

Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis¹

Alex J. Bowers^{2,3}

Teachers College, Columbia University
Bowers@tc.edu

ABSTRACT:

School personnel currently lack an effective method to pattern and visually interpret disaggregated achievement data collected on students as a means to help inform decision making. This study, through the examination of longitudinal K-12 teacher assigned grading histories for entire cohorts of students from a school district (n=188), demonstrates a novel application of hierarchical cluster analysis and pattern visualization in which all data points collected on every student in a cohort can be patterned, visualized and interpreted to aid in data driven decision making by teachers and administrators. Additionally, as a proof-of-concept study, overall schooling outcomes, such as student dropout or taking a college entrance exam, are identified from the data patterns and compared to past methods of dropout identification as one example of the usefulness of the method. Hierarchical cluster analysis correctly identified over 80% of the students who dropped out using the entire student grade history patterns from either K-12 or K-8.

Keywords: cluster analysis, longitudinal studies, data analysis, grades (scholastic), curriculum, dropout research, dropout characteristics, graduation, data driven decision making.

INTRODUCTION

Data driven decision making (3DM), has recently emerged in the literature as a powerful means through which teachers and school leaders are able to gather together around student and school-level data to inform decision making and tailor instruction and resource allocation to students and classrooms (Copland, Knapp, & Swinnerton, 2009; Halverson, Grigg, Prichett, & Thomas, 2007; Ikemoto & Marsh, 2007; Raths, Kotch, & Carrino-Gorowara, 2009; Wayman & Stringfield, 2006a). To date, much of the research on 3DM has identified the practice of creating dialogue around student standardized test scores, which has been shown to increase professional communities of practice in schools, help teachers adjust to changing school needs, and allow school and district leaders to direct the limited resources of a school district to the instructional issues most relevant for their teachers (Bowers, 2008; Honig & Coburn, 2008; Park & Datnow, 2009). However, schools are flooded with data, from test scores, to teacher assigned grades, periodic formative and summative assessments, attendance, discipline records, and more (Bernhardt, 2004; Creighton, 2001a). While some of the 3DM literature has urged school leaders to leverage all forms of data in schools in service to improve student achievement (Bernhardt, 2004), much of the research to date has focused on standardized test scores. One often-overlooked form of data collected daily in schools is teacher-assigned grades (Bowers, 2009). It has been argued that in the U.S. we have a dualistic assessment system, one based on standardized tests that reports to administrators and policy makers, and another based on grades that reports to students, parents and teachers (Farr, 2000). The purpose of this study is to combine the two emerging research domains of 3DM and the usefulness of teacher-assigned grades using a novel form of data mining, patterning and visualization known as hierarchical cluster analysis (HCA), to provide school leaders, researchers and policy makers a method to make better informed decisions in schools earlier, using data already collected on students.

¹ This document is a preprint of a manuscript published in the journal *Practical Assessment, Research & Evaluation* (PARE). Citation: Bowers, A.J. (2010) Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis. *Practical Assessment, Research & Evaluation* (PARE), 15(7), 1-18. <http://pareonline.net/pdf/v15n7.pdf>

² Alex J. Bowers (bowers@tc.edu); Teachers College, Columbia University; Bowers@tc.edu; 525 W. 120th Street, New York, New York 10027. ORCID: 0000-0002-5140-6428, ResearcherID: C-1557-201

³ At the time of original publication in 2010, the author was affiliated with the University of Texas at San Antonio.

Teacher-Assigned Grades as Useful Data in Schools

When considering teacher-assigned grades as useful data in schools, much of the research on grades has maligned grades as a poor assessment of academic knowledge, and has urged teachers to forgo grades for other forms of more standardized or aligned assessments (Brookhart, 1991; Carr & Farr, 2000; Hargis, 1990; Kirschenbaum, Napier, & Simon, 1971; Shepard, 2006; Wilson, 2004). Termed “hodge-podge” or “kitchen sink” grading, surveys of teachers have repeatedly found that teachers award students grades for a variety of factors, including academic knowledge, attendance, participation, and behavior (Brookhart, 1991; Cizek, Fitzgerald, & Rachor, 1995-1996; Cross & Frary, 1999; McMillan, 2001). Nevertheless, teachers have historically been resistant to efforts to reform grading practices (Cizek, 2000), and instead award grades for these variety of factors, all while standardized testing pressures have increased in addition to, rather than in replacement of, all of the past forms of assessment going on in schools daily (Farr, 2000). Indeed, while administrators have indicated that they privilege standardized test scores over other forms of data (Guskey, 2007), little criterion validity has been shown for test scores as they relate to overall student school or life outcomes (Rumberger & Palardy, 2005), whereas teacher-assigned grades have a long history of predicting overall student outcomes, such as graduating or dropping out (Bowers, 2010).

An emerging line of research has begun to ask why teacher-assigned grades are predictive of overall student outcomes, but are a weak indicator of academic knowledge when compared to standardized test scores (Bowers, 2009, 2010; Lekholm & Cliffordson, 2008; O'Connell & Sheikh, 2009). This research has suggested that about 25% of the variance in grades is attributable to assessing academic knowledge (grades and test scores historically correlate at 0.5), but that the other 75% of teacher-assigned grades appear to assess a student's ability to negotiate the social processes of school (Bowers, 2009). Termed a Success at School Factor (SSF), teachers appear to award grades as an assessment of student performance in the institution of schooling, awarding higher grades for participation, behavior, and attendance which in the end appears to be a fairly accurate assessment of overall student outcomes, such as graduating on time (Bowers, 2009, 2010). For school leaders, who have the unique authority to look longitudinally across the system (Bowers, 2008), this research indicates that teacher-assigned grades could be a useful type of data for 3DM, especially when it comes to early determinations of possible overall student outcomes, such as dropping out of school. Indeed, the vast majority of the dropout literature indicates that early school district identification of the students most at risk of dropping out (as early as late elementary and middle school) may be the most effective means of designing and implementing interventions (Alexander, Entwisle, & Kabbani, 2001; Balfanz, Herzog, & MacIver, 2007;

Rumberger, 1995). This is one of the main purposes of data driven decision-making; using data already collected in schools to help drive decisions on improving specific school, teacher and student outcomes.

Visualizing Data for 3DM in Schools

If grades are useful to help teachers and school leaders engage in 3DM, what are the best ways of going about examining the data and determining where, when, and how students may be overly challenged with the system and therefore would need additional resources and opportunities to improve? The use of cross-sectional means and standard deviations for schools, classrooms and subgroups of students has been long proposed (Creighton, 2001b; Konold & Kauffman, 2009). However, aggregated descriptive statistics give only an overview of the central tendency of a sample, obscuring the actual trends in individual student achievement that may provide the clues to inform teachers and school leaders that a student has shifted from on-track performance to significantly challenged with school. An alternative is to inspect every data element individually for each student, but for schools with hundreds or thousands of students, understanding and interpreting trends becomes impossible. As a third option, some researchers have proposed that single student course failures could be used for this purpose, since early failure in reading or mathematics has been shown to be highly predictive of student schooling outcomes (Allensworth & Easton, 2005, 2007; Balfanz et al., 2007). Indeed, much of the past research has focused on using logistic regression to predict the likelihood of dropping out of school given if a student has failed a core course (Alexander et al, 2001, Allensworth & Easton, 2005, 2007; Balfanz et al. 2007). However, this issue returns to the problem of reducing the rich set of data represented by individual student achievement trends to aggregated means and fitted regression slope equations that are generalizable to the population, but less useful for making data driven decisions for individual students and schools. Additionally, depending upon single course failures may be too late for many students, since the negative effects of failure place the students farther and farther behind before the organization can recognize the problem and devise a solution. The goal should be to interrupt a decline in achievement early, before it results in future course failure, especially if an early but small decline is predictive of major future challenges with school years later (Bowers, 2010). However, educators engaged in 3DM currently lack an effective means to disaggregate data while still providing a predictive context for that data. Cizek (2000) provides a caution for proponents of 3DM for teachers and school leaders in today's schools:

It's an unfortunate irony: At no other time have educators, parents, students and policymakers had so much assessment information with

which to make sense of educational reform; at the same time, these groups also receive little guidance regarding what the information means, its quality or what to do with it. Measurement specialists should not be surprised, if, in the face of assessment overload, educators rely increasingly on intuition or arbitrarily pick and choose from discrepant assessment results when they make important educational decisions. (p.17)

Consequently, while much of the 3DM literature has focused on the use of data systems, data rooms, and discussion of student scores (Halverson et al., 2007; Hamilton et al., 2009; Wayman, Cho, & Johnston, 2007; Wayman & Stringfield, 2006b), few studies to date have proposed and tested methods that would allow schools to not only inspect their student's data, but allow practitioners to understand the complexities of the data, analyze longitudinal trends, and make predictions based on their school's past performance. In this study, I adapt innovations from the broader data mining literature to propose the use of hierarchical cluster analysis (HCA) and heatmaps as a novel means to pattern and interpret longitudinal trends in student data, such as teacher-assigned grades, which I use here to illustrate the method. As an example of a small set of hypothetical data using simplified and extreme values to initially demonstrate the differences in patterns, Figure 1A presents an unordered list of hypothetical student non-cumulative grade point averages (GPA) from just grades 9-12 in which an A=4, B=3, C=2, D=1, F=0. For just eight students, this table demonstrates the complexities of the data. From this data, it is difficult to tell one student apart from another. Imagine not eight students for Figure 1A, but hundreds or thousands, and not just for the four years of high school but all thirteen years K-12. Such tables of data become uninterpretable, and lead to the types of garbage-can decision making (Cohen, March, & Olsen, 1972; March, 1997) that Cizek (2000) warns of in the quote above as practitioners become overwhelmed with the size and longitudinal nature of the dataset. As noted above, focusing instead on measures of central tendency or inferential statistics, such as the mean or logistic regression, also does not address the issue, since the goal is to address the individual needs of each student based on their performance to date in the system provided to them.

The broader data mining literature provides a way to bring order and a means to analyze all of the data without aggregating the data, displaying each individual's information patterned and displayed in a way that allows for interpretation of large longitudinal datasets. Known as hierarchical cluster analysis (HCA), this multivariate statistical method uses a series of nested correlation calculations, or distance measures, to reorder a dataset such that "clusters" of data patterns are closest to each other in a

list (Anderberg, 1973; Hubert, Köhn, & Steinley, 2009; Rencher, 2002; Romesburg, 1984; Sneath & Sokal, 1973). As an example, the table in Figure 1A discussed above presents the non-cumulative grades of eight students for four years of high school, ordered by student number. Figure 1B uses HCA to reorder the list, such that longitudinal data patterns that are most similar are closest to each other in the list, such that students 3, 7 and 5 are proximal to each other, while students 6, 1, and 2 are also proximal to each other, but further away from 3 and 7. HCA also provides a means to draw what is known as a cluster tree, or a dendrogram ("dendro" from the Greek meaning "roots or tree"). Based on the distance calculations, here uncentered correlation using an average linkage clustering algorithm (see Methods and Appendix A), the cluster tree on the left in Figure 1B visually represents the similarity or dissimilarity of each of the data patterns, with shorter horizontal lines indicating more similarity, longer horizontal lines indicating more dissimilarity in patterns. Vertical lines connect the closest rows to form the clusters. Thus, each data row is "clustered" by similarity, such that the previously unordered list is reordered with the most similar data patterns closest to each other.

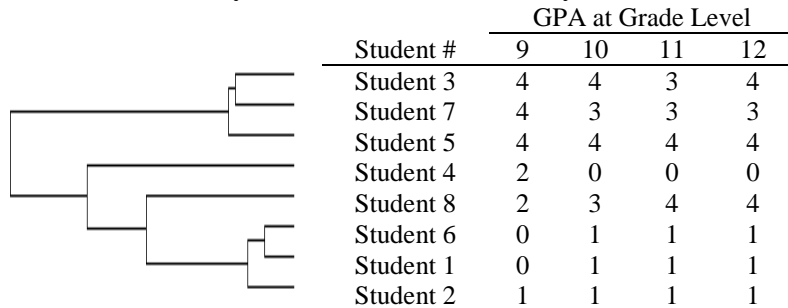
An additional more recent innovation in the data mining literature has been the use of a heatmap with HCA (Eisen, Spellman, Brown, & Botstein, 1998; Weinstein et al., 1997). Tables of numbers or data are difficult for the human eye to interpret, however as stated in the HCA literature, humans are very adept at identifying and interpreting patterns of colors. A heatmap takes advantage of this difference, transforming each data point from a number or symbol (such as a grade) into a block of color, in which a hotter color indicates a higher score (such as red), a cooler color indicates a lower score (such as blue), and a neutral color indicates a central score (such as grey). Figure 1C extends the above HCA example to a heatmap, such that the order of the student list places students with similar grade patterns proximal to each other, the cluster tree indicates the calculated amount of similarity or dissimilarity, and the heatmap allows for the visual inspection and interpretation of the longitudinal grading histories (see Fig. 1). By examining Figure 1C, the longitudinal data patterns for an ordered list of students based on the similarity in grades is made much more obvious. Here, the highly graded students cluster near the top, while the most dissimilar students are in the center (longest horizontal lines in the cluster tree) and the low graded students cluster near the bottom. In addition, heatmaps can also contain dichotomous data as an extension of the main map, here in Figure 1C represented by a black box indicating that a student either dropped out or took the ACT (Fig. 1C, right). For data driven decision-making, from this type of data patterning through HCA and visualization with a heatmap, large longitudinal datasets can be examined without resorting to aggregating the data to overall means, and preserving each student's individual set of data while

Figure 1 An Example of Hierarchical Cluster Analysis (HCA) with non-cumulative Grades

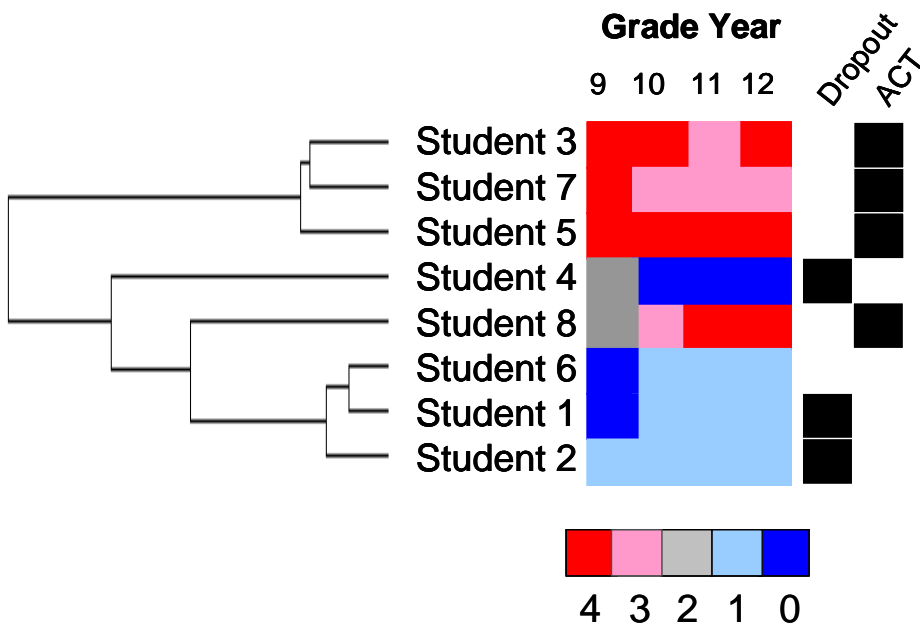
A. An unordered list of hypothetical student non-cumulative GPAs, grades 9-12

Student #	GPA at Grade Level			
	9	10	11	12
Student 1	0	1	1	1
Student 2	1	1	1	1
Student 3	4	4	3	4
Student 4	2	0	0	0
Student 5	4	4	4	4
Student 6	0	1	1	1
Student 7	4	3	3	3
Student 8	2	3	4	4

B. The same list, ordered by hierarchical cluster analysis, with a dendrogram.



C. Clustergram of hypothetical student non-cumulative GPAs, ordered by hierarchical cluster analysis and displayed with a heatmap of student grades.



allowing for pattern recognition, longitudinal analysis, and identification of specific clusters of students based on their performance in the system to date.

Central Aim of the Study

Thus, the central aim of this study is to adapt hierarchical cluster analysis and heatmaps for use with teacher assigned grades for data driven decision-making. The method will be tested and demonstrated with a small sample of data, and the study will explore what the analysis and visualization method can and cannot do for 3DM in schools. The research question of interest here is that as just one example of the usefulness of the method for 3DM, to what extent do student grades cluster through HCA into patterns that identify which students are most at risk of dropping out of school.

METHOD:

Sample and District Context

The entire longitudinal grading histories for the entire class of 2006 for two districts, District A and District B, were collected from the permanent paper file records from both school districts whether or not each student graduated on time in either school district from two cohorts of students. Although multiple school districts were assessed for inclusion in the study, the two districts included in the end were both willing to participate in the study, and had retained achievement records for both students who had graduated and students who had dropped out. Districts A and B are located within the same United States industrial Mid-West state, are within 10 miles of each other, in close proximity to a major metropolitan area, and share a contiguous border. Due to requirements imposed for confidentiality of students, schools and school districts, district specifics are intentionally left vague.

District A is categorized as a mid-sized central city by the United States census, with less than 3000 students enrolled in two elementary schools, one middle school and one high school. In 2006, district demographics included a student population that was about 70% economically disadvantaged, 50% Hispanic, 30% white, and 15% African American (NCES, 2006). District B is categorized as an urban fringe mid-sized city by the United States census, serving fewer than 3000 students who were enrolled in three elementary schools, one middle school and a high school. In 2006 district demographics included a student population that was about 50% economically disadvantaged, 50% white, 20% Hispanic, 15% and African American (NCES, 2006).

Data Collection

The entire longitudinal grading histories of each student in the sample were recorded from the districts' permanent paper file records, copies of report cards, from kindergarten through grade 12 in June of 2006. A student was included in the sample if the student had entered the district at any time on-track to graduate in June of 2006, whether or not the student eventually graduated. This resulted in a sample size of $n=188$.

Grades for each student in each subject at each grade level were recorded. Courses were categorized into subjects based on each district's curriculum guidelines and report cards, such that subject categories included mathematics, English, speaking, writing, reading, spelling, handwriting, science, social studies, foreign language, government, economics, music, physical education, health, computers, study skills, art, life skills and family skills. Letter grades for each subject at each grade level were converted into the following numeric grading scale: A = 4.0, A- = 3.666, B+ = 3.333, B = 3.0, B- = 2.666, C+ = 2.333, C = 2.0, C- = 1.666, D+ = 1.333, D = 1.0, D- = 0.666, E or F = 0. Mean non-cumulative grade point averages (GPA) for each grade level were calculated by calculating the mean GPA for all subjects within each grade level. For each student, other variables were also recorded, such as gender, student transfer into or out of the districts, if the student took the ACT college entrance exam, if the student had graduated on time, or if the student had dropped out prior to graduating. Grades at the high school level were recorded for each semester, denoted as S1 (semester 1) or S2 (semester 2).

Hierarchical Cluster Analysis

Cluster analysis is a descriptive statistical analysis that brings empirically defined organization to a set of previously unorganized data (Anderberg, 1973; Eisen *et al.*, 1998; Jain & Dubes, 1988; Lorr, 1983; Rencher, 2002; Romesburg, 1984; Sneath & Sokal, 1973). There are two types of clustering, supervised and unsupervised. Supervised clustering begins with a defined set of assumptions about the categorization of the data, while unsupervised clustering assumes nothing about the categorization and is designed to statistically discover the underlying structure patterns within the dataset (Kohonen, 1997), a procedure well suited to discovering the underlying patterns within student data in education. While there are many types of unstructured cluster analyses (Anderberg, 1973; Hubert *et al.*, 2009; Lorr, 1983; Romesburg, 1984; Sneath & Sokal, 1973), this study focuses on hierarchical cluster analysis due to the procedure's ability to discover a taxonomic structure within a dataset efficiently (Lorr, 1983; Rencher, 2002; Romesburg, 1984; Wightman, 1993) and its proven use in past studies (Bowers, 2007; Cleator & Ashworth, 2004; Quackenbush, 2006).

Hierarchical clustering provides a way of organizing cases based on how similar the values for the list of variables are

for each case. A brief discussion of the HCA method is provided here while an in-depth presentation of the HCA method used here is provided in Appendix A. In hierarchical clustering, each case is first defined as an individual cluster, a series of numbers for each variable on that case. As an example, this could be a single student's grades in all subjects from grade K through 12. As recommended in the HCA literature (Romesburg, 1984), all data was standardized here through z-scoring to prevent overweighting in the subsequent similarity matrix. A distance measure was then calculated for each case, creating a similarity/dissimilarity matrix. For this study, uncentered correlation was used as the distance measure (see Appendix A). A clustering algorithm was then applied in an iterative fashion at each level of clustering such that the two most similar cases were first joined into a cluster based on how similar the pattern of numbers were for both cases, here using the average linkage clustering algorithm (see Appendix A). This continued in a hierarchical fashion as similar cases were joined to clusters and clusters were themselves joined to similar clusters, until the clustering algorithm defined the entire dataset at the highest hierarchical level as one cluster (Anderberg, 1973; Eisen et al., 1998; Lorr, 1983; Rencher, 2002; Romesburg, 1984; Sneath & Sokal, 1973). Thus, when complete, cases that were previously organized just as a pseudo-random descriptive list, organized alphabetically or by student numbers, were placed nearby other cases in the list with which they had a high similarity, aiding in visualization and identification of empirically defined patterns previously unknown within the dataset. This does not change the data for each case, but merely reorders the cases into clusters based on the similarity of each case's data vector, aiding dataset-wide pattern analysis and interpretation. For an in-depth review of this method, please see Bowers (2007) or Romesburg (1984).

Missing Data

Unlike many of the datasets described in the data mining literature above, education datasets that include all students from a school, cohort or district are notorious for issues with missing data. For the dataset described here of all grades in all subjects K-12 for two cohorts of students, all student cases included missing data for two reasons. First, not all students take all of the same subjects, especially at the high school level. For any one student at the high school level, that student's pattern of course taking differed from many other students. Thus, one student at any one grade level may have data for a subject such as music, but a different student may not have chosen to take that subject at that grade level. Second, many students dropped out of school before the end of grade 12, or transferred into or out of either district, leaving multiple grade levels with no data for that student. These two missing data issues are inherent with these types of district or cohort-wide datasets, and cannot be avoided in

education data. Fortunately, average linkage, as the clustering algorithm, helps to address this missing data issue. In average linkage the distance measure between two cases is the mean pairwise distances between all items contained in the two cases, here uncentered correlation. Hence, if a student drops out and is thus missing data for the later grades the algorithm uses the average of the pairwise distances between that student and the next student to compare for possible cluster inclusion. Thus, students with missing data due to dropout are weighted in their distance measure towards the earlier grade levels that include data for the grades they obtained. Rather than a problem, this provides additional structure within the dataset, as students who dropout at similar times will have similar levels of missing and present data, and thus be weighted similarly in the clustering algorithm and distance measure and pattern together more often, dependent on their grades, which is the overall purpose of the clustering method. While there are other methods to deal with this type of missing data, such as imputing, this study focuses on detailing the overall method and providing a single initial example of its usefulness for education. Thus, while of interest, a discussion and analysis of alternative missing data procedures must be left for future studies.

Clustergrams

To date, while few studies in education use clustering, those that have describe their clustering results in many varied ways (Janosz, LeBlanc, Boulerice, & Tremblay, 2000; Sireci, Robin, & Patelis, 1999; Wightman, 1993; Young & Shaw, 1999). One way to help visualize the organization of the data by hierarchical clustering is to draw a cluster tree, sometimes referred to as a dendrogram (Eisen et al., 1998; Lorr, 1983; Romesburg, 1984). A cluster tree is generated from the similarity matrix outlined above. For each iteration of the clustering algorithm, a line is drawn in the dendrogram as a graphical representation for each case. For each iteration of the algorithm, the cluster tree "grows" as the first level of clusters is connected to other clusters hierarchically, until the entire dataset is represented as a single cluster. Thus, within a cluster tree, clusters of cases and clusters of clusters can quickly be identified by the closeness of lines corresponding to cases and linked to other cases. The unit length of the horizontal line indicates similarity of patterns, the distance in the data space between the two clusters is in the units of the measure, with a shorter line denoting higher similarity.

While clustering provides order to the unordered list, visualization of the data patterns is also important, and one relatively recent innovation in high-dimensionality data visualization is a heatmap (Eisen et al., 1998; Weinstein et al., 1997). A heatmap takes tables of clustered numbers, which the human mind can not easily interpret for pattern recognition, and converts the table into blocks of color,

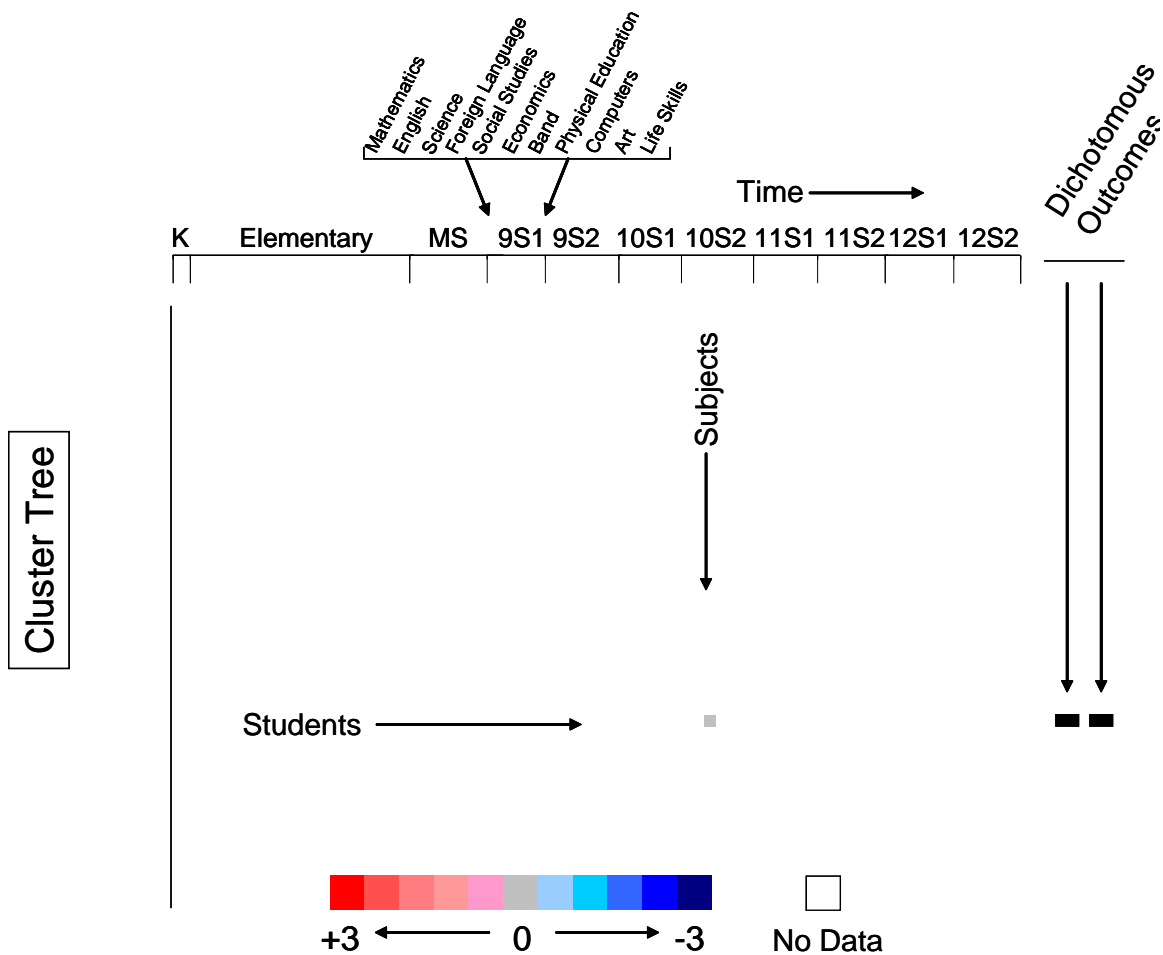


Figure 2: Clustergram Template.

aiding the human eye in visualizing patterns within clustered data and combining these blocks with a dendrogram creating a clustergram (Eisen et al., 1998). For cluster analysis in fields such as the natural sciences, it has become standard to pattern analyze large sets of data and display both a heatmap together with a dendrogram to visualize the patterns within the data and determine if specific patterns align with overall participant outcomes. In addition, while traditional statistical program packages do include clustering algorithms, such as SAS (using PROC CLUSTER) and SPSS, software has been written that can calculate and draw these types of clustergrams (DeHoon, Imoto, Nolan, & Miyano, 2004; Eisen & DeHoon, 2002; Eisen et al., 1998) For this study, publicly available online clustering software (Vilo, 2003) was used to cluster the data, create the heatmap, and cluster tree.

The combination of the cluster analysis, cluster tree, and heatmap, creates the clustergram (see Fig. 2). In the

clustergram, the overall z-scored data for each case is unchanged, but is merely reordered for categorization and pattern interpretation based on how similar each case's data vector is to each other case's data vector. For the clustergrams presented here, student cases are represented as each row. The columns represent a repeating pattern of subjects at each grade level, from more core subjects to more non-core subjects reading from left to right. Thus, for each student in the dataset (each row of data), one can find that student's assigned grade in a subject at any one specific grade level (each column of data). However, a vast table of numbers (here, 188 student rows with 169 subject columns across all grade levels K-12) would be uninterpretable. Following the recommendations for the creation of a clustergram (Eisen et al., 1998; Weinstein et al., 1997), the z-scored grades data were converted into a heatmap, such that any one student's grade in any one specific subject at any one specific grade level is represented as a single color block. The color gradient ranges for these representative

color blocks in the heatmap ranges from a more intense, “colder”, blue for grades -3 standard deviations below the mean, to grey for grades at the mean, to a more intense, “hotter”, red for grades +3 standard deviations above the mean, with missing data represented in white. In this way, rather than a massive table of numbers, a horizontal line of varying color blocks based on that student’s grades represents each student’s grade vector across their time in the school district (see Fig. 2). The hierarchical cluster analysis orders the position of each student within the dataset based on data similarity, thus placing similar lines of data next to each other in the heatmap, allowing for visual interpretation of clusters. Combining the heatmap with the cluster tree allows one to interpret the calculated similarity of each hierarchical cluster in combination with the actual data for every case and every data point. One can then “zoom in” on specific clusters of interest either by eye or by using software to examine the figures more closely.

In addition, clustergrams may also include a final set of data for each case’s data row, in which overall categorical covariates are displayed at the end of the heatmap but were not included in the clustering algorithm calculations (van’tVeer et al., 2002; vandeVijver et al., 2002). For the example presented here, student categorical variables included dropout, if the student took the ACT exam, was female, or had attended district A. In such categorical representations, the presence of the variable for a student’s case is represented by a shaded bar, while the absence is represented in white (Fig. 2, right). When combined with the clustering, heatmap and cluster tree, the categorical variable listing provides the reader with overall information on each student’s case, such as dropping out, patterned in relation to other students with similar data patterns, aiding interpretation of clusters of students and clusters of clusters. Overall, this disaggregated data visualization technique of the clustergram, in which all of the data across the entire dataset is patterned and displayed allows one to examine all of the data together, patterned and disaggregated. In many ways, rather than aggregate data using averages or other measures of central tendency, HCA combined with a clustergram allows for overall data pattern interpretation without the loss of individual student data and variability to aggregation.

Clustergram X-Axis Subject Order

The order of subject columns on the X-axis in the clustered heatmap is as follows reading from left to right: K (kindergarten) - mathematics, speaking, writing, reading; grades 1-4 – mathematics, reading, writing, spelling, handwriting, science, social studies; grade 5 – mathematics, reading, English, spelling, handwriting, science, social studies; grade 6 – mathematics, reading, English, spelling, handwriting, science, social studies, music, physical education, art; grade 7 – mathematics, English, science,

social studies, music, physical education, health, art; grade 8 - mathematics, English, science, social studies, music, physical education, study skills, art; grade 9 semester 1 (9S1) – mathematics, English, science, social studies, foreign language, government, economics, music, physical education, computers, art, life skills, family skills. Grades 9 semester 2 through grade 12 semester 2 repeat the grade 9 semester 1 pattern.

FINDINGS:

An Example of Hierarchical Clustering: HCA using longitudinal grade histories

The main goal of this study is to present hierarchical cluster analysis (HCA) and visualization techniques as a useful method for the organization and pattern analysis of large sets of school and district data to aid data driven decision making (3DM). The study design and the hierarchical clustering and visualization clustergram methods are adapted from the data mining literature detailed above (Eisen et al., 1998; van’tVeer et al., 2002; vandeVijver et al., 2002; Weinstein et al., 1997). Briefly, the study design consists first of a hierarchical cluster analysis and display of a large number of different assessments on each case in the dataset. Here, teacher assigned grades for each student in two cohorts from every subject and every grade level. Second, the cluster pattern is compared to an overall outcome of interest, here student dropout, to assess if the cluster patterns of the assessment align with the overall outcome patterns. Third, other categorical covariates are compared to the cluster pattern, such as gender. Fourth, the hypothesis is that when clustered using the entire longitudinal K-12 grading histories of entire cohorts of students, teacher assigned grades should predict overall student outcomes, such as dropping out or taking the ACT, by clustering students into identifiable clusters based only on their grades.

Figure 3 presents a clustergram that displays the results of the hierarchical cluster analysis and visualization. Since data visualization techniques that simultaneously display each disaggregated data point for the entire dataset and the analysis are rare in education, the Figure 3 clustergram at first glance appears overly complex. However, it consists of three main segments (for a detailed explanation of all of the elements of the figure, please refer to the methods). The center “heatmap” displays the z-scored teacher assigned grades in every subject at every grade level for each student in the dataset. Student cases are the rows. Each column is a specific subject at each grade level, moving from more core courses on the left of each grade level (such as mathematics, English and science) to more non-core courses to the right within each grade level (such as music and art). Each student’s grade for each subject at each grade level is represented by a color block that ranges from a more intense

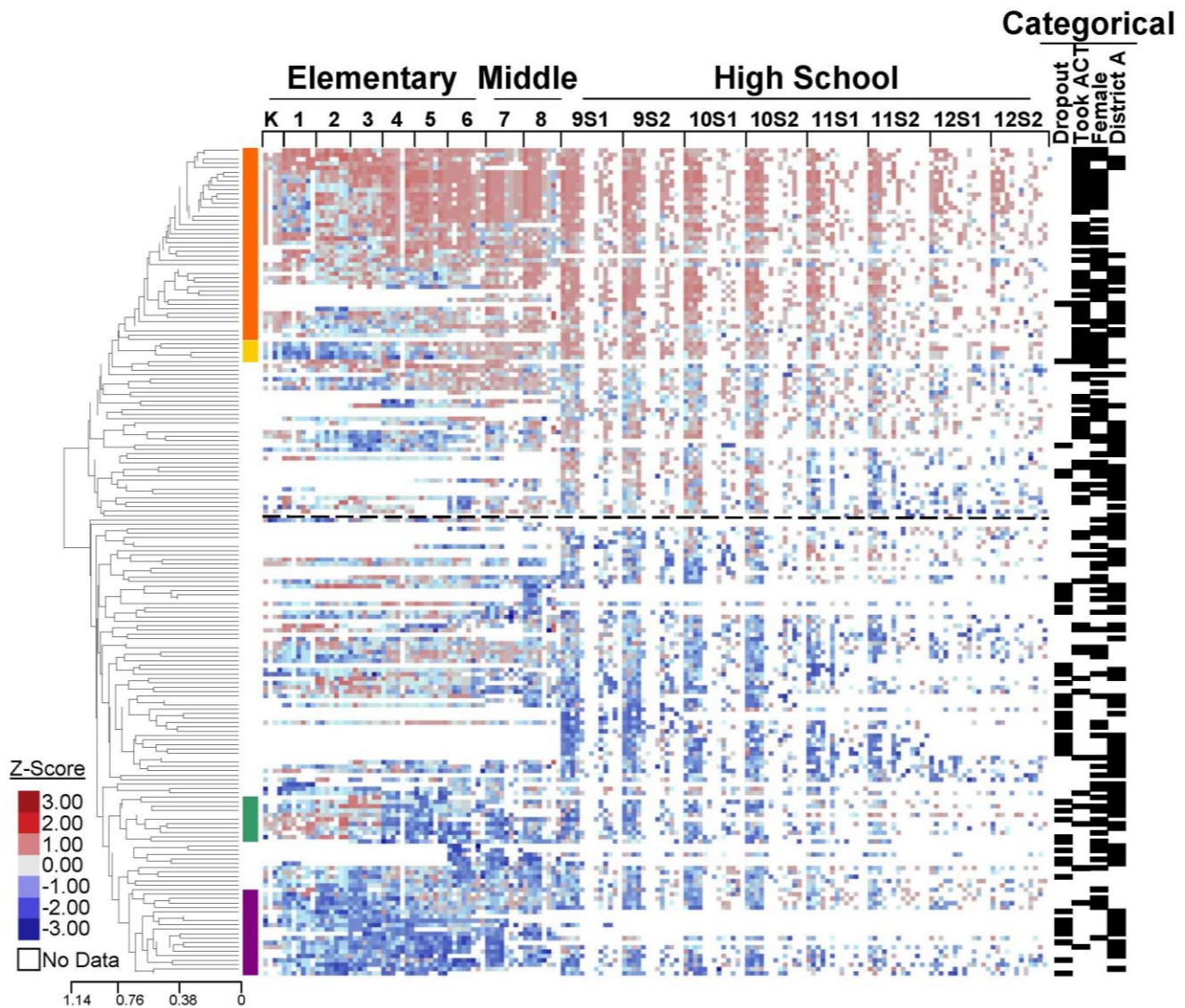


Figure 3: Hierarchical cluster analysis of K-12 student subject-specific grades identifies student dropout. Hierarchical cluster analysis of student subject-specific grades pattern into two main clusters, those who receive generally high grades throughout K-12 and generally graduate on time, and those who receive generally low grades throughout K-12 and dropout more often. Each student is aligned along the vertical axis, with subjects by grade-level aligned along the horizontal axis. Z-scored student grades are represented by a heatmap, with higher grades indicated by an increasing intensity of red, lower grades indicated by an increasing intensity of blue, the mean indicated by grey, and white indicates no data (center). Hierarchical clusters are represented by a cluster tree (left). Black bars represent dichotomous categorical variables for each of the categorical variables listed (right). The dashed black line through the center of the heat map indicates the division line between two major clusters in the full dataset (center). Grade level is indicated along the top horizontal axis (center top). Within each high school grade level two separate semesters are represented, semester 1 (S1) and semester 2 (S2). Subjects are ordered left to right within each grade level from core-subjects to non-core subjects (*see methods*). Four vertical colored bars between the cluster tree and the heatmap (left) denote four sub-clusters detailed in Fig 5.

blue for low grades, to grey for grades close to the mean, to a more intense red for high grades. A white block represents missing data for students in any subject at any grade level. As can be seen from the heatmap, students who transferred into the school districts in elementary or middle school have streaks of white in their row, while students who either transferred out or dropped out have streaks of white extending out through high school (Fig. 3, center). In addition, the clustergram displays the subject enrollment patterns of all students in the dataset, especially at the high school level. Student rows within each grade level have blocks of data to the left within each specific grade level, indicating grades and enrollment in core courses, but also display a more dispersed pattern to the right within a grade level, indicating grades in a variety of non-core courses.

The HCA has reordered the students, from a list ordered alphabetically by last name when the data was collected, to a list ordered by the similarity of each student's longitudinal K-12 grading history pattern. Students who received similar patterns of grades are placed proximal to each other in the list. This clustering is presented in the cluster tree (Fig. 3, left). Cluster similarity is represented by shorter length horizontal lines, and more dissimilar clusters are represented by longer lines, with the two overall largest clusters denoted by the single connection on the cluster tree on the far left (Fig. 3, left) as well as the horizontal dotted black line across the heatmap (Fig. 3, center). The clustering is also evident from the heatmap as students with similar longitudinal grade patterns are clustered together. To maintain confidentiality, student names and identification numbers are not included in the clustergram. However, if this analysis was performed within a school district in which confidentiality was maintained, student names or identification numbers would be listed to the left of each row in the heatmap. Display software, such as a word processor or image viewer, could then be used to zoom in on specific student's patterns.

The final component of the Figure 3 clustergram is the categorical variable listing for each student (Fig. 3, right). As stated above, each student is represented by a row of clustered grade data patterns across the heatmap. On the far right, the categorical data for each student's row of data is presented for if the student dropped out, took the ACT, was female, or attended district A. A black bar indicates the presence of the variable for that student. To aid in reading the figure, the reader may wish to place a blank sheet of paper over the columns of categorical data, and move the paper to the right, revealing one column at a time. In this way, one can compare the categorical variables to the overall clustered pattern to aid interpretation.

As an example of the usefulness of hierarchical cluster analysis and visualization with educational data, K-12

subject-specific grade cluster patterns are informative in identifying student dropout. Figure 3 shows that the sample of students clustered into two main large clusters (Fig. 3, center, dotted black line) in which students generally received high grades throughout their schooling career and graduated on time (Fig. 3 center, upper cluster) or generally received overall low grades near the mean and dropped out more often (Fig. 3 center, lower cluster). Of the students in the lower cluster, 38% of them dropped out of school as compared to only 6% in the upper cluster. When viewed as a percentage of all of the students who dropped out, 88.6% of the dropouts clustered into the low-grading cluster (Fig. 3, center, lower cluster; right dropout category). The opposite pattern occurred in the upper cluster that reflects higher achievement and college preparation. The upper cluster contained few students who dropped out but did contain the majority of students who took the ACT college entrance exam and were female (Fig. 3 center, upper cluster; right, dropout, took ACT, and female categories). Only a slight difference existed between the upper and lower clusters by which of the two districts the students attended (Fig. 3 right, district A category), and this slight difference between the two clusters by district enrollment was confirmed with a chi-square analysis ($\chi^2(1, N=186) = 3.97, p=0.046$). As an early identification method for student dropout, student grade clustering also performed well when the data was reclustered from only K-8 (93.9% of dropouts clustered into the lower cluster) and K-6 (63.0% of dropouts clustered into the lower cluster) (Fig. 4 A & B).

Cluster analysis of course grades also provides an attractive avenue for identifying time points for early instructional intervention by exploring specific student grade cluster patterns. As an example, four individual course grade clusters are identified in Figure 3 between the heatmap and the cluster tree (Fig. 3, left, vertical colored solid bars). These individual grade clusters are informative for dropout identification as each cluster identifies specific patterns of student grades from early elementary throughout the rest of the student's time in the school system (Fig. 3 left; high-high, orange bar; low-high, yellow bar; high-low, green bar; low-low, purple bar). For example, the high-low cluster (Fig. 3, green vertical bar) starts elementary with relatively high grades, but then the grades begin to fall by grade 4 with a high percentage of dropout. This is in contrast to the low-high cluster (Fig. 3, yellow vertical bar) in which the students started elementary school with relatively low grades, but then their grades rose over time with all students in the cluster graduating.

Figure 5 displays a plot of the mean non-cumulative grade point average (GPA) for these four clusters across all subjects for each grade level. While the high-high cluster of students received an "A-" average (near 3.5 GPA) throughout their career in the system with 97.7% graduating on time (Fig. 5, orange), students in the low-low cluster

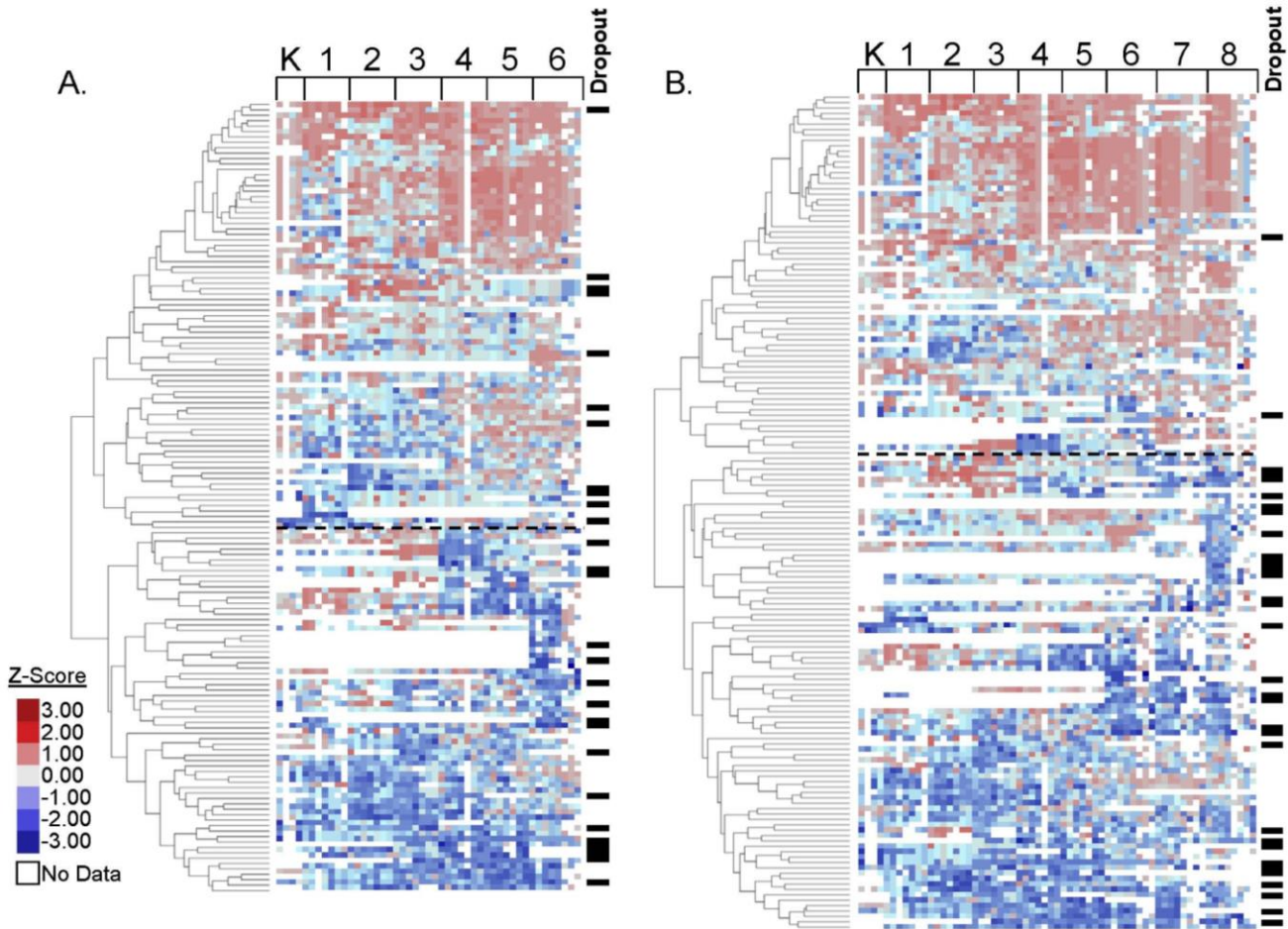


Figure 4: Hierarchical cluster analysis of K-6 and K-8 student subject-specific grades identifies students at risk of dropout and severely challenged by school. Student subject-specific grades were clustered K-6 (A) and K-8 (B). For each clustergram, each student is aligned along the vertical axis, with subjects by grade-level aligned along the horizontal axis. Z-scored student grades are represented by a heatmap, with higher grades indicated by an increasing intensity of red, lower grades indicated by an increasing intensity of blue, the mean indicated by grey, and white indicates no data (center A & B). A cluster tree (left A & B) represents hierarchical clusters. Black bars represent dropout status to the right of each heatmap. The dashed black line through the center of the heatmap indicates the division line between two major clusters in the dataset (center A & B). School and grade-level is indicated along the top horizontal axis (center top). Subjects are ordered left to right within each grade level from core-subjects to non-core subjects (*see methods*).

quickly fell in GPA during early elementary to a C+ average (2.0 to 2.5 GPA) with 40% dropping out (Fig. 5, purple). In contrast to these two groups, the low-high cluster received low grades early but then rose in GPA over time with 100% of the students graduating (Fig. 5, yellow). In addition, the high-low cluster received B+ GPAs up until grade 3 (similar to the high-high cluster) and then fell into a pattern similar to the low-low cluster with GPAs near a C+, with 45% dropping out (Fig. 5, green). For this dataset, these cluster

patterns suggest that early trends in teacher assigned grades appear to be somewhat unstable until grade 4. However, after grade 4, examining specific cluster patterns in this way appears to provide useful information on overall student performance at specific grade levels patterned with students performing similarly.

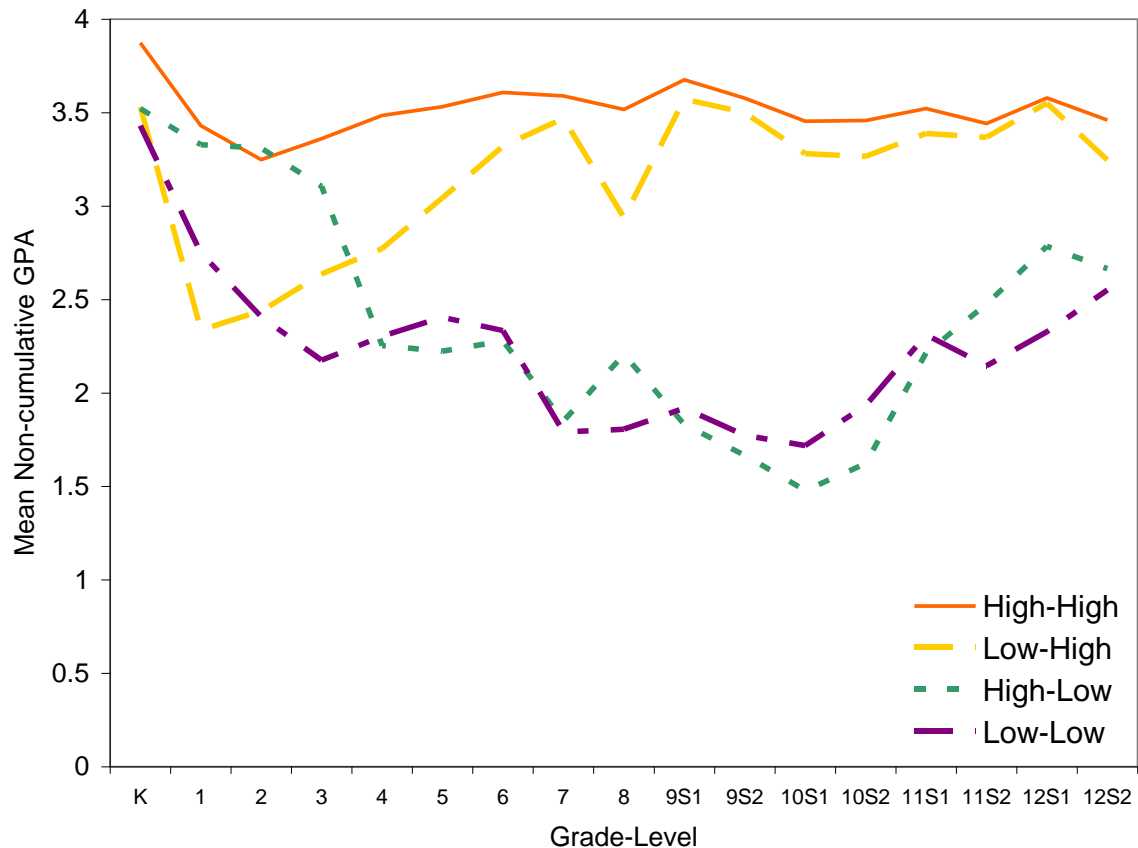


Fig 5 Mean non-cumulative GPA trends, K-12, for four sub-clusters from the hierarchical cluster analysis

Comparison to Past Dropout Identification Literature

Throughout the dropout identification literature the goal is to find a “flag” that accurately identifies students who will ultimately dropout of school (Balfanz et al., 2007; Gleason & Dynarski, 2002). Such flags should provide a means for educators to not only identify which students are at risk of dropping out, but also possible time points, subjects, or areas of schooling through which educators could intervene to help a student graduate. To date, the data on identifying these flags has been mixed (Hammond, Linton, Smink, & Drew, 2007). Previously, to identify flags as variables associated with high risk of dropping out, researchers have first employed a variety of methods to analyze the data, such as linear and logistic regression, determined that a specific variable is significant, and then calculated the percentage of students who dropout who also possess the nominated flag or combination of flags. As an example, using multiple regression Gleason & Dynarski (2002) were able to identify 43% of the students who eventually dropped out using a variety of high school level variables obtained from student surveys, such as family on public assistance, sibling

dropout, high absenteeism, external locus of control, among many others. At the middle school level using the same method, Gleason & Dynarski accurately identified only 23% of the students who eventually dropped out. Recently, Balfanz et al. (2007) identified a combination of flags at the grade 6 level using logistic regression. They were able to identify 60% of the students in their sample who eventually dropped out before graduating from high school. These grade 6 flags included low attendance, unsatisfactory behavior, and failures in math and English. In comparison to this literature, for this dataset, the hierarchical cluster analysis presented here identified student dropouts from only one type of data already collected in schools, teacher assigned grades, and it appears to be an improvement over these past methods. Using K-12 and K-8 data, the cluster analysis identified 88.6% and 93.9% of the students who dropped out, respectively, an apparent improvement over past methods. In addition, hierarchical cluster analysis of K-6 grade data identified 63.0% of the students who dropped out. This is comparable to the grade 6 data of Balfanz et al. (2007).

DISCUSSION:

The central purpose of this study is to introduce hierarchical cluster analysis and pattern visualization methods from the data mining literature and demonstrate the method's utility through one example, identification of student dropout from student K-12 longitudinal grades. For educational data, the method provides a useful and interesting means to visualize and assess an entire disaggregated data history pattern for a student in comparison with every other student's data pattern in a sample. The clustergram allows for the visualization and interpretation of every data point. Each student's data pattern is proximal in the clustergram to students with similar patterns, facilitating system-wide analysis and identification of specific clusters in the dataset. As an example application of the usefulness of cluster analysis with education data, hierarchical cluster analysis of longitudinal student grades in every subject, K-12, provides an interesting avenue to examine assessment patterns to aid in data driven decision-making, and identifying overall student outcomes, such as dropping out or taking the ACT. In comparison to past methods of dropout identification, hierarchical cluster analysis of student grades for this dataset appears to be comparable to past methods.

The application of hierarchical cluster analysis and visualization to education data

This study details the application of HCA and visualization of subject-specific teacher assigned grades. While there is disagreement in the data mining literature over which distance measure and clustering algorithm are best for different applications (Quackenbush, 2006), the uncentered correlation and average linkage methods were chosen here based on their known ability to provide distinctive clusters to provide an initial example application of the method. The question of which clustering method is most useful and efficient with this type of data is of interest, but it is outside the scope of this study. While it is not the purpose of this study to review all types of cluster analysis, future work will focus on comparing distance measures and clustering algorithms to improve the method. Such additional types of distance measures could include Euclidean and city-block distance while comparative clustering algorithms could include k-means and self-organizing maps (Frey & Dueck, 2007; Romesburg, 1984), to name just a few.

The use of cluster analysis in much of the data mining literature has focused on the identification and classification of specific patterns in the data that will predict future participant outcomes (Dolled-Filhart et al., 2006; Kallioniemi, 2002; Lu et al., 2005; Quackenbush, 2006; vandeVijver et al., 2002). This has required clustering in both dimensions, across cases and across potential predictors, in an effort to narrow the number of variables that identify overall case outcomes. For this study, I argue that both dimensions are clustered; students are clustered

hierarchically using the average linkage algorithm, while grades are clustered chronologically and by an ordered repeating pattern from core subjects to non-core subjects. While a subset of subject grades that identify overall course dropout is of interest, and will be explored in future research, the object here is to aid in the identification of potentially useful student data patterns for 3DM. As detailed here with the analysis of specific sub-clusters of students, such as the high-low and low-high clusters, ordering the grades dimension by time allows for the examination of student data trends from early elementary, through high school. While preliminary, the results presented here with subjects and grade-levels ordered chronologically suggest that student grade patterns are somewhat unstable prior to grade four. However, the period between grade 4 and grade 8 seems to be critical in terms of grade patterns when examining overall student performance, such as dropout.

Identification of Dropouts

As an initial example of the usefulness of cluster analysis and visualization for 3DM, I now turn to a discussion of the results of the HCA and visualization for early identification of student dropouts. This study has come to a rather obvious finding; students with generally low grades throughout their career in school dropout. A main critique of this study is that this is already known. However, because literature already exists that demonstrates that student grades are useful in helping to identify who may dropout, this type of data and student outcome provides a useful platform from which to evaluate hierarchical cluster analysis in comparison to past methods.

Past methods of identifying students who may dropout of school have been overly reliant on regression analysis, which inherently aggregates data to the overall means within the dataset. The method here of using HCA retains the disaggregated data for each student, patterns each student's data together with similar student data trends, and allows for interpretation and identification of groups of student patterns that are associated with dropping out. For the dataset examined here, these overall patterns, which appear to become much more stable after grade 4, are as effective as past methods up to the grade 6 level, and the results suggest that the method may be an improvement using higher grade level data. In addition, the analysis here included only one type of data, grades, and this type of data is already present in most schools for every student at every grade level and subject. Rather than collect even more types of data, the results of this study suggests that through the use of these types of pattern analysis and visualization techniques, data that we currently collect in schools but often ignore can be repurposed for 3DM and examined longitudinally to aid in identifying early which students are most challenged by school.

Past research has demonstrated that teacher assigned grades are useful for identifying students who may dropout (Bowers, 2009, 2010). However, to date, the literature using grades to identify dropouts has been problematic in four main ways. First, it is overly concentrated on course failures in core courses such as English and mathematics (Allensworth, 2005; Allensworth & Easton, 2005, 2007; Balfanz et al., 2007; Hammond et al., 2007), a point at which a student has already experienced the deleterious impact of the beginnings of school failure. The findings presented here cluster analyzed the entire grading scale across all subjects, including both core and non-core subjects to capture grading patterns that to date have gone unexamined. Second, many of the past studies required the use of multiple variables in addition to grades, including attendance, and unsatisfactory behavior (Balfanz et al., 2007; Hammond et al., 2007). For the dataset analyzed here, this study suggests that grades alone are very useful for identifying student dropouts when analyzed with hierarchical cluster analysis. Third, these past studies have also overly focused on single grade levels, such as the grade 6 study by Balfanz et al. (2007). The dropout process is a longitudinal “life course perspective” in which student challenges with school slowly build over time (Alexander et al., 2001; Allensworth & Easton, 2007; Finn, 1989; Jimerson, Egeland, Sroufe, & Carlson, 2000), a phenomenon that is important to address in identifying students at risk of school failure, before failure occurs. This aspect of the dropout process is highly amenable to study using hierarchical cluster analysis of longitudinal student grades. Fourth, these past studies have overly relied on achievement in core courses (Allensworth & Easton, 2007; Balfanz et al., 2007; Hammond et al., 2007). This emphasis makes the assumption that core academic knowledge, as represented in core course subject grades such as English and mathematics, is exclusively representing academic knowledge and that little information can be obtained from non-core subject achievement, such as in music, physical education, or art. Excluding non-core course achievement information ignores the wealth of data collected on students that when analyzed longitudinally aids in the identification of students at risk of dropping out of school. Thus, hierarchical cluster analysis of longitudinal student grade patterns addresses these issues in identifying students who may dropout of school.

In addition, the hierarchical cluster analysis and visualization method, detailed here as a clustergram, provides additional information about students that past regression analyses do not. While both types of methods provide information for identification of overall student outcomes prior to those outcomes, the clustergram displays the entire set of data analyzed for every case in the dataset, patterned in a way that aids overall interpretation. This is in stark contrast to regression analyses that aggregate data and report overall parameter estimates. Much like a medical x-

ray, the clustergram provides a unique way to “look inside” each student’s entire history of achievement, and examine that history in context with other students who have performed in a similar manner through pattern analysis. The interpretation of these data patterns for 3DM is then aided through this type of pattern analysis, and helps point to possible areas and timing for future interventions.

As one example for 3DM, examining the clustergram in Figure 3 provides a means to assess the types of courses that students enroll in throughout their career K-12 and analyze the patterns of course taking and curriculum present for different clusters of students. As can be seen in Figure 3, columns of contiguous data patterns begin in the heatmap at grade 9 semester 1, as students take a majority of core courses (core courses are to the left in each column, non-core to the right). These patterns are especially interesting when considering that at the student-level the data are not a sample but rather entire cohorts of students. Thus, interesting and informative patterns in the types of courses taken can be observed. Here, the high-high cluster at the top of the heatmap in Figure 3 appears to take mostly core courses, until curriculum dispersion in grade 12, as their data spreads out across the different types of courses. For the students in the low-low cluster near the bottom of the heatmap, this type of curriculum dispersion begins earlier in grade 11, with these students taking fewer core subjects than other clusters.

As another example for 3DM, the change in Figure 5 for the high-low group occurred as a change between grade 3 and 4 from an average B grade to a C+. Without knowledge of these longitudinal grading history patterns, this type of change may be overlooked in most schools. However, as demonstrated in Figure 3 and 5, students in the green high-low cluster dropped out at an increased rate. The argument here, is that for 3DM using teacher assigned grades as data already collected in schools, knowledge of this seemingly small change in the data pattern in elementary school provides the information to target this narrow window of time to provide these students with additional support before they begin to experience course failure as they reach middle and high school. While this proof-of-concept study included two cohorts from two districts and found similar data patterns across both districts, the next step of this work will be to analyze multiple cohorts over time from the same district to assess the stability of individual grade cluster patterns. If specific patterns are predictive from one cohort to the next within the same district this would indicate schooling or teacher-level effects on student achievement that would be very informative for within district 3DM.

Suggested Citation:

Bowers, A.J. (2010) Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis. *Practical Assessment, Research & Evaluation* (PARE), 15(7), 1-18. <http://pareonline.net/pdf/v15n7.pdf>

REFERENCES:

- Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, 103(5), 760-822.
- Allensworth, E. M. (2005). Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004. Retrieved July 7, 2006, from www.consortium-chicago.org/publications/p75.html
- Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of High School graduation*: Consortium on Chicago School Research at the University of Chicago.
- Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year*. Chicago: Consortium on Chicago School Research.
- Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.
- Balfanz, R., Herzog, L., & MacIver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.
- Bernhardt, V. (2004). *Data analysis for continuous school improvement*. Larchmont: Eye on Education.
- Bowers, A. J. (2007). *Grades and data driven decision making: Issues of variance and student patterns*. Michigan State University, East Lansing.
- Bowers, A. J. (2008). Promoting Excellence: Good to great, NYC's district 2, and the case of a high performing school district. *Leadership and Policy in Schools*, 7(2), 154-177.
- Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, 47(5), 609-629.
- Bowers, A. J. (2010). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *Journal of Educational Research*, 103(3), 191-207.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36.
- Carr, J., & Farr, B. (2000). Taking steps toward standards-based report cards. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 185-208). Norwood: Christopher-Gordon Publishers.
- Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, 19(2), 16-33.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995-1996). Teachers' assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment*, 3(2), 159-179.
- Cleator, S., & Ashworth, A. (2004). Molecular profiling of breast cancer: Clinical implications. *British Journal of Cancer*, 90, 1120-1124.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 17(1), 1-25.
- Copland, M. A., Knapp, M. S., & Swinnerton, J. A. (2009). Principal leadership, data, and school improvement. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook of data-based decision making in education* (pp. 153-172). New York, NY: Routledge.
- Creighton, T. B. (2001a). Data analysis and the principalship. *Principal Leadership*, 1(9), 52-57.
- Creighton, T. B. (2001b). *Schools and data: The educator's guide for using data to improve decision making*. Thousand Oaks: Corwin Press.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.
- D'haeseleer, P. (2005). How does gene expression clustering work? *Nature Biotechnology*, 23, 1499-1501.
- DeHoon, M. J. L., Imoto, S., Nolan, J., & Miyano, S. (2004). Open source clustering software. *Bioinformatics*, 20(9), 1453-1454.
- Dolled-Filhart, M., Ryden, L., Cregger, M., Jirstrom, K., Harigopal, M., Camp, R. L., et al. (2006). Classification of breast cancer using genetic algorithms and tissue microarrays. *Clinical Cancer Research*, 12, 6458-6459.
- Ein-Dor, L., Zuk, O., & Domany, E. (2006). Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proceedings of the National Academy of Sciences*, 103(15), 5923-5928.
- Eisen, M. B., & DeHoon, M. (2002). *Cluster 3.0 Manual*. Palo Alto, CA: Stanford University.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95, 14863-14868.
- Farr, B. P. (2000). Grading practices: An overview of the issues. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 1-22). Norwood: Christopher-Gordon Publishers.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59(2), 117-142.

- Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315(5814), 972-976.
- Gleason, P., & Dynarski, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Students Placed at Risk*, 7(1), 25-41.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholder perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 26(1), 19-27.
- Halverson, R. R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership*, 17(2), 159-194.
- Hamilton, L., Halverson, R. R., Jackson, S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). *Using student achievement data to support instructional decision making* (No. NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Hammond, C., Linton, D., Smink, J., & Drew, S. (2007). *Dropout risk factors and exemplary programs: A technical report*. Clemson, S.C.: National Dropout Prevention Center.
- Hargis, C. H. (1990). *Grades and grading practices: Obstacles to improving education and helping at-risk students*. Springfield: Charles C. Thomas.
- Honig, M. I., & Coburn, C. E. (2008). Evidence-based decision making in school district central offices. *Educational Policy*, 22(4), 578-608.
- Hubert, L. J., Köhn, H.-F., & Steinley, D. L. (2009). Cluster analysis: A toolbox for MATLAB. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 444-512). Thousand Oaks, CA: SAGE Publications Inc.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.), *Evidence and decision making: The 106th yearbook of the National Society for the Study of Education, Part 1* (pp. 105-131). Malden, Mass: Blackwell Publishing.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice Hall.
- Janosz, M., LeBlanc, M., Boulerice, B., & Tremblay, R. E. (2000). Predicting different types of school dropouts: A typological approach with two longitudinal samples. *Journal of Educational Psychology*, 92(1), 171-190.
- Jimerson, S. R., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, 38(6), 525-549.
- Kallioniemi, A. (2002). Molecular signatures of breast cancer - Predicting the future. *New England Journal of Medicine*, 347(25), 2067-2068.
- Kirschenbaum, H., Napier, R., & Simon, S. B. (1971). *Wad-ja-get? The grading game in American education*. New York City: Hart Publishing Company.
- Kohonen, T. (1997). *Self-organizing maps*. New York: Springer.
- Konold, T. R., & Kauffman, J. M. (2009). The No Child Left Behind Act: Making decisions without data or other reality checks. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook of data-based decision making in education* (pp. 72-86). New York, NY: Routledge.
- Lekholm, A. K., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation*, 14(2), 181-199.
- Lorr, M. (1983). *Cluster analysis for social scientists: Techniques for analyzing and simplifying complex blocks of data*. San Francisco: Jossey-Bass, Inc.
- Lu, J., Getz, G., Miska, E. A., Alvarez-Saaverda, E., Lamb, J., Peck, D., et al. (2005). Micro-RNA expression profiles classify human cancers. *Nature*, 435(9), 834-838.
- March, J. G. (1997). Understanding how decisions happen in organizations. In Z. Shapira (Ed.), *Organizational decision making* (pp. 9-32). Cambridge: Cambridge University Press.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- NCES. (2006). Common Core of Data. Retrieved December 18, 2006, from <http://nces.ed.gov/ccd/>
- O'Connell, M., & Sheikh, H. (2009). Non-cognitive abilities and early school dropout: Longitudinal evidence from NELS. *Educational Studies*, 35(4), 475-479.
- Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making. *School Leadership and Management*, 29(5), 477-494.
- Quackenbush, J. (2006). Microarray analysis and tumor classification. *The New England Journal of Medicine*, 354(23), 2463-2475.
- Raths, J., Kotch, S. A., & Carrino-Gorowara, C. (2009). Research on teachers using data to make decisions. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook of data-based decision making in education* (pp. 207-221). New York, NY: Routledge.
- Rencher, A. C. (2002). *Methods in multivariate analysis* (2nd ed.). Hoboken: John Wiley & Sons, Inc.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32(3), 583-625.

- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, 42(1), 3-42.
- Shen, R., Ghosh, D., Chinnaiyan, A., & Meng, Z. (2006). Eigengene-based linear discriminant model for tumor classification using gene expression microarray data. *Bioinformatics*, 22(21), 2635-2642.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (Fourth ed., pp. 623-646). Westport, CT: Praeger Publishers.
- Sireci, S. G., Robin, F., & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education*.
- Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: W.H. Freeman.
- van'tVeer, L. J., Dai, H., vandeVijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., et al. (2002). Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature*, 415, 530-536.
- vandeVijver, M. J., He, Y. D., van'tVeer, L. J., Dai, H., Hart, A. A. M., Voskuil, D. W., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *The New England Journal of Medicine*, 347(25), 1999-2009.
- Vilo, J. (2003). Expression profiler - EPCLUST. 2005, from <http://ep.ebi.ac.uk/EP/EPCLUST/>
- Wayman, J. C., Cho, V., & Johnston, M. T. (2007). *The data-informed district: A district-wide evaluation of data use in the Natrona county school district*. Austin: The University of Texas at Austin.
- Wayman, J. C., & Stringfield, S. (2006a). Data use for school improvement: School practices and research perspectives. *American Journal of Education*, 112(4), 463-468.
- Wayman, J. C., & Stringfield, S. (2006b). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education*, 112(4), 549-571.
- Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J., Kohn, K. W., et al. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science*, 275(5298), 343-349.
- Wightman, L. F. (1993). *Clustering United States law schools using variables that describe size, cost, selectivity, and student body characteristics*. Newtown, PA: Law School Admission Council.
- Wilson, M. (2004). *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education*. Chicago: The University of Chicago Press.
- Young, S., & Shaw, D. G. (1999). Profiles of effective college and university teachers. *The Journal of Higher Education*, 70(6), 670-686.
- Zapala, M. A., & Schork, N. J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences*, 103(51), 19430-19435.

APPENDIX A

Following the recommendations of the literature cited above, hierarchical cluster analysis was performed in this study as follows. First, student course grades were categorized by subject for each semester and grade level from grades K through 12 as discussed in the methods. Second, grades were converted from letter grades to a five-point scale (0-4). Third, the data matrix Y was obtained which contained the data for both cohorts of students with every subject specific grade, K-12, in which y'_i is an observation vector corresponding to each student case, and y_j is a column corresponding to subject specific numeric grades, converted from letter grades as detailed above. Fourth, each y_j was normalized through z-scoring, so that the data in the entire matrix Y was replaced with z-scores based on the means of each subject specific and grade-level specific column, y_j . This step is recommended to control for overweighting in the clustering algorithm by arbitrary cases (Rencher, 2002; Romesburg, 1984). Fifth, the similarity distance matrix was generated. The distance measure employed was uncentered correlation which is commonly used in hierarchical clustering (Eisen & DeHoon, 2002; Romesburg, 1984) and is represented by the following equations:

$$r(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\sigma_x^{(0)}} \right) \left(\frac{y_i}{\sigma_y^{(0)}} \right)$$

Equation 1

in which

$$\sigma_x^{(0)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}$$

Equation 2

$$\sigma_y^{(0)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i)^2}$$

Equation 3

The uncentered correlation function, $r(x_i, y_i)$ defined in Equation 1 is highly similar to the Pearson product moment correlation, except that it assumes that the mean is 0 for every series even when it is not, through the use of a modified standard deviation (σ) (equations 2 and 3) for data vectors for two separate cases, x_i and y_i . This is important when considering two vectors that have the same shape but are separated by a constant value, and thus offset from each other. The Pearson correlation (a centered correlation) would be the same for these two vectors, namely 1, while the uncentered correlation for these two vectors would not be 1 (Anderberg, 1973; Eisen & DeHoon, 2002). This is valuable when examining the similarity of trends of student grade patterns over time, since if the trend of two students was the same, yet they were always offset by one letter grade, the Pearson product moment correlation would deem the two similar. The use of uncentered correlation helps to address this issue in the similarity matrix. Furthermore, it should be noted that the choice of which distance measure is “best” for any particular application is under contention (Anderberg, 1973; Ein-Dor, Zuk, & Domany, 2006; Eisen & DeHoon, 2002; Eisen et al., 1998; Jain & Dubes, 1988; Lorr, 1983; Lu et al., 2005; Romesburg, 1984; Shen, Ghosh, Chinnaiyan, & Meng, 2006; Sneath & Sokal, 1973; vandeVijver et al., 2002; Weinstein et al., 1997; Zapala & Schork, 2006). Hence, while the question of which distance measure performs best with education data is of interest, it is outside the scope of this study.

The sixth step is to apply a clustering algorithm iteratively to the distance matrix. The use of the average linkage algorithm here is due to its demonstrated success in the past in identifying patterns predictive of overall outcomes from similar types of datasets (Bowers, 2007; D'haeseleer, 2005; Eisen et al., 1998; Quackenbush, 2006; Romesburg, 1984). For average linkage, Equation 4, if $r(x_i, y_i)$ is equal to Equation 1, uncentered correlation, the distance between any two clusters A and B is defined as the average distance of the total number of cases within both clusters, $n_A n_B$, between the total number of cases in cluster A, n_A , and the total number of cases in cluster B, n_B , such that:

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} r(x_i, y_j)$$

Equation 4

where the sum is over all of x_i in A and all of y_j in B. Equation 4 is applied iteratively over the distance matrix, as the two vectors with the smallest distance are joined into the first cluster and the matrix is updated with the average linkage of the vectors from Equation 4. This process iterates over the matrix hierarchically, clustering similar clusters to similar cases and other clusters, until the entire dataset is finally clustered into one final cluster.