji} = \max(D_j - d(j,i),0)$.

Step 3: Calculate the total gain obtained by selecting object i as the next medoid: $G_i = \sum_j C_{ji}$.

Step 4: Choose as the next medoid the not yet selected object i which maximizes G_i

Continue the process until K initial medoids have been assigned (Kaufman, 102-103).

During the SWAP phase of the algorithm, the clustering is improved by comparing the effects of "swapping" all pairs of objects (i,h) for which i is a selected medoid and h is not. Note that the value of a clustering is measured by the sum of dissimilarities between each object and the most similar medoid. So, we need to determine the effect on the sum of dissimilarities made by swapping each pair of objects (i,h) and then decide which pair, if any to swap (Kaufman, 103).

The calculations of the effect of a swap between medoid i and unselected point h is calculated as follows:

Step 1: Consider a non-selected object j and calculate its contribution C_{ijh} to the swap:

a. If j is more dissimilar from both i and h than from another medoid, C_{ijh} is zero.
b. If j is not further from i than from any other medoid, i.e. $d(j,i) = D_j$, then two situations must be considered:
    i. If j is closer to h than to the second closest medoid, i.e. $d(j,h) < E_j$, then in this case $C_{ijh} = d(j,h) - D_j$.
    ii. If j is at least as distant from h as from the second closest medoid, i.e. $d(j,h) \geq E_j$, then $C_{ijh} = E_j - D_j$.
c. If j is more distant from medoid i than from at least one of the other medoids, but closer to h than to any of them, then $C_{ijh} = d(j,h) - D_j$. Note, this D_j will not be equal to $d(j,i)$.

Step 2: Calculate the total effect of a swap by adding the contributions of C_{ijh}:

$$T_{ih} = \sum_j C_{jih}$$

Next, it is decided whether to carry out a swap by selecting the pair (i,h) which minimizes T_{ih}. If the minimum T_{ih} is negative, the swap is carried out and the SWAP algorithm returns to Step 1. If the minimum T_{ih} is positive or zero, then the average dissimilarity cannot be decreased by making a swap and so the algorithm ends (Kaufman, 103-104).

Like the K-means algorithm, PAM is guaranteed to converge to a locally optimal solution, though not necessarily a globally optimal one. However, by using the BUILD algorithm to choose the initial medoids, the probability of not finding globally optimal clusters is very small. The main desirable difference between K-means and PAM is that PAM minimizes average dissimilarities rather than sums of squares of distances. This greatly reduces the impact of outliers on the formation of the clusters, ensuring a more robust clustering. Also, the cluster medoids chosen in the PAM algorithm provide representative examples, or prototypes, which are highly useful in many situations. Finally, PAM is able to compute clusters using only the dissimilarities between data objects without regard to their "position" in *n*-dimensional space. Methods like K-means employing objective functions featuring SSE are simpler to compute and take less time to run, but advantages of using PAM far outweigh the computational costs.

Agglomerative Nesting

Agglomerative nesting (AGNES) is another approach to clustering that differs from the partitioning methods described above. The AGNES algorithm starts by considering each point an individual cluster, and then at each step merging the closest pair of clusters until there is one large cluster at the end. Because it would not be useful to only display the final output, the entire process is displayed graphically as a tree-like diagram called a dendrogram (Tan, 515). The sample dendrogram shown below displays the example clustering of four 2-dimensional points. Note the height of the connecting bar indicates the dissimilarity of the two previously linked clusters.

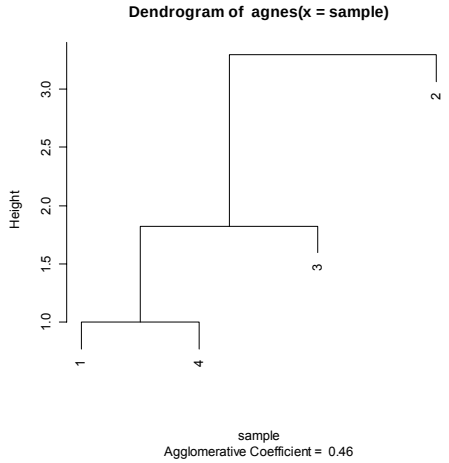

Figure 2.1 AGNES Dendrogram

The variations within the AGNES program come from the chosen definition of "closest clusters." Cluster proximity is usually defined with a specific ideal cluster in mind. For example, several agglomerative nesting techniques use a proximity function focusing on a graph-based view of clusters. Others look at clusters as being represented by a prototype, such as a centroid, as in K-means or PAM (Tan, 517). The two most widely used methods for determining proximity are the group average technique, sometimes called "unweighted pair-group average method" (UPGMA), which employs the former proximity based clustering, and Ward's method, which employs the latter prototype based clustering (Kaufman, 203).

The default version of AGNES run by the statistical program R used in this study determines the proximity of clusters by averaging the pairwise proximity among all the pairs of points in the different clusters (UPGMA). For two clusters C_i and C_j of sizes m_i and m_j, respectively, their proximity is expressed by the following equation (Tan, 522):

Equation 2.8 Group Average Cluster Proximity

$$proximity(C_i, C_j) = \frac{\sum_{\mathbf{x} \in C_i, \mathbf{y} \in C_j} proximity(\mathbf{x}, \mathbf{y})}{m_i * m_j}$$

Ward's method is an alternative approach that defines proximity between two clusters as the increase in Sum Squared Error (SSE) that results from these clusters being merged (Tan, 517). Thus, Ward's method has the same objective function as that of K-means (see Equation 2.6).

Although Ward's method is computationally similar to K-means, and therefore theoretically well known, in practice UPGMA has been shown to be more reliable and robust. One advantage to using UPGMA over Ward's method is that the dissimilarity between merging clusters remain statistically consistent. In other words, as the sample size increases, the dissimilarity calculated by Ward's method blows up to infinity, but UPGMA remains stable. Also, Ward's method only performs well if the natural clusters are of equal diameter and contain an equal number of objects, whereas UPGMA has been shown to be effective in a wide variety of sampling situations (Kaufman, 243).

Because AGNES employs a nearest neighbor approach, it is highly susceptible to minute changes in the positions of the objects, i.e. noise. Even the methods using the most stable proximity measures are not nearly as robust as more globally optimizing methods such as K-means or PAM. AGNES tends to make good local decisions about merging clusters at each

stage, but once that choice to merge has been made, it cannot be undone at a later time. Thus, there is no sense of optimizing a global objective function (Tan, 526). In practice, AGNES is often used to determine the number of naturally occurring clusters in a data set. Then, another, more robust method is used to actually enumerate these clusters.

Dimension Reduction

The standard clustering algorithms AGNES and PAM described above work very well with data in low dimensional settings, but several problems arise applied to high dimensional data sets. The standard geometric properties of density and proximity are intuitively simple in two or three dimensions, but behave strangely in higher dimensional settings. Because AGNES and PAM rely on these properties as their foundation for clustering, these algorithms cannot generate meaningful clusters with high-dimensional data. For example, the volume of a hypersphere with radius r is proportional to r^d, where d is dimension. Therefore, as d increases, unless the number of data points in the hypersphere also increases exponentially, its density tends to zero. Also, proximity between points tends to become uniform in high-dimensional spaces. With more attributes, or dimensions, contributing to the distance between two points, there are many more ways for points to be equally distant (Tan, 572).

Even though density is more uniform in a high-dimensional setting, clusters become very unstable. To illustrate this phenomenon, we examine unit balls in several dimensions. In one dimension, a "ball" with $r = 1$ is a line segment of length 2. The outer 10% of the radius comprises two portions of length .1 on either end of the line segment (in red). This outer 10% "shell" makes up 10% of the total length of the ball.

Figure 2.2 1-D Ball

In 2 dimensions, a ball with radius 1 has area equal to π, while the shell made of the outer 10% of the radius has an area equal to $\pi - \pi(.9)^2$. Therefore, the proportion of area that the outer 10% of the 2-D Ball makes up is 17%.

$$\frac{\pi - \pi(.9)^2}{\pi} = 1 - (.9)^2 = 1 - .81 = .17$$

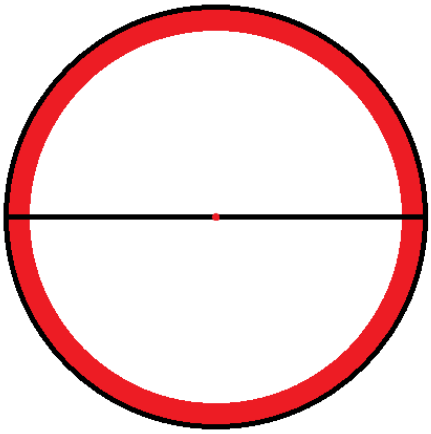

Figure 2.3 2-D Ball

In general, for any *n*-dimensional Euclidean ball of radius *r*, the shell made from the outer 10% of the radius has a volume proportional to r^n. The proportion of the *n*-dimensional volume contained in the outer 10% shell can be found using the following equation:

Equation 2.9 Proportion of Volume in Outer 10% of n-D ball

$$P_{shellvolume} = \frac{a_n r^n - a_n (.9r)^n}{a_n r^n} = 1 - (.9)^n$$

Thus, for large dimensional spaces, most of the volume of an *n*-D ball is contained in the outer rim. This makes clusters based on Euclidean balls highly unstable. As one can see in the figure below, when most of the data is along the outer rim of a cluster, a small change in the center of the cluster results in a large number of points being reassigned.

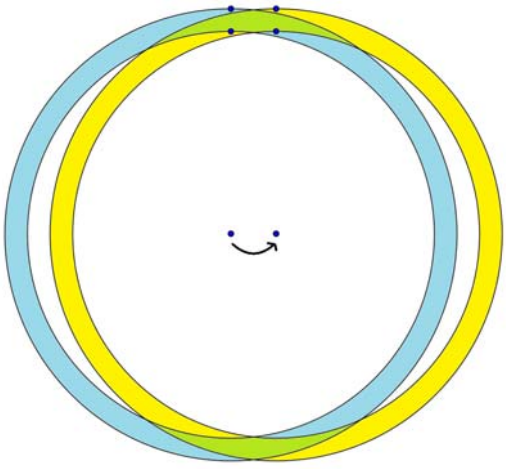

Figure 2.4 Cluster instability in high dimensions

Because of these inherent problems trying to find clusters in high dimensional data sets, usually a procedure is carried out to reduce the dimension of the data set while somehow retaining most of the pertinent information. Each of these methods reduces dimensionality by building new attributes from linear combinations of the original attributes.

White Box/Expert Reduction

The simplest and most widely used method of dimension reduction involves using "expert" knowledge to combine relevant attributes or discard ones known to be unimportant. For example, instead of considering attendance for each class as a separate attribute, one could combine individual attendance points to form several "Weekly Attendance" vectors or even a single "Average Attendance" or "Total Attendance" vector. This method requires many choices to be made by someone who has prior expert knowledge about the data set. Unfortunately, this allows for bias and human error to more heavily influence later cluster analysis. Also, by using prior knowledge to reduce dimension, one might overlook interesting and unexpected relationships in the data.

Singular Value Decomposition

The goal of dimension reduction is to find the projection from *m*-dimensional Euclidean space ($\Re^m$) to a lower dimensional space (i.e. $\Re^1$ to $\Re^5$ or so) preserving as much interesting information as possible. For most Data miners, "interesting patterns" come from data points that are varied. For example, if every student has the same score on an assignment, that assignment does not tell us much about different groups of students. Therefore, our mathematical goal for dimension reduction would be to find the best projection preserving greatest variance.

Proposition 2.1: Let **A** be an $n \times m$ matrix with standardized column vectors. Let $\vec{x} \in \Re^m$ have length 1. The projection $\Re^m \to \Re^1$ given by $a \mapsto a \cdot \vec{x}$, where a is a row of **A**, that maximizes the variance of $a \cdot \vec{x}$ occurs when $\vec{x}$ is an eigenvector of $\mathbf{A}^T\mathbf{A}$ corresponding to the largest eigenvalue λ_1.

Proof:

Because each column of **A** has been standardized, $mean(a \cdot \vec{x}) = 0$. Therefore, $\text{var}(a \cdot \vec{x}) = E\left|a \cdot \vec{x}\right|^2 = \left\|a \cdot \vec{x}\right\|_2^2$. Then if we want to maximize the variance of $a \cdot \vec{x}$, we need to maximize the following expression:

$$(\mathbf{A}\vec{x})\cdot(\mathbf{A}\vec{x}) = (\mathbf{A}\vec{x})^T(\mathbf{A}\vec{x}) = (\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x}).$$

So our goal could be written as the following expression:

Equation 2.10 Maximizing Variance

$$\max_{|x|=1}(\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x})$$

Because $\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x}$ is a smooth function on a compact subset of $\Re^m$, we can use Lagrange multipliers to find the critical points. So, introducing λ as an $(m + 1)$th variable, we want to take the partial derivatives of the following equation.

Equation 2.11 Lagrange Multiplier method for maximizing variance over first vector

$$\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x} - \lambda(|\vec{x}| - 1) = 0$$

Let us write $\mathbf{A}^T\mathbf{A} = (c_{ij})$, where $c_{ij} = c_{ji}$. Then we can expand Equation 11 into

$$(x_1 \quad \ldots \quad x_m)\begin{pmatrix} c_{11} & \ldots & c_{1i} & \ldots & c_{1m} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \ldots & c_{ii} & \ldots & c_{im} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \ldots & c_{mi} & \ldots & c_{mm} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} - \lambda(x_1^2 + \ldots + x_m^2 - 1)$$

$$= (x_1 \quad \ldots \quad x_m)\begin{pmatrix} c_{11}x_1 + \ldots + c_{1i}x_i + \ldots + c_{1m}x_m \\ \vdots \\ c_{1i}x_1 + \ldots + c_{ii}x_i + \ldots + c_{im}x_m \\ \vdots \\ c_{1m}x_1 + \ldots + c_{im}x_i + \ldots + c_{mm}x_m \end{pmatrix} - \lambda(x_1^2 + \ldots + x_m^2 - 1)$$

$$= \sum_{i,j} c_{ij}x_i x_j - \sum_{j=1}^{m} \lambda x_j^2 + \lambda$$

$$= 2\sum_{i<j} c_{ij}x_i x_j + \sum_{i=1}^{m} c_{ii}x_i^2 - \sum_{j=1}^{m} \lambda x_j^2 + \lambda$$

Then, for each $i \in (1,\ldots,n)$,

$$\frac{\partial}{\partial x_i}\left(2\sum_{i<j} c_{ij}x_i x_j + \sum_{i=1}^{m} c_{ii}x_i^2 - \sum_{j=1}^{m}\lambda x_j^2 + \lambda\right) = 0 \Rightarrow$$

$$2\sum_{j=1}^{m} c_{ij}x_i - 2\lambda x_i = 0 \Rightarrow$$

$$\sum_{j=1}^{m} c_{ij}x_i = \lambda x_i$$

and

$$\frac{\partial}{\partial \lambda}\left(\sum_{i,j} c_{ij}x_i x_j - \sum_{j=1}^{m}\lambda x_j^2 + \lambda\right) = 0 \Rightarrow$$

$$|\vec{x}|^2 = 1$$

Therefore, $\mathbf{A}^T\mathbf{A}\vec{x} = \lambda\vec{x}$, meaning $\vec{x}$ is an eigenvector of $\mathbf{A}^T\mathbf{A}$ with eigenvalue λ.

Note that

$$|\mathbf{A}\vec{x}| = \sqrt{\mathbf{A}\vec{x}\cdot\mathbf{A}\vec{x}} = \sqrt{\vec{x}^T\left(\mathbf{A}^T\mathbf{A}\right)\vec{x}} = \sqrt{\vec{x}^T\lambda\vec{x}} = \sqrt{\lambda\vec{x}^T\vec{x}} = \sqrt{\lambda}\,.$$

Therefore, to maximize variance, we should choose the eigenvector corresponding to the largest eigenvalue, λ_1. □

We have found the principal component that maximizes the variance of $a\cdot\vec{x}$, but now we need to find the next vector that retains the maximum amount of variance while being orthonormal to this principal component, $\vec{v}_1$.

Proposition 2.2: Fix $\vec{v}_1, \lambda_1$, where $\mathbf{A}^T\mathbf{A}\vec{v}_1 = \vec{v}_1\lambda_1$, with $|\vec{v}_1| = 1$, to be the optimal variance preserving vector described above. Then, the vector $\vec{x}$ that maximizes $\text{var}(a\cdot\vec{x})$ subject to the constraints $|\vec{x}| = 1$ and $\vec{x}\cdot\vec{v}_1 = 0$ is an eigenvector corresponding to the second largest eigenvalue, λ_2, of $\mathbf{A}^T\mathbf{A}$.

Proof:

As in Proposition 2.1, we see that $\text{var}(a\cdot\vec{x}) = (\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x})$, and so we can use Lagrange multipliers with two constraints to find its critical points. We need to take the partial derivatives of the following equation:

Equation 2.12 Lagrange Multiplier method for maximizing variance over second vector

$$\vec{x}^T(\mathbf{A}^T\mathbf{A})\vec{x} - \lambda(|\vec{x}| - 1) - \mu(\vec{v}_1^T\vec{x}) = 0$$

Using the same steps found in Proposition 2.1, we can show that this equation is equivalent to:

$$2\sum_{i<j} c_{ij}x_i x_j + \sum_{i=1}^{m} c_{ii}x_i^2 - \sum_{j=1}^{m} \lambda x_j^2 + \lambda - \sum_{i=1}^{m} \mu v_{1i}x_i = 0$$

Then, for each $i \in (1,\ldots,n)$,

$$\frac{\partial}{\partial x_i}\left(2\sum_{i<j} c_{ij}x_i x_j + \sum_{i=1}^{m} c_{ii}x_i^2 - \sum_{j=1}^{m} \lambda x_j^2 + \lambda - \sum_{i=1}^{m} \mu v_{1i}x_i\right) = 0 \Rightarrow$$

$$2\sum_{j=1}^{m} c_{ij}x_i - 2\lambda x_i - \mu v_{1i} = 0 \Rightarrow \quad ,$$

$$2(\mathbf{A}^T\mathbf{A})\vec{x} - 2\lambda\vec{x} - \mu\vec{v}_1 = 0$$

$$\frac{\partial}{\partial \lambda}\left(\sum_{i,j} c_{ij}x_i x_j - \sum_{j=1}^{m} \lambda x_j^2 + \lambda - \sum_{i=1}^{m} \mu v_{1i}x_i\right) = 0 \Rightarrow \quad ,$$

$$|\vec{x}|^2 = 1$$

and

$$\frac{\partial}{\partial \mu}\left(\sum_{i,j} c_{ij}x_i x_j - \sum_{j=1}^{m} \lambda x_j^2 + \lambda - \sum_{i=1}^{m} \mu v_{1i}x_i\right) = 0 \Rightarrow$$

$$\vec{v}_1^T\vec{x} = 0$$

Therefore,

$$2(\mathbf{A}^T\mathbf{A})\vec{x} - 2\lambda\vec{x} - \mu\vec{v}_1 = 0 \Rightarrow$$
$$2\vec{v}_1^T(\mathbf{A}^T\mathbf{A})\vec{x} - 2\lambda\vec{v}_1^T\vec{x} - \mu\vec{v}_1^T\vec{v}_1 = 0 \Rightarrow$$
$$2(\mathbf{A}^T\mathbf{A}\vec{v}_1)^T\vec{x} - 2\lambda\vec{v}_1^T\vec{x} - \mu\vec{v}_1^T\vec{v}_1 = 0 \Rightarrow$$
$$2(\lambda_1\vec{v}_1)^T\vec{x} - 2\lambda\vec{v}_1^T\vec{x} - \mu\vec{v}_1^T\vec{v}_1 = 0 \Rightarrow \quad ,$$
$$2\lambda_1\vec{v}_1^T\vec{x} - 2\lambda\vec{v}_1^T\vec{x} - \mu\vec{v}_1^T\vec{v}_1 = 0 \Rightarrow$$
$$0 - 0 - \mu\cdot 1 = 0 \Rightarrow$$
$$\mu = 0$$

and so $2\mathbf{A}^T\mathbf{A}\vec{x} - 2\lambda\vec{x} = 0 \Rightarrow$
$\mathbf{A}^T\mathbf{A}\vec{x} = \lambda\vec{x}$.

Once again, $\vec{x}$ is an eigenvector with eigenvalue λ. Then eigenvector $\vec{v}_2$ that will maximize variance will be the one with the second largest eigenvalue, λ_2. □

We can repeat the process in Proposition 2.2 of finding new unit vectors that maximize the variance of $a \cdot \vec{x}$ while being orthogonal to all preceding vectors. This set $(\vec{v}_1,\ldots,\vec{v}_m)$ forms a new orthonormal basis for $\Re^m$. Note that if **A** is a matrix of full rank, then each eigenvalue λ_i will be nonzero. Now, define vectors $\vec{u}_i := \frac{\mathbf{A}\vec{v}_i}{\sqrt{\lambda_i}} \in \Re^n$. Each vector $\vec{u}_i$ has length 1, as $|\mathbf{A}\vec{v}_i| = \sqrt{\lambda_i}$. Also, for each $i \neq j$,

$\vec{u}_j \cdot \vec{u}_i = \frac{\mathbf{A}\vec{v}_j}{\sqrt{\lambda_j}} \cdot \frac{\mathbf{A}\vec{v}_i}{\sqrt{\lambda_i}} = \frac{1}{\sqrt{\lambda_j\lambda_i}}\vec{v}_j(\mathbf{A}^T\mathbf{A}\vec{v}_i) = \frac{1}{\sqrt{\lambda_j\lambda_i}}\vec{v}_j\lambda_i\vec{v}_i = \sqrt{\frac{\lambda_i}{\lambda_j}}\vec{v}_j\vec{v}_i = 0$. Therefore, $(\vec{u}_1,\ldots,\vec{u}_m)$ is also an orthonormal set. We can extend this set to an orthonormal basis for $\Re^n$, $(\vec{u}_1,\ldots,\vec{u}_n)$.

Definition 2.1: Let **A** be an $n \times m$ matrix of full rank. Define **U** to be an $n \times n$ matrix with row vectors $(\vec{u}_1,\ldots,\vec{u}_n)$ and **V** to be an $m \times m$ matrix with row vectors $(\vec{v}_1,\ldots,\vec{v}_m)$ as given above. Let **S** be an $n \times m$ diagonal matrix with each *ith* diagonal entry equal to $\sqrt{\lambda_i}$. These values are known as the singular values. Then the **Full Singular Value Decomposition** of the matrix **A** is given by:

Equation 2.13 Full Singular Value Decomposition (SVD)

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

For $(\vec{e}_1,\ldots,\vec{e}_n)$ the standard basis for $\Re^n$ the following diagram shows the linear transformations represented by the matrix multiplication in the SVD:

$$\begin{array}{ccc} \underset{\vec{e}_1,\ldots,\vec{e}_m}{\Re^m} & \overset{\mathbf{A}}{\rightarrow} & \underset{\vec{e}_1,\ldots,\vec{e}_n}{\Re^n} \\ \mathbf{V} \uparrow\downarrow \mathbf{V}^T & & \uparrow \mathbf{U} \\ \underset{\vec{v}_1,\ldots,\vec{v}_m}{\Re^m} & \underset{\mathbf{S}}{\rightarrow} & \underset{\vec{u}_1,\ldots,\vec{u}_n}{\Re^n} \end{array}$$

Figure 2.5 SVD Transformations

Factor Interpretation of Singular Value Decomposition

We can interpret the columns of **V** (or the rows of $\mathbf{V}^T$) as new factors that are linear combinations of the original attributes ordered in such a way that preserves the maximum data variance. Each diagonal value of **S** gives a measure of how much variance is captured by the associated vector $\vec{v}_i$. The rows of **U** multiplied by the eigenvalues λ_i give each data point's

coordinates in the new basis $(\vec{v}_1,\ldots,\vec{v}_m)$. Note that because $(\vec{v}_1,\ldots,\vec{v}_m)$ is an orthonormal set, the distances between the data points have been preserved by transformation (Skillicorn, 55).

For dimension reduction purposes, the diagonal values of **S** are often plotted as in the following figure:

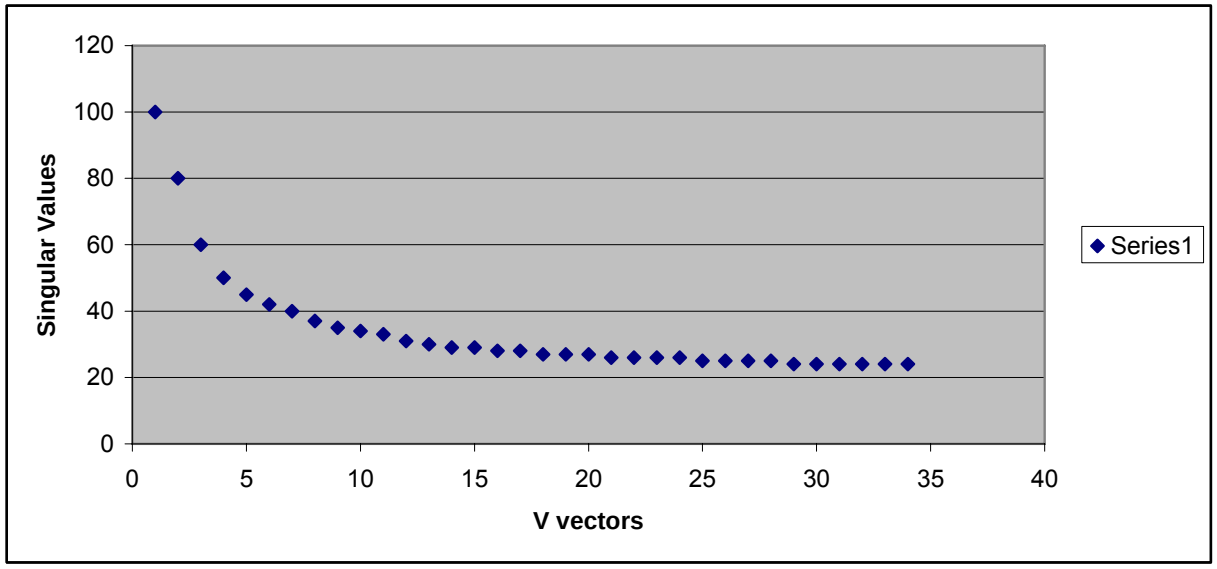

Figure 2.6 Singular Values

One would assume that the smaller singular values represent vectors that capture mostly noise and not much of the relevant data patterns. Then one can safely truncate the matrix $\mathbf{V}^T$, cutting off the last *m-r* columns and retain most of the variation. Traditionally, the number of columns of $\mathbf{V}^T$ that are retained is chosen by examining the singular value plot (given above) for an "elbow," or a value where the rate of change goes from a sharp decline to a more gradual one. In this case, the elbow occurs between $\vec{v}_4$ and $\vec{v}_6$. Then it would be reasonable to truncate $\mathbf{V}^T$ to an $m\times 5$ matrix.

In practice, the decision to truncate to a specific number of dimensions is often based more on trial and error than any nice theoretical cutoff. Most real data sets do not produce singular values with a nice elbow, as seen in the example from Fall 2010's student scores shown below.

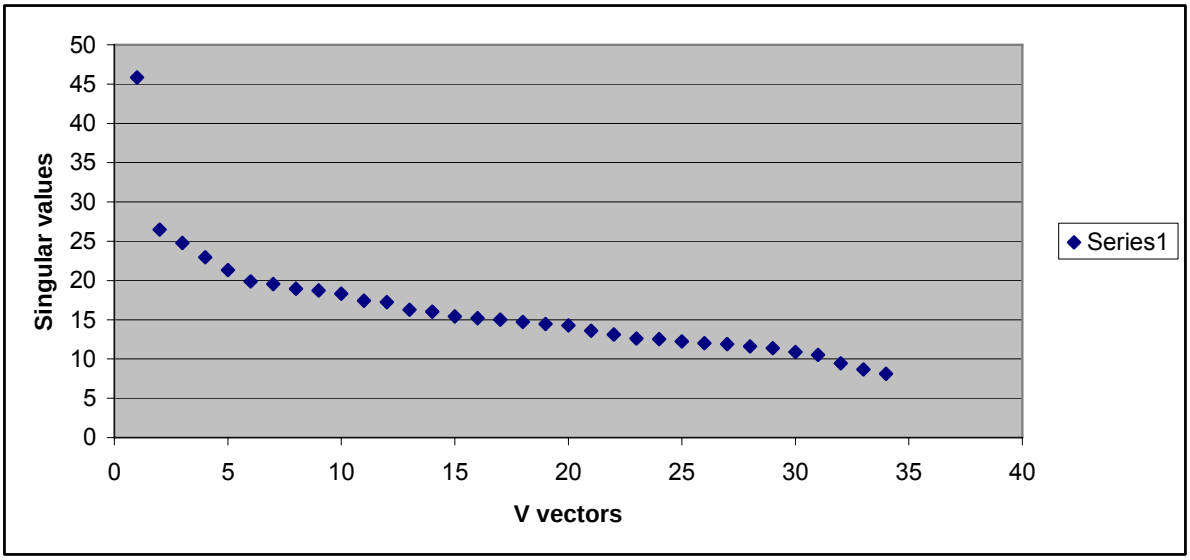

Figure 2.7 Singular Values from Fall 2010 Data

The reduced-dimension SVD of a matrix **A** can be given by the following equation:

$$\mathbf{A}_r = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r^T,$$

where $\mathbf{A}_r$ is still an $n \times m$ matrix, but $\mathbf{U}_r$ has been truncated to be an $n \times r$ matrix, $\mathbf{V}_r^T$ is now an $r \times m$ matrix, and $\mathbf{S}_r$ is still diagonal with r nonzero entries. The traditional clustering algorithms are then applied to the entries of $\mathbf{U}_r$, which are the coordinates in the reduced space given by $span(\vec{v}_1, \ldots, \vec{v}_r)$.

While Singular Value Decomposition is an elegant and computationally manageable way to reduce dimension while retaining important information, there are some limitations. Most importantly, SVD does not preserve sparseness. If **A** is a very large matrix with mostly zero entries (i.e. sparse), it is still possible to perform many calculations by storing the values and positions of the nonzero entries of **A** rather than storing the whole matrix. However, the **U** and **V** matrices of the SVD of a sparse matrix **A** will likely not be sparse matrices and thus too large to manage. Therefore, this method will not be computationally feasible for many data sets, including document data and website analysis.

Principal Component Analysis

Principal Component Analysis (PCA) is an approach more commonly used by statisticians to identify a new set of attributes that captures the variability of the data. As with

SVD analysis, PCA matrix decomposition to find new vectors that more efficiently capture the variation in the data. However, unlike SVD, PCA relies on the covariance of the different attributes (see Equation 2.3) to determine which attributes are redundant and which ones can safely be combined while retaining the maximum amount of variation in the data. The covariance matrix, **S**, of the *m* by *n* matrix **D,** has entries s_{ij} defined as follows:

Equation 2.14 Covariance matrix

$$s_{ij} = \text{cov}(\mathbf{d}_{*i}, \mathbf{d}_{*j})$$

Covariance matrices are examples of positive semidefinite matrices, defined to be matrices **M** such that, for any $\mathbf{x} \in \Re, \mathbf{xMx}^T \geq 0$. Positive semidefinite matrices have some nice properties, including that all of the eigenvalues are non-negative, which allow for the use of simpler and more efficient decomposition algorithms. This has contributed to its widespread use in the realm of Statistics. However, PCA only takes into account the covariance of the attributes and not the mean of each variable. Singular Value Decomposition is an equivalent analysis to PCA that incorporates this important information (Tan, 702).

Chapter 3 - Finding Patterns in Student Behavior

After researching various clustering methods, it was decided to try a combination of Singular Value Decomposition (SVD), AGglomerative NESting (AGNES), and Partitioning About Medoids (PAM) algorithms to form student groups. First, the 30+ dimensional data set was reduced to one of only four or five dimensions by truncating the orthogonalized coordinate matrix created by the SVD. Then, the statistical package R's algorithm AGNES was used to roughly determine the number of natural clusters in the data set. However, the AGNES algorithm uses a nearest neighbor approach to pair up sets of points, making it unstable and unsuitable for determining the composition of the cluster sets. Therefore, with the appropriate number of clusters known, the PAM program was run to demarcate the student clusters. Although clustering about centroids is easier to work with theoretically, using medoids has the advantage of being less influenced by outliers. These methods were chosen for their simplicity and because the algorithms are clearly understood. After initial testing, it was concluded that using more sophisticated methods such as SemiDiscrete Decomposition (SDD) or density based clustering algorithms was not necessary to establishing clear and useful student clusters.

Spring 2008: Trial Run

To test the effectiveness of these clustering methods and to practice using the procedures, a trial run was held during the summer of 2008 using data from the Spring 2008 semester. Only 99 students were enrolled in the Studio College Algebra class that semester, so the data sample size was too small to include in the study, but ideal for a trial run. It was later determined that advisors had been placing struggling students into the Studio section of the course, so the data sample was not representative of the typical College Algebra student population. During this practice trial, the researcher discovered that there was no clear place to truncate the SVD matrices so that most of the important information was kept, but the noise filtered out. Thus, several clustering trials were performed, using 3, 4, 5, and 6 dimensions of data. The AGNES graphs indicated there were likely 4 or 5 natural clusters, including a group of outliers. Because the SVD, AGNES, and PAM algorithms seemed to work well together to cluster students into

natural groups, plans were made to use these algorithms on student data collected over the next four semesters.

Fall 2008

Before forming student clusters in Fall of 2008, several decisions needed to be made about the various parameters and data sources. First, the researcher needed to choose the student scores (data vectors) to be included in the SVD analysis. In order to make a timely analysis which would benefit students as early in the semester as possible, the researcher decided to use only data collected from the first four weeks of the semester, which included problem scores from the first examination. The Studio College Algebra course instructors collect more scores than a typical class, making it possible to have over 30 assignment scores from the first four weeks alone (See Appendix A: Clusters, Fall 2008). Also, the online homework system has been designed to collect copious amounts of data for research, so every student interacting with the system is recorded. It was theorized that student persistence in completing online homework assignments successfully might contribute to cluster formation, so the number of attempts each student made until they reached a score of at least 90% were recorded. A readiness test covering basic algebra skills and a pretest covering the course material also were administered at the beginning of the semester. Then, for the Fall 2008 semester, the data vectors included for initial SVD analysis were sixteen Exam 1 problem scores, four Written Assignment scores, two Studio assignment scores, ten Attendance points, five Online Homework scores, Inverse time to 90% (ITN) on each of the five online homework assignments, Readiness Test, and Pretest scores.

The matrix **V** formed by the SVD process revealed the contributions of the original assignment vectors in the new orthogonalized system. Each row of **V** can be read as a linear combination of the original assignments. The assignments with higher coefficients contributed more weight to the new vectors, while those with coefficients close to zero did not. By examining the coefficients in matrix **V**, it was determined that the ITN numbers and the readiness and pretest scores did not contribute greatly to the student variation (See Appendix B: Data Analysis, Initial Vectors, Fall 2008). Also, it was known that in subsequent semesters, the pretest and readiness tests would be replaced by an online Math Placement exam to be

administered online during summer enrollment. Thus, the ITN numbers and readiness and pretest scores were removed from the clustering data set.

After the data was collected and organized into a 332 by 37 matrix, the column entries were centered and normalized by assignment using standardization formula found in Chapter 2 (see Equation 2.5). This was done to ensure that every assignment was counted equally in its contributions to the student groups. The researcher considered weighting some scores by multiplying a scale factor to the more "important" assignments. However, introducing an expert opinion on which assignments were more significant than others without quantitative evidence contradicted the goal of using data mining to cluster students. Thus it was decided not to scale the assignments.

After centering and normalizing the 332 by 37 matrix, the program Matlab was used to decompose the matrix **A** into the Singular Value matrices **U**, **S**, and **V**. The matrix **U** contained the coordinate entries in the new vector system, the diagonal matrix **S** the dilation values ordered from highest to lowest, and **V** the linear combinations of the original vectors making up the new vectors (See Chapter 2 for more detail). The dilation matrix **S** quantifies the amount of point variation each new vector captures. Usually, one uses the values of the matrix **S** to determine where to truncate the matrix decomposition. The first 5 diagonal values of the **S** matrix were 59.07, 30.95, 26.38, 25.52, and 23.914. The remaining values decreased slowly, so there was no clear "elbow" at which to make the cut. However, considering the problems that occur with higher dimensional data sets, the decision was made to truncate the decomposition at the 4th dimension so that roughly 65% of the data points in a cluster would be contained in the interior of the cluster sphere (See Chapter 2 for theoretical justification).

After truncation, the matrices **U**, **S**, and $\mathbf{V}^{\mathbf{T}}$ had dimensions 332 by 4, 4 by 4, and 37 by 4, respectively. One could either form clusters based on the coordinates in U or on the stretched coordinates in **US**. By using the coordinates in **US**, the cluster spheres would be compressed; that is, less variation in the direction of first vectors would be included in the same cluster. If we assume that the first vector captures the general success of each student, then a narrower variation of student success levels would be included in each cluster. Because the researcher wanted to capture personality traits and attitudes about mathematics in the clusters as well as overall success in the class, she did not want small changes in grades to affect clustering.

Therefore, the decision was made to group students based on their coordinates in the matrix U alone.

Next, using the statistics program R, an AGNES dendrogram was produced from the truncated 337 by 4 dimensional matrix U to help identify the proper number of clusters. The AGNES dendrogram below seems to indicate that the data set naturally contains either 3 or 4 clusters with a set of outliers.

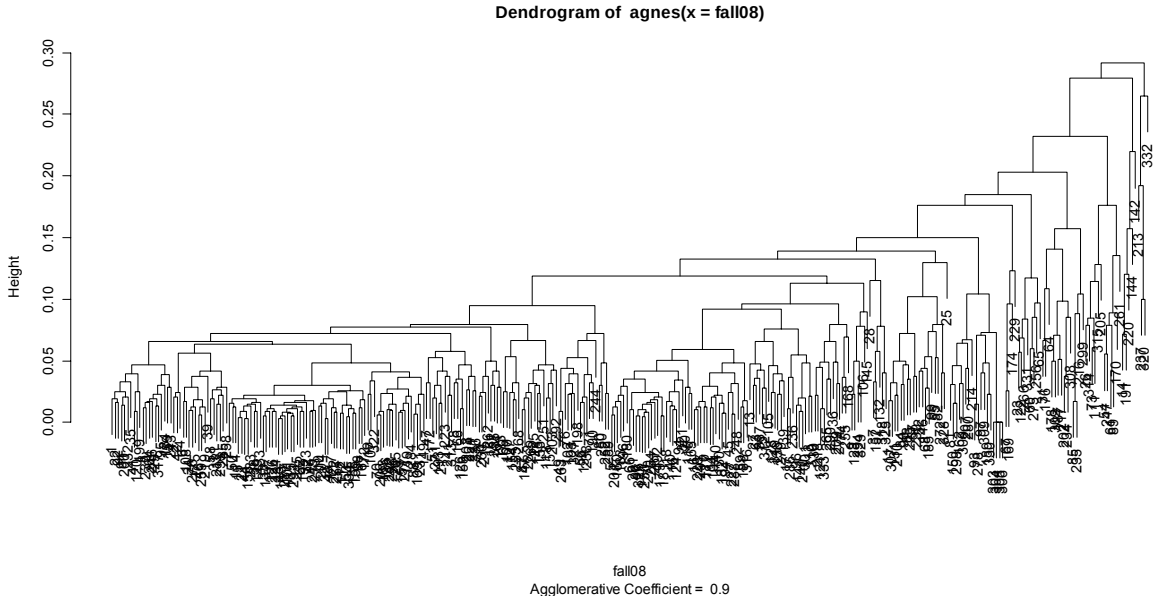

Figure 3.1 AGNES Output: Fall 2008

Because it was unclear whether there were four or five natural clusters in the data set, the PAM program was run twice (with k=4, and with k=5) and the resulting clusters were compared. The chart below shows that Group 1 and Group 2 in both SVD4 and SVD5 contained mostly the same students. Likewise, Group 4 from SVD4 and Group 5 from SVD5 contained the same students (in red). The biggest difference between the two clustering setups was that Groups 3 and 4 in SVD5 were each divided evenly between two different groups in SVD4 setup (in blue). So, three of the clusters look stable, but it is unclear whether to have two more distinct groups, or divide these students among the remaining clusters.

Table 3.1 Comparison Between Four Student Clusters and Five Student Clusters

SVD 5/SVD 4	1	2	3	4
1 (OA)	96	2	15	6
2 (E)	2	52	10	4
3 (UA)	1	15	21	1
4 (SS)	32	3	23	0
5 (RM)	8	4	9	28

To help make this decision, we compare the graphs of the two cluster schemes. The graph below shows the projection of the 4 dimensional points given by **U** onto the first two components:

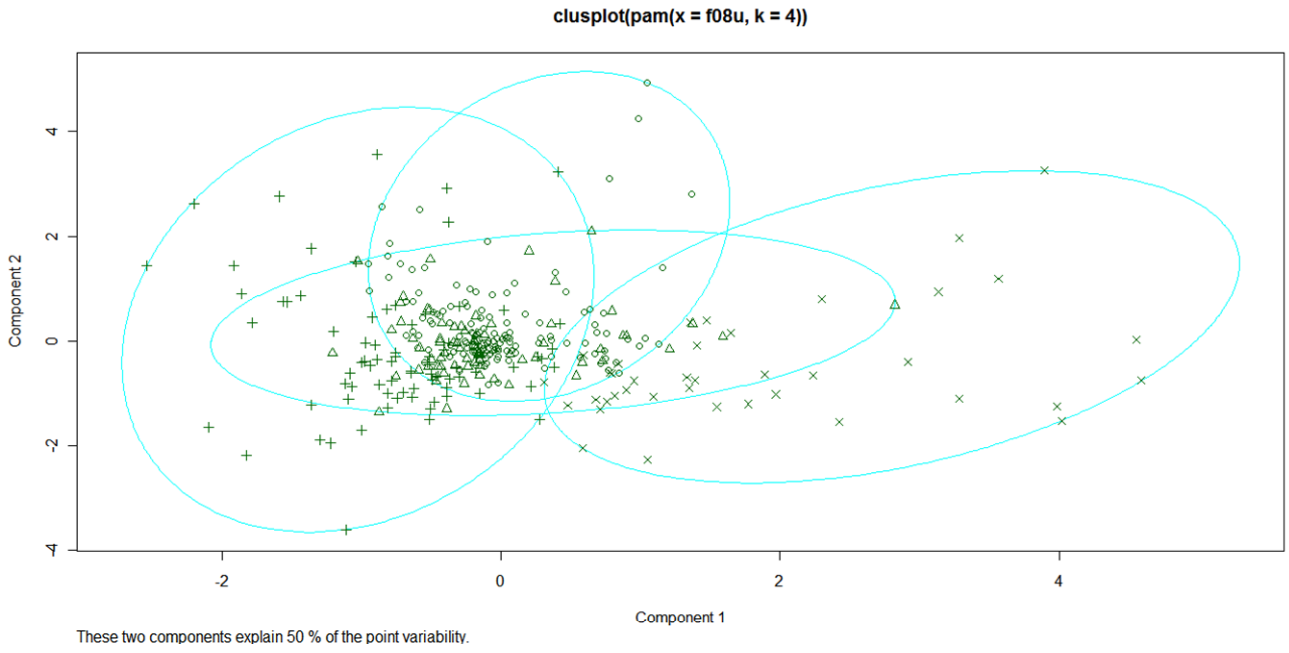

Figure 3.2 PAM Groups Fall 2008: 4 Clusters

Key: Circles- Group 1

Triangles- Group 2

Plus signs- Group 3

X's- Group 4

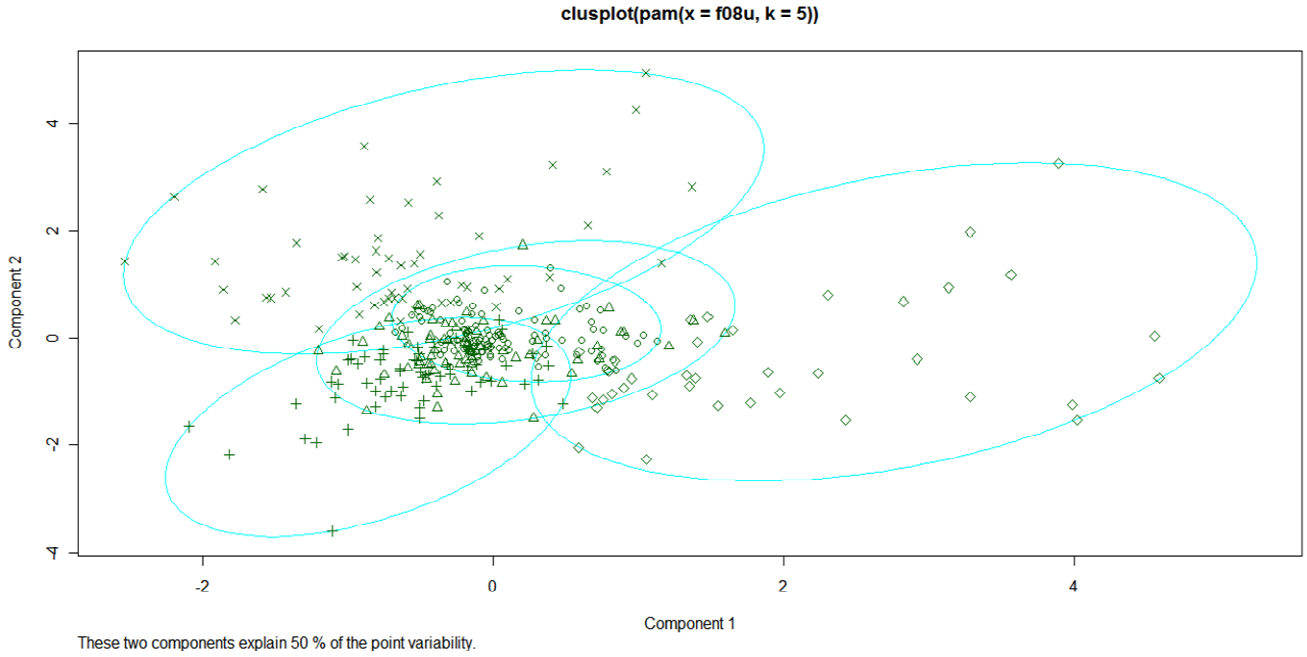

Figure 3.3 PAM Groups Fall 2008: 5 Clusters

Key: Circles- Group 1

Triangles- Group 2

Plus signs- Group 3

X's- Group 4

Diamonds- Group 5

Note that Group 4 from SVD4 and Group 5 from SVD5 are almost identical in the graphs. However, Groups 3 and 4 in SVD5 seem to be a more natural split on the left side of the graph than Groups 1 and 3 in SVD4. Also, Groups 1 and 2 are both tighter in SVD5. This information leads to the conclusion that the data more naturally supports five clusters. Once chosen, this number of clusters remained constant for all following semesters. (See Appendices Fall 2008 PAM4 and Fall 2008 PAM5 for larger graphs).

Spring 2009

In general, class makeup from the fall and spring semesters of any given year are very different from one another. Students who enroll in the fall sections of College Algebra are

usually entering first year students who have taken their most recent math course less than one year prior to enrolling. In the spring semester, many more students are upper classmen who have either delayed taking their college math requirements or have dropped/failed previous attempts. To compare data from fall and spring semesters, the assumption was made that while concentrations may change, the basic characteristics of student groups would not be drastically different. In order to determine the most useful groupings, two variations of the SVD/PAM process were run and compared.

In the first trial, the exact same procedure used in Fall 2008 was applied to the data from Spring 2009. Of the five resulting groups, two were large (77 and 60) and the remaining three were roughly the same size and small (24, 18, 19). These groups represented the highest, second lowest, and lowest scoring students, respectively. The two large groups represented students who had a B average grade on exam 1. One of the large groups did relatively well on the other assignments and maintained their B grade, while the other group did not score well on class assignments and dropped future exam scores. (For more data, see Appendix A: Clusters, Spring 2009)

One could make the assumption that the underlying character traits of students in a given school year do not change. In this case, it would be reasonable to use the same attribute vectors to group students in both the fall and spring semesters. Because there are many more students in the fall, the researcher made the choice to use the attribute vectors from Fall 2008 in forming the groups from Spring 2009. The SVD process broke down the matrix A into 3 component matrices **U**, **S**, and **V**. Remember, **U** represented the coordinates in the new orthogonal space, **S** was the dilation matrix, and **V** represented how the new vector components were derived from the original ones. Then, the matrix **V** from Fall 2008 should be incorporated into the matrix decomposition for Spring 2009.

First, the data from Spring 2009 was organized so the assignment and attendance scores matched up with those from Fall 2008. The problems from Exam 1 were very similar those from the previous semester, but were in a different order. For example, problem 16 in the spring semester was comparable to problem 11 from the fall, so the matrix $\mathbf{A_s}$ was rearranged to have the problem 16 scores in the problem 11 slot. Next, only two written homework assignments were collected in the first month of classes in Spring 2009, compared to four in the fall. Averages of the students' first two homework scores were used to fill in the missing data from

assignments 3 and 4. Finally, iClicker scores from the appropriate class sessions were used in lieu of attendance points (See Appendix A: Clusters, Spring 2009).

One would wish to find the coordinates of the new students in the V vectors from Fall 2008. So, if

Equation 3.1 Singular Value Decomposition

$$A_s = U_s S_s V_s^T$$

Then, $$U_s = A_s [S_s V_s^T]^{-1}$$

Adjustments needed to be made, as $S_s V_s^T$ was not a square matrix and thus not invertible. However, $\mathbf{S_s}$ was a 332*37 diagonal matrix, so truncating the last 295 rows containing only zeros to make $\mathbf{S_{trunc}}$ did not alter the information the matrix contained. Therefore, $S_{trunc} V_s^T$ included the same information as $S_s V_s^T$ and was an invertible square matrix. Finally, the 198*37 matrix $\mathbf{A_s}$ was multiplied to this matrix to obtain a matrix **U`** with new student coordinates in the Fall 08 orthogonalized vectors. Five groups were then pulled out using PAM with the first 4 coordinates of each row in **U`**.

The sizes of the groups using the Fall 2008 vectors were less segregated between large and small than those using only spring data. The largest group (60 students) had the highest Exam 1 average, as well as the highest scores on all other assignments. The remaining four groups ranged in size from 42 to 30. One group of 30 clearly contained the low scoring students, while the other three groups had Exam 1 differences of 3 to 4 points between them.

Several factors were considered in the decision to use the student clusters from the second grouping method. First, the V vectors from Fall 2008 were formed using a much larger and varied sample of students, so ideally they would be more representative of the true population. Also, the standard deviations of subsequent exam scores suggested that the second method did a slightly better job of predicting future success. If we compare the normalized averages of the group scores from each category (Exam 1, Studio, Attendance, Online Homework, and Written Homework), we see many more similarities between the Fall 08 groups and Spring 09 groups formed by using the Fall 08 vectors than the alternative.

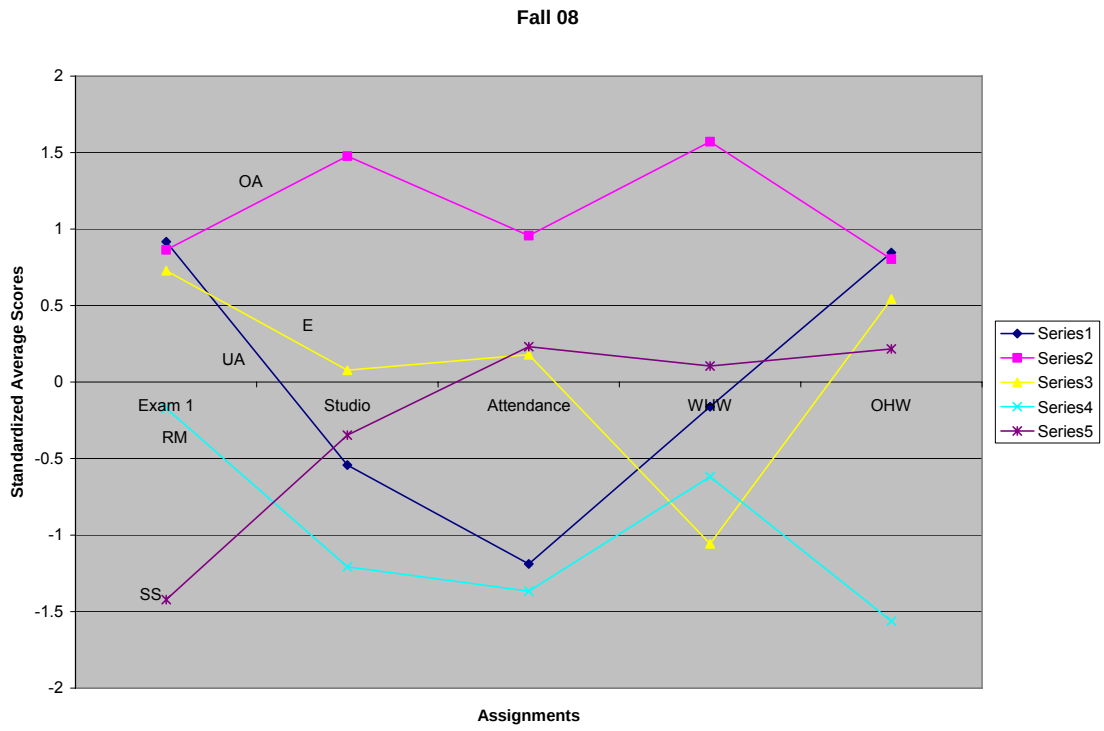

Figure 3.4 Average Cluster Scores for Fall 2008

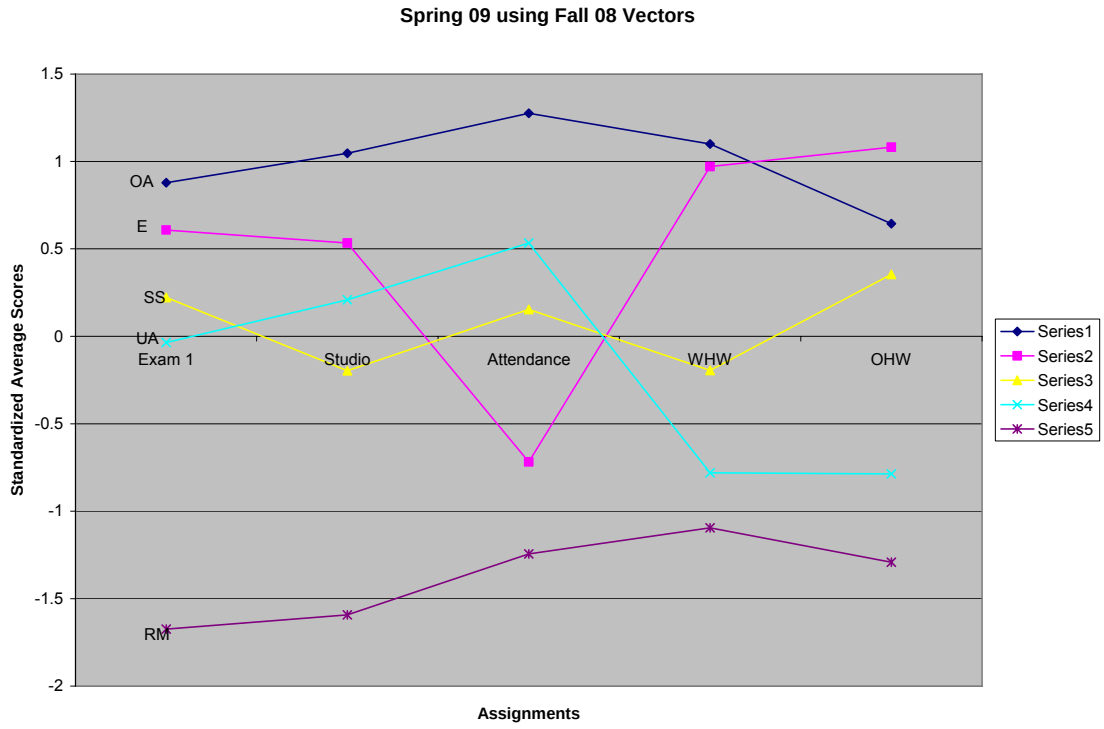

Figure 3.5 Average Cluster Scores for Spring 09: Using Fall 2008 base vectors

Note that both of the above charts have a clear group that is performing poorly in every category (RM) and one that is doing very well in every category (OA). Group E does moderately well in all categories except one, and Group UA are inconsistent in their performance.

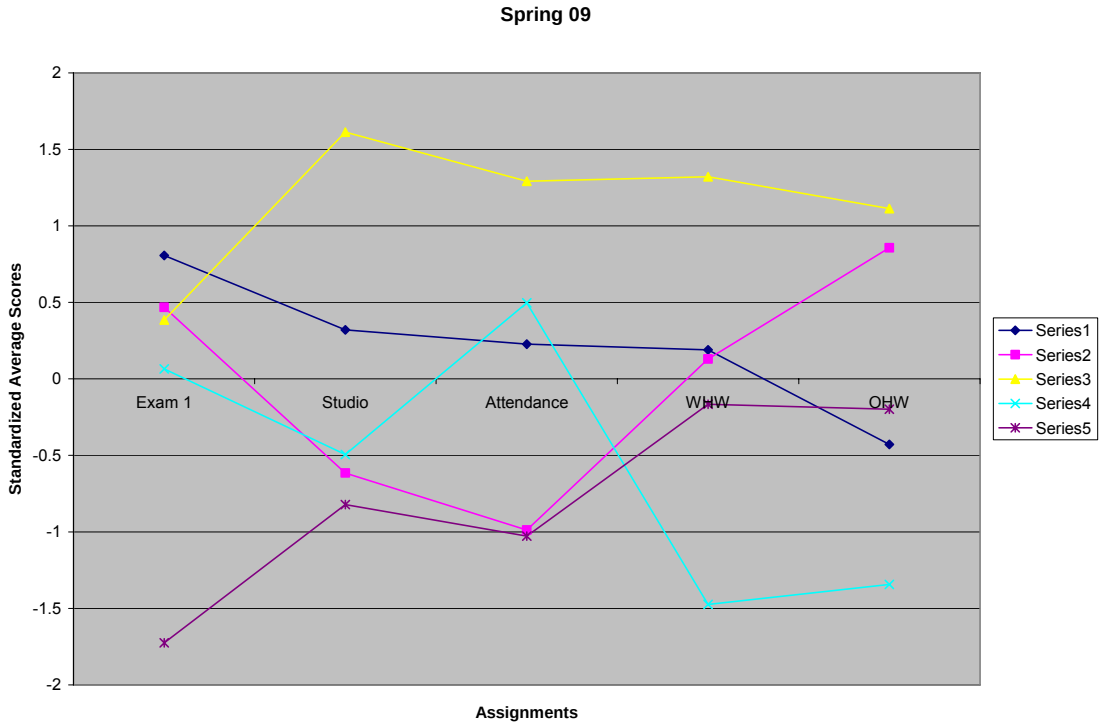

Figure 3.6 Average Cluster Scores for Spring 2009: Using Spring 2009 Base Vectors

Although the graph from Spring 2009 may look similar to that from Fall 2008, on closer inspection, there are some striking differences. The group that had the lowest score on Exam 1 in Spring 2009 also had the lowest attendance and studio scores, unlike the SS group in Fall 2008 who had above average scores in all other categories. Also, the group with the highest scores in all the assignment categories did not have a high Exam 1 score.

Fall 2009 and Spring 2010

In the Fall semester of 2009, data was collected from 362 Studio College Algebra students and analyzed using the same techniques as in the previous fall. This time, with the parameters already set by previous trials, fewer steps were necessary to form student clusters.

The 362 by 33 matrix was processed into its Singular Value Decomposition. Then, the U matrix was truncated at the fourth column, leaving a 362 by 4 matrix of coordinates in the new orthogonal vector system. The points in this matrix U were then clustered using the PAM algorithm with K=5 medoids (See Appendix A: Clusters, Fall 2009).

The Spring 2010 student clusters were formed using the same steps used in Spring 2009. That is, for each of the 130 students from the spring semester, his or her coordinates in the Fall 2009 SVD vector system were found, forming the matrix U. Then the PAM clustering program with k=5 was run on the 130 by 4 truncated matrix U.

White Box Clusters

In order to compare the SVD clustering method to a more standard classification scheme, the researcher developed comparison clusters based on a "white box" grouping scheme. First, the student scores were organized into 5 category vectors: Exam 1, Studio, Attendance, Written Homework, and Online Homework. Then the scores were normalized and then clustered using the PAM program with the same number of clusters that was used in the SVD method (k=5). The biggest advantage of clustering students in this way was that the behavior characteristics of each cluster were readily determined by the coordinates of its members. For example, White Box Group 4 from Spring 2009 had medoid coordinates Ex 1: -.6428, Studio: .53418, Attendance: .1954121, Written homework: .52231, Online homework: .6042. Therefore, White Box Group 4 contained students who scored highly on their written assignments, had better than average attendance, but performed poorly on Exam 1. By comparing the SVD cluster groups to these White Box cluster groups, the researcher could determine if the new classification method would prove to be more predictive of student success and better describe student attitudes and interests (See Appendix A: Clusters, Fall 2008). The results of this comparison can be found in the next chapter.

Chapter 4 - Describing Student Clusters

The clustering algorithms and data mining analysis produced five distinct groups of students in Studio College Algebra. However, these data mining procedures merely identified groups of students who showed similar patterns of behavior without revealing what attributes made each group distinctive. Further quantitative and especially qualitative analysis was needed to determine the characteristics of these groups. In order to help differentiate instruction, the researcher sought to provide profiles of each student cluster. These descriptions would need to contain information about how to identify members of a certain cluster and what characteristics each members would likely possess.

The quantitative analysis described in this chapter provided ways to depict the similar patterns of behavior expressed by the student clusters. The researcher examined the composition of the Singular Value Decomposition (SVD) vectors and average assignment scores to develop a general picture of how cluster members performed in the class. In addition, the researcher tracked student behavior from later in the semester and their final grades to determine the academic success of the various student clusters. Comparing average scores from semester to semester provided a reasonable way to demonstrate that the students from each semester fell into similar behavior patterns and could be grouped into the same clusters.

Quantitative analysis could illuminate which actions made members of each group similar to each other and distinct from members of the other groups. However, to understand the attitudes, beliefs, and motivation behind these behaviors, the researcher needed to employ qualitative research methods. Select representatives of each student cluster were interviewed about their reactions to College Algebra and mathematics, career goals, study habits, understanding of the concepts, and application methods. The interview transcripts were then examined for patterns of opinion and belief.

Using results from both the quantitative and the qualitative analysis, the researcher was able to compile profiles of each student cluster and give them illustrative names. Although the full names and descriptions will be given later, for now we will label the five groups OA, E, UA, SS, and RM. The remainder of this chapter narrates in more detail the quantitative and qualitative analysis, the group profile descriptions, and the methods employed to ensure reliability and validity.

Quantitative Analysis

The quantitative analysis of the clusters began with examining the vectors of the matrices formed by the Singular Value Decomposition (SVD) process. Then, the average assignment scores from members of each were calculated and compared. Cluster behavior was tracked throughout the semester to determine if clusters could be used to predict likelihood of success, failure, or dropping the course.

During the SVD process, the original vectors representing assignment and exam scores were transformed into new orthogonal vectors that captured the most variation of the points. The columns of matrix V were new vectors and each entry showed how these vectors were related to the original scores. For example, the value in position (1, 6) of matrix V shows how strongly in the positive or negative direction Exam 1 problem 6 influenced the new vector 1. If a column in V had many values whose absolute value was large (>.2), then those assignment scores captured a great deal of student variation. Therefore, the assignment had a large impact in determining the SVD clusters. Note from the table below that attendance points from the end of the first month contributed highly to group placement, while studio assignment scores did not. The significant problems from Exam 1 were the least complex examples of each type of problem: procedures, graphing and slope, and applications (See Appendix B: Data Analysis, V Vectors).

Table 4.1 Highly Contributing Assignments

	Exam 1 Problems	Written Homework	Online Homework	Studio Assignments	Attendance (Days)
Fall 08/ Spring 09	1, 2, 5, 7, 8, 9, 10, 12, 13(11)	1, 4	2	1	5, 6, 7, 9
Fall 09/ Spring 10	1, 2, 3, 6, 7, 9, 11, 12, 13, 15	1	1, 2, 4, 5		3, 4, 5, 6, 7

One can use the medoids of the five groups from each semester to get a general picture of which assignments the typical student scored well on and which assignments he or she did not. First, the coordinates of the medoids in the new vectors were examined. For example, the medoid of Group OA from Spring 2009 had very low negative coordinates for Vector 3 and low negative coordinates for the other three vectors. Then, because Vectors 1 through 4 had negative contributions from Exam Problems 1 and 2, this indicates students from Group OA scored highly

on those problems. From this information, broad descriptions of each group's success up to the first exam can be determined. The following group summaries were based entirely on these medoids. A full description can be found in the appendix (See Appendix B: Data Analysis, Medoid Coordinates).

The coordinates for the medoids of Group OA were such that most assignments had positive scores. Although there was some variation from semester to semester, these students scored highly in most categories, particularly studio and attendance. They excelled on problems from the first exam that were standard and had been covered in class. They also scored highly on applied problems, unlike students in the other groups. Group OA students did well on problems focusing on procedures that had not been worked through previously. The exceptions to these good scores were nonstandard applied problems and some graphing problems.

Students in Group E scored highly on assignments from the first two weeks of class, but then their grades dropped. On Exam 1, they did well with the procedural problems and few applied problems learned at the very beginning of the semester. Their weakest areas were graphing and nonstandard applications. Group UA's medoid coordinates indicated that its members had sporadic attendance records and did not complete or submit many of their homework assignments. These students did not perform mathematical procedures well on Exam 1, but did score highly on graphing problems.

Students in Group SS had high attendance rates and did well on their written homework and studio assignments. They struggled with online homework. Their performance on Exam 1 indicated that they had trouble with nonstandard problems and all types of applied problems. The medoid coordinates for Group RM were extreme, indicating that it is a group consisting of outliers, and thus might be less cohesive than the others. These students had mostly low scores in all categories, most noticeably in studio and applied exam problems.

In addition to examining individual assignment scores for each cluster, the average scores from the first four weeks of class in several categories were computed and compared. These categories consisted of Attendance, Written Homework, Online Homework, Studio, and Examination One. Students in the OA group consistently scored higher than any other group in all categories. Cluster E students usually had the second highest scores in all categories, except for attendance where they usually came in third. Students in Group UA had fairly good Exam 1 scores (depending on the semester, they had the highest or third highest averages). However,

their performance in all other categories was second to lowest. Behavior in the SS groups was slightly different in the Fall and Spring semesters. In Fall 2008 and Fall 2009, this group had the lowest Exam 1 average score, but was the middle or second highest group for all other categories. In the Spring, SS students scored in the middle for most categories. Finally, students in the RM group had the worst scores for every category with few exceptions, where they had the second lowest scores.

Qualitative Analysis

Examining the quantitative data provided clues to how each student cluster behaved, but could not illuminate why these students exhibited these behaviors. In order to understand the motivations behind these actions and determine if members of student clusters have similar beliefs as well as similar behaviors, the researcher posed these questions directly to students through interviews.

A preliminary round of interviews was conducted during the fall semester of 2008. These and all other interviews were conducted with procedures approved by the Institutional Review Board (IRB). Specifically, students were offered $10 for their time, but no extra credit. These interviews were designed to test the effectiveness of the first interview protocol and determine which questions would most successfully bring out students' thoughts (See Appendix C: Interview Protocols, Fall 2008). The interviews were conducted not knowing which group the interviewee belonged to in order to help the interviewer preserve impartiality.

The first interview protocol was separated into three sections. The first section asked questions about the students' opinions, reactions to the course, and work habits. The questions were kept intentionally vague and open ended to capture the widest range of student responses. The second part of the interview was designed to assess conceptual understanding of functions, their representations, and their applications. The third section of the interview assessed problem solving abilities by having students retake part of a recent exam and explain their thought processes. An opinion survey was added to the end of the interview. Approximately 110 students received invitations to be interviewed, 18 students scheduled a meeting, and 13 students were interviewed.

The first round of interviews was not very successful in determining the characteristics of each group because of several flaws in the scheduling process and the protocol design. After the

interviews had been conducted, the researcher discovered that only three of the five groups were represented by the interviewed students. Also, many of the students did not know how to respond to the vague questions. There were not enough follow up questions to elicit student opinions. The conceptual knowledge part of the interview was also too general. Students were asked to think of ideas and definitions on the spot and often froze or simply responded, "I don't know."

Even though the clusters' qualitative characteristics were not fully uncovered by the first round of interviews, a general picture was formed by their responses. Members of groups OA and UA seemed to really enjoy the course, and students in Group E were ambivalent. These students seemed to prefer algebraic manipulations and procedures to the applied problems in the studios (See Appendix C: Interview Protocols, Fall 2008).

Several changes were made to the interview protocol and student recruitment for subsequent interview appointments. Because so few students replied to appointment requests during the fall interviews, the students were recruited more aggressively in the spring. The first round of students were invited to come in for interviews earlier in the semester, three days after taking their first exam. Initially, the fifteen most representative students from each group were sent emails. These students were chosen because they were closest to the medoids of their respective groups. Every student who replied to the initial email was sent a follow up email suggesting an appointment date and requesting a confirmation. Each scheduled student was also sent a reminder email the day before the scheduled appointment.

In total, nine students were interviewed from March 5 to March 13. After the second exams, 20 students from each group were contacted, including those students who did not respond in the first round. Ten students were interviewed during this round. A colleague of the researcher confirmed that each cluster was represented by a least three interviewees. The researcher conducted sixteen of the nineteen interviews, and another colleague conducted the remaining three. This was done mainly to ensure the reliability of the interview protocols.

Using experience gained from the Physics Education Research Workshop held in fall of 2008, the interview protocol was rewritten (See Appendix C: Interview Protocols, Spring 2009). The first thirteen questions focused on the students' opinions and attitude toward Studio College Algebra and mathematics. Students were asked to describe their previous experiences in mathematics classes, how they felt about their current class and progress, and how they expected

to use their math skills in their future careers. Students also were asked to describe their homework and study habits. The questions for the first section were more structured than those of the previous interview protocol and included more follow-up inquiries. Interviewers were instructed to seek clarifications and extensions of student responses. Some of the questions from the survey section were incorporated into the interviews as well, allowing the students to explain their responses.

The section on conceptual understanding was heavily modified. In the initial protocol, students were asked to draw on memory and construct responses without cues, which led to blank looks and frustration. For this next round, the interviewer provided three examples of representations of the same function. The first was a t-chart with eight inputs and their output values. Next, the same polynomial function was given in both standard and factored form. Finally, the student was presented with a graph of the function. Initially, the student was asked to tell the interviewer what they knew about what was written on the sheets of paper, one at a time. Then, the student was presented with all three pictures and asked to make connections between. If the student had not yet made the observation that all three pictures were representations of the same function, the interviewer guided them to make that connection and then justify their conjecture.

To demonstrate their knowledge of applications of functions, students were then given a graph showing a scatter plot and a linear regression model. They were asked to interpret the model and extract the important information (See Appendix C: Interview Protocols, Conceptual Handouts). Because they had material to work with, students were much more forthcoming and revealed more about their conceptual understanding of functions. When they were later prompted to give more examples of functions or situations where a function might be useful, many students were able to provide insightful responses. Finally, the survey portion of the interview was discarded, and its questions incorporated into the protocol.

All nineteen interviews were transcribed and coded without knowing to which group the students belonged. Using Miles and Huberman's Qualitative Data Analysis as a guide, the transcripts were coded in the following manner: the interviews were read through once to identify general themes and common responses. During the second reading, these responses were classified more concretely and organized under the categories of positive, neutral, and negative responses (See Appendix D: Coding Scheme). During the third reading, the transcripts

were marked with response labels, and more labels were created as necessary. The transcripts were marked again during a final reading, with the two markings later compared for consistency. Finally, the interviewed students were identified by their group number, and the coded responses were organized by group. Each group's qualitative characteristics could be reasonably determined by coded responses that occurred in multiple interviews (See Appendix E: Grouping Chart, SVD).

Profiles of the Groups

Using the insight gained from the quantitative and qualitative analysis, the researcher compiled the following profiles of each group. Included with each profile is a breakdown of the groups' average assignment scores from each semester. The charts are color coded to provide a third dimension to the chart and help compare these scores to other groups in that semester:

Top Group
2nd Highest
Middle
2nd Lowest
Bottom Group

Note that each student cluster is only being compared to other clusters in that semester. Also, all the total number of points attainable for each exam was 80, with the Final Examination worth 160 points.

Group OA (OverAchievers):

Average Final Grade: 3.24 Percent of Students: 33%

This is a group of hard working students who have a positive attitude about mathematics and a good work ethic. Their average Composite ACT score is 22.38 and their Math ACT score is 21.44, so they are well prepared to take the course. These students do well on their first exam, but what most separates them from everyone else is that they also do very well on all their other assignments and attend almost every class session. The Exam 1 problems on which these students excel are the standard problems that are most focused on in class, as well as the graphing and conceptually based problems. These students continue to do well on the subsequent exams and earn an average of A or B on the final. Very few of these students drop the course and the percentage of students that earn a C or better is 97.3%. A separate set of pre and post exams showed that these students grew the most conceptually out of all of the groups.

Student interviews revealed that these students think that mathematics is used often to solve problems and has applications in art, to explain phenomena, and to improve society in general. Most think math is useful to themselves and their future careers, but are not sure if the specific skills they learned in College Algebra will apply to their careers. By far, they think the most useful part of Studio College Algebra is recitation. The instructors are helpful and they appreciate being able to go over homework problems in class. These students enjoy the overall course and its structure, even though they often struggle to understand the concepts. During the interview, their comments about the course were mostly positive. The students appreciate the online homework and studio assignments as well as lectures. The technological aspects of the course are also appreciated, including the convenience of online homework and that the lecture notes are posted online. That said, these students often have problems understanding how the studio part of the course corresponds to the general class goals and become frustrated with online homework glitches.

These students study and work on their homework from 1-3 hours per week. They seek help from a variety of sources, including friends, the instructor, the textbook, and class notes. They make sure to do their homework before recitation and use the online homework hints often.

When asked to describe different representations of a function during interviews, these students demonstrated solid conceptual understanding of functions, their characteristics and applications. Their use of mathematical vocabulary was good to fair. They were also able to make connections between three different representations of the same function, either on their own or with prompting. Although these students could identify an increasing trend in a linear regression model, most students in this group mistakenly identified this trend as representing an "average" of the dependent variable.

Table 4.2 OverAchiever Average Scores

OverAchievers	Fall 08	Spring 09	Fall 09	Spring 10
Composite ACT	22.2	22.74	23.595506	
Math ACT	21.7	21.83	23.033708	
Percent of Students	35.31%	30.30%	34.25%	27.69%
Average Exam 1 Score	72.12	69.08	69.39	63.41
Studio	17.24	19.275	19.49	17.22
Attendance	9.36	8.81	7.26	7.08
WHW	51.33	30.38	20.39	18.71
OHW	47.9	44.39	47.93	45.53
Average Exam 2 Score	58.90678	67.49	58.48	58.912
Average Exam 3 Score	59.923077	60.85	65.6	61.5
Final Exam Score	104.3	112.03	115.17	114.242
Grade in Course	2.898	3.38	3.47	3.30
% C or better (of completed)	96.61%	98.33%	97.56%	96.97%
% Completed Course	99.16%	100% (tie)	99.19%	91.67%

Group E (Employees):

Average Final Grade: 2.62 Percent of Students: 24.7%

These students tend to treat the Studio College Algebra course like a low paying job. They do only what they think is expected of them, then are "paid" for their efforts with a passing grade. These students enter the course reasonably well prepared with an average Composite ACT score of 22.6 and Math ACT score of 21.8. Students in this group perform fairly well on the first exam, averaging a B grade. They score well on standard problems that are reviewed many times in class, but not so well with problems that require innovative reasoning. Their exam performance remains steady throughout the course, staying in the B/C range. They attend most classes at the beginning, but their attendance drops as the semester continues. Students in Group 2 complete and turn in most of their assignments, but they do not score as highly as those students in Group 1. Not many of these students are likely to drop out of the course (3%), but only 88.2% of the total group earn a C or better. A separate set of pre and post exams showed that these students demonstrated moderate conceptual growth.

When interviewed about the Studio College Algebra course and mathematics in general, these students said that mathematics is very important for solving problems and explaining how things work. However, they do not believe that everyone needs to learn mathematical skills, including themselves. Their confidence in their own mathematical abilities is low, they dislike math in general, and so they are trying to just "get through the class." Although they do not enjoy mathematics, these student have generally positive or ambivalent opinions about the Studio College Algebra course. They particularly enjoy using Excel in the studio sessions and the integration of other types of technology in the course, such as iClickers. These students think recitations and lectures are the most helpful parts of the course because the instructors are knowledgeable and supportive. Their least favorite part of the course is the Online Homework.

Students in this group spend an average of 1 to 2 hours a week studying or doing homework. If they have questions, they are likely to ask a friend or go to the instructor. These students take notes during lecture and refer to them often later. These students try to get their homework done before recitation so they can ask questions. During interviews, when these students were asked to perform tasks to demonstrate their knowledge, they were very dependent

on prompting. The students were able to make connections between different representations of functions, but only after the interviewer gave hints. Their range in vocabulary use was wide, as was their ability to name specific functions. The students often performed memorized procedures without any justification. They could not describe any situations in which functions would be used. However, the students could interpret a linear regression model fairly well.

Table 4.3 Employee Average Scores

Employees	Fall 08	Spring 09	Fall 09	Spring 10
Composite ACT	23.3	21.65	21.0581395	
Math ACT	22.3	20.61	19.9069767	
Percent of Students	20.18%	15.15%	30.30%	35.38%
Average Exam 1 Score	70.86	66.03	58.6	62.2
Studio	13.6	17.633	16.17	16.02
Attendance	8.6	7.29	6.99	6.4
WHW	25.9	29.6	16.97	17.17
OHW	45.3	47.21	43.27	44.36
Average Exam 2 Score	57.9	50.1	45.62	50.93
Average Exam 3 Score	60.8	50.28	56.34	54.71
Final Exam Score	104.2	101.9	98.24	94.7
Grade in Course	2.53	2.87	2.55	2.74418605
% C or better (of completed)	86.15%	93.33%	87.16%	90.70%
% Completed Course	95.59%	100% (tie)	99.09%	93.48%

Group UA (UnderAchievers):

Average Final Grade: 2.13 Percent of Students: 13.7%

Students in this group are well prepared and intelligent, but are bored and frustrated with the material presented in College Algebra. They tend to drop the course or underperform. Their average Composite ACT score is 23.29 and their Math ACT score is 22.34, which is the highest of all the groups. However, their first exam scores are in the middle of the groups, a high C. They do well with applied problems and questions focusing on graphing, but they are sloppy with procedures and nonstandard algorithms. Their scores drop for subsequent examinations, ending with a low C for the Final Exam. These students have decent attendance scores and fairly high online homework scores, but do not perform as well on the other assignments, especially the written homework. Their attendance drops after the first few weeks. Almost 13% of these students drop the course, and only 73% earn a C or better. Also, only half of the students in this group took the pre and post exams that measure conceptual growth, and those that did showed very little increase in conceptual knowledge.

During interviews, students in this group thought mathematics was useful for a variety of reasons, especially to solve problems. However, they believed that math was only useful to some people, and although many expressed confidence in their mathematical abilities, they did not enjoy math. They thought math was not personally useful and that they would not be using math in the future careers. These students thought the class was easy and the emphasis on review boring. However, they still expressed frustration with different aspects of the course, including their own performance. Studio and recitation were their favorite parts of the course and they enjoyed using Excel and other forms of technology. The recitation and lecture instructors were helpful. Their least favorite parts of the course were lecture and homework. Most students thought there should be less homework assigned, and disliked having to do so many applied problems. They also thought the examinations were too tough, mostly because they had to justify their answers.

These students sought help from written sources, like the textbook, notes, and online homework hints as well as from friends. They did not seek help from the instructor or tutors. Some students confessed to not studying much. They worked on homework anywhere from 1 to

5 hours a week. When asked to demonstrate their knowledge during the interviews, these students were very good at describing functions and making connections between different representations. However, they did not always use proper mathematical vocabulary, and were not able to come up with examples of functions or uses for functions. Their interpretations of a linear regression model were insightful and accurate.

Table 4.4 UnderAchiever Average Scores

UnderAchievers	Fall 08	Spring 09	Fall 09	Spring 10
Composite ACT	24	22.41	22.033333	
Math ACT	23.83	20.55	20.933333	
Percent of Students	12.70%	18.18%	12.15%	19.23%
Average Exam 1 Score	72.6	58.78	53.01	57.32
Studio	12	16.6	14.99	9.98
Attendance	7.2	8.24	6.79	5.13
WHW	34.57	18.93	10.26	11.08
OHW	42.07	35.17	19.97	38.22
Average Exam 2 Score	60.71	51.03	38.8	43.1
Average Exam 3 Score	59.08	44.31	50.4	50.21
Final Exam Score	106.86	86.22	84.91	85.47
Grade in Course	2.64	2.06	1.86	1.83
% C or better (of completed)	75.68%	80.00%	65.71%	73.91%
% Completed Course	86.05%	97.22%	79.55%	92.00%

Group SS (Sisyphean Strivers):

Average Final Grade: 1.53 Percent of Students: 17.6%

Students in this group have a good attitude toward mathematics, try hard to succeed in the course, but still perform poorly. These students are generally not well prepared for the course, with an average Composite ACT score of 20.92 and Math ACT score of 19.27. These students attend class regularly and turn in all of their homework assignments. Despite this, their first examination scores are among the lowest of the five groups (C/D), and they maintain this low average throughout the semester. They are particularly weak with applied problems, although their mastery of procedures and algorithms are good. These students are less likely to drop the class than their test scores predict, with a rate of 7.5%. They continue to struggle heavily and 55% of them end up earning a C or better in the course. The pre and post exams that measure conceptual growth showed that these students did demonstrate moderate conceptual gains.

When interviewed these students revealed that they like math in general. They think mathematics has a variety of practical applications, and is especially useful in explaining how things work, in creating art, and solving problems. The students think they will be able to use the skills they learned in Studio College Algebra in their careers, but their confidence in their own abilities are low, and they expect to struggle to understand new ideas and solve problems. Although the class is review for most of them, they have to work hard to understand the concepts, fulfill deadlines, and submit their homework to the appropriate place. They enjoy most aspects of the course, especially recitation. These students also think the online homework is a very helpful practice, because they can redo the problems as often as they like. Because the problems change with every other attempt, the students get even more opportunities to practice the underlying procedures. The students tend to have trouble understanding the website, and feel that posting the lecture notes online discourages them from coming to class. By far, their least favorite part of the course is lecture, partly because of the intimidating class size. Although they enjoy working with a partner in Studio, they do not understand how that part of the class fits in with the rest of the curriculum.

Students spend an average of 3 to 4 hours a week on homework and studying. They usually get help from their friends when they have problems understanding the material, and are

not likely to ask instructors. These students spend a lot of time and effort completing their homework, shooting for scores of 100% and using several sources of information to help. During the interviews, students in this group demonstrated fairly solid conceptual understanding of the material, much more than their exam performances would indicate. They were especially adept at interpreting the applied linear regression model. They also seemed to pull out more information from graphs and pictures than from the formulas and charts.

Table 4.5 Sisyphean Striver Average Scores

Sisyphean Strivers	Fall 08	Spring 09	Fall 09	Spring 10
Composite ACT	20.37	21.68	19.857143	
Math ACT	18.64	20.16	18.238095	
Percent of Students	17.20%	21.21%	16.02%	14.62%
Average Exam 1 Score	51.06	61.69	43.6	53.9
Studio	12.5	15.3	16.09	12.95
Attendance	8.6	7.95	7.26	6.3
WHW	37.13	22.5	17.57	14.63
OHW	42.07	42.52	44	40
Average Exam 2 Score	53.39	52.7	35.93	44.56
Average Exam 3 Score	40.38	48.5	51	47
Final Exam Score	68.4	90.8	80.56	98.93
Grade in Course	1.388	2.21	2.18	1.94
% C or better (of completed)	48.15%	78.57%	40.00%	72.22%
% Completed Course	93.10%	100% (tie)	86.21%	94.74%

Group RM (Rote Memorizers):

Average Final Grade: 1.24 Percent of Students: 10.6%

These students do not like math, do not enjoy Studio College Algebra. Most members of this group will drop or fail the course. Their Composite and Math ACT are on the low of the scale, being 21 and 20.6 respectively. These students are likely to attend class the first day, and to turn in their first assignments. Then they stop coming to class and do not turn in further homework or studio assignments. Their Exam 1 average is around 50%, and those students that do not drop the course fail the subsequent exams as well. Although they perform poorly on most of the exam problems, these students struggle the most with graphing and interpreting graphs. This group has a drop-out rate of 13% and only 51% of those students who do complete the course earn a C or better. Those that took the pre and post tests measuring conceptual growth demonstrated that they learned very little.

These students had overwhelmingly negative views about mathematics, and their opinions got worse after taking Studio College Algebra. Their group was the only one whose negative comments during interviews outnumbered their positive or neutral comments. Students in this group believe that mathematical ability is inherited or intrinsic, not learned, and that only some people (not them) need to know mathematics. They have very little confidence in their abilities and doubt they will ever have to use skills they learn in Studio College Algebra in their careers. In particular, students expressed a dislike of graphs and fractions. This group thinks that recitation is the most helpful part of the course, because they get help with their homework. They all expressed frustration understanding the concepts and they struggle with completing their assignments. Some students enjoyed working with a partner in Studio and thought the online homework was convenient. Most did not like lecture due to its size. Online homework was not popular either, mostly because of the different due dates and the fact that the problems change after every second attempt.

These students all had tutors, who helped them complete their homework and study for exams. If they needed help understanding a problem, they went to their friends or tutor, but not the instructor or text. One student admitted to not doing her homework until after most of the problems had been done in recitation. These students' level of conceptual knowledge was very

low. Students did not volunteer much information, although they were able to make some connections between different functions after prompting. They attempted to commit actions like distributing without any reason why they should do so. Their use of vocabulary was fair to poor. These students were not able to interpret a situation involving linear regression.

Table 4.6 Rote Memorizer Average Scores

Rote Memorizers	Fall 08	Spring 09	Fall 09	Spring 10
Composite ACT	22.12	19.22	22.631579	
Math ACT	21.61	17.28	21.526316	
Percent of Students	14.50%	15.10%	7.18%	3.07%
Average Exam 1 Score	62.6	40.3	50.26	23.5
Studio	10.27	10.833	1.92	1
Attendance	7.08	6.89	2.76	3.56
WHW	30.13	17.12	7.4	10.6
OHW	24.41	31.907	25.27	25
Average Exam 2 Score	57.91	33.83	42.19	10
Average Exam 3 Score	46.07	29.85	49.69	6
Final Exam Score	92.1	62.11	86.42	12
Grade in Course	1.51	1.14	1.05	0
% C or better (of completed)	58.14%	57.14%	39.13%	0.00%
% Completed Course	87.76%	93.33%	88.46%	75.00%

Identifying Clusters from Following Semesters

Preliminary interviews from Fall 2008 and interviews from Spring 2009 suggested that the qualitative traits from the groups extend from semester to semester. Because interviewing students every semester would be time consuming and costly, a method was needed to match new groups to previously identified groups using only quantitative data. This could be done by looking at group size, performance on Exam 1, and an analysis of the SVD vectors and the position of the group medoids.

Group OA was the most straightforward group to identify. Because this group had the largest size for the first two semesters, it would likely continue to be a large group. Also, this group should have the highest or second highest Exam 1 score. All of the other homework, studio, and attendance scores should be the highest as well. The medoid of the group should be in a position so that the SVD vectors indicated positive scores on most of the assignments.

Because the size of Group E varied from Fall 2008 to Spring 2009, this was not a good indicator for subsequent semesters. The Exam 1 score should be the second or third highest, and all of the other assignment and attendance averages should be in that higher middle range as well. The best way to identify Group E was through SVD vector analysis; this group did well on the standard exam problems (both conceptual and procedural), but did not do well on problems whose solutions had not been worked out in class beforehand.

The size of Group UA should be relatively small, somewhere between 10 and 20 percent. The Exam 1 score and ACT data should be middle to high, but their other attendance and homework scores should be low. The SVD analysis should show that this group does well with conceptual and applied problems, but has trouble accurately carrying out procedures.

Groups SS and RM both had low Exam 1 scores, but they were distinguishable by other factors. Group SS was usually slightly larger than Group RM. More importantly, Group SS had decent scores on other assignments and was likely to have earned the second or third highest attendance average, while Group RM should have the lowest. In addition, analysis of the SVD coordinates should show that Group RM does poorly with graphing problems.

"White Box" Clustering

Clustering techniques such as AGNES, PAM, and SVD analysis are referred to as "black box" techniques, where a program looks for patterns in a data set without preconceived targets or external direction. "White Box" data mining techniques refer to programs that look for patterns that fit a theoretical framework or use expert guidance to modify results. Often, data miners use white box clustering methods to form an alternate grouping scheme. This is done in an attempt to measure the validity and utility of the black box groups, as standard statistical measures do not apply to most data mining scenarios.

Traditionally, a teacher defines a student's progress based on how well they performed in different categories of assignments. For example, a teacher might say someone is a "good student" if they perform well on the examinations, turn in all of their homework assignments, and regularly attend class. A "struggling student" might have similar attendance and homework points, but do poorly on the exam. The White Box grouping scheme described in Chapter 3 mimics this natural teacher-based grouping. Rather than relying on algorithms to identify the linear combinations of scores that best capture student variation, the White Box approach grouped students based on their scores in predetermined categories: Exam 1 total, Studio total, Attendance total, Written homework total, and Online homework total (See Chapter 3 for more detail on the process of forming the White Box groups). The aim of grouping students using the SVD method was to develop a way to determine personality traits and work habits of different types of students, and then predict their success in learning the material and passing the course. By comparing the SVD groups to the White Box groups, the researcher could evaluate the relative advantage of using the SVD method in achieving these research objectives.

First, the personality traits and work habits of students in the SVD groups were compared to those of the students reorganized into White Box groups. Using the same interviews previously described in this chapter, the response codes for each student were reassigned to the White Box group to which the student belonged (See Appendix E: Grouping Chart, White Box). Five of the interviewees belonged to White Box Group 1, six to White Box Group 2, two to White Box Groups 3 and 5, and four to White Box Group 4. The code charts reveal that the White Box groups were much less cohesive than the SVD groups, as derived from the repetition of student responses. A coded response was considered "repeated" if 50% or more of the students in a given group made the statement during their interview. In total, 35.8% of the coded

responses in the SVD grouping scheme were repeated, while only 23.1% of those in the White Box grouping system were repeated. When all students were put into one large group, only 14.14% of comments were repeated by more than half of the students. Looking more carefully at the grouping charts, one sees that many of the responses in the White Box groups contradict each other. For example, in White Box Group 1, Studio, Recitation, Lecture, and Online Homework all appear simultaneously as the most helpful part of the course and as the least helpful part of the course. Similarly, one of the students in White Box Group 3 reported spending more than 5 hours a week studying and completing homework assignments, while the other admitted to never studying and spending less than 1 hour a week on homework. While there were a few inconsistencies in the SVD Grouping chart, they occurred much less often and were not as blatant. This indicates that the SVD grouping system is much more effective at pulling out students with similar opinions and study habits. However, the difference might be exaggerated by the choice of student interview subjects. Students were invited to participate in interviews based on their proximity to the medoids of the SVD groups, not the White Box groups.

Another indicator of a relevant and successful grouping is the ability to make predictions about future student behavior. Because only scores from the first 4-5 weeks of the semester were used in both grouping processes, we can compare the relative behavior of the two grouping schemes measured by examinations and final grades to assess which groups were more cohesive throughout the semester. In general, the White Box grouping scheme was able to identify the highest and lowest performing students better than the SVD scheme. For example, in Fall 2008, the highest scoring White Box group of 60 students earned an average final grade of 3.133 with standard deviation .76, while the OA group of 119 had an average final grade of 2.89 with standard deviation .78. Also, in the same semester, the lowest scoring White Box group contained 15 out of 19 dropouts, with the remaining 61 students earning an average final grade of 1.06. These low performing students were divided fairly evenly between the SS and RM groups in the SVD scheme, with those groups of students earning final grades of 1.39 and 1.51, respectively. In the semesters following Fall 2008, the highest and lowest groups for both grouping schemes had mostly the same students and earned similar average scores on their final exams and in the course. However, the White Box groups had slightly lower standard deviations for their overall course grades, while SVD groups had lower standard deviations for their Final Examination Scores.

The two grouping schemes differ the most when comparing the middle two or three groups. In general, the White Box scheme had one high performing group, one or two very low performing groups, and two or three groups in the middle that acted similarly in all categories. The following charts show the average final grades for each of the groups in the two grouping schemes. Notice the averages for the White Box grouping scheme do not fit a linear model as well as those for the SVD grouping scheme. The White Box chart for Fall 2008 (Series 1) shows three middle groups with roughly the same final grade, while Fall 2009 (Series 3) has one high group, two middle, and two low.

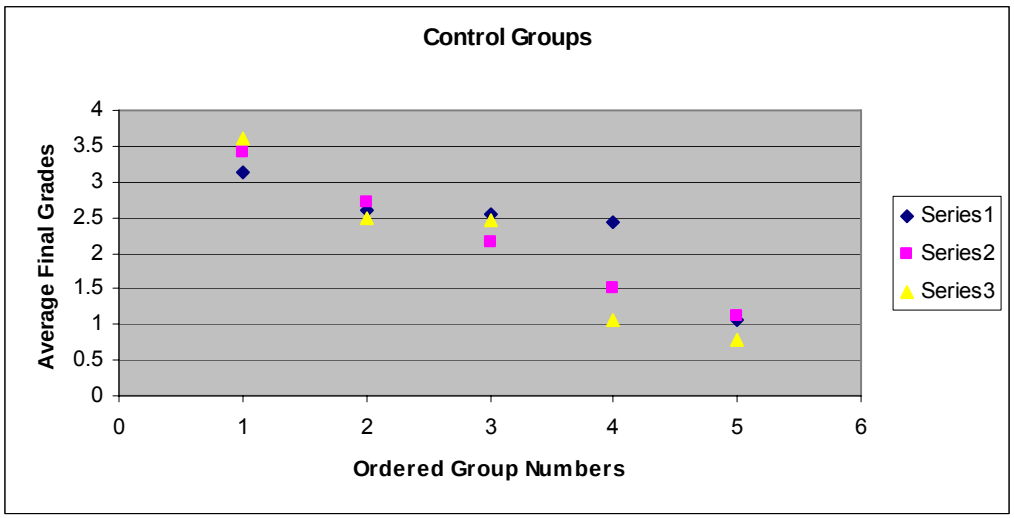

Figure 4.1 White Box Group Average Final Grades

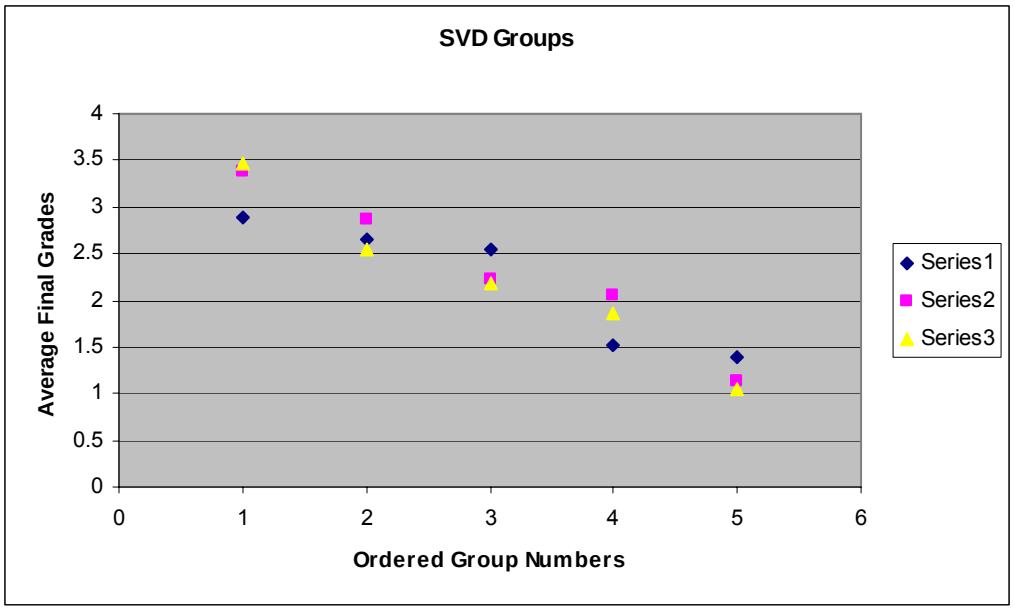

Figure 4.2 SVD Group Average Final Grades

When linear regression models were applied to each of the semesters' groups' grades, the R-squared values for the White Box Scheme's models were substantially lower than those for the SVD Scheme's models in the Fall semesters, and roughly equal in Spring 2009.

Table 4.7 R-Squared Values for Linear Regression Fit of Average Final Grades

R-Squared Values	White Box	SVD
Fall 2008	0.7791	0.8925
Spring 2009	0.9921	0.9666
Fall 2009	0.9346	0.9652

This indicates that the SVD grouping scheme distinguishes the middle groups from each other better than the White Box grouping scheme in their grades as well as their attitudes.

Reliability and Validity of Student Attitude and Behavior Groups

When using Data Mining methods, there is no standard method for assessing the validity and reliability of clusters using traditional statistics. However, measures to support good data collection and analysis practices were included in the research protocol. Each semester, the same simple, robust method was used to identify student clusters. Although a program like

AGNES is useful to estimate the number of natural clusters in a data set, it is highly susceptible to being altered by random noise. This occurs because it is based on grouping "nearest neighbor" points together, so a tiny shift may result in drastically different clusters. Once the number of clusters is known, the program PAM is a much more reliable grouping algorithm. In this program, medoids rather than centroids were used as the ideal "middle" of the clusters. This reduces the influence of outliers, as described in Chapter 3.

To strengthen the reliability of the interviews, a protocol was created using guidelines from the PER Workshop and Miles and Huberman's Qualitative Data Analysis. Each interview was conducted by following the protocol while being recorded for transcription. Also, each interview was coded twice independently by the researcher, with the coding compared for reliability. Finally, several interview transcripts were coded independently by another graduate student and compared to the original coding. The second set of codes was a subset of the original set, indicating the two coders did not have drastically different interpretations of the text. However, because the additional coder was not as familiar with the coding scheme as the researcher, her marks were not as detailed or numerous.

Data was collected over four semesters from students in the same large lecture class to help ensure consistency over time. In total, scores were collected from 1027 students. The students from the fall and spring semesters of the same academic year were compared using coordinate vectors from the much larger and more diverse fall class. Because students from the academic year had the same coordinate vectors, one could test the reliability of the clusters by comparing medoid coordinates. In each case, at least 2 of the 4 medoid coordinates matched up by being either both positive, both negative, or both between -.01 and .01 (close to zero). Comparing average college aptitude test scores, exam scores, final grades, and dropout rates of clusters within a semester show that their rankings stay relatively constant from semester to semester (See Appendix B: Data Analysis, Medoid Coordinates).

The main goal of the study was to determine if student attitudes about mathematics and studio college algebra can be derived from data analysis of their academic behavior. One could then use this data to enhance learning and predict future behavior. The validity of the academic scores collected from each student was secured by gathering data from the field. In compliance with IRB, the data pulled together from each student was only what would normally be recorded

during the course of establishing a final grade. Thus, the students were unaware of their involvement in an educational study.

Confirming the validity of the student attitude clusters was more complicated. The cluster descriptions were created using two independent methods- student interviews and analysis of student scores. The interviews were conducted and coded without knowing to which group the students belonged. The representative student interviewees were not connected to the interviewer in any way, and were assured that their responses would remain anonymous. After evaluating a trial set of interviews, the researcher was able to identify and edit leading, vague, or otherwise unsuitable questions. Due to time and budget constraints, interviews were only recorded and analyzed for one semester leading to concerns that other semesters' groups could have differing or disparate opinions. Quantitative data was used to fill in these gaps.

Chapter 5 - Conclusions

By applying several standard Data Mining techniques to student-generated scores in Studio College Algebra, five distinct groups of similarly performing students were identified. Quantitative and qualitative interview analysis determined that these students not only exhibited analogous behaviors, but also that they had related beliefs, work habits, levels of conceptual understanding and likelihood of success. The make-up of student clusters remained relatively stable from semester to semester. To help identify and characterize members of each student cluster, the researcher assigned descriptive names and general profiles to each group.

Hypotheses

Upon revisiting the original research goals, the researcher found the results of the student clustering mostly affirmed the hypotheses, which are restated below:

1) Patterns and similarities in student behavior can be efficiently and accurately identified using standard Data Mining techniques known as clustering algorithms.
2) College Algebra students' attitudes and beliefs about mathematics can be revealed by examining their behavior in the course.
3) This information can be used to develop effective mathematics placement strategies, identify students in need of intervention, and improve freshman retention.

The application of basic Data Mining techniques was successful in identifying patterns in student behavior. Distinct clusters of students exhibited similar behavior in the class, and these clusters were reasonably stable from semester to semester. Initially, the selection and applications of the clustering techniques involved many decisions on the part of the researcher. This involved some trial and error, and required at least two semesters of active experimentation. However, once these choices were made and the procedure for identifying the student clusters was established, the algorithms could be run quickly and efficiently. By using information recorded as graded assignments from the first four weeks of class, an instructor can have the results of the clustering analysis by the fifth week of the semester.

Interviews with members of each student cluster confirmed that the distinct groups were composed of students with similar beliefs about mathematics, ways of solving problems, and attitudes about Studio College Algebra. In summary, members of the OverAchiever group had positive views of mathematics and their own abilities to succeed. They enjoyed Studio College Algebra and had a strong conceptual understanding of functions and their applications. Students from the Employee group viewed Studio College Algebra as a hurdle to overcome before they could graduate from Kansas State University. They did not particularly enjoy or dislike the Studio course, but cannot stand math in general. These students relied heavily on other people for help because they have a low opinion of their own abilities to understand math concepts. UnderAchievers were well prepared for the course, and confident in their own aptitude for learning mathematics. However, they found the course frustrating and boring, and were the most likely group to drop out of the class. Members of the Sisyphean Striver group struggled heavily to succeed in Studio College Algebra. Although these students liked math, enjoyed many aspects of the course, and demonstrated solid conceptual understanding during interviews, they did not perform well on exams. In contrast, students from the Rote Memorizer group had uniformly negative opinions about math, their academic abilities, and Studio College Algebra. They tended to put forth minimal effort and relied on tutors and friends to help them with their homework. Consequently, the Rote Memorizer group was the lowest performing cluster.

In addition to uncovering beliefs and personality traits, clustering students better enables instructors to predict the probability of success of students in their courses. For example, 97.6% of the members of the OverAchiever group completed the course, and 98.3% of those students passed with a C or better. Even though members of the Employee group had at an average final grade that was .6 points below that of the OverAchiever group, 93.8% of those that finished the course earned a C or better. Members of the UnderAcheiver cluster had the second highest probability of dropping the course, with only 78.8% of its members finishing. However, of those group members finishing, 82% of them were able to earn at least a C. Rote Memorizers and Sisyphean Strivers had very similar low passing rates for those students who completed the course, with 58.9% and 57.7%, respectively. Members of the Rote Memorizer group, however, were much more likely to drop the course, with only 69% finishing the semester, while Sisyphean Strivers had a completion rate of 85.4%.

Attempting to confirm the final hypothesis by applying insight gained from cluster analysis is an ongoing process. The next section of this chapter describes and evaluates the researcher's application this cluster analysis to differentiating instruction.

Differentiating Instruction

After identifying and uncovering the characteristics of student clusters, the researcher attempted several strategies to differentiate instruction in Studio College Algebra. First, the researcher sought to improve placement by working with advisors to enroll students belonging to different clusters in the appropriate mathematics course. Currently, advisors suggest enrollment in a math course based on the incoming student's ACT scores and high school courses. Clustering information provides additional valuable input for choosing not only the level of mathematics, but also the type of course best suited for each student. Next, the researcher targeted the Sisyphean Striver group for a mid-semester intervention in testing strategies. Because members of this group were hardworking but inefficient studiers, the researcher felt these students would benefit most from expert guidance. Reactions from students, reports from advisors, and collaboration with colleagues have influenced other ongoing attempts to differentiate instruction, which will be described in more detail in this section.

Placement

Kansas State University has begun to differentiate instruction of College Algebra by offering two different versions of the course. The Traditional College Algebra course covers the fundamental properties of functions and has been taught the same way for the last 25 years. The traditional course is structured so that students meet with an instructor twice a week in a large lecture setting, and then once a week with a teaching assistant for recitation. The content of the course features basic functions, their properties, and rules for correctly manipulating those functions. Students are required to submit a weekly written homework assignment covering short applied problems and a weekly online homework assignment assessing procedural fluency.

Studio College Algebra is a new version of the course that was first offered in the Fall 2007 semester. In this course, one hour of lecture a week is replaced by a studio session in the computer lab. The following chart summarizes the weekly course schedule:

Table 5.1 Studio College Algebra Course Schedule

Class	Lecture	Studio	Recitation
Day of the Week	Wednesday	Thursday/Friday	Monday/Tuesday
Class Size	300-400	45-55	22-27
Description	The course coordinator lectures for 50 minutes, stopping 2 or 3 times to assess understanding with iClicker questions.	Students work in pairs through a guided computer lab assignment involving scenarios using real data.	Students review questions about the homework with a recitation instructor.
Motivation	The students are introduced to a new concept.	The students wrestle with applying the concept to a real-life situation. They learn about the advantages and disadvantages of applying this concept.	After having time to work with the concept in several ways, the student can ask for help with points of confusion.

Studio College Algebra is designed to prepare students for careers in business, agriculture, the social sciences, and fields where they will be interpreting data sets. Therefore, the lessons and assignments emphasize modeling with functions and other common applications, especially in the lab portion of the class. Less time is spent studying the specific properties of these functions beyond demonstrating how different properties affect ways in which the function can be applied to describing data sets. Even though more students are likely to pass Studio College Algebra than Traditional College Algebra, preliminary studies have shown these students are just as likely to earn a C or better in General Calculus as students who pass Traditional College Algebra (Bennett, et. al., 7).

Students in different clusters showed widely diverse reactions to Studio College Algebra and had varying degrees of success passing the course. For example, members of the OverAchiever and Sisyphean Striver groups very much enjoyed Studio College Algebra and the

lab portions of the class. Although they often did not see the connection between the labs and the rest of the course, the students enjoyed relating the concepts about functions to applications relevant to their lives. However, OverAchievers were very successful in the course and were able to earn high grades, while Sisyphean Strivers struggled heavily and often failed despite their hard work. Members of the Employee cluster were ambivalent about Studio College Algebra. While they did not dislike the course, they did not find the content or applications particularly stimulating. Most students earned B's or C's and few dropped out or failed. Rote Memorizers actively disliked the applied aspects of the course. They were confused by seeing mathematics applied to scenarios that did not lead to one correct "solution," and were much more comfortable manipulating equations. Members of this group were the most likely to drop out of the course, and only 58% of those who did finish the term earned a C or better. Interestingly, members of the UnderAchiever group tended to enjoy the lab assignments where they saw new material and more complex scenarios. In general, though, these students found the class boring, and many dropped the class. Students in the UnderAchiever group were challenged by the open-ended assignments, but were frustrated on examinations when they had to justify solutions that seemed obviously true. The average final grade for members of this group was a 2.1, which did not reflect their level of conceptual understanding.

This information about students can be used to more effectively place them into the proper math course. In the Spring 2010 semester, the researcher met with the Arts and Sciences Advising Team to describe the student cluster research and its applications to placement. Each member of the committee received the profiles of student clusters that were discussed in Chapter Four. Included at the bottom of each profile were the following placement suggestions:

OverAchiever Placement Suggestion: Although these students are likely to do well in whatever course they are placed in, they really seem to enjoy Studio College Algebra.

Employee Placement Suggestion: These students enjoy Studio College Algebra and will benefit from learning ways that mathematics is useful through working with programs like Excel. However, they would likely perform just as well in Traditional College Algebra.

UnderAchiever Placement Suggestion: These students should be given more challenging material to keep them interested and motivated. Most of them have already mastered the material presented in College Algebra, and so are ready to take Trigonometry or some other higher-level course.

Sisyphean Striver Placement Suggestion: These students enjoy taking Studio College Algebra and seem to benefit from the class structure and increased time with instructors. They should be given extra opportunities, however, to improve their problem solving skills and mathematical independence by participating in Problem Solving Workshops run by the Mathematics department.

Rote Memorizer Placement Suggestion: These students do not like Studio College Algebra and do not benefit from the extra contact with instructors. They should be placed in Traditional College Algebra where there is less focus on applied problems and using technology. They might also benefit from working with qualified tutors to improve their performance on homework assignments and attendance.

Interestingly, the advisors immediately recognized the student profiles as typifying many of the advisees they had worked with over the years. The members of the advising committee agreed to distribute the profiles and placement suggestions to other faculty members who work with the Dean's office to advise incoming students. Preliminary reports show that using these profiles reduced the numbers of Rote Memorizers and UnderAchievers placed into Studio College Algebra. Cluster analysis of students in Studio College Algebra during the Fall 2010 semester shows that Rote Memorizers consist of only 6.7% of the students. During the four semesters before these profiles were distributed, 10.6% of the students fell into the Rote Memorizer group. Similarly, in the Fall 2010 semester, only 6.4% of the Studio College Algebra students belonging to the UnderAchiever group. This was down from 13.7% for the other four semesters (See Appendix A: Fall 2010). It remains to be seen how augmented advising will affect overall student performance in College Algebra (both Traditional and Studio). In addition, follow-up interviews with advisors should be conducted to determine how they implemented the student profiles and placement suggestions and whether this information was useful.

Intervention for Sisyphean Strivers

Sisyphean Strivers spend a lot of time and effort trying to understand the concepts in Studio College Algebra. They attend almost all of their classes, complete every assignment, and report that they spend at least three to four hours a week studying. When members of this group were interviewed, they showed solid conceptual understanding of functions and their applications. Unfortunately, these students were unable to demonstrate this knowledge during written examinations. Because of their determination and work ethic, the researcher felt students in this cluster would be ideal participants in an academic intervention.

In the Fall 2009 and Spring 2010 semesters, the researcher designed and conducted weekly workshops to help students in the Sisyphean Striver group learn to improve their performance on written examinations. The workshops were scheduled for Tuesday evenings during the time set aside for examinations so that no student would have academic conflicts. While students in the SS group were the intended beneficiaries of the intervention, everyone in Studio College Algebra was invited to participate. The main goals of these Problem Solving Workshops were to reduce testing anxiety by introducing students to the structure of exam questions and to improve the effectiveness of their study habits. At the beginning of each session, students were given handouts with problems linking concepts from that week to previous exam questions. Students then worked in groups with minimal guidance from the instructor to solve the problems on the handout. This collaborative learning was meant to foster confidence among the students so that they did not need to rely on an expert to tell them how to solve each problem. Each exam question was preceded by one or two leading questions highlighting the main ideas and techniques used to solve the problem. The following example is an excerpt from a Problem Solving Workshop covering power functions. Note: the starred problems have appeared on a previous semester's exam.

1) Let $f(x) = 5.7x^{1.7}$ be a power function.

a. What is the coefficient?

b. What is the power?

c. What is the output of the function if the input is x = 0? When the input is x = 2?

d. Is the function increasing or decreasing? How do you know?

***2) Suppose the number of on-duty injuries associated with a certain occupation could be modeled by the function $I(x) = 124.8x^{-.346}$, where *x* represents the numbers of years after 1990.

a. Find *I*(7) and label your answer.

b. Between the years 1995 and 2000, are the number of injuries increasing or decreasing? How do you know?

Despite the researcher's best efforts to design a workshop that would increase the confidence of students and help them learn to demonstrate their knowledge on examinations, results of implementing this intervention were inconclusive. On average, only one or two students attended each workshop, and so they were unable to work in groups as the researcher intended. Several students expressed a desire to attend the workshops, but were unable to come due to work and personal commitments. However, it should be noted that every student who attended more than two workshops earned at least a C in the course. These preliminary findings indicate that if scheduling and attendance issues could be resolved, the workshops might be an effective resource for Sisyphean Strivers.

Attempt at Automated Placement

As a result of the low attendance at the Problem Solving Workshops, the researcher began exploring methods of identifying members of clusters earlier in the semester. Even though the clustering analysis could be implemented by the fifth week of the semester, it seemed that this was still too late to affect student behavior. Implementing an intervention at Week Five had little impact on already established student study habits. Therefore, in order to introduce students to effective support structures and accurately place them into the appropriate mathematics course, members of student clusters must be identified before the beginning of the semester.

Because no student behavior data can be collected before the course begins, clustering techniques are not appropriate for identifying cluster members in this situation. One can instead try procedures for classifying students into the established behavioral groups using "white box" methods. The OverAchiever, Employee, UnderAchiever, Sisyphean Striver, and Rote Memorizer groups provide a framework for employing the traditional mathematics education

research method of using an expert-designed instrument to classify students into groups. Using distinctive responses from the interviews, the researcher built the following list of statements:

- I want to learn more about using computer spreadsheets like Excel.
- Being good at mathematics is something that a person is born with, like being left-handed.
- It is very important to me that I attend a small class where the instructor can keep track of my progress.
- If I don't know how to do a problem, looking back at my notes or the textbook is helpful.
- I usually only understand a new concept after working with a friend or a tutor.
- I anticipate using math in my future career.
- I'm pretty confident in my mathematical skills.
- If I miss class, I can learn the material on my own or with a tutor.
- Math classes can be fun.

These statements were added to the end of the Summer 2010 mathematics placement examination. Every incoming student was asked to rate whether he or she agreed with each statement using a five-point scale. These statements were designed to distinguish members of one group from another by examining their responses, and determine the probability that a given student belongs to a particular cluster.

Several classical and modern methods for constructing a classification scheme based on student responses to these statements exist. The researcher decided to explore the use of Classification and Regression Trees (CART), and random forests possibly simpler and streamlined classification systems to the traditional method of using statistical linear classifiers. Briefly, CART analysis determines the most efficient way to divide students into two relatively homogenous subgroups based on a response to one of the questions. The CART program then repeats the process until all students have been divided into groups containing only members of a single behavior cluster (Breiman, 21). Like the AGNES and PAM algorithms, implementing the CART program involves assigning several parameters and choosing between many techniques that affect the final output. Unfortunately, after running several configurations of the algorithm

on the student data from the Fall 2010 semester, the researcher was unable to find a decision tree that would correctly classify more than 50% of the students.

This result suggests that either that the instrument of classification is faulty, or that students do not have formed opinions about mathematics and Studio College Algebra until they have attended several weeks of class. While this classification attempt did not successfully identify members of student clusters before the start of the semester, it did lead to several possible avenues of further study.

Raising Online Homework Standards

Members of the Employee group are distinguished from the other students by the attitude that Studio College Algebra is like a low paying job. These students do what they feel is required of them and then are "paid" with a passing grade. One can conjecture that members of this group would respond well to raising minimum standards. A fellow researcher at Kansas State University, Bill Weber, tested this conjecture by raising the minimum accepted score for the online homework assignments. Starting in the Fall 2009 semester, students' online homework scores were only recorded if they earned at least 50% of the available points on a given assignment. Any score below 50% counted as a score of zero.

Interestingly, members of the OverAchiever and Sisyphean Striver groups responded most positively to this change in expectations. After the raised standards were enacted, significantly fewer members of these groups scored less than 50% on Online Homework Assignments. This change had the opposite effect for members of the targeted Employee group, however. From this data, one could surmise that these students only complete the tasks they feel are reasonable, and would rather not attempt an assignment than not be rewarded for minimal effort (Weber, 39).

Extensions and Future Research

The difficulty in classifying students before the start of the semester suggests that before attending college, their opinions about mathematics and university mathematics courses are naïve or unformed. However, the cohesion of the behavior clusters identified after the fourth week of classes, and the low attendance at Problem Solving Workshops indicate that students' attitudes and behaviors are well established by the end of the first month of the semester.

Interesting extensions of student clustering research would be investigating what events occur to give rise to these habits and opinions and how these attitudes and behaviors change over time.

Together with a colleague at Kansas State, the researcher has been looking at the stability of these clusters throughout the semester. Drew Cousino, a graduate student in the mathematics department at Kansas State University, is developing a system of tracking student behavior through Bayesian analysis. With this algorithm, one can track shifts in behavior over time by determining the probability that a given student is acting like a member of an established behavior cluster after each assignment. The following figure is an example of a chart tracking a single member of the OverAchiever group during the course of the Fall 2009 semester. The bars are color coded so that each color represents the probability that the student belongs to a behavior cluster based on their accumulated actions up to that point.

Table 5.2 Color Key for Bayesian Behavior Graphs

	OverAchiever
	Employee
	UnderAchiever
	Sisyphean Striver
	Rote Memorizer

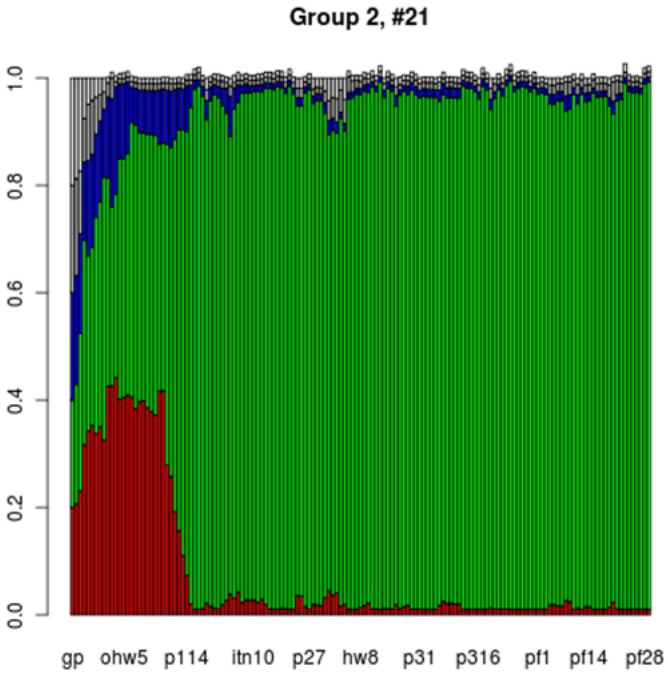

Figure 5.1 Behavior patterns of a student in Group OA

Each bar along the x-axis represents a homework assignment score or examination problem score, presented in the order in which they were graded. The green bars represent the likelihood that this student is a member of the OA group, that is, that they are behaving the way other members of the OA group are behaving. One can see in this chart that this student established his or her behavior pattern by the end of the first examination, and acted consistently like an OverAchiever for the rest of the semester.

This next chart shows the actions of a student who was originally clustered into the UnderAcheiver group, but then changed his or her behavior after the first examination. For most of the remainder of the semester, this student behaved like an OverAchiever.

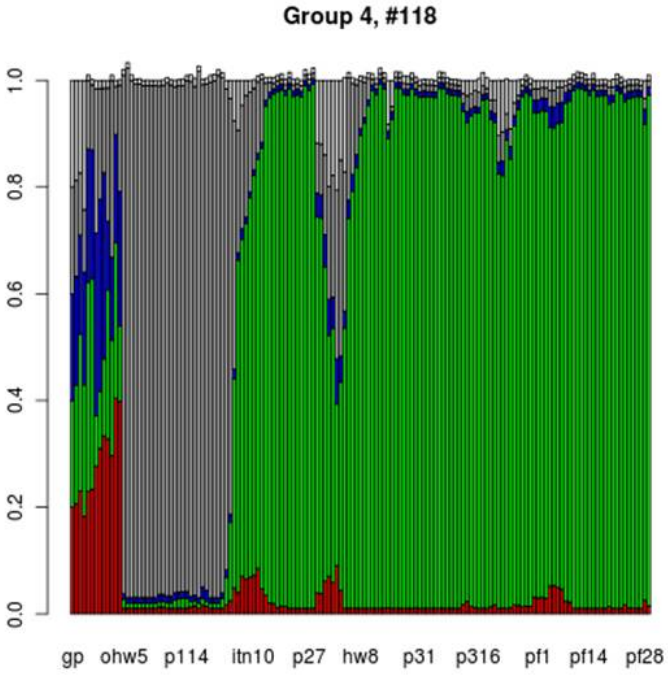

Figure 5.2 Behavior patterns of a student in Group UA

This preliminary analysis indicates that student behavior groups might not be stable throughout the semester. If one tracks each student's behavior with the Bayesian analysis described above, one could form final clusters of students based on their likelihood of belonging to a behavior group by the end of the semester. The chart below compares the composition of the student behavior clusters identified by the first four weeks of student scores with the composition of groups formed by Bayesian analysis using scores from the entire semester. Note

that only those students who completed the final examination and the course are represented in the table.

Table 5.3 Stability of Behavior Groups (Fall 2009)

		Baysian Groups After Final Exam				
		OA	E	UA	SS	RM
Behavior Clusters Identified after first 4 weeks	OA	**73**	11	4	11	0
	E	21	**38**	6	13	1
	UA	4	0	**10**	2	0
	SS	5	0	0	**30**	1
	RM	2	0	0	0	**4**

According to this analysis, the Employee cluster was the least stable, with only 62.5% of identified members of this group behaving as Employees by the end of the semester. Fortunately, 51% of the students who changed behavior patterns ended the semester performing like OverAchievers. In general, 13.8% of students altered their behavior to be most like members of the OverAchiever group. Sadly, the most stable group was the Sisyphean Striver cluster, indicating that most members of this group were not able to learn how to demonstrate their conceptual knowledge on examinations and move to another group.

This preliminary analysis presents many interesting avenues of future inquiry. For example, what events cause students to change their behavior patterns? Many members of the Employee group become engaged in the material part way through the semester and begin participating in class more, earning higher scores on examinations, and generally behaving more like a member of the OverAchiever group. Did a change in attitude cause this shift in activity, or was there an external motivating factor? Individual Bayesian display graphs can identify students who experienced a change in behavior but are unable to explain the cause. Qualitative analysis of student interviews could help researchers gain insight into the influences behind changing student behavior. Eventually, instructors may be able use this information to inspire positive changes in the efforts of their own students.

Other future areas of research involve extending cluster analysis to students beyond those enrolled in Studio College Algebra. For example, do students exhibit these patterns of behavior only in mathematics courses, or do students act the same way in all of their courses? This topic can be explored in two different ways: by tracking the same students in different classes, or by

performing cluster analysis on student populations in different courses. In the first case, one could reveal the stability of groups across disciplines. Does a student belonging to the OverAchiever group excel in Studio College Algebra because he or she enjoys mathematics or because he or she is always a good student? By performing cluster analysis on students in other classes, one could ascertain if other academic disciplines are characterized by the same behavior patterns.

Most students in Studio College Algebra are in their first year of studies. By examining student performance in more advanced courses, one could study how behavior patterns change as a result of years spent in college. Different techniques would be required to study students in higher level classes, as the smaller class size prevent researchers from being able to collect sufficient quantitative data. Likewise, studying student behavior in a smaller institution, like a four-year liberal arts college, would require new research methods. However, this diverges from the goal of providing timely information about students in a large lecture class in order to effectively differentiate instruction. By applying cluster analysis to adapting instruction of large first-year courses, instructors can offer the same advantages of a small class: personalized, effective education.

Bibliography

Baker, Ryan S.J.d. and Yacef, K. "The State of Educational Data Mining in 2009: A Review and Future Visions." *Journal of Educational Data Mining.* Volume 1, Issue 1, October 2009: 3-17.

Bennett, A., Manspeaker, R., Natarajan, R., and Paulhus, J. "Studio College Algebra at Kansas State University," *Moving Forward: Innovations in Introductory Collegiate Mathematics*, Haver, W.E. and S.L. Ganter (Eds.), MAA Reports, Washington, DC: Mathematical Association of America, 2010.

Breiman, Leo, Jerome Friedman, Richard Olshen, and Charles Stone. *Classification And Regression Trees.* New YorK: Chapman & Hall, 1984.

Committee on Developments on the Science of Learning. *How People Learn: Brain, Mind, Experience, and School.* Bransford, John, et. al. National Research Council. Washington, D.C.: National Academy Press, 2000.

"First International Conference on Data Educational Data Mining." *Journal of Educational Data Mining.* 27 March 2010. <http://www.educationaldatamining.org/EDM2008/>

Hall, Tracey, Nicole Strangman, and Anne Meyer. "Differentiated Instruction and Implications for UDL Implementation." *National Center on Accessible Instructional Materials.* 14 Jan. 2011 National Center on Accessing the General Curriculum. 14 March 2011.

Kaufman, Leonard, and Peter J. Rousseeuw. *Finding Groups in Data: an Introduction to Cluster Analysis*. New York: Wiley, 1990.

Lohr, Steve. "Netflix Awards $1 Million Prize and Starts a New Contest." *Bits: Business, Innovation, Technology, Society*. The New York Times. 21 Sep. 2009. Web. 1 Apr. 2011.

Miles, Matthew B., and A. Michael Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*. Second Edition. Thousand Oaks: SAGE Publications, 1994.

Office of Planning and Analysis. "Academic Year Tuition and Required Fees." Kansas State University. 1 Apr. 2011. Web. < http://www.k-state.edu/pa/statinfo/reports/tuition/ksutuition.pdf>

Office of Planning and Analysis. "Big Twelve Longitudinal Retention Survey." Kansas State University. 1 Apr. 2011. Web. <http://www.k-state.edu/pa/statinfo/retention/ethnicity.pdf>

Office of Planning and Analysis. "Student Retention and Graduation Rates." Kansas State University. 1 Apr. 2011. Web. <http://www.k-state.edu/pa/statinfo/factbook/student/retention.pdf>

Parker, Melanie. "Placement, Retention, and Success: A Longitudinal Study of Mathematics and Retention." *Journal of General Education.* Volume 54.1 2005: 22-40.

Pedersen, G. L. "Academic performance and demographic variables in predicting success in college algebra and graduation rates in an urban multi-campus community college." (Florida). Ed.D. dissertation, Florida Atlantic University, 2004.

Prosser, Michael, and Keith Trigwell. *Understanding Teaching and Learning: The Experience in Higher Education*. New York: Society for Research into Higher Education and Open University Press, 1999.

Reinsberg, René. "Netflix." *Technology Review*. Massachusetts Institute of Technology, Feb 2010. Magazine. 1 Apr. 2011.

Skillicorn, David. *Understanding Complex Datasets: Data Mining with Matrix Decompositions*. New York: Chapman & Hall/CRC, 2007.

Subban, Pearl. "A Research Basis Supporting Differentiated Instruction." *Australian Educational Researcher*. 2006. Australian Association for Research in Eduation. 21 March 2011. <http://www.aare.edu.au/06pap/sub06080.pdf>

"Table 409: Average size and scores of eighth-grade mathematics classes and Index of Teachers' Emphasis on Mathematics Homework (EMH), by country." *Digest of Education Statistics.* US. Department of Education. 22, March 2011. <http://nces.ed.gov/programs/digest/d08/tables/dt08_409.asp>

Tan, Pan-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. New York: Pearson Education, 2006.

Tomlinson, Carol Ann, and Jay McTighe. *Integrating Differentiated Instruction and Understanding by Design: Connecting Content and Kids*. Alexandria: Association for Supervision and Curriculum Development, 2006.

Weber, William J. "Effects of Requiring Students to Meet High Expectation Levels Within an On-Line Homework Environment." (Kansas). Ed.D. dissertation, Kansas State University, 2010.

Wormeli, Rick. *Differentiation: From Planning to Practice, Grades 6-12*. Portland: Stenhouse Publishers, 2007.

Appendix A - Clusters

This appendix organizes the cluster information from four semesters: Fall 2008, Spring 2009, Fall 2009, and Spring 2010. For each semester, average scores on assignments are given for both SVD and White Box grouping schemes. Also, PAM and AGNES outputs are displayed.

Fall 2008

SVD Groups

Table A.1 Fall 2008 SVD Group Averages

Fall SVD 37	1.000	2.000	3.000	4.000	5.000
Name	OA	E	UA	SS	RM
Size	119	68	43	58	49
Ave Composite ACT	22.202	23.038	24.032	20.378	22.128
Ave Math ACT	21.657	22.302	23.839	18.649	21.605
Average Exam 1 Score	72.126	70.868	72.605	51.069	62.592
StDev Exam 1 Score	4.106	5.096	5.948	11.365	11.874
# who took Exam 2	119	65	36	55	44
Average Exam 2 Score	58.907	57.953	60.714	53.392	57.909
StDev Exam 2 Score	15.683	19.326	14.191	18.188	15.275
# who took Exam 3	117	63	34	52	38
Average Exam 3 Score	59.923	60.810	59.088	40.385	46.079
StDev Exam 3 Score	12.027	13.407	15.094	13.727	18.535
# who took Final Exam	118	63	35	48	35
Average Final Exam Score	104.314	104.222	106.857	68.438	92.057
StDev Final Exam Score	19.252	22.050	24.060	21.754	22.393
Average Grade in Course	2.898	2.538	2.649	1.389	1.512
StDev Grade in Course	0.789	1.091	1.296	0.920	1.009
% C or better (of completed)	96.610%	86.154%	75.676%	48.148%	58.140%
% completed course	99.160%	95.588%	86.047%	93.103%	87.755%

Table A.2 Fall 2008 Component Averages for SVD Groups

Comp Aves for SVD Grps	Exam 1	Studio	Attendance	WHW	OHW
1 (OA)	72.126	17.244	9.370	51.336	47.903
2 (E)	70.868	13.610	8.603	25.897	45.306
3 (UA)	72.605	12.000	7.256	34.570	48.316
4 (SS)	51.069	12.509	8.655	37.138	42.074
5 (RM)	62.592	10.276	7.082	30.133	24.410

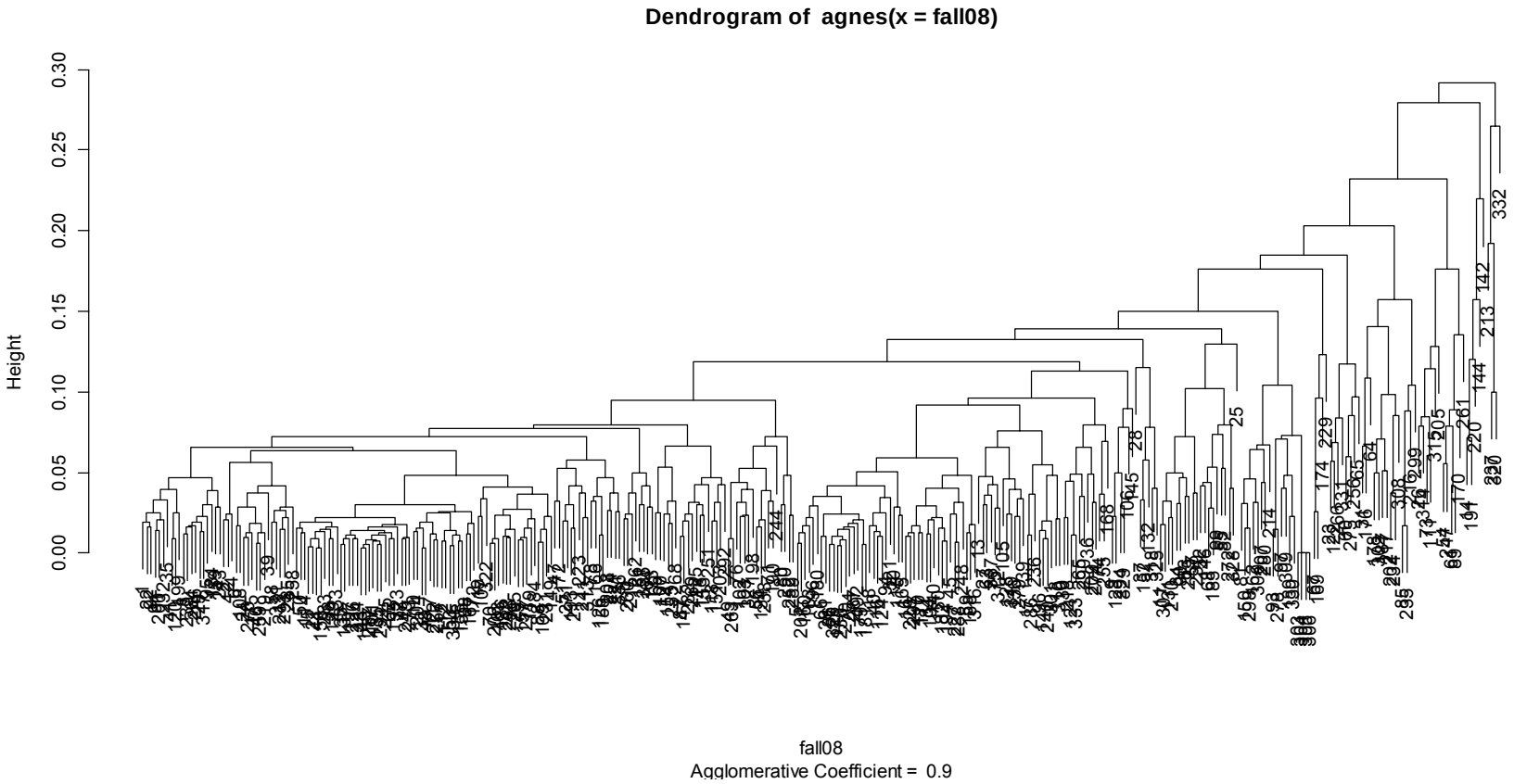

Figure A.1 Fall 2008 AGNES Dendrogram

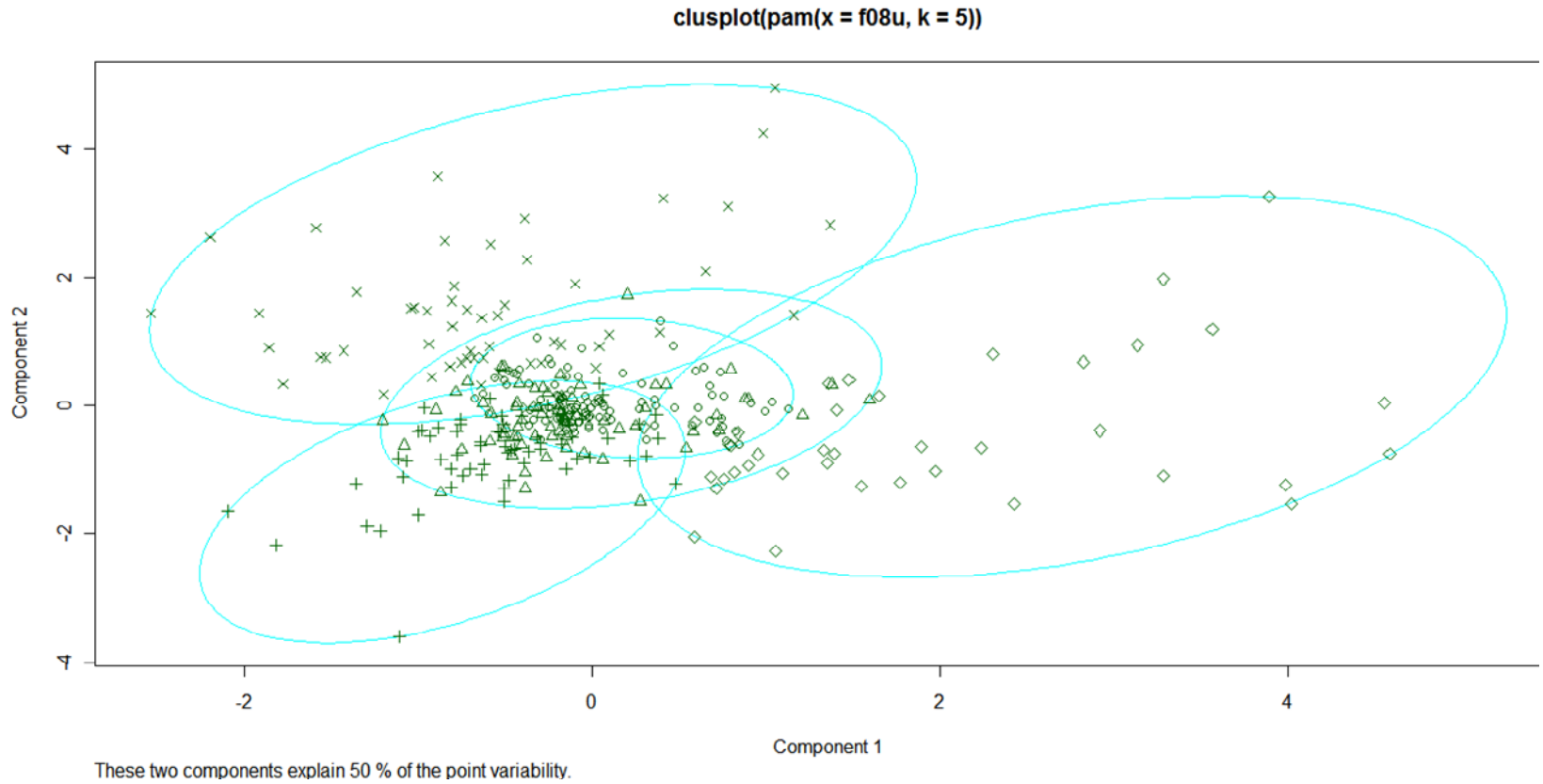

Figure A.2 Fall 2008 PAM plot

White Box Groups

Table A.3 Fall 2008 White Box Group Averages

White Box	1	2	3	4	5
Size	60	64	42	94	76
Ave Composite ACT	22.694	22.846	23.091	21.608	21.923
Ave Math ACT	21.980	22.269	22.667	20.797	21.039
Ave Exam 1 Score	73.067	70.219	69.857	67.968	56.368
StDev Exam 1 Score	4.079	5.079	6.831	7.699	14.629
# who took Exam 2	59	62	41	91	63
Average Exam 2 Score	60.966	56.067	61.366	57.596	54.810
StDev Exam 2 Score	14.061	22.376	14.818	16.081	16.336
# who took Exam 3	58	61	41	90	54
Average Exam 3 Score	63.000	59.541	55.780	54.067	41.907
StDev Exam 3 Score	11.541	11.879	13.462	13.821	16.904
# who took Final Exam	60	62	41	90	46
Average Final Exam Score	108.58	104.73	98.68	93.69	79.04
StDev Final Exam Score	19.42	25.61	21.67	21.74	27.77
Average Grade in Course	3.13	2.59	2.55	2.43	1.07
StDev Final Grade	0.77	1.15	1.11	0.81	0.89
% C or better (of completed	98.33%	84.38%	80.95%	90.00%	36.07%
% completed course	100.00%	100.00%	100.00%	95.74%	80.26%

Table A.4 Fall 2008 Component Averages for White Box Groups

Comp Aves for White Box Grps	Exam 1	Studio	Attendance	WHW	OHW
1	73.067	20.450	9.750	54.492	48.502
2	70.219	17.289	8.891	24.828	47.141
3	69.857	6.440	8.619	36.214	48.367
4	67.968	15.777	8.766	50.154	45.949
5	56.368	8.178	6.724	24.237	28.553

Comparison

Table A.5 Fall 2008 SVD and White Box Group Comparison

SVD /WB groups	1	2	3	4	5
1 (OA)	56	5	8	50	0
2 (E)	1	39	15	2	11
3 (UA)	0	11	14	7	6
4 (SS)	2	8	5	21	22
5 (RM)	1	1	0	11	36

Spring 2009

SVD Groups

Table A.6 Spring 2009 SVD Group Averages

Fall SVD 37	1	2	3	4	5
Name	OA	E	RM	UA	SS
Size	60	30	30	36	42
Ave Composite ACT	22.739	21.652	19.222	22.409	21.680
Ave Math ACT	21.826	20.609	17.278	20.545	20.160
Average Exam 1 Score	69.083	66.033	40.300	58.778	61.690
StDev Exam 1 Score	7.827	8.802	16.028	8.445	9.285
# who took Exam 2	59	30	30	34	42
Average Exam 2 Score	67.491	50.100	33.833	51.030	52.725
StDev Exam 2 Score	9.527	20.174	19.797	13.487	16.966
# who took Exam 3	58	29	22	32	40
Average Exam 3 Score	60.845	50.276	29.857	44.313	48.500
StDev Exam 3 Score	16.237	18.723	17.517	16.847	18.210
# who took Final Exam	59	27	26	31	36
Average Final Exam Score	112.034	101.900	62.111	86.219	90.778
StDev Final Exam Score	26.046	25.475	29.543	30.242	27.187
Average Grade in Course	3.383	2.867	1.138	2.057	2.214
StDev Grade in Course	0.804	0.973	1.090	0.846	1.122
% C or better (of complete	98.33%	93.33%	57.14%	80.00%	78.57%
% completed course	100.00%	100.00%	93.33%	97.22%	100.00%

Table A.7 Spring 2009 Component Averages for SVD Groups

Comp Aves for SVD Grps	Exam 1	Studio	Attendance	WHW	OHW
1 (OA)	69.083	19.275	8.806	30.383	44.392
2 (E)	66.033	17.633	7.289	29.600	47.217
3 (RM)	40.300	10.833	6.889	17.017	31.907
4 (UA)	58.778	16.597	8.241	18.931	35.158
5 (SS)	61.690	15.298	7.952	22.500	42.524

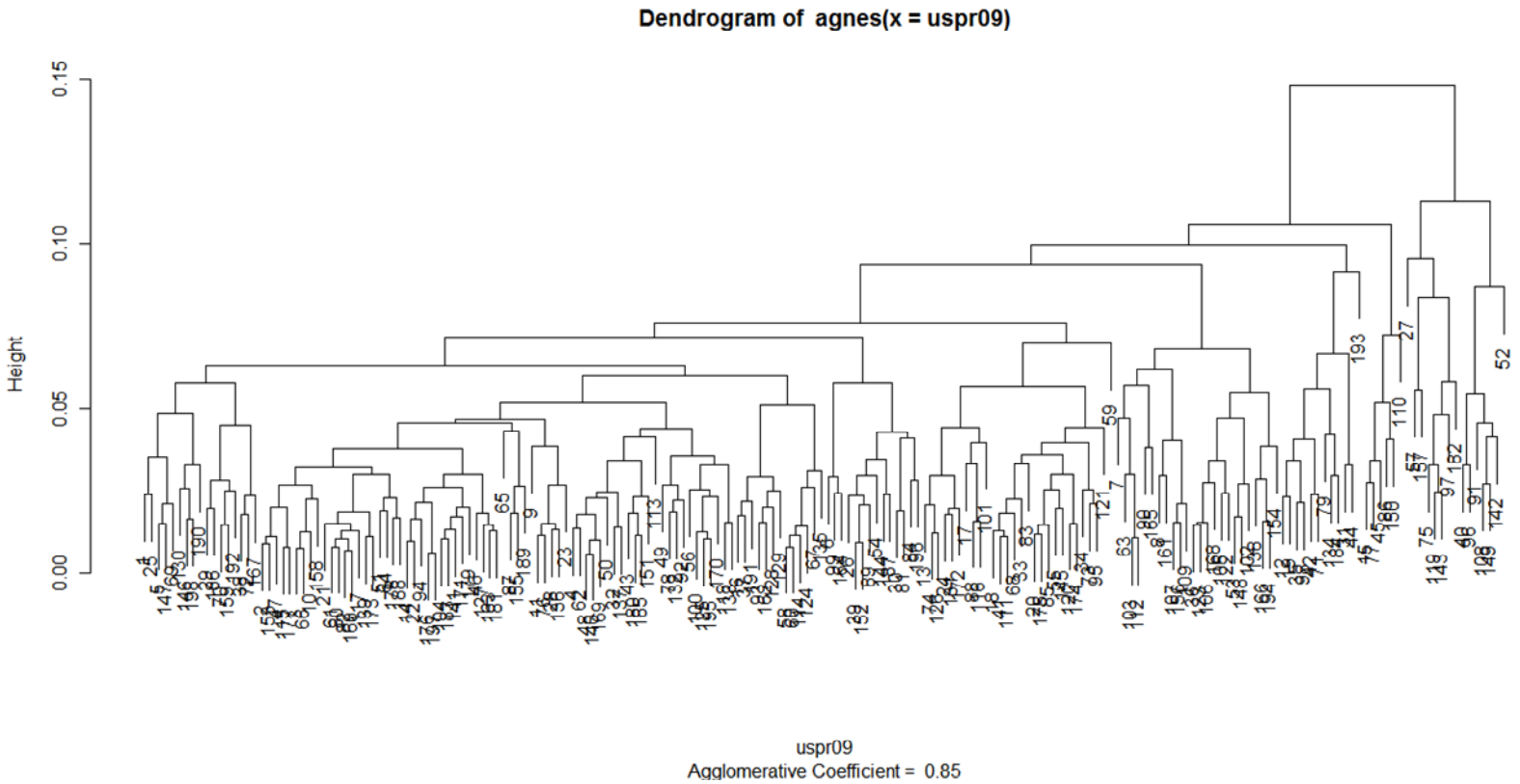

Figure A.3 Spring 2009 AGNES Dendrogram

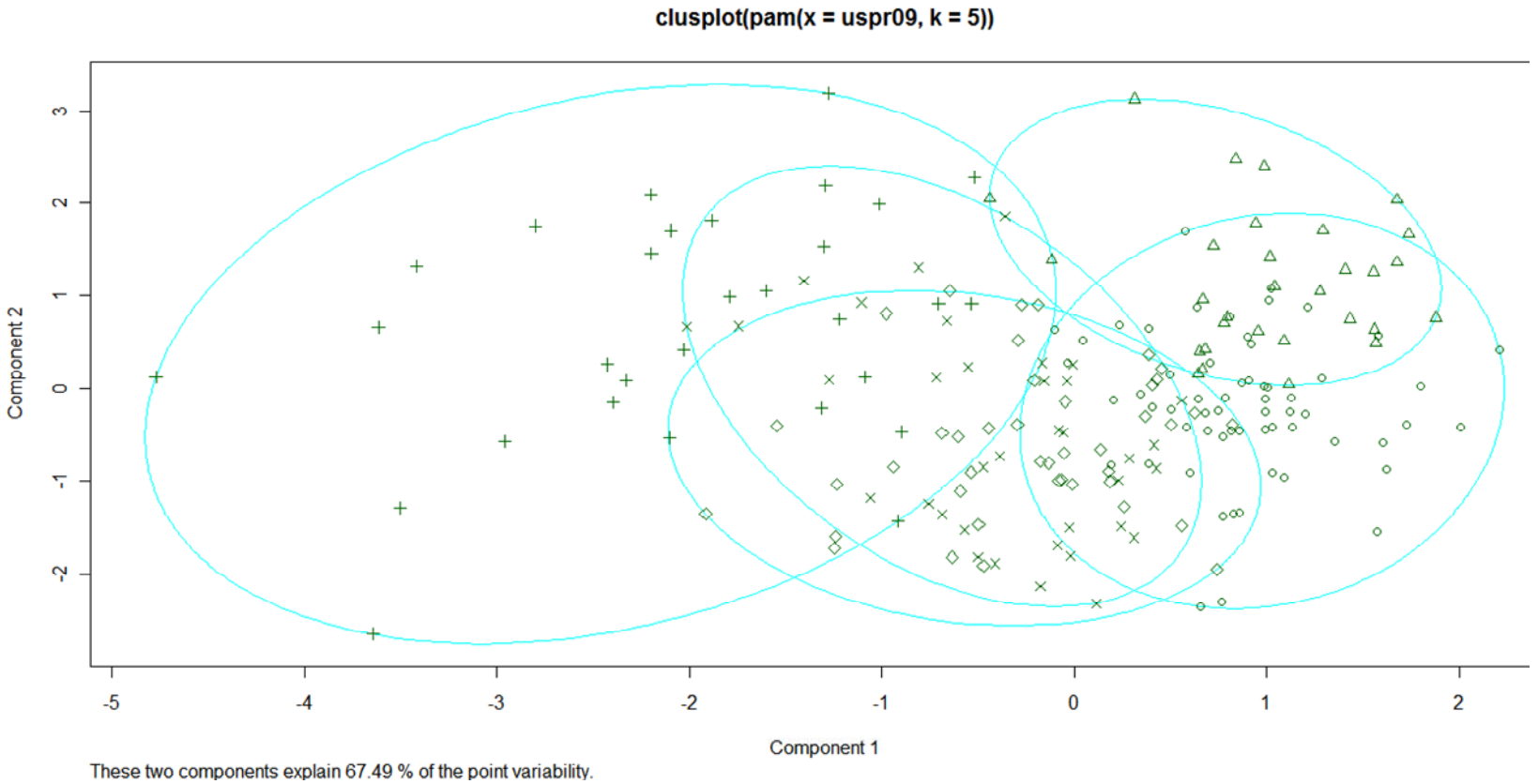

Figure A.4 Spring 2009 PAM plot

White Box Groups

Table A.8 Spring 2009 White Box Group Averages

White Box	1	2	3	4	5
Size	33	67	29	50	19
Ave Composite ACT	23.100	23.612	22.267	20.568	19.818
Ave Math ACT	22.350	22.878	20.067	18.946	18.000
Ave Exam 1 Score	67.970	71.030	53.966	52.560	44.579
StDev Exam 1 Score	6.473	5.158	15.776	8.744	16.331
# who took Exam 2	33	66	26	48	17
Average Exam 2 Score	58.879	65.258	44.423	44.667	35.824
StDev Exam 2 Score	15.159	12.495	19.441	17.669	18.925
# who took Exam 3	31	65	24	46	14
Average Exam 3 Score	54.000	60.662	38.917	41.391	36.929
StDev Exam 3 Score	18.719	15.658	19.576	17.042	18.524
# who took Final Exam	25	59	24	44	17
Average Final Exam Score	102.120	101.051	91.680	92.159	90.706
StDev Final Exam Score	24.764	29.318	27.856	26.316	29.529
Average Grade in Course	2.727	3.418	1.500	2.160	1.118
StDev Final Grade	1.039	0.678	1.171	0.934	1.166
% C or better (of completed)	93.94%	100.00%	57.14%	78.00%	41.18%
% completed course	100.00%	100.00%	96.55%	100.00%	89.47%

Table A.9 Spring 2009 Component Averages for White Box Groups

Comp Aves for WB Grps	Exam 1	Studio	Attendance	WHW	OHW
1	67.970	15.167	8.566	16.091	39.236
2	71.030	19.724	8.607	31.612	47.527
3	53.966	9.052	5.552	17.086	37.800
4	52.560	19.280	8.433	29.540	44.838
5	44.579	10.632	7.491	11.921	14.300

Comparison

Table A.10 Spring 2009 SVD and White Box Group Comparison

SVD /WB groups	1	2	3	4	5
1 (OA)	1	42	3	9	1
2 (E)	2	16	4	8	0
3 (RM)	1	0	9	9	11
4 (UA)	12	2	5	10	7
5 (SS)	13	7	8	14	0

Fall 2009

SVD Groups

Table A.11 Fall 2009 SVD Group Averages

Fall SVD 33	1	2	3	4	5
Name	SS	OA	E	UA	RM
Size	58	124	110	44	26
Ave Composite ACT	19.857	23.596	21.058	22.033	22.632
Ave Math ACT	18.238	23.034	19.907	20.933	21.526
Average Exam 1 Score	43.603	69.395	58.609	53.068	50.269
StDev Exam 1 Score	10.261	5.132	7.212	14.378	19.046
# who took Exam 2	50	121	109	38	17
Average Exam 2 Score	35.939	58.483	45.623	38.806	42.188
StDev Exam 2 Score	14.152	12.146	13.452	18.258	19.292
# who took Exam 3	47	119	103	30	16
Average Exam 3 Score	51.000	65.605	56.350	50.400	49.688
StDev Exam 3 Score	16.881	12.319	15.060	19.725	21.010
# who took Final Exam	48	121	105	33	12
Average Final Exam Score	80.563	115.174	98.248	84.909	86.417
StDev Final Exam Score	26.985	18.564	22.980	35.488	41.890
Average Grade in Course	2.176	3.472	2.556	1.857	1.045
StDev Grade in Course	1.126	0.793	1.008	1.353	1.430
% C or better (of completed)	40.00%	97.56%	87.16%	65.71%	39.13%
% completed course	86.21%	99.19%	99.09%	79.55%	88.46%

Table A.12 Fall 2009 Component Averages for SVD Groups

Comp Aves for SVD Grps	Exam 1	Studio	Attendance	WHW	OHW
1 (SS)	43.603	16.095	7.004	17.578	43.997
2 (OA)	69.395	19.492	7.256	20.391	47.935
3 (E)	58.609	16.168	6.991	16.968	43.268
4 (UA)	53.068	14.989	6.790	10.261	19.973
5 (RM)	50.269	1.923	2.760	7.404	25.273

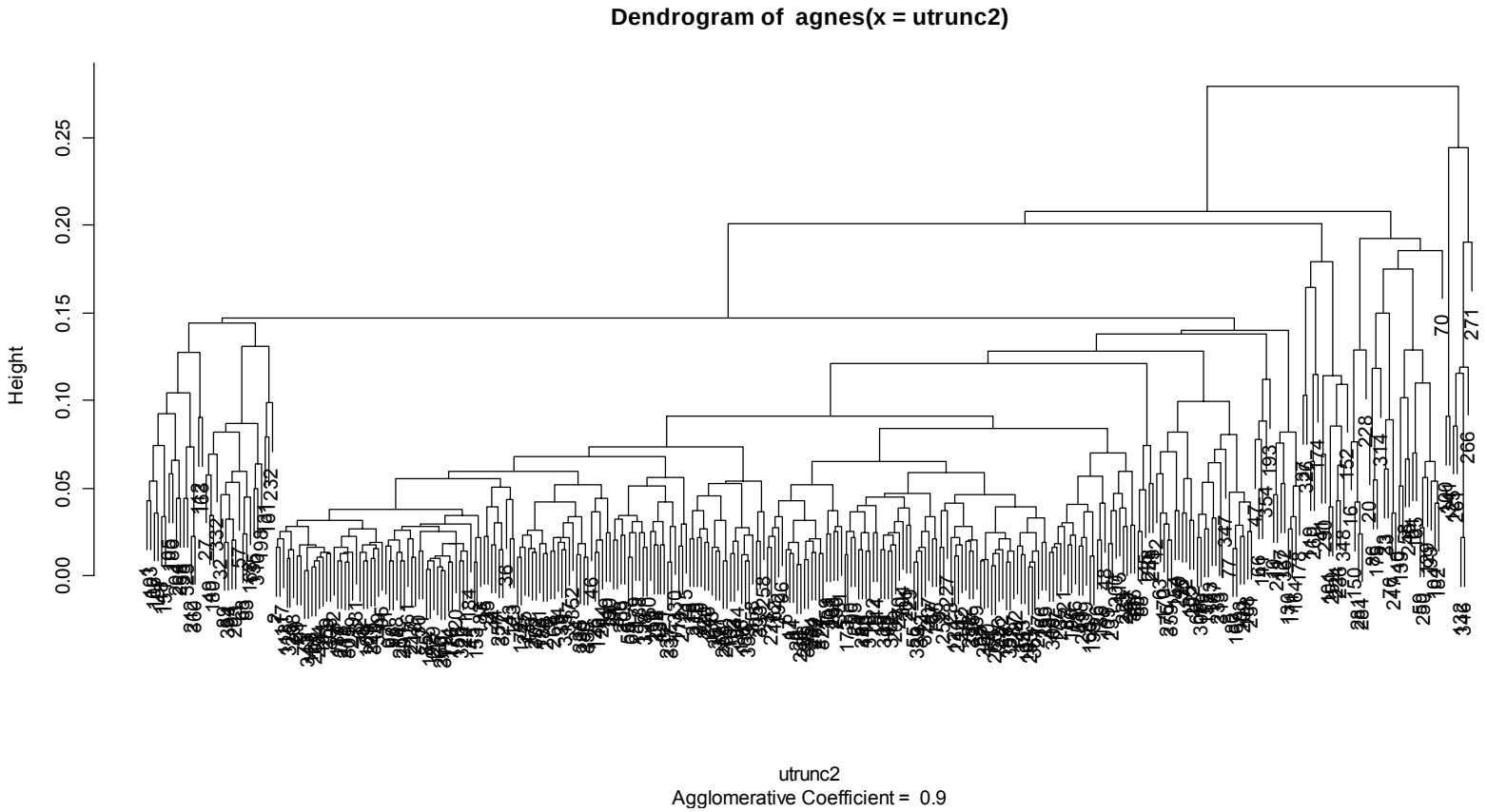

Figure A.5 Fall 2009 AGNES Dendrogram

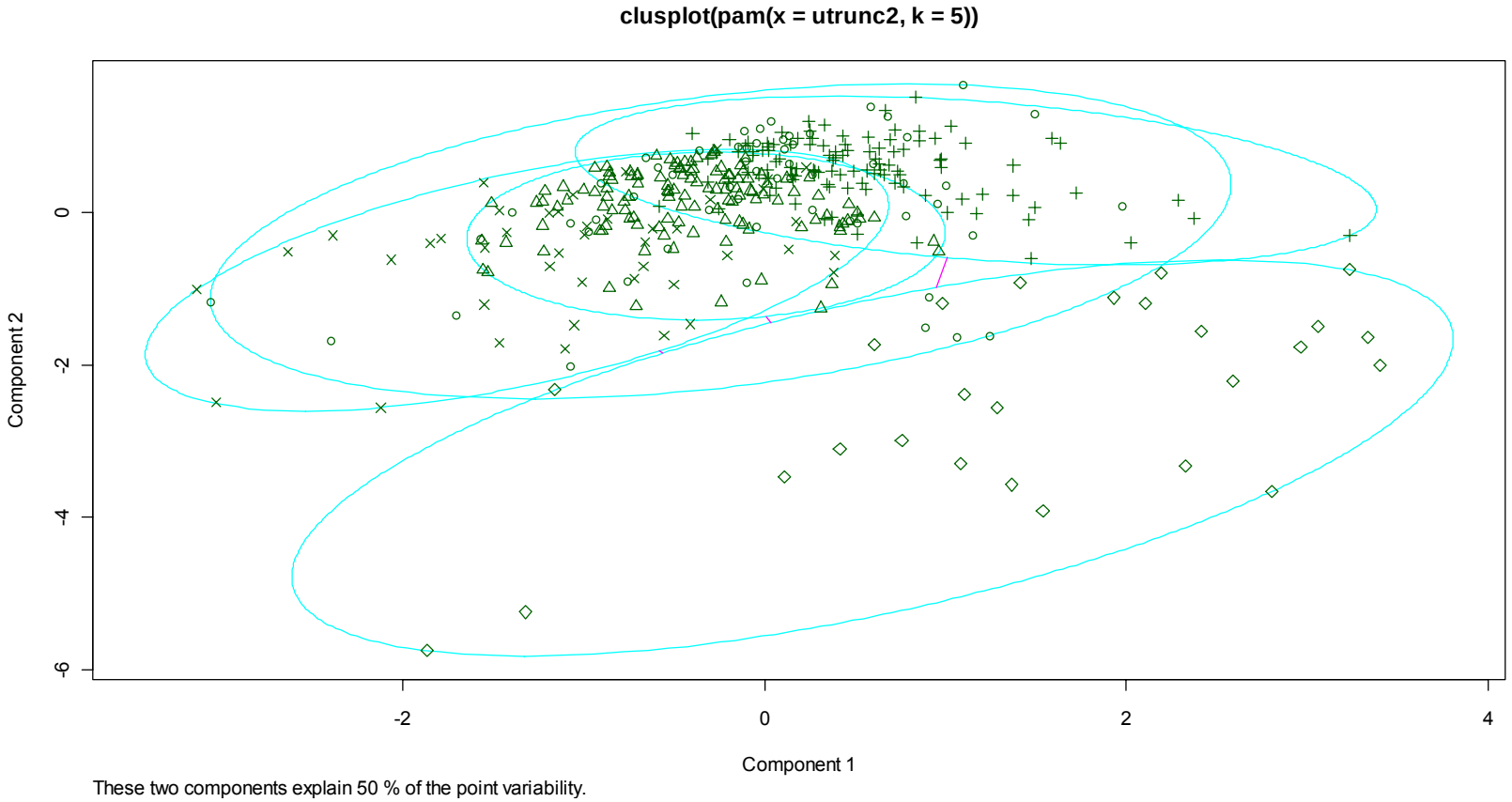

Figure A.6 Fall 2009 PAM plot

White Box Groups

Table A.13 Fall 2009 White Box Group Averages

White Box	1	2	3	4	5
Size	41	125	75	98	25
Ave Composite ACT	19.962	23.315	21.459	21.284	22.375
Ave Math ACT	19.115	22.652	20.525	19.838	20.750
Ave Exam 1 Score	41.512	69.136	60.947	53.622	45.783
StDev Exam 1 Score	12.842	4.954	9.412	8.969	17.901
# who took Exam 2	35	123	71	95	14
Average Exam 2 Score	31.129	59.273	42.721	43.348	41.538
StDev Exam 2 Score	16.128	10.840	15.717	12.963	20.472
# who took Exam 3	28	120	67	88	12
Average Exam 3 Score	38.929	67.050	54.955	55.591	50.417
StDev Exam 3 Score	19.187	10.093	15.350	15.044	21.786
# who took Final Exam	28	125	66	91	9
Average Final Exam Score	65.750	116.208	97.258	93.582	72.889
StDev Final Exam Score	34.459	17.339	25.193	21.842	44.823
Average Grade in Course	1.063	3.600	2.449	2.489	0.789
StDev Final Grade	0.914	0.622	1.022	0.981	1.273
% C or better (of completed)	48.39%	99.19%	86.96%	85.11%	33.33%
% completed course	78.05%	99.20%	92.00%	95.92%	91.30%

Table A.14 Fall 2009 Component Averages for White Box Groups

Comp Aves for WB Grps	Exam 1	Studio	Attendance	WHW	OHW
1	41.512	11.268	6.427	5.988	21.820
2	69.136	20.580	7.380	20.984	48.467
3	60.947	10.733	6.760	16.500	44.652
4	53.622	20.082	7.102	18.755	40.405
5	45.783	1.348	2.457	4.978	23.022

Comparison

Table A.15 Fall 2009 SVD and White Box Group Comparison

SVD /WB groups	1	2	3	4	5
1 (SS)	9	2	12	34	1
2 (OA)	0	101	19	4	0
3 (E)	7	21	35	46	1
4 (UA)	25	1	4	14	0
5 (RA)	0	0	5	0	21

Spring 2010

SVD Groups

Table A.16 Spring 2010 SVD Group Averages

Fall SVD 33	1	2	3	4	5
Name	UA	SS	E	OA	RM
Size	25	19	46	36	4
Ave Composite ACT					
Ave Math ACT					
Average Exam 1 Score	57.320	53.895	62.196	63.417	23.500
StDev Exam 1 Score	11.357	7.944	14.886	8.534	7.188
# who took Exam 2	21	18	45	34	3
Average Exam 2 Score	43.095	44.556	50.933	58.912	10.000
StDev Exam 2 Score	18.014	16.561	19.565	13.574	11.136
# who took Exam 3	19	16	45	33	2
Average Exam 3 Score	50.211	47.000	54.711	61.500	6.000
StDev Exam 3 Score	20.735	17.974	19.346	13.523	8.485
# who took Final Exam	18	15	44	33	1
Average Final Exam Score	85.471	98.933	94.705	114.242	12.000
StDev Final Exam Score	31.209	25.789	34.520	16.967	0.000
Average Grade in Course	1.82609	1.94444	2.74419	3.30303	0
StDev Grade in Course	1.15413	1.25895	1.19708	0.88335	0
% C or better (of completed)	73.91%	72.22%	90.70%	96.97%	0.00%
% completed course	92.00%	94.74%	93.48%	91.67%	75.00%

Table A.17 Spring 2010 Component Averages for SVD Groups

Comp Aves for SVD Grps	Exam 1	Studio	Attendanc	WHW	OHW
1 (UA)	57.320	9.980	5.130	11.080	38.224
2 (SS)	53.895	12.947	6.303	14.632	40.000
3 (E)	62.196	16.054	6.402	17.174	44.357
4 (OA)	63.417	17.222	7.076	18.708	45.525
5 (RM)	23.500	1.000	3.563	10.625	25.000

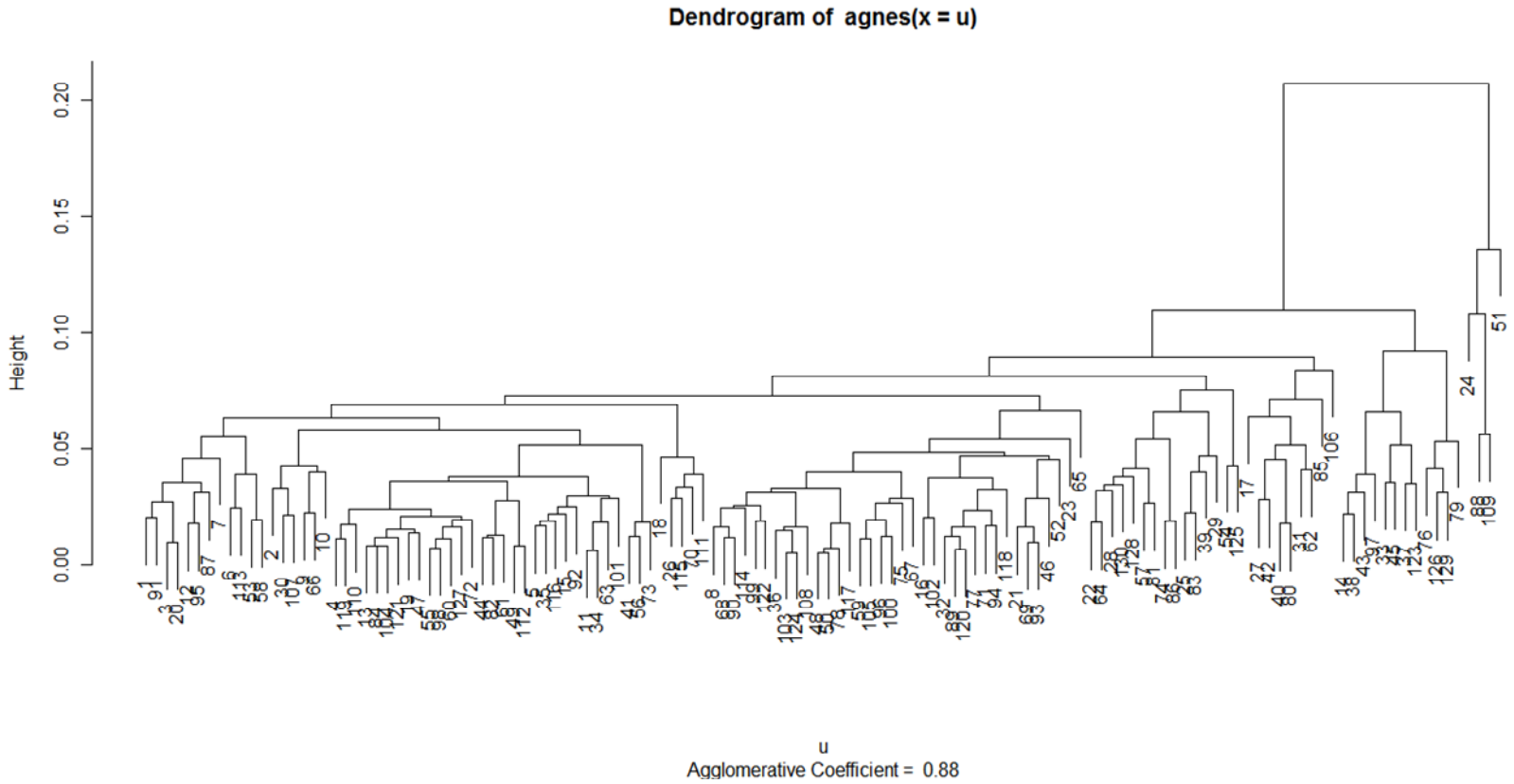

Figure A.7 Spring 2010 AGNES Dendrogram

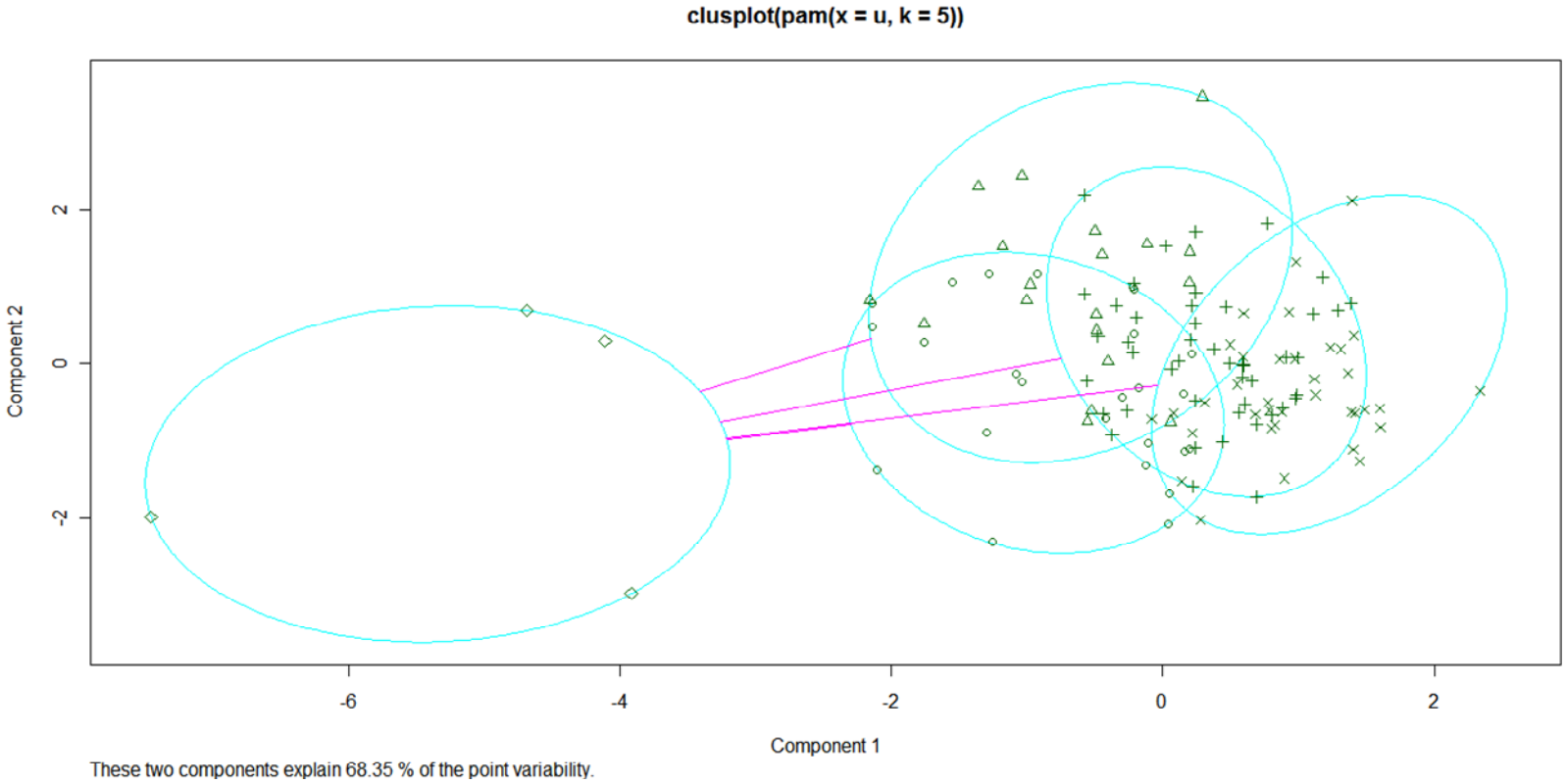

Figure A.8 Spring 2010 PAM plot

White Box Groups

Table A.18 Spring 2010 White Box Group Averages

White Box	1	2	3	4	5
Size	49	33	19	15	14
Ave Composite ACT					
Ave Math ACT					
Ave Exam 1 Score	63.0204	54.697	75.789474	49.0667	44.7143
StDev Exam 1 Score	8.21454	11.4414	4.1039134	11.4796	15.2236
# who took Exam 2	49	32	19	13	8
Average Exam 2 Score	52.9592	44.375	69.105263	36.6154	28.5
StDev Exam 2 Score	14.0163	18.1601	10.671326	15.3977	25.202
# who took Exam 3	49	31	19	11	5
Average Exam 3 Score	55.449	51.7097	69.710526	43.9091	16.4
StDev Exam 3 Score	16.3275	16.6277	7.3848866	18.9866	30.4105
# who took Final Exam	48	29	19	10	4
Average Final Exam Score	103.604	89.4828	120.63158	89.1	33.75
StDev Final Exam Score	24.5355	28.1179	20.393928	24.3285	59.7683
Average Grade in Course	2.95745	2.23333	3.8421053	1.46154	0.54545
StDev Final Grade	0.90787	1.0063	0.5014599	1.19829	0.9342
% C or better (of completed)	97.87%	86.67%	100.00%	53.85%	27.27%
% completed course	95.92%	90.91%	100.00%	86.67%	78.57%

Table A.19 Spring 2010 Component Averages for White Box Groups

Comp Aves for WB Grps	Exam 1	Studio	Attendance	WHW	OHW
1	63.020	16.908	6.724	20.112	47.631
2	54.697	14.167	6.970	12.576	37.691
3	75.789	22.947	7.421	21.763	49.442
4	49.067	7.800	4.367	14.633	42.687
5	44.714	0.643	3.250	1.964	24.107

Comparison

Table A.20 Spring 2010 SVD and White Box Group Comparison

SVD /WB groups	1	2	3	4	5
1 (UA)	5	7	1	5	7
2 (SS)	6	7	0	5	1
3 (E)	18	10	12	3	3
4 (OA)	20	9	6	1	0
5 (RM)	0	0	0	1	3

Fall 2010

SVD Groups

Table A.21 Fall 2010 SVD Group Averages

Fall SVD 33	1	2	3	4	5
Name	SS	OA	UA	E	RM
Size	71	128	20	70	21
Ave Composite ACT					
Ave Math ACT					
Average Exam 1 Score	54.972	70.211	57.950	62.100	33.762
StDev Exam 1 Score	7.323	5.944	11.830	7.900	10.723

Table A.22 Fall 2010 Component Averages for SVD Groups

Comp Aves for SVD Grps	Exam 1	Studio	Attendance	WHW	OHW
1 (SS)	54.972	17.725	8.331	19.401	44.345
2 (OA)	70.211	19.547	8.292	19.102	46.077
3 (UA)	57.950	14.200	6.421	12.425	17.380
4 (E)	62.100	11.879	6.469	16.050	43.220
5 (RM)	33.762	11.024	6.179	8.595	24.062

Appendix B - Data Analysis

Table B.1 Fall 2008 Trial V Vectors

Exam 1	Problems:						
1	2	3	4	5	6	7	8
-0.0931446	0.127756	-0.020122	0.052501	-0.133728	0.245594	-0.049521	0.0413895
-0.11629143	0.100817	0.007135	0.025736	0.04744	0.30072	-0.026535	0.0675214
-0.14308697	0.036509	-0.013401	0.15722	0.230119	-0.316634	-0.100284	-0.2073237
-0.13019854	0.11002	-0.135904	0.0704	0.033367	-0.289797	0.082335	-0.1169879

Exam 1 Continued							
9	10	11	12	13	14	15	16
-0.01758345	0.368744	-0.179355	0.416099	-0.092017	0.099959	-0.144149	0.1367354
-0.13672515	-0.110484	-0.331851	0.26355	-0.245418	-0.05666	0.063836	0.0152399
0.012445704	-0.046978	0.066491	0.249777	0.232997	0.14527	0.066135	0.1177162
0.2401694	0.011147	0.082893	0.103245	-0.032718	-0.142218	-0.035884	0.3441482

Written Homework			
WHW 1	WHW 2	WHW 3	WHW 4
-0.0439013	0.126087	-0.357214	-0.172823
0.32254425	-0.307113	0.251018	0.128136
0.12510554	-0.057934	0.192713	-0.006145
-0.07009467	0.136586	-0.111621	0.33098

Studio	
ST 1	ST 2
0.199724	-0.140027
0.0562	-0.08696
0.05963	-0.037659
0.068059	-0.272512

KEY:	<-.2	-0.2>X>-0.1	
	>.2	.1>X>.1	
	Fall 08 only		
Highly Contributing Score			

Attendance									
AT 1	AT 2	AT 3	AT 4	AT 5	AT 6	AT 7	AT 8	AT 9	AT 10
0.010071257	0.232614	0.223082	-0.154723	0.010122	0.071965	-0.142165	0.0409439	-0.026913	0.01205
0.081088185	0.018217	-0.122219	0.158103	-0.213525	-0.127573	0.092238	-0.1065933	-0.206603	0.311498
0.0878264	0.12008	0.175248	0.092538	-0.123753	0.198048	0.184454	-0.1520981	0.219605	0.070715
0.25342298	-0.134426	-0.021702	-0.083265	-0.13618	-0.372816	0.084811	0.0366409	0.00717	-0.078917

Online Homework				
OHW 1	OHW 2	OHW 3	OHW 4	OHW 5
-0.05318851	0.210139	-0.001149	0.032149	-0.162785
-0.08395434	-0.123083	-0.077738	0.00621	0.031139
0.23785467	-0.051289	0.027279	0.334388	-0.236332
-0.28473928	-0.050264	0.023822	0.050055	0.161805

Inverse time to 90%				
ITN 1	*ITN 2*	*ITN 3*	*ITN 4*	*ITN 5*
-0.00063635	*-0.13176*	*0.050981*	*0.051764*	*0.010169*
0.05225074	*-0.01841*	*-0.10032*	*-0.00792*	*0.004224*
-0.17892418	*-0.09287*	*0.132473*	*0.022128*	*-0.09181*
0.091818053	*-0.05418*	*-0.05577*	*0.037516*	*0.089552*

Readiness Test	***Pretest***
-0.035919	*-0.02554*
0.0126463	*-0.00034*
0.0102129	*-0.04398*
-0.047696	*-0.00659*

V Vectors

Fall 2008/Spring 2009

Table B.2 Fall 2008 V Vectors

Fall 08	***EX 1.1***	***EX 1.2***	EX 1.3	EX 1.4	***EX 1.5***	EX 1.6	***EX 1.7***	***EX 1.8***	***EX 1.9***	***EX 1.10***	EX 1.11	***EX 1.12***
Spring 09	***EX 1.1***	***EX 1.2***	EX 1.3	EX 1.4	***EX 1.5***	EX 1.6	***EX 1.7***	***EX 1.8***	***EX 1.9***	***EX 1.10***	EX 1.16	***EX 1.12***
	0.0984	-1.333	0.0893	-0.171	0.1848	-0.0632	0.1941	-0.0367	0.1018	-0.241	0.369	-0.1926
	-0.1279	-0.1014	-0.0501	-0.2011	0.2288	-0.15	0.1626	0.1298	-0.1288	-0.2108	0.0735	-0.4634
	-0.1677	-0.0527	-0.1709	0.1403	-0.3261	0.0002	-0.1004	-0.1274	-0.2995	-0.0545	0.1889	0.1853
	-0.1507	-0.1369	0.0634	0.1377	-0.2367	0.0802	-0.2559	-0.1891	0.0008	0.1006	-0.0035	-0.0978

Fall 08	***EX 1.13***	EX 1.14	EX 1.15	EX 1.16
Spring 09	***EX 1.11***	EX 1.14	EX 1.13	EX 1.15
	-0.1497	0.1966	-0.0192	-0.0252
	-0.1052	-0.1341	0.0326	0.0034
	-0.2323	-0.0399	-0.0898	0.0295
	-0.1561	0.1855	0.4113	0.116

HW 1	HW 2	HW 3	***HW 4***
0.4147	-0.0022	0.0787	0.2216
-0.3952	-0.0836	-0.1319	0.1273
-0.2373	-0.0328	-0.0332	0.205
-0.0416	-0.153	0.2671	0.1538

ST 1	ST 2
-0.3227	0.0238
0.0671	0.0456
-0.1141	0.0302
0.0949	-0.3302

Fall 08	AT 1	AT 2	AT 3	AT 4	***AT 5***	***AT 6***	***AT 7***	AT 8	***AT 9***	AT 10
Spring 09	AT 1	IC1	AT 2	AT 3	***IC 2***	***AT 4***	***AT 5***	***IC 3***	***AT 6***	***IC 4***
	-0.1446	0.0707	0.103	-0.0843	-0.1346	-0.0242	0.0659	-0.0792	-0.1383	0.0723
	0.0191	-0.2677	-0.0754	-0.0875	0.2447	0.3165	-0.1917	0.0105	-0.0454	-0.0223
	0.0565	-0.1814	-0.2129	-0.0083	-0.171	-0.189	-0.2169	0.1926	-0.1164	0.2738
	-0.1611	-0.0937	-0.0194	0.2535	0.2036	0.1795	0.1836	-0.009	-0.1641	-0.1375

OHW 1	***OHW 2***	OHW 3	OHW 4	OHW 5
0.0812	-0.0119	0.0191	-0.1582	-0.015
-0.0424	-0.0972	0.0273	0.0184	-0.0226
0.1727	0.2571	0.0429	-0.1614	-0.0548
0.0031	-0.1213	0.0496	0.0377	0.0367

KEY:	<-.2	-0.2>X>-0.1
	>.2	.1>X>.1
	Fall 08 only	
Highly Contributing Score		

Table B.3 Fall 2009 V Vectors

Fall 09	***EX 1.1***	***EX 1.2***	***EX 1.3***	EX 1.4	EX 1.5	***EX 1.6***	***EX 1.7***	EX 1.8	***EX 1.9***	EX 1.10	***EX 1.11***	***EX 1.12***
Spring 10	***EX 1.1***	***EX 1.2***	***EX 1.3***	EX 1.4	EX 1.5	***EX 1.6***	***EX 1.7***	EX 1.8	***EX 1.9***	EX 1.10	***EX 1.12***	***EX 1.13***
	0.1587	0.0711	0.1476	0.0288	-0.1809	0.3623	-0.0526	-0.037	-0.0304	0.1072	0.1386	0.5537
	-0.3486	0.4657	-0.0972	-0.0859	-0.0038	0.2071	-0.2108	0.1036	0.2153	-0.0728	0.2653	-0.3747
	0.0375	-0.1263	0.2146	0.3258	0.1944	-0.2584	-0.4704	-0.0648	0.1247	0.1429	0.1018	-0.0761
	0.2118	-0.0361	0.0041	0.0718	-0.0631	-0.0506	0.2022	-0.1245	-0.0857	-0.0795	-0.0036	-0.1449

Fall 09	***EX 1.13***	EX 1.14	***EX 1.15***	EX 1.16
Spring 10	***EX 1.11***	EX 1.14	***EX 1.15***	EX 1.16
	0.3138	-0.0624	-0.1196	-0.042
	-0.0113	-0.0347	0.0661	-0.0076
	0.1743	-0.0469	-0.021	0.1651
	-0.1835	-0.2845	0.5621	0.0773

HW 1	HW 2
0.2254	-0.1471
0.2157	-0.0879
0.2045	-0.0403
0.2201	-0.07

ST 1	ST 2
-0.098	0.0799
0.0417	-0.1218
0.0265	0.0786
-0.3866	0.1027

AT 1	AT 2	***AT 3***	***AT 4***	***AT 5***	***AT 6***	***IC 1***	IC2
0.0694	0.0664	-0.1454	0.2074	-0.1057	0.1384	-0.1566	0.0904
0.123	-0.007	-0.1412	0.0075	-0.1864	0.1203	0.1126	0.0761
0.3389	0.0844	0.1263	-0.069	0.1004	-0.08	-0.1232	0.1063
0.0682	0.0428	0.001	-0.1511	0.0658	-0.1038	-0.064	-0.0458

OHW 1	***OHW 2***	OHW 3	***OHW 4***	***OHW 5***
0.1509	-0.0977	0.1036	-0.1119	0.2046
0.0615	-0.2243	0.0541	-0.1253	0.0515
0.1369	-0.121	0.1086	0.1547	-0.288
0.2585	-0.1073	0.0523	-0.2184	0.1477

KEY:	<-.2	-0.2>X>-0.1
	>.2	.1>X>.1
Highly Contributing Score		

Types of Exam Problems

Table B.4 Exam Problem Descriptions

Exam 1	Standard procedural problems	Nonstandard procedural problems	Graphing/ Slope problems	Standard Applied problems	Nonstandard Applied problems
Fall 2008	1, 3, 8, 9, 10	2, 4	5, 6, 7, 15	12, 13, 14, 16	11
Spring 2009	1, 3, 8, 9, 10	2, 4	5, 6, 7, 13	11, 12, 14, 15	16
Fall 2009	1, 3, 8, 9, 10	2, 4	5, 6, 7	11, 12, 13, 16	14, 15
Spring 2010	1, 3, 8, 9, 10	2, 4	5, 6, 7	11, 12, 13, 16	14, 15

Medoid Coordinates

Key:

Key:
Very High High Low Very Low
***Note: Coordinates are considered "matched" if
Both are positive
Both are negative
Both are between -.01 and .01

Table B.5 OverAchiever Medoid Coordinates

Fall 2008 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*2/4*
-0.0434	0.0278	0.0191	-0.0233	
Spring 2009 Medoid Coordinates				
V1	V2	V3	V4	
-0.004	-0.0158	-0.0376	-0.0104	

Fall 2009 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
0.0434	0.0161	-0.0111	-0.027	
Spring 2010 Medoid Coordinates				
V1	V2	V3	V4	
0.0183	0.0013	0.0408	-0.0118	

Table B.6 Employee Medoid Coordinates

Fall 2008 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*2/4*
-0.0278	-0.0447	-0.0606	-0.0345	
Spring 2009 Medoid Coordinates				
V1	V2	V3	V4	
0.0109	-0.0462	0.002	-0.0364	

Fall 2009 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
0.0204	-0.0185	0.006	0.035	
Spring 2010 Medoid Coordinates				
V1	V2	V3	V4	
0.01	-0.0164	0.0013	-0.0083	

Table B.7 UnderAchiever Medoid Coordinates

Fall 2008 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
-0.028	0.0001	0.0143	0.0461	
Spring 2009 Medoid Coordinates				
V1	V2	V3	V4	
0.0178	0.0276	0.0692	0.0041	

Fall 2009 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
-0.0436	0.0065	-0.1082	0.0144	
Spring 2010 Medoid Coordinates				
V1	V2	V3	V4	
0.0141	-0.0009	-0.0486	0.0016	

Table B.8 Sisyphean Striver Medoid Coordinates

Fall 2008 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
0.0254	0.0691	-0.0345	0.0052	
Spring 2009 Medoid Coordinates				
V1	V2	V3	V4	
-0.0089	0.0405	-0.0134	0.0104	

Fall 2009 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*3/4*
-0.01	0.0643	0.0366	-0.0137	
Spring 2010 Medoid Coordinates				
V1	V2	V3	V4	
-0.0018	0.0314	0.0112	0.0301	

Table B.9 Rote Memorizer Medoid Coordinates

Fall 2008 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*2/4*
0.0168	-0.0506	0.0662	-0.0058	
Spring 2009 Medoid Coordinates				
V1	V2	V3	V4	
-0.0007	-0.0091	0.0124	0.0229	

Fall 2009 Medoid Coordinates				*Matched:*
V1	V2	V3	V4	*2/4*
-0.0945	0.1061	0.0503	-0.0141	
Spring 2010 Medoid Coordinates				
V1	V2	V3	V4	
-0.0645	0.0576	-0.136	0.0397	

Medoid Position

By examining the medoid coordinates in the new vector system and the contributions of the original vectors to each new vector, we can pull out assignments for which members of each group typically performed well or poorly. You may wish to refer to the types of exam problems for each semester, found in the V Vectors section of this appendix.

Key:
Very High High Low Very Low

Table B.10 OverAchiever Significant Assignments

Fall 2008 Assignment Scores											
1.2	1.3	1.5	1.7	1.8	1.9	1.12	1.14	1.15	1.16		
HW 1	HW 3	ST 1	ST 2	AT 1	AT 2	AT 3	AT 5	AT 6	AT 9	AT 10	OHW 4

Spring 2009 Assignment Scores											
1.1	1.2	1.3	1.4	1.5	1.9	1.10	1.11	1.12	1.13	1.16	
HW 1	HW 2	ST 1	IC 1	AT 2	AT 3	IC 3	AT 6	OHW 1	OHW 2	OHW 4	OHW 5

Fall 2009 Assignment Scores									
1.2	1.5	1.6	1.7	1.8	1.10	1.11	1.12	1.13	1.15
HW 1	ST 1	AT 3	AT 4	AT 5	AT 6	IC 2	OHW 5		

Spring 2010 Assignment Scores												
1.1	1.3	1.4	1.5	1.7	1.9	1.10	1.11	1.12	1.15	1.16		
HW 1	HW 2	ST 1	AT 1	AT 2	AT 4	AT 6	IC1	OHW 1	OHW 2	OHW 3	OHW 4	OHW 5

Table B.11 Employee Significant Assignments

Fall 2008 Assignment Scores													
1.1	1.2	1.3	1.5	1.8	1.9	1.10	1.13	1.15	1.16				
HW 1	HW 2	HW 4	ST 1	AT 2	AT 3	AT 4	AT 6	AT 7	AT 8	AT 9	AT 10	OHW 2	OHW 3

Spring 2009 Assignment Scores											
1.1	1.2	1.7	1.8	1.12	1.13	1.14	1.16				
HW 1	HW 2	ST 1	IC 1	AT 3	IC 2	AT 4	AT 6	OHW 1	OHW 2	OHW 4	OHW 5

Fall 2009 Assignment Scores										
1.1	1.2	1.3	1.4	1.5	1.7	1.8	1.9	1.12	1.13	1.14
HW 1	HW 2	ST 1	ST 2	AT 2	AT 4	IC 1	OHW 1	OHW 2	OHW 4	OHW 5

Spring 2010 Assignment Scores							
1.1	1.3	1.4	1.8	1.9	1.12	1.13	1.15
HW 2	ST 1	ST 2	AT 2	IC 1	OHW 2		

Table B.12 UnderAchiever Significant Assignments

Fall 2008 Assignment Scores										
1.1	1.2	1.4	1.5	1.7	1.8	1.10	1.13	1.15		
HW 1	HW 2	HW 3	ST 1	ST 2	AT 2	AT 3	AT 4	AT 5	AT 9	OHW 2

Spring 2009 Assignment Scores											
1.1	1.2	1.3	1.9	1.10	1.11	1.16					
HW 1	HW 2	ST 1	ST 2	IC 1	AT 2	IC 2	IC 3	AT 6	OHW 1	OHW 2	OHW 4

Fall 2009 Assignment Scores							
1.3	1.4	1.5	1.6	1.7	1.10	1.11	1.13
HW 1	AT 1	AT 3	At 3	OHW 1	OHW 2	IC 1	IC 2

Spring 2010 Assignment Scores						
1.2	1.3	1.4	1.5	1.6	1.7	1.13
HW 1	AT 1	AT 3	AT 4	AT 5	OHW 4	OHW 5

Table B.13 Sisyphean Striver Significant Assignments

Fall 2008 Assignment Scores									
1.2	1.4	1.5	1.7	1.8	1.10	1.12	1.14		
HW 1	HW 4	ST 1	AT 2	AT 3	AT 5	AT 6	AT 10	OHW 1	OHW 2

Spring 2009 Assignment Scores							
1.2	1.4	1.5	1.7	1.8	1.12	1.13	
ST 1	AT 3	AT 4	AT 5	AT 6	AT 10	OHW 2	OHW 4

Fall 2009 Assignment Scores									
1.1	1.2	1.3	1.4	1.6	1.7	1.8	1.9	1.11	1.12
HW 1	AT 1	AT 3	AT 6	IC 1	IC 2	OHW 1	OHW 2	OHW 3	OHW 5

Spring 2010 Assignment Scores							
1.2	1.5	1.7	1.9	1.12	1.13	1.14	1.15
HW 1	ST 1	OHW 1	OHW 2	OHW 3	OHW 4		

Table B.14 Rote Memorizer Significant Assignments

Fall 2008 Assignment Scores													
1.3	1.4	1.5	1.7	1.8	1.9	1.12	1.13						
HW 1	HW 3	HW 4	ST 1	AT 2	AT 3	AT 5	AT 6	AT 7	AT 9	AT 10	OHW 1	OHW 2	OHW 4

Spring 2009 Assignment Scores								
1.1	1.2	1.3	1.4	1.5	1.7	1.8	1.11	1.13
HW 1	HW 2	ST 2	AT 1	IC 1	AT 2	AT 6	OHW 1	OHW 2

Fall 2009 Assignment Scores							
1.1	1.2	1.5	1.7	1.9	1.11	1.12	1.13
ST 1	ST 2	AT 1	AT 4	IC 1	OHW 1	OHW 2	OHW 5

Spring 2010 Assignment Scores									
1.1	1.2	1.3	1.4	1.6	1.10	1.11	1.12	1.13	1.15
ST 1	ST 2	AT 1	AT 4	IC 1	OHW1	OHW 4	OHW 5		

Appendix C - Interview Protocols

Fall 2008

1. *Prepare for the interview at least 5 minutes before the scheduled time. Unlock the conference room (Reta has the key) and leave the door open. Set out the IC Recorder, two copies of the Informed Consent form and a pad of paper for students to write or draw on as needed when they answer the questions. Have a calculator and a copy of the student's recent Studio College Algebra Exam available.*
2. *When the student arrives, introduce yourself and welcome the student by name. Close the door to the conference room. Ask for permission to record the interview. If permission is granted, start the recorder.*
3. *Explain the purpose of the interview:*

We are interviewing students in Studio College Algebra to better describe the characteristics of students enrolled in the class. This is prompted by a desire to understand how different students react to certain aspects of the course, how they set about learning the material, and their level of conceptual understanding. The general goal is to use this information to improve teaching and assessment. This interview should take approximately 20-45 minutes. Your participation is completely voluntary and your grade will not be affected by your answers in this interview. You will receive $10 for your time for participating in this interview and you may also benefit by improvements in instruction in mathematics and by having a chance to go over the most recent exam an instructor. In the event we include any of your comments in a discussion or publication about our findings, your privacy will be maintained by the use of a pseudonym. We have two copies of an Informed Consent Form for you to sign, one for our records and one for you to keep.

4. *Have them read and sign the form. If they decline to sign the form, thank them for their time and terminate the interview. Otherwise sign and date the form as witness and then proceed to the questions below.*

5. **Background/Attitude Questions**. *Stay aware of the time and try not to let this section exceed 20 minutes so you have time for the rest of the material. In the (unusual) event that a student wants to spend more than 20 minutes on this, explain politely that you need*

to get to some additional questions and promise them they will have a chance to make more comments at the end.

A. Describe your feelings towards mathematics at the beginning of the semester as you entered into this course.

B. What is your view about mathematics? Learning mathematics?

C. How do you usually study for math assessments? Did you study differently for assessments in this course? Explain.

D. If you get stuck on a problem or have trouble understanding a concept, what do you do? *Ask them to explain why they choose to seek help or not.*

E. How did you prepare for the most recent exam? *If necessary, ask for elaboration: memorizing formulas, going through previous exams, redoing online homework, etc.*

F. How much time outside of class do you normally spend each week on College Algebra-related work? *Ask them to specify which activities make up this time.*

G. How did you study for the online assessments in this course?

H. Did you utilize the written help tutorials in the online homework assignments?

   a. Do you feel the written tutorials were beneficial? Why and/or how were they beneficial?

   b. What changes would you suggest to be made to the written tutorials to make them more beneficial?

   c. In the future, would you be more or less likely to view written tutorials for assistance on assessments or other course work?

I. Did you utilize the video help tutorials in the online homework assignments?

   a. Do you feel the video tutorials were beneficial? Why and/or how were they beneficial?

   b. What changes would you suggest to be made to the video tutorials to make them more beneficial?

   c. In the future, would you be more or less likely to view video tutorials for assistance on assessments or other course work?

J. What did/didn't you like about the extra credit assignment?

   a. What suggestions would you make in order to improve the assignment?

   b. How did you decide which problems to complete?

K. What are your future career goals?

a. Do the online assessments cover information that you feel is important to know for your future? Explain why/why not.

L. What aspect of the algebra class (lecture, recitation, written homework, online homework, studio) have you found most helpful? *Ask them to explain why this has been helpful.*

M. What aspect of class (lecture, recitation, written homework, online homework, studio) have you found least helpful? *Ask them to explain what the problems with this aspect of the class are.*

N. I will now hand over short survey for you to complete about your confidence levels and learning environment preferences. Please read the directions and feel free to ask questions. *Answer any questions the student has about the survey. After they finish, ask if they have any comments they want to make about their answers.*

6. **Concept Questions**. *This section should take 5 – 10 minutes.*

Now I want to ask you a few questions about some basic mathematical concepts.

A. What is a function?

B. What are the different ways you know to represent a function? (*Ask for up to 3 representations, or until the student runs out of ideas*)?

C. Can you explain how [definition 1] and [definition 2] are related? *Ask this question for a particular pair of definitions.*

D. Can you give me a few examples of how functions are useful? *If the student gets stuck, ask them to name a few specific functions and describe their important characteristics*.

7. **Problem Solving**. *This section should take 15 minutes or less*.

I will now hand you a copy of your previous exam. The exam was designed to encourage students to use several different methods of problem solving. So that I can get a better understanding of your thought process, please explain how you approached and worked through each circled problem. Feel free to make notes on the paper, and please talk through your approach out loud for the recorder.

A. *After the student has explained how they attempted each of the three circled problems, remind them that they can ask other questions they have about the exam.*

8. **Other comments.**

   Are there any comments or questions you would like to make about learning algebra? *Ask follow-up questions or provide answers (if you know the answers) as appropriate.*

9. *Thank the student for participating. Let them know they are always welcome to email any additional comments or suggestions for the course to rbm001@math.ksu.edu.*
10. *Stop the recorder.*
11. *Fill out the receipt. Remember to put the wrap around cover behind the receipt. You need to get the student's address and social security number. Since we are paying the student, we are legally obligated to get their social security number. Be sure they sign the receipt. Once the receipt is signed, given them a $10 bill and thank them again. Place one copy of the receipt in the envelope with the money and leave the other receipt in the receipt book.*
12. *Listen to the interview on the recorder and write up your notes. Turn the consent form, your notes, and whatever the students wrote on their pad in to Rachel Manspeaker. Transfer the recording to the computer system and erase the IC Recorder.*

Spring 2009

1. *Prepare for the interview at least 5 minutes before the scheduled time. Unlock the conference room (Reta has the key) and leave the door open. Set out the IC Recorder, two copies of the Informed Consent form and a pad of paper for students to write or draw on as needed when they answer the questions. Have a calculator and a copy of the student's recent Studio College Algebra Exam available.*
2. *When the student arrives, introduce yourself and welcome the student by name. Close the door to the conference room. Ask for permission to record the interview. If permission is granted, start the recorder.*
3. *Explain the purpose of the interview*:

We are interviewing students in Studio College Algebra to better describe the characteristics of students enrolled in the class. This is prompted by a desire to understand how different students react to certain aspects of the course, how they set about learning the material, and their level of conceptual understanding. The general goal is to use this information to improve teaching and assessment. This interview should take approximately 20-45 minutes. Your participation is completely voluntary and your grade will not be affected by your answers in this interview. You will receive $10 for your time for participating in this interview and you may also benefit by improvements in instruction in mathematics and by having a chance to go over the most recent exam an instructor. In the event we include any of your comments in a discussion or publication about our findings, your privacy will be maintained by the use of a pseudonym. We have two copies of an Informed Consent Form for you to sign, one for our records and one for you to keep.

4. *Have them read and sign the form. If they decline to sign the form, thank them for their time and terminate the interview. Otherwise sign and date the form as witness and then proceed to the questions below.*
5. **Background/Attitude Questions**. *Stay aware of the time and try not to let this section exceed 20 minutes so you have time for the rest of the material. In the (unusual) event that a student wants to spend more than 20 minutes on this, explain politely that you need*

to get to some additional questions and promise them they will have a chance to make more comments at the end.

A. What do you think mathematics is all about? Is it important? Why do we spend so much time learning about math?

B. Describe your feelings towards mathematics at the beginning of the semester as you entered into this course. Have they changed at all over the last few months? (*Follow up questions: How?, Why/ Why not?, etc.*)

C. Describe your experience with Studio College Algebra so far this semester.

a. Have you enjoyed the class?

b. Are you doing well?

c. Is it what you expected?

D. How much time outside of class do you normally spend each week on College Algebra-related work? *Ask them to specify which activities make up this time.*

E. If you get stuck on a problem or have trouble understanding a concept, what do you do? *Ask them to explain why they choose to seek help or not.*

F. How do you usually study for math assessments?

Did you study differently for assessments in this course? Explain.

G. How did you prepare for the most recent exam? *If necessary, ask for elaboration: memorizing formulas, going through previous exams, redoing online homework, etc*.

H. How did you study for the online assessments in this course?

I. The homework assignment covering section 2.5 was different from the rest. You got to chose which type of problem you wanted to do- either an Agriculture, Business, Education, or Social Science problem.

a. Which problem did you chose? How did you decide?

b. What suggestions would you make in order to improve the assignment?

J. What are your future career goals?

a. Which assessments cover information that you feel is important to know for your future? Explain why/why not.

K. What aspect of the algebra class (lecture, recitation, written homework, online homework, studio) have you found most helpful? *Ask them to explain why this has been helpful*.

L. What aspect of class (lecture, recitation, written homework, online homework, studio) have you found least helpful? *Ask them to explain what the problems with this aspect of the class are.*

M. What suggestions do you have for improving the course?

6. **Concept Questions**. *This section should take 5 – 10 minutes.*

Now I want to ask you a few questions about some basic mathematical concepts.

Give the students handouts showing different representations of a function.

A. Please describe what you see on these papers.

i. What do you know about these figures?

ii. Do you see any connections between them?

B. These are all different representations of a specific function. What else do you know about functions?

i. Can you give more examples of types of functions?

ii. What are some of the important characteristics of these functions?

C. Here is an example of one way functions are useful. *Hand the student an example of a linear regression model.*

i. Can you describe to me what information this is telling us?

ii. Is this helpful? How? What does this tell us that we didn't know before?

iii. Do you know any other uses for functions?

7. **Problem Solving**. This *section should take 15 minutes or less*.

I will now hand you a copy of your previous exam. The exam was designed to encourage students to use several different methods of problem solving. So that I can get a better understanding of your thought process, please explain how you approached and worked through each circled problem. Feel free to make notes on the paper, and please talk through your approach out loud for the recorder.

A. *After the student has explained how they attempted each of the three circled problems, remind them that they can ask other questions they have about the exam.*

This concludes our interview. (*Thank the student for participating*) Before you leave, I'd like to know if there were any questions I should have asked you, but I missed. *Ask follow-up questions or provide answers (if you know the answers) as appropriate.*

9. *Thank the student, again, for participating. Let them know they are always welcome to email any additional comments or suggestions for the course to rbm001@math.ksu.edu.*
10. *Stop the recorder.*
11. *Fill out the receipt. Remember to put the wrap around cover behind the receipt. You need to get the student's address and social security number. Since we are paying the student, we are legally obligated to get their social security number. Be sure they sign the receipt. Once the receipt is signed, given them a $10 bill and thank them again. Place one copy of the receipt in the envelope with the money and leave the other receipt in the receipt book.*
12. *Listen to the interview on the recorder and write up your notes. Turn the consent form, your notes, and whatever the students wrote on their pad in to Rachel Manspeaker. Transfer the recording to the computer system and erase the IC Recorder.*

Notes on Interview:

Student: _____

Date/Time: _____

Interviewer: _____

Definitions Provided:
Problems Provided:
Problems Attempted:
Comments on the interview:

Conceptual Handouts

Functions

Input x	Output $f(x)$
-3	0
-2	-1.5
-1	0
0	3
1	6
2	7.5
3	6
4	0

Figure C.1 Function: Ordered Pairs

$$f(x) = -\frac{1}{4}(x+3)(x+1)(x-4)$$

$$f(x) = -\frac{1}{4}(x^3 - 13x - 12)$$

Figure C.2 Function: Algebraic Representation

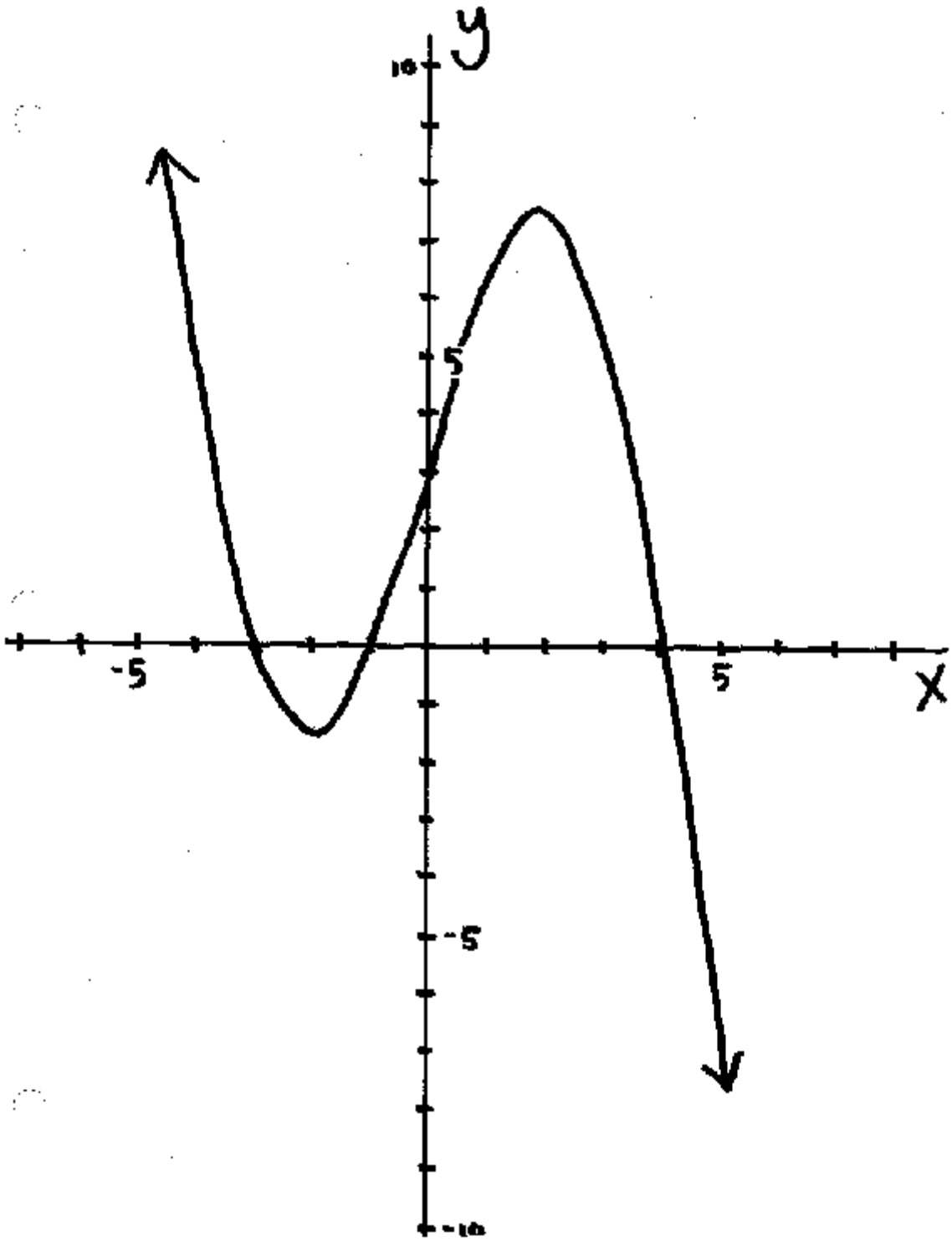

Figure C.3 Graphical Representation

Application: Tornado Sightings

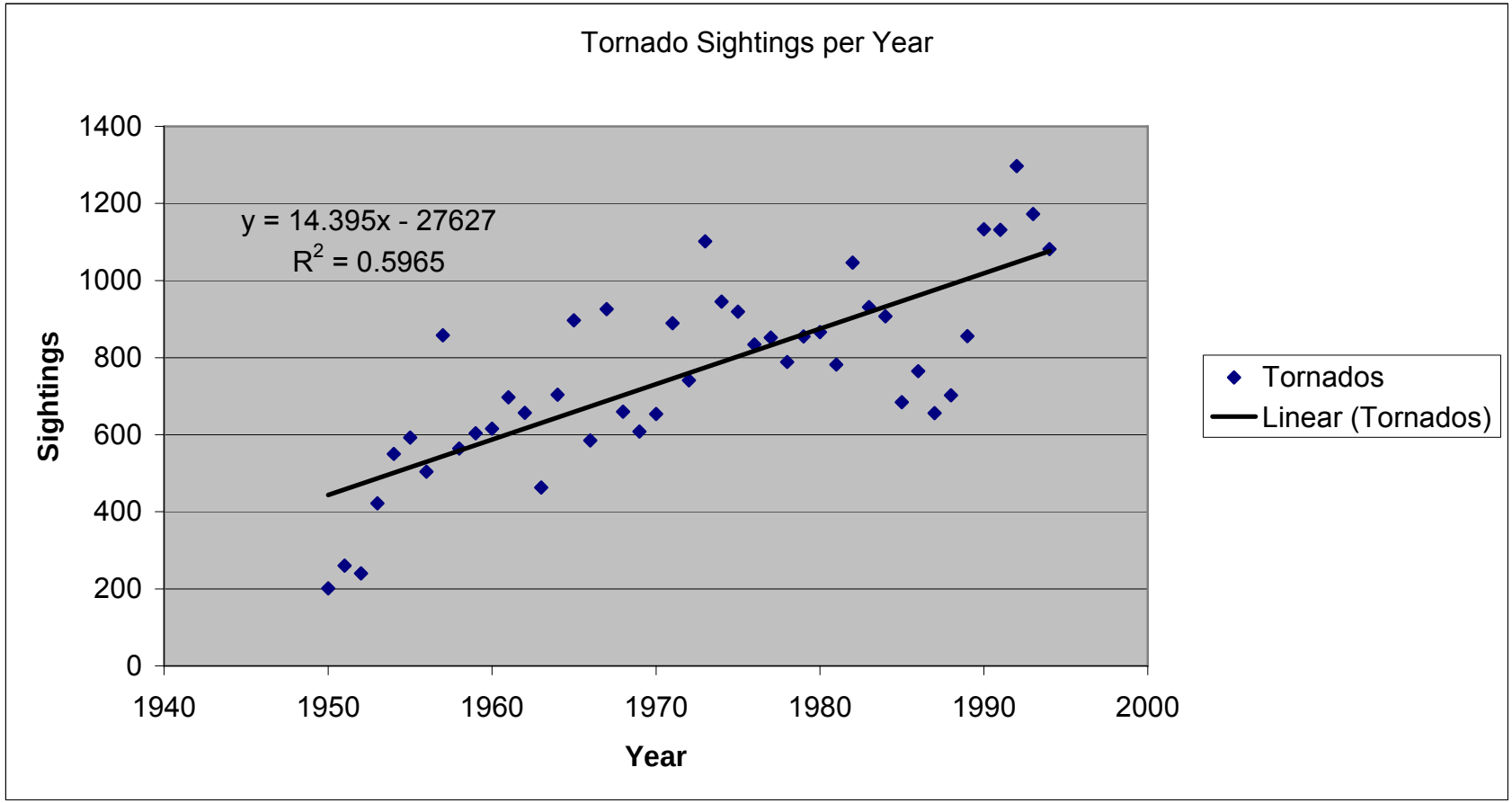

Figure C.4 Application of Functions: Linear Regression Model

Appendix D - Coding Scheme

Table D.1 Coding Scheme

Research Questions	Positive comments: Purple	Neutral Comments-Black	Negative comments: Red
1) How does the student feel about mathematics and their ability to do and understand math? **A (affect)**	PREV: student has had positive previous experiences in math classes CHANGE: because of SCA, the student has improved their opinion of math MATH: student enjoys mathematics in general CONF: the student is confident in their mathematical ability USE.SELF: mathematics is useful in their personal life USE.SOC: the use of mathematics benefits society in general USE.JOB: the student anticipates using mathematics in their future career SCA.USE.JOB: skills learned in SCA will be helpful in a future career SCA.USE.SELF: skills learned in SCA will be useful in their personal life STU.USE.JOB: skills learned in the studio portion of SCA will be helpful in a future career CRIT: learning and doing mathematics enhances one's critical thinking EXPLAIN: math is used to explain natural phenomena SOLVE: math is useful for solving problems ART: math has intristic/artistic merit	PREV: the student is ambivalent about their previous experiences in math classes CHANGE.NONE: SCA has not affected the student's opinion of mathematics WORK: student treats math classes like a low paying job: "show up and do what they say" NUMB: math is about numbers USE.SOME: math is only useful to some people in certain situations	IDK.#: the number of times the student says "I don't know," "I don't remember," or some other expression of not understanding during the entire interview PREV: the student had negative previous experiences in math classes CHANGE: because of SCA, the student's opinion of math has deteriorated MATH: the student dislikes math in general CONF: the student expresses lack of confidence in their mathematical ability FRAC: student expresses dislike of fractions GRAPH: student expresses dislike of graphs INTRIN: talent for mathematics is an intrinsic ability USE.SELF: mathematics (or things learned in SCA) will not be used it the student's life SCA.USE.SELF: student doubts they will ever use skills learned in SCA in their personal life USE.JOB: the student thinks they will not need to use math in their future career SCA.USE.JOB: the student doubts they will ever use skills

Research Questions	Positive comments: Purple	Neutral Comments- Black	Negative comments: Red
2) What is the student's reaction to the course? **R (reaction)**	**underlined: most helpful** SCA: students likes the course in general SCA.CONN: skills learned in SCA are being used in other courses REVIEW: the class is mostly review, to the relief of the student EASY: the class is easier than expected, to the relief of the student SCHED: student enjoys the variety of teaching techniques and classroom styles LEC: lecture is helpful LEC.NOTES: posting lecture notes online is helpful LEC.ICLICK: the student likes using the iclickers LEC.INS: the lecture instructor is helpful OHW: online homework is helpful OHW.CHANGE: the student appreciates how changing problems on the online homework builds understanding OHW.REDO: the student appreciates being able to redo the online homework multiple times OHW.CONVEN: student finds turning homework in online convenient REC: recitation is helpful REC.HW: the student appreciates going over homework problems in recitation REC.SIZE: the student likes the small size of recitation classes REC.INS: the student thinks his or her instructor is helpful in recitation	REVIEW: the class is mostly review, but the student is ambivalent SCA: the student is ambivalent about the course in general INT.ALG: the student has taken intermediate algebra REPEAT: the student has taken CA at a previous time LEC.ICLICK: the student mentions using iclickers, but is ambivalent	**underlined: least helpful** SCA: the student does not like the course in general REVIEW: the class is mostly review, which bores the student HARD: the course is too hard STRUG: student expresses general frustration or struggling EASY: the course is too easy SCHED: the SCA schedule is confusing HW.T: the number of different times to turn in assignments is confusing DEADLINE: the student admits to having problems meeting deadlines HW.LESS: there is too much homework HW.MORE: there should be more homework EXAM: the exams are too difficult EXAM.JUST: the student does not like needing to justify their answers on exams LEC: lecture is not helpful LEC.SIZE: the lecture is too large (and intimidating) LEC.NOTES: class notes available online discourages class attendance LEC.ICLICK: the student does not like using iclickers in lecture LEC.INS: the lecture instructor is not helpful REC: recitation is not helpful REC.INS: the recitation instructor is not helpful REC.HW: only covering homework problems is boring and not helpful BOOK: the textbook is not helpful STU: studio is not helpful STU.CONN: the student does not see any connection between studio and the rest of the course

2) What is the student's reaction to the course? R (reaction)	STU.APP: the student appreciates applying math techniques to real life situations in studio STU: studio is helpful STU.CONN: skills learned in studio are being used in other courses STU.PARTNER: the student likes being paired with a partner in studio STU.EXCRED: student likes having the oppurtunity to earn extra credit through studio		STU.PARTNER: the student does not like being paired with a partner in studio STU.EXCRED: too much extra credit is available through studio assignments OHW: online homework is not helpful OHW.CHANGE: the student does not like the changing problems in the online homework OHW.COMP: the student has formatting or technical problems with the online homework WEBSITE: the website is not well organized or helpful WHW.WP: the student does not like word problems

Research Questions	Positive comments: Purple	Neutral Comments- Black	Negative comments: Red
3) What does the student do in order to succeed? **E (effort)**	HW.BEFORE: student attempts their homework assignment before recitation OHW.VIDEO: student uses video help to do online homework OHW.HINTS: student uses online hints to help with online homework OHW.100%: the student always aims for a score of 100% on their online homework STDY.OLDEX: student studies uses old exams found online	LEC.NOTES: the student takes notes during lecture TUT: student meets regularly with a tutor HW.#: the number of hours students spends doing their homework HW.NOTES: student uses class notes to complete homework HW.FRIEND: student does their homework with a friend HW.TUT: student does their homework with a tutor OHW.#: the number of hours student spends doing their online homework OHW.NOTES: student uses class notes to do online homework OHW.TUT: the student goes through their online homework with a tutor OHW.WRITE: student writes online homework solutions on paper first OHW.VID: student uses videos from the internet to help with online homework OHW.BOOK: student uses the textbook to help with online homework OHW.FRIEND: student does their online homework with a friend WHW.#: the number of hours student spends doing their written homework HELP.FRIEND: student gets help from friends outside of class HELP.INS: student gets help from class instructors HELP.VID: student gets help from reviewing lecture videos HELP.BOOK: student looks up information in textbook when confused STDY.#: the number of hours a student studies before each exam STDY.NOTES: student studies for exams with the help of class notes STDY.BOOK: student studies for exams with the help of the textbook STDY.HW: student studies for exams with the help of old HW problems STDY.FRIEND: student studies for exams with friends STDY.TUT: student studies for exams with a tutor	OHW.NONE: the student admits to not completing online homework on a regular basis HELP.NONE: the student admits to not seeking help when they have a problem HW.NONE: the student admits to not completing homework assignments on a regular basis HW.REC: the student does not attempt any homework problems until after recitation STDY.NONE: the student admits to not studying for exams REDO: student was not aware they had the option to redo online homework LEC.VIDEO: student was not aware that videos of the lectures were available online STU.NONE: student does not regularly attend studio

Research Questions	Positive comments: Purple	Neutral Comments- Black	Negative comments: Red
4) What is the student's conceptual understanding of "function"? **C (concept)**	VOCAB.# : the student correctly uses 10 or more mathematical terms CHART.DRAW: student sketches a rough graph from the points on the chart CHART.ZEROS: student identifies the zeros on the chart CHART.Y-INT: student identifies the y-intercept on the chart EQ.DEG: student identifies the degree of the polynomial from the equation EQ.ZEROS: student identifies the zeros on the factored equation EQ.Y-INT: student identifies the function's y intercept on the distributed equation EQ.LC: student identifies the function's leading coefficient from the equation EQ.DIS: student attempts to distribute the factors of the polynomial to justify the two forms being equal EQ.DRAW: student sketches a rough graph using the equations GRAPH.Y-INT: the student identifies the y intercept on the graph GRAPH.ZEROS: the student identifies the function's zeros on the graph GRAPH.TP: student identifies the turning points of the graph GRAPH.LC: student identifies the sign of the leading coefficient from the graph " " --> " " : the student makes and verifies connections between two different representations FUNC.ALL3: the student made connections between representations without prompting FUNC.EQ: the student made connections between the two equations without prompting FUNC.#: student was able to name 3 or more other types of functions	VOCAB.#: the student correctly uses 5 to 10 or more mathematical terms CHART.PATTERN: student looks for a pattern in the outputs. EQ.DIS: attempted to distribute without justification FUNC.#: student was able to name 1 to 2 other types of functions FUNC.USE.#: student gives 1 other use for a function " " --> " " : student makes connections between two different representations but does not support FUNC.ALL3: the student made connections between representations after prompting FUNC.EQ: the student made connections between the two equations after prompting TORN.INC: student claims the number of tornado sightings increase every year TORN.FIT.AVE: student describes the line of best fit as the "average" of the points TORN.FIT.TREND: student identifies the line of best fit as showing a general trend, but expresses this idea vaguely/poorly	IDK.#: the number of times the student says "I don't know," "I don't remember," "I have no idea," or some other expression of not understanding during the content portion of the interview VOCAB.#: the student correctly uses less than 5 mathematical terms VOCAB.INCORR.#: the number of mathematical terms the student uses incorrectly EQ.DIS: student makes a mistake distributing --> : the student makes an incorrect connection between different representations FUNC.ALL3: even after prompting, the student does not make a connection between the different representations FUNC.#: student was not able to name any other type of functions FUNC.USE.#: student is unable to give other uses for functions TORN.FIT.PRESENT: the student incorrectly noted the line of best fit as being able to predict current values

	FUNC.DESC: the student was able give descriptions of the other functions FUNC.USE.#: student gives 2 or more other uses for functions TORN.INC: the student remarks that in general, the number of tornado sightings have increased TORN.FIT.TREND: the student accurately described the line of best fit as showing a general trend TORN.JUST: student describes possible reasons behind the increasing trend of tornado sightings TORN.FIT.FUTURE: the student accurately described the line of best fit as being able to predict future values R.SQ: the student correctly identified the meaning of the R-squared value		R.SQ: the student did not know the meaning of the R squared term

Appendix E - Grouping Chart

SVD Groups

Table E.1 OverAchiever Interview Comments

Research Questions	Group OA: 4 interviewees		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.USE.SELF	2A.CHANGE. NONE	A.CHANGE
	4A.SOLVE	2A.NUMB	A.CONF
	4A.USE.SOC	A.USE.SOME	A.MATH
	A.ART		A.PREV
	A.CONF		A.SCA.USE. JOB
	A.EXPLAIN		A.SCA.USE. SELF
	A.MATH		A.STRUG
	A.PREV		A.USE.JOB
	A.SCA.USE. JOB		
	A.SCA.USE. SELF		
	A.USE.JOB		
≥ 50% repeat	0.272727273	0.666666667	0
2) What is the student's reaction to the course?	2R.LEC		2R.HW.T
	2R.LEC. NOTES		2R.OHW. COMP
	2R.REVIEW		2R.STU
	3R.REC		**2R.STU**
	3R.REC.HW		3R.STRUG
	3R.SCA		4R.STU.CONN
	3R.STU		R.EXAM
	4R.OHW		R.HW.LESS
	4R.REC		**R.LEC**
	4R.REC.INS		R.LEC.SIZE
	R.EASY		**R.OHW**
	R.LEC		R.OHW. CHANGE
	R.LEC.ICLICK		R.SCA
	R.OHW. CONVIEN		
	R.OHW.REDO		
	R.REC.SIZE		
	R.STU.APP		
	R.STU.EXCEL		
	R.STU. EXCRED		
	R.WHW		
	R.WHW		
≥ 50% repeat	0.476190476		0.461538462

Research Questions	Group OA: 4 interviewees		
3) What does the student do in order to succeed?	2E.OHW. HINTS	2.E.HELP. FRIEND	E.OHW.REDO
	3E.HW. BEFORE	2E.HELP. BOOK	E.STDY.NONE
	3E.STDY. OLDEX	2E.HELP.INS	
	E.OHW.100%	2E.HW.1-2	
		2E.HW.3	
		3E.OHW. WRITE	
		E.HELP. NOTES	
		E.HW.NOTES	
		E.OHW.BOOK	
		E.OHW. NOTES	
		E.STDY.2	
		E.STDY.HW	
		E.STDY. NOTES	
≥ 50% repeat	0.75	0.461538462	0
4) What is the student's conceptual understanding of "function"?	2C.CHART-> EQ	2C.CHART. PATTERN	2C.EQ.DIS
	2C.CHART-> GRAPH	2C.FUNC.ALL3	2C.R-SQ
	2C.FUNC. ALL3	2C.FUNC. USE.1	C.TORN.FIT. PRESENT
	2C.FUNC.EQ	2C.TORN.INC	C.VOCAB. INCORR.2
	3C.TORN.FIT. TREND	3C.FUNC.2	
	4C.EQ.DIS	3C.VOCAB.6-8	
	4C.FUNC. DESC	4C.TORN. FIT.AVE	
	C.CHART-> EQ->GRAPH	C.FUNC.3	
	C.EQ.ZEROS		
	C.EQ->GRAPH		
	C.FUNC.USE.3		
	C.GRAPH.TP		
	C.GRAPH. ZEROS		
	C.TORN. FIT.FUT		
	C.TORN.INC		
	C.VOCAB.10		
50% repeat	0.411764706	0.875	0.5

Table E.2 Employee Interview Comments

Research Questions	Group E: 4 interviewees		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.CHANGE	2A.CHANGE. NONE	2A.MATH
	2A.CRIT	3A.USE. SOME	3A.CONF
	2A.SCA.USE. JOB	A.NUMB	4A.WORK
	3A.SOLVE		A.PREV
	4A.EXPLAIN		A.SCA.USE. JOB
	A.ART		A.USE.SELF
	A.CONF		
	A.SCA.USE. SELF		
	A.STU.USE. JOB		
≥ 50% repeat	0.555555556	0.666666667	0.5
2) What is the student's reaction to the course?	2R.LEC	R.REPEAT	**2R.OHW**
	2R.REC		R.EASY
	2R.REC.HW		R.HARD
	2R.REC.INS		R.HW.MORE
	2R.SCA		R.OHW
	3R.STU		R.OHW. COMP
	4R.STU. EXCEL		R.REC.HW
	R.LEC		R.STRUG
	R.LEC.ICLICK		**R.STU**
	R.LEC.NOTES		R.STU. EXCRED
	R.REVIEW		
	R.SCHED		
	R.STU		
	R.STU.APP		
≥ 50% repeat	0.5	0	0.1
3) What does the student do in order to succeed?	2E.HW. BEFORE	2E.OHW. NOTES	E.STDY.NONE
	2E.OHW. HINTS	2E.OHW. WRITE	
	3E.STDY. OLDEX	2E.STDY. NOTES	
	E.OHW.100%	3E.HELP. FRIEND	
		3E.HELP.INS	
		3E.HW.1-2	
		E.HELP.TUT	
		E.HW.5	
		E.LEC.NOTES	
		E.OHW.1	
		E.STDY.TUT	
		E.TUT	
		E.WHW.1	
≥ 50% repeat	0.75	0.461538462	0

Research Questions	Group E: 4 interviewees		
4) What is the student's conceptual understanding of "function"?	2C.EQ.ZEROS	2C.CHART. PATTERN	2C.FUNC. USE.0
	2C.GRAPH.TP	2C.TORN. FIT.TREND	C.FUNC.0
	2C.TORN.FIT. TREND	2C.TORN.INC	C.VOCAB.2
	3C.FUNC.3-4	3C.EQ.DIS	
	C.EQ.DEG	4C.FUNC.ALL3	
	C.EQ.DIS	C.CHART. DRAW	
	C.EQ--> GRAPH	C.FUNC.EQ	
	C.FUNC.DESC	C.TORN.FIT. AVE	
	C.FUNC.EQ		
	C.FUNC.USE.2		
	C.GRAPH. DEG		
	C.GRAPH.LC		
	C.GRAPH. Y-INT		
	C.GRAPH. ZEROS		
	C.R-SQ		
	C.TORN.FIT. FUTURE		
	C.TORN.JUST		
	C.VOCAB.15		
50% repeat	0.222222222	0.625	0.333333333

Table E.3 UnderAchiever Interview Comments

1) How does the student feel about mathematics and their ability to do and understand math?	2A.CONF	2A.NUMB	2A.CONF
	2A.STU.USE. JOB	4A.USE.SOME	2A.SCA.USE. JOB
	4A.SOLVE	A.CHANGE. NONE	2A.USE.SELF
	A.CRIT		2A.WORK
	A.EXPLAIN		3A.MATH
	A.SCA		A.CHANGE
	A.SCA.USE. JOB		A.PREV
	A.STU.APP		A.REVIEW
	A.STU.USE. JOB		
≥ 50% repeat	0.1	0.333333333	0.125
2) What is the student's reaction to the course?	2R.EASY	2R.INT.ALG	2R.EXAM
	2R.LEC	R.SCA	2R.WHW.WP
	2R.REC.INS		3R.HW.LESS
	2R.STU		**3R.LEC**
	3R.REC		4R.STRUG
	3R.REVIEW		R.EXAM.JUST
	3R.STU		R.HW.MORE
	3R.STU.EXCEL		R.LEC.INS
	R.LEC.ICLICK		**R.OHW**
	R.LEC.INS		R.OHW.COMP
	R.LEC.NOTES		**R.REC**
	R.OHW		R.REC.HW
	R.OHW.REDO		R.SCHED
	R.REC.HW		**R.STU**
	R.REC.SIZE		R.STU.CONN
	R.SCA		**R.WHW**
	R.STU.CONN		
≥ 50% repeat	0.235294118	0	0.1875
3) What does the student do in order to succeed?	3E.STDY. OLDEX	2E.HELP. BOOK	E.STDY.NONE
	4E.OHW. HINTS	2E.HW.1	
	E.HW. BEFORE	2E.HW.2	
		2E.HW.3-5	
		2E.OHW. WRITE	
		2E.STDY.HW	
		2E.STDY. NOTES	
		5E.HELP. FRIEND	
		E.HELP. NOTES	
		E.HW.FRIEND	
		E.OHW.BOOK	
		E.OHW. NOTES	
≥ 50% repeat	0.666666667	0.083333333	0
Research Questions	**Group UA: 5 interviewees**		

4) What is the student's conceptual understanding of "function"?	2C.CHART. DRAW	2C.TORN.FIT. AVE	2C.FUNC.0
	2C.CHART. ZEROS	2C.TORN.INC	2C.FUNC. USE.0
	2C.CHART-> GRAPH	3C.FUNC. USE.1	2C.TORN.FIT. PRESENT
	2C.EQ.DEG	4C.FUNC.ALL3	
	2C.EQ.DIS	C.CHART. DRAW	3C.VOCAB.1-3
	2C.EQ.TP	C.CHART. PATTERN	C.EQ.DIS
	2C.EQ-> GRAPH	C.CHART-> GRAPH	C.FUNC.ALL3
	2C.FUNC.3-5	C.EQ.DIS	C.R-SQ
	2C.GRAPH. DEG	C.TORN.FIT. TREND	
	2C.GRAPH. Y-INT		
	2C.GRAPH. ZEROS		
	2C.R-SQ		
	3C.FUNC.EQ		
	3C.TORN.FIT. FUTURE		
	C.CHART. Y-INT		
	C.CHART-> EQ->GRAPH		
	C.EQ.DRAW		
	C.EQ.LC		
	C.EQ.ZEROS		
	C.GRAPH.LC		
	C.GRAPH.TP		
	C.TORN.FIT. TREND		
	C.TORN.INC		
	C.VOCAB.11		
≥50% repeat	0.083333333	0.222222222	0.142857143

Table E.4 Sisyphean Striver Interview Comments

Research Questions	Group SS: 3 interviewees		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.EXPLAIN	3A.CHANGE. NONE	A.STRUG
	2A.SCA.USE. JOB	A.NUMB	A.CONF
	2A.USE.SOC	A.USE.SOME	A.SCA.USE. JOB
	3A.SOLVE		
	3A.MATH		
	A.ART		
	A.CONF		
	A.USE.JOB		
	A.USE.SELF		
≥ 50% repeat	0.555555556	0.333333333	0

Research Questions	Group SS: 3 interviewees		
2) What is the student's reaction to the course?	2R.REC	R.REPEAT	2R.DEADLINE
	2R.STU. PARTNER		2R.HW.T
	3R.EASY		**2R.LEC**
	3R.REVIEW		2R.LEC. NOTES
	R.LEC		2R.STU
	R.LEC.ICLICK		2R.WEBSITE
	R.OHW		3R.STRUG
	R.OHW		3R.STU.CONN
	R.OHW. CHANGE		R.BOOK
	R.OHW.HINTS		R.EXAM.JUST
	R.OHW.REDO		R.LEC.ICLICK
	R.REC.INS		R.LEC.SIZE
	R.SCA.CONN		R.OHW.COMP
	R.SCHED		R.REC.HW
	R.STU		R.SCHED
	R.STU. EXCRED		**R.STU**
	R.WHW		R.STU. PARTNER
≥ 50% repeat	0.235294118	0	0.470588235
3) What does the student do in order to succeed?	3E.STDY. OLDEX	2E.HELP. FRIEND	E.LEC.VIDEO
	E.OHW.100%	2E.HW.3-4	
		2E.STDY. NOTES	
		E.HELP.VID	
		E.HW.NOTES	
		E.LEC.NOTES	
		E.OHW.BOOK	
		E.OHW. NOTES	
		E.OHW.VID	
		E.OHW.WRITE	
		E.STDY.3	
≥ 50% repeat	0.5	0.272727273	0
4) What is the student's conceptual understanding of "function"?	2C.EQ.DIS	2C.FUNC.ALL3	C.R-SQ
	2C.GRAPH. Y-INT	2C.FUNC. USE.1	C.VOCAB. INCORR.2
	2C.GRAPH. ZEROS	2C.TORN.FIT. AVE	C.FUNC.0
	2C.TORN.FIT. FUTURE	2C.VOCAB.5-6	
	2C.TORN.FIT. TREND	C.CHART. DRAW	
	2C.TORN.INC	C.CHART. PATTERN	
	2C.TORN.JUST	C.EQ.DIS	
	3C.FUNC.EQ	C.FUNC.2	
	C.CHART. ZEROS	C.FUNC.EQ	
	C.CHART-> GRAPH	C.TORN.INC	
	C.EQ.DEG		
	C.EQ-GRAPH		
	C.FUNC.5		
	C.FUNC.ALL3		
	C.GRAPH. DEG		
	C.GRAPH.TP		
	C.R-SQ		
	C.VOCAB.11		
50% repeat	0.444444444	0.4	0

Table E.5 Rote Memorizer Interview Comments

Questions	Group RM: 3 interviewees		
1) How does the student feel about mathematics and their ability to do and understand math?	A.PREV	3A.USE.SOME	2A.CHANGE
	A.USE.SELF	A.CHANGE.NONE	2A.SCA.USE.JOB
	A.USE.SOC		2A.SCA.USE.SELF
			3A.GRAPHS
			3A.INTRIN
			3A.MATH
			4A.CONF
			A.FRAC
≥ 50% repeat	0	0.5	0.875
2) What is the student's reaction to the course?	2R.EASY	2R.INT.ALG	2R.LEC.SIZE
	2R.REC		2R.OHW.CHANGE
	2R.STU		3R.STRUG
	2R.STU.PARTNER		R.HW.T
	3R.REC.HW		**R.LEC**
	R.LEC		R.OHW
	R.LEC.INS		R.OHW.COMP
	R.OHW		**R.REC**
	R.OHW.CONVEN		R.REC.INS
	R.REC.SIZE		R.STU
	R.STU		**R.STU**
	R.STU.APP		R.STU.CONN
≥ 50% repeat	0.416666667	1	0.230769231
3) What does the student do in order to succeed?	2E.OHW.HINTS	2E.HELP.TUT	E.HW.REC
	4E.STDY.OLDEX	2E.OHW.FRIEND	
		2E.STDY.TUT	
		3E.HW.1-2	
		3E.TUT	
		E.HELP.FRIEND	
		E.HW.>5	
		E.HW.TUT	
		E.OHW.BOOK	
		E.OHW.WRITE	
		E.STDY.1	
		E.STDY.FRIEND	
		E.STDY.HW	
≥ 50% repeat	1	0.384615385	0
4) What is the student's conceptual understanding of "function"?	2C.FUNC.DESC	2C.EQ.DIS	2C.R-SQ
	2C.TORN.INC	2C.FUNC.ALL3	2C.VOCAB.INCORR.2-4
	C.CHART->GRAPH	2C.FUNC.USE.1	C.FUNC.ALL3
	C.TORN.FIT.FUT	2C.TORN.INC	C.VOCAB.1
		2C.VOCAB.7-8	
		C.CHART.DRAW	
		C.EQ->GRAPH	
		C.FUNC.2	
		C.FUNC.EQ	
		C.GRAPH->EQ	
		C.TORN.FIT.TREND	
≥50% repeat	0.5	0.45454	0.5

White Box Groups

Table E.6 White Box Group 1 Interview Comments

Research Questions	Group 1 - IIIII		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.ART	2A.NUMB	2A.SCA.USE.JOB
	2A.CRIT	3A.NO.CHANGE	A.CONF
	2A.EXPLAIN	3A.USE.SOME	A.MATH
	2A.MATH		A.PREV
	2A.SCA.USE.JOB		A.STRUG
	2A.USE.SOC		A.USE.SELF
	3A.CONF		A.WORK
	3A.SOLVE		
	3A.STU.USE.JOB		
	A.CHANGE		
	A.USE.JOB		
	A.USE.SELF		
	A.USE.SOLVE		
≥ 50% repeat	0.230769231	0.666666667	0
2) What is the student's reaction to the course?	2R.LEC	R.INT.ALG	2R.STU
	2R.LEC.ICLICK		**2R.STU**
	2R.REC		2R.STU.CONN
	2R.SCHED		3R.REC.HW
	2R.STU		3R.STRUG
	2R.STU.EXCEL		R.BOOK
	3R.EASY		R.DEADLINE
	3R.REC.INS		R.EXAM.JUST
	4R.REVIEW		R.HW.MORE
	R.LEC		R.HW.T
	R.LEC.NOTES		**R.LEC**
	R.OHW		R.LEC.NOTES
	R.OHW		R.LEC.SIZE
	R.OHW.CHANGE		**R.OHW**
	R.OHW.HINTS		R.OHW.COMP
	R.OHW.REDO		**R.REC**
	R.REC		R.STU.CONN
	R.SCA		R.STU.PARTNER
	R.STU		R.WHW.WP
	R.STU.EXCRED		
	R.WHW		
≥ 50% repeat	0.142857143	0	0.105263158

Research Questions	Group 1 - IIIII		
3) What does the student do in order to succeed?	2E.OHW.HINTS	2E.HW.1	E.STDY.NONE
	2E.STDY.OLDEX	2E.OHW.1	
	E.HW.BEFORE	2E.OHW.NOTES	
	E.OHW.100%	2E.OHW.WRITE	
		2E.WHW.1-2	
		4E.STDY.NOTES	
		5E.HELP.FRIEND	
		E.HELP.INS	
		E.HELP.NOTES	
		E.HW.5	
		E.HW.FRIEND	
		E.HW.NOTES	
		E.LEC.NOTES	
		E.STDY.3	
		E.STDY.HW	
≥ 50% repeat	0	0.133333333	0
4) What is the student's conceptual understanding of "function"?	2C.CHART.ZEROS	2C.CHART.DRAW	C.FUNC.USE.0
	2C.EQ.ZEROS	2C.CHART.PATTERN	C.R-SQ
	2C.GRAPH.DEG	2C.FUNC.USE.1	C.VOCAB.3
	2C.GRAPH.TP	2C.TORN.FIT.AVE	
	2C.GRAPH.ZEROS	2C.TORN.INC	
	2C.VOCAB.11	5C.FUNC.ALL3	
	3C.CHART->GRAPH	C.EQ.DIS	
	3C.EQ.DEG	C.FUNC.2	
	3C.EQ.DIS	C.FUNC.EQ	
	3C.EQ->GRAPH	C.TORN.FIT.TREND	
	3C.FUNC.3-5	C.VOCAB.5	
	3C.FUNC.EQ		
	3C.R-SQ		
	3C.TORN.JUST		
	4C.GRAPH.Y-INT		
	4C.TORN.FIT.FUTURE		
	4C.TORN.FIT.TREND		
	C.CHART.DRAW		
	C.EQ.LC		
	C.EQ.TP		
	C.FUNC.EQ		
	C.GRAPH.LC		
	C.TORN.INC		
≥ 50% repeat	0.47826087	0.090909091	0

Table E.7 White Box Group 2 Inteview Comments

Research Questions	Group 2- IIIIII		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.SCA.USE.JOB	3A.NUMB	2A.CONF
	2A.SCA.USE.SELF	3A.USE.SOME	2A.WORK
	2A.USE.SELF	4A.CHANGE.NONE	A.CHANGE
	3A.EXPLAIN		A.MATH
	4A.SOLVE		A.PREV
	4A.USE.SOC		A.SCA.USE.JOB
	A.ART		A.SCA.USE.SELF
	A.CONF		A.STRUG
	A.CRIT		A.USE.JOB
	A.MATH		
	A.PREV		
	A.USE.JOB		
≥ 50% repeat	0.25	1	0
2) What is the student's reaction to the course?	2R.LEC		2R.STU
	2R.LEC		3R.STRUG
	2R.STU.APP		**3R.STU**
	3R.OHW		3R.STU.CONN
	3R.REC		R.EASY
	3R.REVIEW		R.EXAM
	3R.STU		R.HW.LESS
	3R.STU.EXCEL		R.HW.T
	4R.REC		**R.LEC**
	4R.REC.HW		R.LEC.SIZE
	4R.REC.INS		**R.OHW**
	4R.SCA		R.OHW.CHANGE
	R.EASY		R.OHW.COMP
	R.LEC.ICLICK		R.SCA
	R.LEC.NOTES		R.STU.EXCRED
	R.OHW.CHANGE		
	R.OHW.CONVEN		
	R.OHW.REDO		
	R.REC.SIZE		
	R.STU.EXCRED		
	R.WHW		
	R.WHW		
≥ 50% repeat	0.409090909		0.2

Research Questions	Group 2- IIIIII		
3) What does the student do in order to succeed?	2E.OHW.100%	2E.HELP.BOOK	E.OHW.REDO
	2E.OHW.HINTS	2E.HW.3	E.STDY.NONE
	5E.HW.BEFORE	3E.HELP.FRIEND	
	5E.STDY.OLDEX	3E.HELP.INS	
		3E.OHW.NOTES	
		3E.OHW.WRITE	
		3E.STDY.NOTES	
		4E.HW.1-2	
		E.HELP.NOTES	
		E.HW.NOTES	
		E.LEC.NOTES	
		E.OHW.BOOK	
		E.STDY.2	
		E.STDY.HW	
≥ 50% repeat	0.5	0.428571429	0
4) What is the student's conceptual understanding of "function"?	2C.CHART->EQ	2C.CHART.PATTERN	2C.EQ.DIS
	2C.CHART->GRAPH	2C.EQ.DIS	2C.R-SQ
	2C.EQ.ZEROS	2C.FUNC.USE.1	C.FUNC.0
	2C.FUNC.ALL3	2C.TORN.FIT.AVE	C.FUNC.USE.0
	2C.FUNC.EQ	3C.FUNC.2	C.TORN.FIT.PRESENT
	2C.FUNC.USE.2-3	3C.TORN.INC	C.VOCAB.2
	2C.GRAPH.TP	3C.VOCAB.6-8	C.VOCAB.INCORR.2
	2C.GRAPH.ZEROS	5C.FUNC.ALL3	
	2C.VOCAB.10-15	C.FUNC.EQ	
	3C.EQ.DIS	C.TORN.FIT.TREND	
	3C.TORN.FIT.TREND		
	4C.FUNC.DESC		
	C.CHART->EQ->GRAPH		
	C.EQ->GRAPH		
	C.FUNC.4		
	C.GRAPH.DEG		
	C.TORN.FIT.TREND		
	C.TORN.INC		
	C.TORN.JUST		
≥ 50% repeat	0.157894737	0.4	0

Table E.8 White Box Group 3 Inteview Comments

Research Questions	Group 3-II		
1) How does the student feel about mathematics and their ability to do and understand math?	A.EXPLAIN	A.CHANGE.NONE	2A.CONF
	A.SCA.USE.JOB	A.USE.SOME	A.CHANGE
			A.MATH
			A.SCA.USE.JOB
≥ 50% repeat	0	0	0.25
2) What is the student's reaction to the course?	2R.OHW	R.SCA	**2R.LEC**
	2R.REC		R.HW.LESS
	2R.REC.HW		R.LEC.INS
	R.REC.SIZE		R.LEC.SIZE
	R.STU		R.OHW.CHANGE
			R.WHW.WP
≥ 50% repeat	0.6	0	0.166666667
3) What does the student do in order to succeed?	E.OHW.HINTS	2E.OHW.WRITE	E.STDY.NONE
	E.STDY.OLDEX	E.HELP.FRIEND	
		E.HELP.TUT	
		E.HW.>5	
		E.HW.1	
		E.STDY.TUT	
		E.TUT	
≥ 50% repeat	0	0.142857143	0
4) What is the student's conceptual understanding of "function"?	2C.TORN.FIT.FUTURE	2C.FUNC.USE.1	C.FUNC.0
	C.FUNC.DESC	C.EQ.DIS	C.FUNC.ALL3
	C.TORN.INC	C.FUNC.2	C.R-SQ
		C.FUNC.ALL3	C.TORN.FIT.PRESENT
		C.FUNC.EQ	C.VOCAB.1
		C.GRAPH->EQ	
		C.TORN.FIT.AVE	
		C.TORN.INC	
		C.VOCAB.7	
≥ 50% repeat	0.333333333	0.111111111	0

Table E.9 White Box Group 4 Interview Comments

Research Questions	Group 4- IIII		
1) How does the student feel about mathematics and their ability to do and understand math?	2A.EXPLAIN	2A.NUMB	2A.CONF
	2A.STU.USE.JOB	A.CHANGE.NONE	2A.SCA.USE.JOB
	4A.SOLVE	A.USE.SOLVE	A.MATH
	A.CHANGE	A.USE.SOME	A.REVIEW
	A.CONF		A.USE.SELF
	A.SCA		A.WORK
	A.SCA.USE.SELF		
	A.STU.APP		
≥ 50% repeat	0.375	0.25	0.333333333
2) What is the student's reaction to the course?	2R.REC.INS	2R.REPEAT	**2R.LEC**
	2R.REC	R.INT.ALG	**2R.OHW**
	2R.STU		2R.OHW.COMP
	2R.STU.CONN		2R.SCHED
	3R.STU		4R.STRUG
	R.EASY		R.EXAM
	R.LEC.ICLICK		R.EXAM.JUST
	R.LEC.NOTES		R.HARD
	R.OHW.REDO		R.HW.LESS
	R.REC.HW		R.HW.MORE
	R.REC.INS		R.HW.T
	R.REC.SIZE		R.LEC.ICLICK
	R.SCA		R.OHW
	R.SCA.CONN		R.WEBSITE
	R.STU.EXCEL		**R.WHW**
	R.STU.PARTNER		R.WHW.WP
≥ 50% repeat	0.3125	0.5	0.3125
3) What does the student do in order to succeed?	3E.OHW.HINTS	2E.HELP.BOOK	R.LEC.VIDEO
	4E.STDY.OLDEX	2E.HW.3-4	
		2E.OHW.BOOK	
		2E.OHW.WRITE	
		E.HELP.FRIEND	
		E.HELP.INS	
		E.HELP.TUT	
		E.HELP.VID	
		E.HW.1	
		E.HW.5	
		E.OHW.1	
		E.OHW.VID	
		E.STDY.HW	
		E.STDY.TUT	
		E.TUT	
		E.WHW.1	
≥ 50% repeat	1	0.25	0

Research Questions	Group 4- IIII		
4) What is the student's conceptual understanding of "function"?	2C.EQ.DIS	2C.CHART.PATTERN	2C.FUNC.0
	2C.FUNC.3-5	2C.EQ.DIS	2C.FUNC.USE.0
	2C.FUNC.EQ	2C.FUNC.USE.1	C.EQ.DIS
	C.CHART.DRAW	2C.TORN.FIT.AVE	C.TORN.FIT.PRESENT
	C.CHART.Y-INT	3C.FUNC.ALL3	C.VOCAB.2
	C.CHART.ZEROS	C.CHART.DRAW	C.VOCAB.INCORR.2
	C.CHART->EQ->GRAPH	C.CHART->GRAPH	
	C.EQ.DEG	C.TORN.FIT.TREND	
	C.EQ.DRAW	C.TORN.INC	
	C.EQ.TP	C.VOCAB.6	
	C.EQ->GRAPH		
	C.FUNC.ALL3		
	C.GRAPH.DEG		
	C.GRAPH.TP		
	C.GRAPH.Y-INT		
	C.GRAPH.ZEROS		
	C.R-SQ		
	C.TORN.FIT.FUTURE		
	C.TORN.INC		
≥ 50% repeat	0.157894737	0.5	0.333333333

Table E.10 White Box Group 5 Interview Comments

Research Questions	Group 5- II		
1) How does the student feel about mathematics and their ability to do and understand math?	A.PREV	2A.USE.SOME	2A.CONF
	A.USE.SELF		2A.SCA.USE.SELF
	A.USE.SOC		A.CHANGE
			A.FRAC
			A.GRAPHS
			A.INTRIN
			A.MATH
			A.SCA.USE.SOC
≥ 50% repeat	0	1	0.25
2) What is the student's reaction to the course?	R.EASY	2R.INT.ALG	R.HW.T
	R.LEC		R.OHW
	R.LEC.INS		R.OHW.CHANGE
	R.OHW.CONVEN		R.OHW.COMP
	R.REC		**R.REC**
	R.REC.HW		R.REC.INS
	R.STU		R.STRUG
	R.STU		R.STU
	R.STU.APP		**R.STU**
	R.STU.PARTNER		R.STU.CONN
≥ 50% repeat	0	1	0
3) What does the student do in order to succeed?	2E.STDY.OLDEX	2E.OHW.FRIEND	E.HW.REC
	E.OHW.HINTS	2E.TUT	
		E.HELP.FRIEND	
		E.HELP.TUT	
		E.HW.2	
		E.HW.TUT	
		E.OHW.1	
		E.OHW.BOOK	
		E.STDY.1	
		E.STDY.FRIEND	
		E.STDY.HW	
		E.STDY.TUT	
		E.WHW.1	
≥ 50% repeat	0.5	0.153846154	0
4) What is the student's conceptual understanding of "function"?	C.CHART->GRAPH	C.CHART.DRAW	C.FUNC.ALL3
	C.TORN.INC	C.ED.DIS	C.R-SQ
		C.EQ->GRAPH	C.VOCAB.INCORR.2
		C.FUNC.ALL3	C.VOCAB.INCORR.4
		C.FUNC.INC	
		C.FUNC.USE.1	
		C.TORN.FIT.TREND	
		C.TORN.INC	
		C.VOCAB.8	
≥ 50% repeat	0	0	0